Curtin University Sustainability Policy (CUSP) Institute

Determinants of Public Transport Use in Perth, Western Australia

S Zaung Nau

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

April 2015

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

………………….

S Zaung Nau

24 Apr. 15

Abstract

The main objective of this thesis is to systematically examine primary determinants that explain both spatial and temporal variations in public transportation use in Perth, thus offering a comprehensive and cogent analysis of public transport use in Perth metropolitan suburbs. To that end, the thesis first provides rich descriptions of public transport usage patterns in Perth based on types of patrons and various temporal and spatial factors. Second, it offers a comprehensive and rigorous analysis of use-determinants and develops a robust predictive model based on land use characteristics, urban form, socio-economic conditions and public transport availability factors which can inform policymaking.

This research makes significant theoretical and empirical contributions to our understanding of public transport system. To examine a wide spectrum of public transport use-determinants, this thesis constructs a novel and comprehensive database of public transport in Perth by integrating multiple data-bases on land use characteristics, geo-spatial information, socio-economic conditions, urban form factors, public transport service provisions at the suburb level. Further, it employs the data on revealed preferences derived from smart cards to analyse the variations in temporal and spatial patterns of public transport use in Perth depending on types of patrons, the origin suburbs of their journeys and the day and time of travel. It also has a practical implications for public transport planning for Perth since it speaks to the relative contribution of land use characteristics, socio-economic factors, and public transport service provision in each suburb, which can in turn help to increase public transport usage and promote the role of public transport in Perth.

The research deploys factor analysis to identify latent variables among a wide variety of explanatory variables and then applies the multiple regression analysis to develop a predictive model of public transport use in Perth metropolitan suburbs. The main finding of this research is that the Bus/Ferry service provision density factor is the most important factor in explaining the spatial and temporal variations in public transport use in Perth's metropolitan suburbs, along with socio-economic factors (below $2000 weekly earner dominant income group factor) and land use characteristic (students and mid-aged dominant resident population density factor). Therefore, this thesis points out the importance of addressing the combined influences of land use characteristics, socio-economic factors, and service provision factors in shaping public transport use.

Based on the findings, this study makes some policy prescriptions to improve the public transport system in Perth. First, this study suggests that it is important *to consider the role of university student population density in public policy that plans for sustainable transportation.*

Also, public transport provision should *target the areas where there are a high number of residents whose weekly income is below $2000 for addressing the social and economic disadvantages low-income households encounter in using public transport. There is a need to integrate the mixed land use development and sustainable transportation system because the findings confirm that the presence of high employment densities in the CBD, which is the target area for high provision of public transport service, is highly conductive to higher usage* Lastly, the study *recommends the use of integrated public transport modes and land use planning to increase public transport usage, thus stressing the need to provide more frequent feeder bus service to train stations and more densely distributed bus stops to increase accessibility.*

Acknowledgements

First and foremost I would like to thank my supervisors Professor Jeffrey Kenworthy and Professor Dora Marinova who gave me the freedom to find my own path and words of wisdom when I needed them. Without her support, I could not have completed this project.

I owe a debt of gratitude to Dr Min Ye Paing Hein for his encouragement, support and advice on statistical matters, without which I could not have accomplished detailed analysis. I would also like to thank Mr Jason Turowetz and other friends in University of Wisconsin for their support.

Last but not the least, I would like to express my thanks to:

- Mr Mark Burgess, Mr Wayne Veaney and Mr Lee Sadler (Public Transport Authority) for providing the most important SmartRider and Timetabling datasets,
- Ms Belinda Hayward and Mr. David Chiang (Landgate) for kindly providing me with road centerline and administrative boundary area maps of Western Australia,
- Ms. Janet Iverach (Strategic Policy and Research, Department of Planning) for supporting this project with 2008 employment survey data with detailed GIS map,
- Ms. Katie Norris (Information Consultancy Service, Australian Bureau of Statistics) who provided the estimated resident population, Journey to work data, Correspondence or Concordance files used to convert data for one of geographic area to another and access to Census Table Builder (Pro) to extract the required socio-economic data as necessary,
- Mr. Greg Knight (Curtin University, Business Intelligence and Statistics) for giving access to Curtin Business Intelligence datawarehouse to extract university student population and students (up to year 12) population datasets.

It is dedicated to a place ‘Perth’ which gave me a chance to explore in academic world and ‘Kachin people’ who gave me strength to keep going during the long journey of my study in Australia.

Table of Contents

Table of Figures

Table of Tables

1 Introduction

1.1 Background

The purpose of this research is to empirically test the collective influence of land use characteristics, socio-economic characteristics, urban forms and service provision on public transport use in Perth's metropolitan areas.

Perth [1]is the second fastest growing city in Australia, and it is noteworthy that its population has grown at a rate of 150% between 1973 and 2013, a period of 40 years, Australian Bureau of Statistics (2 April 2014a). According to the "Regional Population Growth, Australia, 2008-09 (Western Australia)" report from the Australian Bureau of Statistics (ABS), the population growth rate of WA in 2008-09 was 3.1%, which was higher than Australia's annual average of 2.5% for the five years leading up to June 2009; it was also the fastest of all Australian states and territories. In June 2009, the Australian Bureau of Statistics (30 March 2010, p. 312-313) estimated the resident population of Western Australia at approximately 2.25 million people, and 74% of the state population (1.7 million) was estimated to live in the state's capital Perth.

Later in 2014, the Australian Bureau of Statistics (2 April 2014b) announced that the estimated population in Western Australia (WA) has reached 2.5 million, of whom 1.9 million reside in Perth (as of June 2013). This population growth rate has gradually accelerated in recent years, with WA's state population growth rate at 3.3% and Perth's at 3.5% between 2012 and 2013 (Australian Bureau of Statistics (2 April 2014a)). Further, the population is projected to reach 4.8 million by 2053 (Australian Bureau of Statistics (2 April 2014a)). Thus, Perth's rapid population growth poses significant challenges for public policy and planning to accommodate increasing demand for housing, transportation, employment, and other community facilities and services.

Sustainability for the future is one of the key things to consider when planning for such a significant population growth and dramatic increase in travel demand. Private car travel contributes significantly to greenhouse gas emissions in Australia. According to the Department of Transport: Government of Western Australia (August 2011) report, car travel contributes to 21% of Perth's greenhouse gas emissions for households and 19% nationwide. Therefore, there is a need to increase public transport usage, as is discussed in the following section.

[1] In this thesis, Perth refers to the state capital of Western Australia and Perth Central Business District (CBD) refers to the area consist of four suburbs such as West Perth, Perth, Northbridge and East Perth, Landgate (2013). Perth suburb refers to one of four suburbs in the CBD.

1.1.1 The Need to Increase Public Transport Usage

Due to the increasing greenhouse gas concentration in the atmosphere, climate changes have been anticipated and warned of globally. In Australia, the Department of Climate Change: Australian Government (2008) has also identified six vulnerable sectors, and begun working with local governments to adapt to appropriate changes. In addition to this issue, fuel prices, interest rates, and mortgage costs are now significant factors in determining the socio-economic and financial circumstances of households within Australian cities Reserve Bank of Australia (2009). Based on data from the Millennium Cities Database for Sustainable Transport, Kenworthy (2007) pointed out that Perth, closely followed by Melbourne, Brisbane, and Sydney, is the 13th highest-ranked city for private transport energy use per capita out of 84 cities around the world. To alleviate greenhouse gas emissions and the pressures of fuel prices and mortgage costs, it is essential to understand what determines residents' travel patterns. This will help in providing better public transport services, which can attract more users.

<u>Environmental Sustainability Needs</u>

In this section, the current status of the greenhouse gas inventory at the nation and state levels are also described and then the status of residential transportation as one of the main contributors to greenhouse gas emissions is discussed, as is its significant growth rate over a 22-year period.

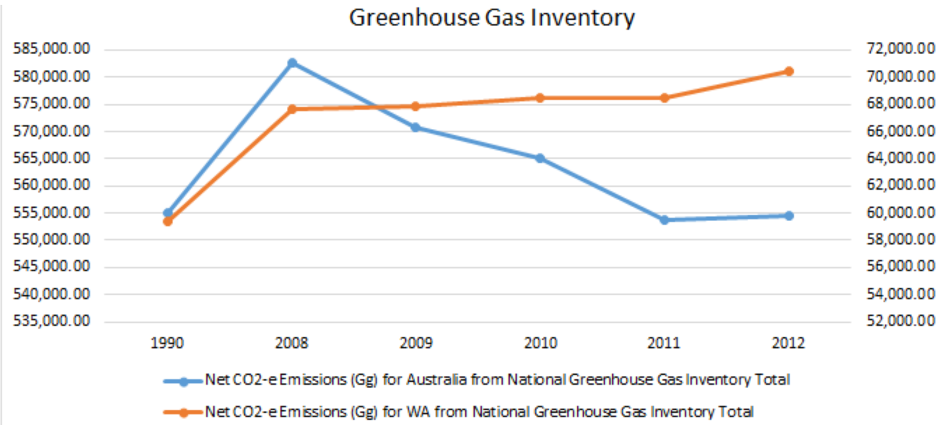


Figure 1: Comparison of National and WA State Greenhouse Gas Inventory

Source of Data from Department of Environment-Australian Government (n.d-a)

As shown in

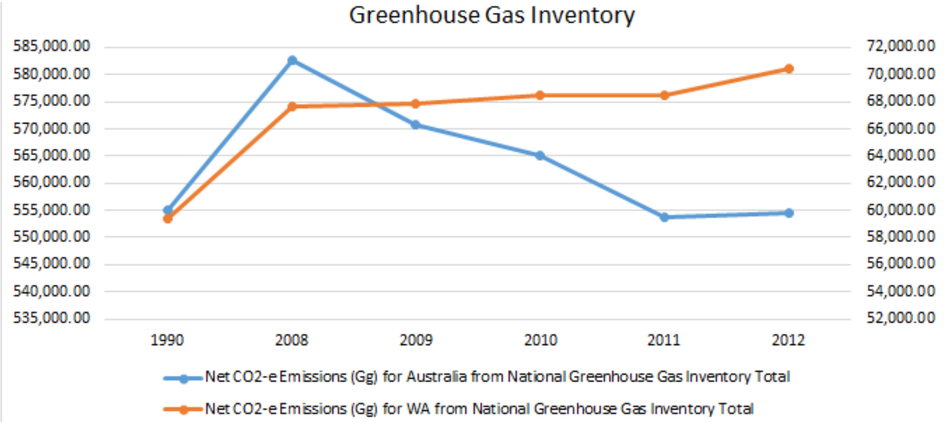


Figure 1, nationwide greenhouse gas emissions increased significantly, from 554,907 Giga grams in 1990 to 582,728 Giga grams in 2008, though they managed to drop back to 554,568 Giga grams in 2012, Department of Environment-Australian Government (n.d-a). However, this was not the case at the Western Australian state level. Here, there was significant 14% increase in greenhouse gas emission between 1990 and 2008, from 59,403 to 67,616 Giga grams, followed by a further increase of 3% by 2012, Department of Environment-Australian Government (n.d-a).

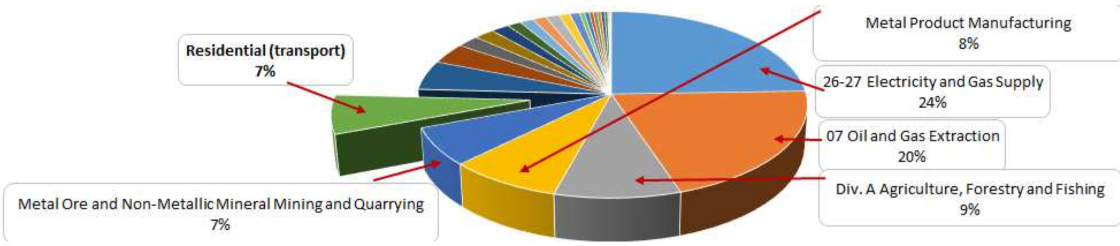


Figure 2: Greenhouse Gas Inventory by Economic Sector in 2012, Western Australia,

Source of Data from Department of Environment-Australian Government (n.d-b)

Figure 2 shows greenhouse gas emissions by economic sector in 2012. Western Australia's residential transport is the sixth highest contributor of all economic sectors to greenhouse gas emissions in that state. Given that approximately 7% of all greenhouse gas emissions in Western Australia are from residential private car use, Department of Environment-Australian Government (n.d-b), shifting from private to public transportation could significantly lower future emissions.

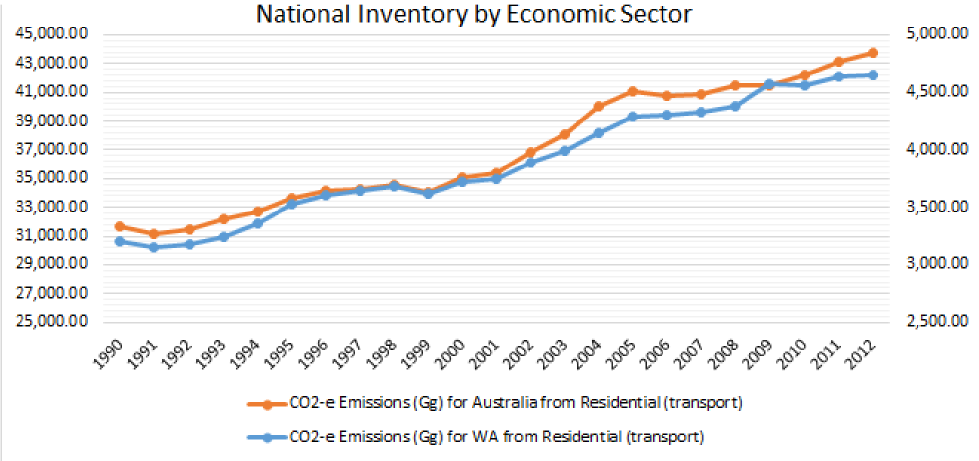


Figure 3: Comparison of National and WA State Greenhouse Gas Inventory by Economic Sector (Residential Transport),

Source of Data from Department of Environment-Australian Government (n.d-b)

As shown Figure 3, the greenhouse gas emissions from private car use have been significantly increasing from 1990 to 2012, at 38% nationwide and 45% in Western Australia, Department of Environment-Australian Government (n.d-b). The Department of Transport: Government of Western Australia (August 2011) suggests that greenhouse gas savings of 270.4kg CO2e per year can be achieved if an alternative form of transportation, such as public transport, is substituted for 1 car trip (20km journey) per week. All of these figures regarding greenhouse gas emissions highlight the important role of shifting from private car travel to public transport use to achieve a reduction and ensure a better future.

Greenhouse gas emissions from residential transportation affects not only environmental sustainability, but also the economy, in that significant increases in private car travel demand creates car congestion, which in turn increases the travel times and costs.

Economic Sustainability Needs

There is a substantial cost to the economy associated with traffic congestion. The Bureau of Infrastructure, Transport and Regional Economics (BITRE) provides the following quotes from Infosheet No 16 – Urban Congestion – The Implication for Greenhouse Gas Emissions:

- Approximately 50% of vehicle kilometres of travel (VKT) is travelled in congested conditions; and

- The rough (order of magnitude) social expenditure for traffic congestion, based on loss of productive time and road user costs, is on the order of $12.8 billion per annum in major cities across Australia.

In Perth, the estimated economic cost of traffic congestion was approximately 1 billion in 2009, and it is expected to more than double (to 2.1 billion) by 2020 (Public Transport Authority (n.d)). Accordingly, it can be argued that this economic imposition significantly impacts the economy of Perth's metropolitan areas. It follows that significant investment in infrastructure is required to improve traffic flows and promote public transport usage as an alternative to increasing motor vehicle congestion. To this end, Western Australian government departments have announced new land use and transportation strategies for Perth's metropolitan areas. These are described in the following section.

Oil depletion is another important reason to promote the availability of the public transport. Dodson (2007) examine oil vulnerability of Australian cities as a result of rising fuel prices and emphasize the negative impact of rising fuel cost on urban social differentiation, segregation or exclusion. This study links transportation structure and social exclusion as high levels of car dependence exist for outer suburbs in which people with low socio-economic status are concentrated. Similarly, Dodson (2006) argue that the gradient of oil vulnerability is a function of socio-economic , transport (public or private) availability and geographical space. For example, greater public transport service availability in the middle suburban areas (closer to the city centre) is connected to lower oil vulnerability while outer suburbs and fringe areas in most of the growth corridors to the north, north-east, east and south-east of the Perth metropolitan area are suffering from higher oil vulnerability. Not only in Australia, but also in other countries, researchers are trying to highlight the need to reduce the an oil consumption. Perl (2010, , pg.353) predict the oil shortage in 2025 for consumption in US transportation because "*anticipated oil production of 26.3bb in 2025 would be 30 percent below the projected 'business-as-usual' consumption of 37.6bb in 2025*".

1.1.2 Land Use and Transport Strategies for Urban Growth in Perth

Public transportation is playing an increasingly important part in Perth's attempts to cope with massive population growth. In particular, there is a need to make its transportation system more attractive and sustainable. The present research examines public transport usage patterns and their relationships with land use, socio-economic and urban form factors. Public transport service provision features are also considered as moderating factors, as they could encourage or discourage the people in WA to use public transportation.

The Department of Planning and Western Australian Planning Commission (August 2010) released “Direction 2031” to lay out medium term land use and transport plans for Perth’s metropolitan suburbs, where the population is expected to reach 3.5 million. To ensure sustainability “Direction 2031” targets a 50% increase in the current average residential density, which translates in 10 dwellings per gross urban-zoned hectare. For new development areas, the target is 15 dwellings per gross urban-zoned hectare. The plan seeks to increase residential density not only in the central, but also the outer sub-regions. Accordingly, it defines the different levels of residential density zones as low, medium, and high density, which have 10, 15, 17 units per gross urban hectare respectively. Three structural elements are identified for an integrated spatial framework: activity centres networks, which would provide a more reasonable distribution of employment and amenities across the city; movement networks, which would integrate public and private transportation to support the activity centres network; and green networks that would preserve biodiversity, natural amenities, and valuable natural resources. Further, “Direction 2031” identifies six sub-regional planning areas—central, north-west, north-east, south-west, south-east and Peel—and emphasises increasing population density in the central sub-region and easing fringe urban development. Its plans for dwelling development and levels of employment project dwelling and employment distributions that reflect a population distribution plan, which will enhance the agglomeration of economic activity and encourage business synergies. Finally, this plan highlights the need to integrate employment, retail, and activity centres such as educational institutions, community facilities, hospitals, medical centres, and libraries with high-frequency public transportation corridors.

Aligned with this “Direction 2031” plan, the Department of Transport (July 2011b) also released a public transport map for Perth for 2031 to project the future of its public transportation network. The plan involves increasing the capacity of existing networks by providing faster feeder bus services to train stations, building more train stations, and expanding the transit network by, for example, adding a light rail infrastructure before 2031, and a rapid transit infrastructure beyond 2031. This plan also highlights the need to integrate train service, road-based rapid transit service, and buses, and to increase service provision significantly in the central sub-region, where higher intensity activities (including population, employment distribution) are projected. According to the Department of Transport (July 2011a, , p.g 6), public transport use is expected to account for:

- “*One in eight of all motorised trips (currently one in fourteen),*
- *One in five motorised trips in the morning peak period (currently one in eight),*
- *Over 30% of peak hour distance travelled (currently around 20%), and*
- *Nearly 70% of all trips to the CBD (currently around 47%)”.*

One significant planned public transport infrastructure development is the metro area express (MAX), a new 22km light rail network scheduled to commence service in 2019. This new network will run from the northern suburb of Mirrabooka to the CBD, then split into two lines: from Victoria park transfer station and QEII Medical Centre, (Government of Western Australia (n.d)). This new network is to provide more frequent and high capacity service in the inner north, central west, and eastern suburbs of Perth by late 2022.

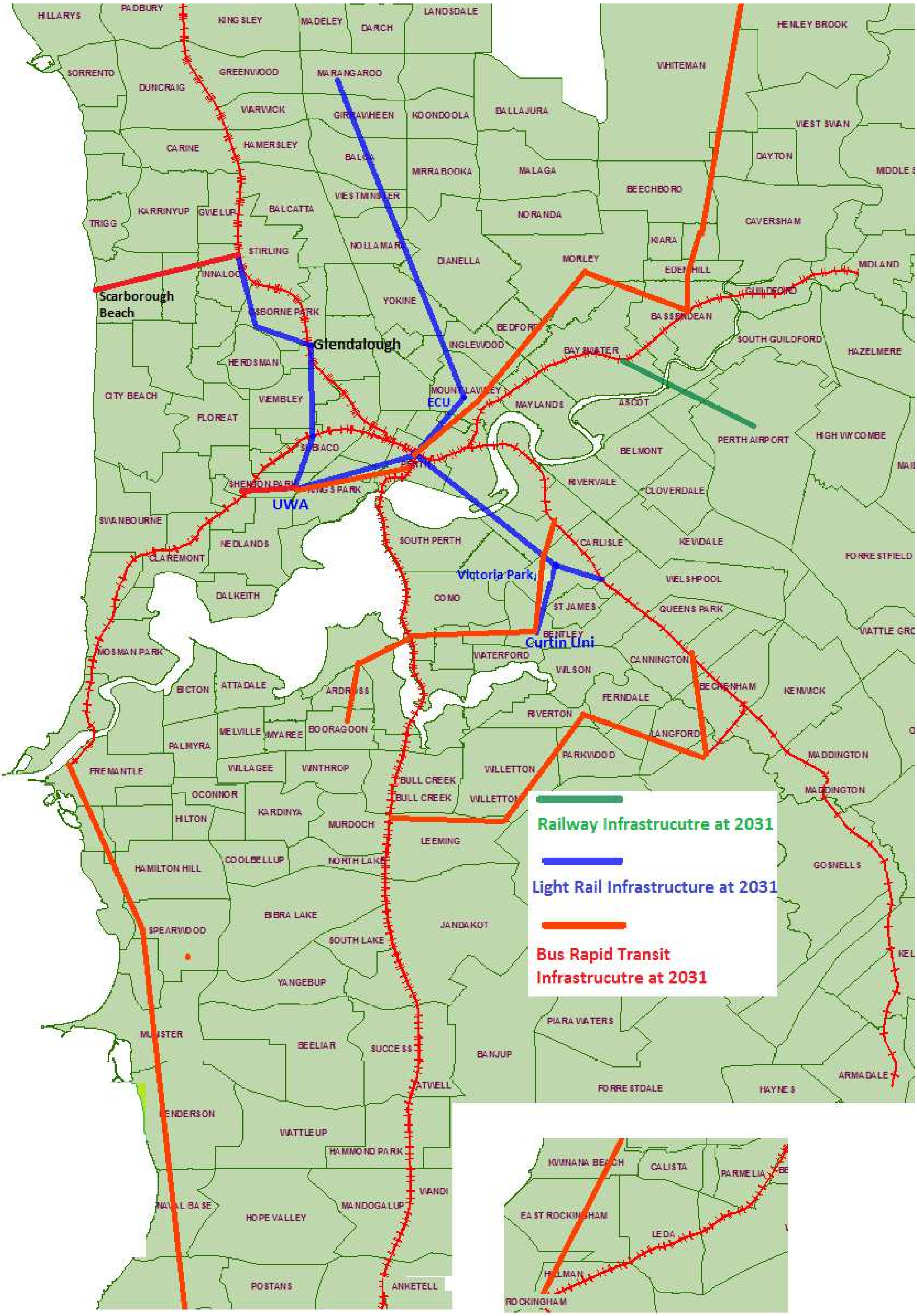


Figure 4: Rapid Transit Infrastructure Combined Stage 1 & Stage 2 Projects

Source of data from (Government of Western Australia (n.d))

Further, the Department of Transport (July 2011a) announced two stage projects to enhance the capacity and infrastructure of public transportation networks. In stage one, the northern suburbs' railways will be extended north to Yanchep, and light rail infrastructures will be developed from Mirrabooka to Perth, and then link the service to the University of Western Australia (UWA) and Curtin University. Some projects being considered for stage two include: a railway to Perth airport from the city centre, a Cannington-to-Fremantle cross-city link via Murdoch, Fremantle to Cockburn Central through Cockburn Coast, Fremantle to Rockingham via Latitude 32 and Kwinana, along with other bus priority projects.

All of these plans indicate the importance of an integrated public transportation network to improving service provision, as well as the need to identify the main determinants of public transport use. This will allow for the most effective and efficient services, which should be based on the current and changing state of these determinants and their relation to the land use development strategies of Western Australia, especially Perth's metropolitan areas.

The Department of Transport (July 2011a) notes that the substantial benefits from increased use of public transport include reduced CBD car parking costs, opportunities to free up car parking spaces for more productive uses, reductions in the travel time and road trauma, and the reduction of harmful environmental impacts, including greenhouse gas emission.

In summary, there is a substantial need to increase public transport use and reduce the increasing demand for private car use that so far has accompanied rapid population growth. This, in turn, will significantly reduce greenhouse gas emissions and promote environmental sustainability, along with the added economic benefits from reducing car congestion. State governments across Australia, the department of planning, the department of transport, and many other governmental departments are planning and implementing integrated land use and transport strategies, such as Directions 2031 – the Perth in 2031 plan. To assist with these strategies and plans, it is essential to understand what the main determinants of public transport use are. This will allow to determine how much service provisions (in terms of frequency and network density) should be enhanced to increase public transport use. Many previous studies have examined the relationships between land use characteristics, urban forms, and socio-economic characteristics and travel demand, as well as the relationship between service provision and public transportation use. In the following section, gaps in previous research are identified. Further, it emphasises the need to investigate how all of these factors collectively produce variations in public transport use, especially in the context of the Perth metropolitan suburbs.

1.2 Statement of the Problem, or ‘Gaps’ in the Research

As described in chapter 2, the interaction of public transport demand and supply is a very complex and dynamic phenomenon with multiple temporal and spatial elements. Spatial variation in demand is one of the cornerstones of public transport modelling. A variety of existing studies deploy various combinations of observed variables to examine the relationship between land use characteristics, urban forms, socio-economic factors, and public transport demand. Even though the general directions of the relationships between a host of selected variables and public transport demand are consistent and robust in many studies, there is a great deal of variance in the magnitudes of these relationships, which is sensitive to the locations researched and the methods used. Much research has been carried out concerning the relationship between land use characteristics, urban forms, and transport behaviour in general, including both private and public transportation ((Boarnet (2001) Cervero (1991), Cervero (2002b), Ewing (1996), McNally (1997), Newman (1996)). Some travel demand research also controls for socio-economic factors in examining the relationship between land use characteristics and travel patterns (Bresson (2004), Cervero (2002b), Hanson (1981), Holtzclaw (2002), Kenworthy (2008a), Kenworthy (1999b), Kenworthy (2008), Kenworthy (2013), Limtanakool (2006), Lin (2008), Paulley (2004), Preston (2001), Sappington (2007) and Stead (2001a)). There is also some research which specifically focuses on public transportation demand’s relationship with land use characteristics and socio-economic factors (White (2004), Bresson (2004), Buehler (2012), Goodwin (1985), Curtis (2011a), Holmgren (2013), Paulley (2006) and White (2009)).

Most of the studies are partial, using various sets of variables to examine the aforementioned relationships. It is more informative to take into account the inter-related nature of the determinants of land use characteristics, urban forms, and socio-economic factors. To fill this specific knowledge gap in the current literature on public transport demand, there is a need to systematically examine the combined influence of these factors, as well as service provision, on public transport use, and to do so on more granular geographical scales and with revealed preference data on travellers’ behaviour.

Based on their international review and evaluation of the literature on relationships between land use characteristics and travel patterns over a 20-year period, Stead (2001b) emphasise the unresolved difficulty of replicating outcomes for comparison in different areas with disparate socio-economic backgrounds. They point out that not only the socio-economic factors, but also the land use characteristics, are interrelated, such that analysing the individual effects of these characteristics on travel patterns is difficult and complex. In addition to identifying gaps in current knowledge, they critique previous studies regarding their data accuracy, reliability, quality, applicability of research methods, and interpretation of findings. More importantly, they stress that the strength of empirical findings and geographical scales of analysis—whether at the regional, city, or neighbourhood levels— need to be considered when identifying policy implications.

Additionally, Willumsen (2011) state that transport demand and supply have very strong dynamic elements, including temporal and spatial factors. He suggests that the spatiality of demand may be the underlying problem in terms of equilibrating demand and supply in transportation modelling. He also states that socio-economic elements, like economic growth, car ownership, public transport usage, and public transport service provision are intertwined in vicious circle. Further, he suggests considering precision and accuracy, decision-making context, and level of detail when evaluating analytical approaches to transport modelling. He also highlights the need for future research on the determinants of public transport use so that sustainable and integrated land use and transportation policies can be made to reduce long term negative effects like urban sprawl and low density developments, for which it is more difficult and expensive to provide good public transportation service.

Moreover, Handy (2005) states the need for future research to include a wider range of explanatory variables, so as to provide a more comprehensive conceptual model of relationships among land use, physical activities, and transportation demand. This need is echoed in Gim’s (2012) meta-analysis of the relationship between density and travel behaviour. He asserts the importance of rigorousness in both research design and technique for discovering the true magnitude of these relationships and producing valid findings. Gim (2012) finds that most of the previous studies of travel behaviour concentrated more on enhancing technical rigorousness than on improving research design. Thus, he urges enhancing rigorousness in terms of limitations in data availability, accuracy, and the granularity of geographical scale. Regarding diligence in research design for data collection, Louviere (2000) highlights the risk of biased results that may occur with stated preference methods.

The limitations of previous research designs mostly relate to the level of inclusion for explanatory variables, data availability, data collection methods, and the granularity of geographical scales. Further, the majority of previous studies used survey or census data collected with stated preference methods to produce hypothetical behaviour analyses. No study has yet looked at comprehensive data from actual trips. Therefore, there is a need to identify the main determinants of public transport usage by using the data collected from revealed preference methods.

Relating to rigorousness of research techniques, multicollinearity between explanatory variables is generally not considered in research modelling transportation demand. Some researchers however use factor analysis to overcome the issue of multicollinearity among observed variables, and also to condense large numbers of these variables into underlying transportation and land use dimensions (e.g. Kenworthy (1986)) or identify patrons’ attitudes and preferences (Kitamura (1997), Mokhtarian (2002) and Shiftan (2008a)). Cervero (1997) and Ewing (2001) use factor analysis to discern dimensions of land use characteristics and urban forms, including density, diversity, and design. They also highlight the need to explore latent variables in terms of multicollinearity among land use characteristics, urban forms, socio-economic characteristics, and public transport service provision, and to specify how these latent variables collectively determine public transport usage.

The present research builds on existing studies while analysing a fuller and richer set of variables to disentangle the complex relationships among determinants of public transport demand. In particular, the study's main objective is to examine the major determinants of public transport use in Perth's metropolitan suburbs. In pursuit of this objective, the research develops the most granular and comprehensive database on public transport usage currently available for Perth. This was done by mining, compiling and merging a wide spectrum of variables covering land use characteristics, urban forms, socio-economic factors, public transport service provision, and public transport usage from a variety of data sources, including smart card data and public transport service timetable data from Perth's public transport authority Transperth; estimated resident population data and census data from the Australian Bureau of Statistics; employment survey data from the WA Department of Planning; road centreline data from Perth's land management authority Landgate; data on students up to year 12 population from the Western Australia Department of Education; university student population data from Curtin University's Office of Strategy and Planning; and rent data from the real estate organisation Real Estate in Western Australia (REIWA). This careful data mining allows for an exhaustive database of revealed preference data to be constructed, allowing for a comprehensive analysis of the combined effects of land use characteristics, urban forms, socio-economic factors, and service provision on temporal and spatial patterns of public transport usage in Perth.

From this database, the study develops a statistical inference model to specify the effects of the aforementioned factors on public transport use density at the suburb-level. To minimize the loss of the data's richness, each explanatory variable is disaggregated carefully, with the aim of capturing important variations. For example, resident population density by age and gender are included separately, rather than aggregated as total resident population density. In addition, service provision is broken into three hours periods on weekdays, Saturdays, and Sundays, instead of aggregating all service provision into week data. Given the number of variables and their interrelationships, factor analysis is applied; this reduces the number of variables included in the analysis by collapsing each variable into an underlying factor. To ensure the rigorousness of the research technique, all assumptions on which multiple-regression is based are thoroughly tested, and required data transformations are performed. Such an analysis can then be used to develop an effective predictive model of public transportation that can be used by planning organisations, such as the Perth Transport Authority. Thus, the model promises to inform the designing of an effective and sustainable public transport system.

Perth's rapid population growth brings the significant challenge to public policy and planning to integrate the mixed land use development and sustainable transportation. Moreover, these integrated land use and sustainable transportation planning is critical to minimize the car dependence from rapidly growing population. Due to having limited resources including budgets for new public transport infrastructure and operational cost, a better understanding on the main determinants of public transport use is essential in planning to integrate mixed land use development and sustainable transportation and to allocate limited public transport service resources with better optimisation for maximum utilisation.

To accommodate this rapid population growth in Perth, the department of Planning and Western Australian Planning Commission (August 2010) developed the "Direction 2031 and beyond" state development plan. This plan clearly outlines the distribution of additional dwelling and employment in six planning sub-regions to meet the needs of the growing population. The state development plan defines the hierarchy of activity centres to provide a more equitable distribution of jobs and amenity throughout the city, namely: capital city, primary centres, strategic metropolitan centres, secondary centres, district centres, neighbourhood centres, local centres and specialised centres. This plan also provides the list of suburbs in which socio-economic activities and infrastructures will be developed in these activity centres.

Along with this state strategic plan for an activity centres network, a public transport network plan for Perth in 2031 was also developed by the Department of Transport (July 2011a). It aims at managing the movements between these activity centres in a more sustainable way and according to it (Department of Transport (July 2011a, , pg. 6), public transport will account for:

- "*one-in-eight of all motorised trips (currently one-in-fourteen);*
- *one-in-five motorised trips in the morning peak period (currently one-in-eight)*
- *over 30% of peak hour distance travelled (currently around 20%); and*
- *nearly 70% of all trips to the CBD (currently around 47%)".*

To facilitate the expected growth in public transport usage, rapid transit infrastructure is proposed to be developed in two stages. Additionally, public transport service capacity is also projected to increase to 390 equivalent railcars (one-car units), 29 light railcars and 2050 buses by 2031.

The funding required by 2031 is estimated as an annual total cost of $4.1 billion, including fleet expansion $1.2 billion and infrastructure capital expenditure $2.9 billion (Department of Transport (July 2011a, , pg.31). The cost benefit analysis used shows that the benefits can be accrued mainly from the lower levels of congestion, reduced CBD car parking costs, reduced road trauma, reduced carbon emissions cost and other reduced environmental impacts. The projected benefit cost ratio[2] is estimated at 1.8 over a 30 year-period and 2.2 over a 40 year-period.

These projected growths in population and public transport usage imply a strong need to conduct rigorous research to identify the determinants of public transport use in Perth so that land use integrated with public transport service provisions can be planned based on influential factors. As such, transport use facilitating increased activity movements by the growing population can be shifted to more sustainable public transport modes. Identifying the determinants of public transport use in Perth plays an important role in planning the integrated land use and transportation systems at strategic, tactical and operational levels.

Based on ~~the~~ travel cost in 2010, the net individual saving is $20 per commuter per day for a typical 15 km peak period work journey to Perth CBD if the passenger makes the shift to public transport. Switching to public transport from car usage can lower net travel cost and bring environment and social benefits. The Bus Industry Confederation (September 2011) states that a 10% shift from car to public transport nationwide would reduce greenhouse gas emission by more than 400,000 tonnes per annum. Additionally, every million passenger kilometres on public transport, instead of private transport such as motor cars, can result in saving 45,000 litres of fuel, Bus Industry Confederation (September 2011). Moreover, the Department of Transport (July 2011b) indicates that public transport patronages, on average, accumulate seven times more incidental exercises than private car users and saving each hour spending behind the wheel of a car can reduce the likelihood of obesity by 6%.

In summation, there is a need to examine the relationship between land use characteristics, urban forms, socio-economic factors, and public transport service provision and public transport use by way of more comprehensive explanatory variables, along with the application of a rigorous research design and techniques that consider multicollinearity, and to base the analysis on actual revealed preference data regarding public transportation use.

[2] Benefit cost ratio means the value for money to the community of the proposed investment.

1.3 Aims of the Research

The main objective of this thesis is to develop a comprehensive understanding of the determinants of public transport usage in the Perth metropolitan suburbs. Whereas prior studies only address partial aspects of this topic, the present research aims for a more complete analysis. This is facilitated, in part, through the use of a more comprehensive dataset than has been previously available. Using these data, the study examines interrelationships among land use characteristics, socio-economic traits, urban form factors and service provision factors, and specify how they collectively influence travel behaviour.

The policy-relevant model developed from this research enhances the understanding of how public transport use patterns in Perth are shaped by various factors. In particular, the ***expected benefits of the model developed in this study for policymaking*** are as follows:

1. The model can assist in predicting shifts in public transport use due to changes in economic factors such as income, car ownership, and average weekly rent.
2. The model can help policymaking concerning increased public transport connectivity and service quality, which are required to accommodate predicted usage, by predicting supply requirements on the basis of a range of socio-economic factors and land use characteristics of individual suburbs.
3. Policy interventions to reduce car dependence and fuel consumption can be analysed to determine how land use characteristics, public transport accessibility, and quality of service can be transformed to increase public transport use, and,
4. More broadly, this research, which introduces and analyses a wide range of potential explanatory factors of public transport use, can help to generate a more widely applicable model for predicting use and identifying factors that can be changed through policy interventions.

The combined influence of the explanatory variables is taken into consideration in developing a statistical model that forecasts public transport use at a very granular spatial level. The main aim in developing this multilevel regression model is to facilitate operational, tactical, and then strategic public transport planning.

The two main objectives of this research can be summarised as follows:

1. To provide a comprehensive descriptive analysis of the public transport usage patterns in Perth's metropolitan suburbs that is based on various temporal and spatial factors, as well as types of patrons.
2. To develop a thorough, rigorous analysis of the determinants of public transport use, and their combined influence on such use, with the intention of constructing a robust demand-forecasting model that can inform policy initiatives.

1.4 Theoretical Framework

Figure 5 is the conceptual theoretical framework for analysing the relationship between public transport usage density and its determinants. The framework illustrates the synergistic impact that all land use characteristics, socio-economic factors, urban forms and service provision have on public transport usage density.

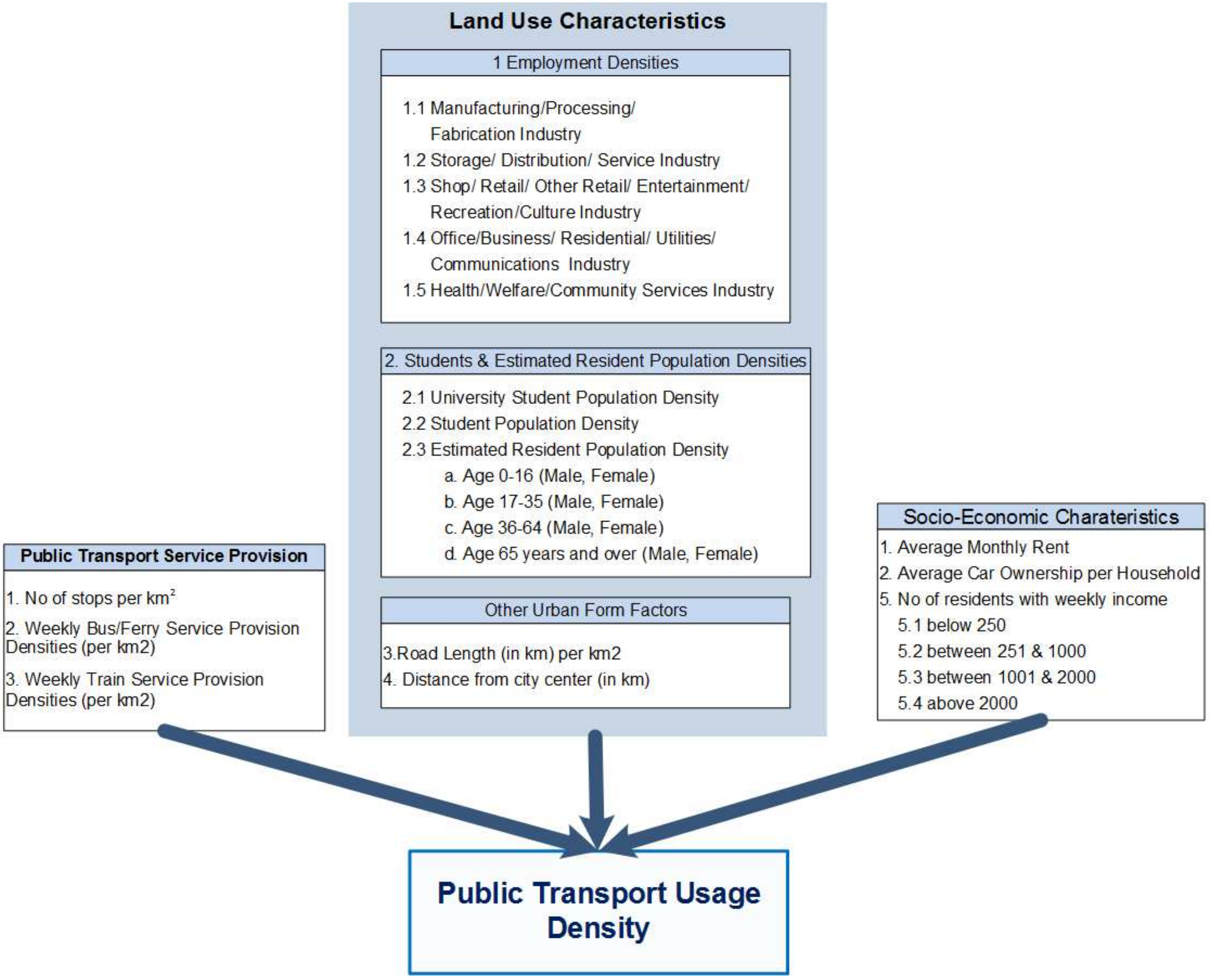


Figure 5: Conceptualised Theoretical Framework

Despite the fact that train speed is significantly faster than that of the bus, the train and bus usages in the Perth metropolitan areas are not much different. According to a report by the Department of Transport (2011a, pg.14), in 2009 bus service accounted for 56% of public transport trips while train usage represented 44% of public transport trips and bus services were used as a major mode of transfer to train services. This implies that train services and bus services in the Perth metropolitan areas are not substitute or competing modes of public transport services but rather complimentary or complementing public services. This complementary nature of public transport is the reason for the relative speeds of train and bus services not to be included in this analysis.

This framework is developed based on the explanatory variables used to analyse public transport usage in previous studies and 2008-2009 data availabilities for the studied area. In the present study, several multilevel regression models are developed based on various combinations of these explanatory variables. This comparative analysis is meant to examine the robustness of each variable's influence, both in terms of its magnitude and (statistical) significance level, when numerous combinations of other observed variables are controlled. As such, it enables a better understanding of the main determinants of public transport use in Perth's metropolitan suburbs. There are other dimensions of mobility and spatial dynamics, such as walkability and bicycle use which fall outside of the scope of this framework because of its focus on public transport. Based on the available data, the developed framework captures the most important aspects of public transport use.

1.5 Brief Description of the Research Design and Definitions of Variables

This research uses a correlational design in which the relationships between explanatory variables and public transport usage density, as the outcome, are analysed. In terms of its temporal dimension, the research design is "cross sectional", which allows for the determination of whether variation in the independent variables (e.g. land use characteristics, urban forms, socio-economic characteristics, and service provision factors) has any relationship with that of the dependent measure (e.g. public transport usage density). In spatial terms, the study takes place at the suburb-level, a unit of analysis that represents the second most granular geographical scale in the Australian census, Australian Bureau of Statistics (2007). All observed variables and public transport usage are aggreated at the suburb-level, and their respective densities are calculated based on the urbanised areas of each Perth metropolitan suburb.

There are two stages of statistical analysis in this quantitative research, factor analysis and multiple regression analysis. First, factor analysis is applied to identify the latent variables among the explanatory variables—based on their correlations—which in turn satisfies the assumption of multicollinearity, such that these latent variables are used in a subsequent multiple regression analysis. Thorough initial checks of all assumptions of regression modelling, along with necessary data transformations are conducted. Then, multiple regression analysis is used to develop a predictive model of public transport usage. As mentioned, a number of regression models can be developed from various combinations of explanatory variables to compare the robustness of their individual influence on the dependant measure. Rigorous cross model validations are also conducted to verify the model's robustness.

The data for this research was obtained from a variety of sources. The variables used in this study are as follows:

- **<u>Spatial variables</u>**

  All density variables are based on the urbanised area of Perth's metropolitan suburbs. The Western Australia Land Information Authority Landgate provided information on the administrative boundary areas of these suburbs. Based on historical imagery data from Google earth, the urbanised areas of each suburb within the administrative boundary were calculated.

- **<u>Dependent variable</u>**

  Public Transport Usage Density

  Public transport trip data in 2009 was extracted from the Transperth SmartRider ticketing system, in the form of revealed preference travel data. Transperth has an electronic system based on the use of a plastic card with a magnetic strip called SmartRider. The system records the start and the end of the journey as each passenger is required to tag-in when boarding and tag-off when disembarking. Public transport usage is measured as the total number of trips generated from origin suburbs during the year of 2009. The total use was then divided by the total urbanised area of each origin suburb to calculate public transport usage density. These trip-data comprise types of patrons, locations of trip origins and destinations, and times of boarding and alighting. This allows the research to include analyses of temporal and spatial variations in public transport use and the different contributions of various types of patrons.

- **<u>Independent variables</u>**

  As the literature review (chapter 2) shows, land use characteristics, urban forms, socio-economic factors, and service provision are highly correlated with travel behaviour. Thus, all of these variables are controlled for this research, and their respective effects are evaluated.

  1. **<u>Land Use Characteristics</u>**

     Density is included as a land use characteristic in this research. It is defined in terms of the number of residents, employees, and students in an area. Further, separate variables with different subcategories are included. Estimated resident population densities, employment densities, and student population densities are measured per square kilometres at suburb level. According to Pushkarev (1976), this type of density measure is known as "density of developed land" and can be defined as the total residential or employment population of a suburb divided by the area within the suburb used for some urban purpose.

Data on **estimated resident populations for 2009** comes from the Australian Bureau of Statistics. The estimated resident population is sub- categorised into four age groups: 0-16, 17-35, 36-64, and 65 years and over, and also partitioned by gender. These subcategories provide a more detailed picture of the differential contribution of each group~~s~~ to variations in public transport use across Perth's metropolitan suburbs.

Data on **employment populations were** extracted from the employment survey (2008), provided by the Department of Planning. Both full time and part time employment are included in this analysis. Further, there are eleven types of industries[3] taken into consideration: primary/rural, manufacturing/ processing/ fabrication, storage/distribution, service, shop/ retail, other retail, office/business, health/ welfare/ community services, entertainment/ recreation/ culture, residential and utilities/ communications. The total number of employees in each industry is agglomerated at the suburb-level on the basis of the location of employment activities.

Data on the **population of students up to year 12 for 2009 were** extracted from Curtin University's business intelligent data warehouse, and validated with the student enrolment dataset provided by the Department of Education of Western Australia. The average of student enrolments in two terms is calculated for each school and then aggregated at the suburb-level based on the school location.

Data on university student populations in 2009 were also extracted from Curtin University's business intelligent data warehouse and validated with student enrolment data in each university's annual reports. University student population data are based on the campus locations where the students were enrolled in 2009.

2. **Urban Form**

The two dependent variables considered as urban form factors in this study are total road length in kilometres and distance from the city centre.

Total road length in kilometres is one of the urban form factors used in this research. It is measured (per unit) in squared kilometres as road network density. This data were

[3] Teleworking is not a common workplace flexibility arrangement in Western Australia. For example, the publication by ABS 6210.5 - Workforce Participation and Workplace Flexibility, Western Australia, October 2010, does not list telecommuting or teleworking as a common mode of flexible workplace arrangement.

generated from the road centreline dataset provided by Landgate and aggregated into total road lengths/kilometres for each suburb.

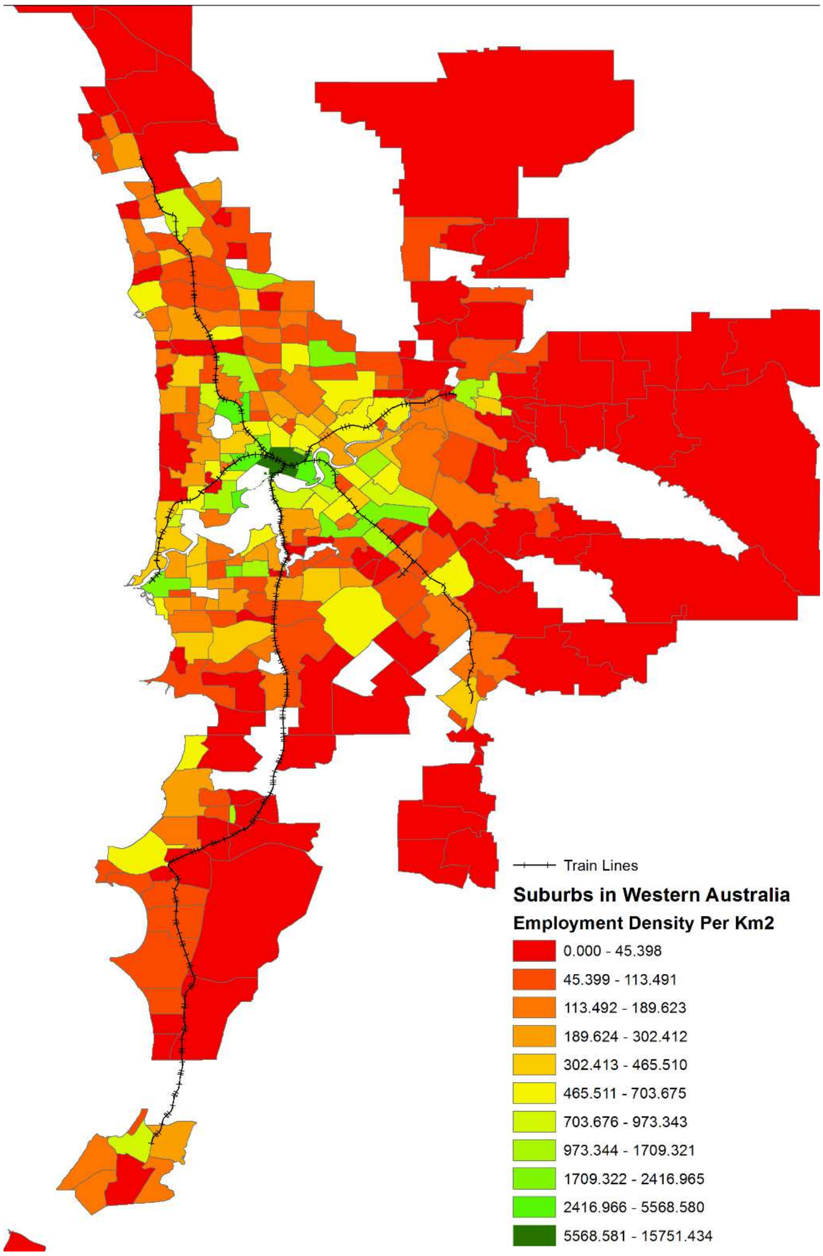


Figure 6: Employment Density per Km^2 Map

Distance from city centre is a second urban form factor that measures the degree of urban sprawl. It is measured as the direct distance between the centre points of the Perth suburb and other suburbs. The centre points of Perth metropolitan suburbs are calculated based on their administrative boundaries. According to Newman et al. (2003, pg.7), "*Perth is very monocentric in terms of the distribution of employment opportunities*". Figure 6 shows that the distribution of employment is mainly located in the Central Business District Area. The distance from the city centre plays an important role in mobility and choice of transport mode. Therefore, distance from city centre is addressed explicitly in this study as one of the urban form factors determining public transport usage.

3. Socioeconomic Factors

Socioeconomic factors have been found to affect travel behaviour. They include income, household expenditure, and car availability. Controlling these variables helps to isolate the relationship of land use characteristics and urban form with public transport use. For example, in examining the relationship between resident population density and public transport use, it is logical to control for income along with car ownership because people who earn higher incomes tend to be less inclined to use public transport, as they are more likely to own a car. Therefore, socio-economic factors need to be considered for their contribution to public transport use.

The number of residents in four different weekly income groups is included as one of three socio-economic factors. Residents are divided into four categories: (1) weekly income below $250 (2) between $250 and $1000 (3) between $1000 and $2000 and (4) above $2000. The number of residents in these four income groups is used instead of average weekly income for each suburb to ascertain how each group contributes differently to public transportation use. Income data were extracted from the 2011 Australian Census.

Average car ownership per household[4] is measured as the total number of registered cars owned or used by household members, or parked in the garage or near the occupied household on the night of the census, and divided by the number of households in a given suburb. These data were also collected from the 2011 Australian Census.

Average monthly rent is included as a measure of housing expenditures. The average monthly rent information for all suburbs in the study was taken from monthly rental data for 2009, published by the Real Estate Institute of Western Australia. Rental data for all types of dwellings are considered in this analysis.

[4] Average car ownership per household is included to measure alternative transport mode availability to public transport. Car leasing and car sharing are uncommon options for Perth residents. New transport options, such as Uber and Lyft are seen as substitute services for taxi rather than common use. Furthermore, their status is questionable in Western Australia because they are illegal in some Australia states (e.g. Victoria, New South Wales and Queensland).

4. Service Provision Factors

Public transport service provision factors have a significant impact on its attractiveness or competitiveness relative to other modes of transportation. According to White (2004), the factors that affect the competitiveness of service quality include frequency of service and distance to the nearest stop at trip origin and destination, as well as accessibility. Thus, these factors should be controlled when determining their relative contributions to variation in public transport use.

Service Provision Density is a measure of total public transport service provision by all modes of travel offered, including train, bus and ferry in the Perth metropolitan area. Accordingly, train, bus and ferry service timetables (provided by Transperth for 2009) are used to aggregate the total service frequencies for three-hour periods on weekdays, Saturdays and Sundays for each suburb. To compute public service population densities, the ratio of total service frequencies to urbanised areas was calculated. Thus, instead of total weekly service frequencies for each suburb, service frequencies were broken into 23 time segments, and different modes of public transport were sub-categorised as variables to examine their respective effects on public transportation use. The findings can assist planners in allocating limited resources in ways that increase service frequency during optimal three- hour-periods.

Average stops per km^2 is used to measure public transport service density and accessibility, which play a key role in attracting patrons. All train, bus, and ferry stop location data files were provided by Transperth. The total number of public transport stops by all modes were counted at suburb level and divided by the urbanised area of the Perth metropolitan suburbs to generate the average stops per km^2 in each suburb.

1.6 Outlines of Chapters

The chapters in this thesis are as follows:

- A literature review of theories and previous research concerning different aspects of the relationship between land use characteristics, urban forms, socio-economic factors, service provision, and travel behaviour—in general or with respect to public transport usage, specifically,
- Detailed descriptions of the origins of multiple data sources based on theoretical considerations and data requirements, as well as the process of integrating and constructing the meticulous data warehouse to examine the synergistic influences of observed variables on public transport usage density,
- Justification of the statistical techniques used to develop a predictive model of public transport usage,
- Documentation of findings from the multilevel regression models, and
- Policy recommendations based on the research findings.

Chapter 1 (this chapter) explains the background of the research problem in the environmental, economic and personal health contexts, and establishes the need for sustainable transportation, especially increases in public transport use. Additionally, the significance of the research is explained. A brief description of the research design and list of variables in this study are also presented in this chapter.

The second chapter (chapter 2) reviews the literature on the effects of land use characteristics, urban form, socio-economic factors, and public transport service provision on travel behaviour. This chapter also provides a historical context for scholarship concerning transport demand modelling in general, especially with respect to three main areas:

- Prior research has examined the determinants of transport demand by using the various combinations of explanatory variables. The numerous empirical findings are described for three categories of observed variables: land use characteristics, socio-economic factors, and public transport service provision factors, and are organized in terms of which variables were reported to be most significant in the studies. This chapter also provides a discussion of how the effects of these variables differ depending on which other variables are controlled.
- The history of transport modelling is briefly described, and two main approaches are reviewed: aggregated analysis and disaggregated analysis, which are applied in the traditional four steps transport modelling paradigm. Further, a brief explanation of existing transport usage models and public transport models is also provided. This is

followed by a description on the transport models currently used in Perth, Western Australia, specifically.

- Recent research that uses public transport smart card data to analyse the temporal/spatial variation and regularity of travel patterns is also reviewed in this chapter.

A summary of the literature review is as follows:

As pioneers of research on the relationship between land use characteristics and travel demand, Newman and Kenworthy (1989) found that density is most strongly correlated with automobile use, as measured in terms of petrol consumption. Later, Cervero (1997) found that population density, land-use diversity, and pedestrian-oriented design generally contribute to decreases in trip rates and increases in non-auto travel. Nevertheless, Zhang (2004) argues that the effects of land use attributes vary depending on trip purpose and how public transport use is measured—either at travel origins or destinations. He emphasizes the composite effect of land use characteristics on transportation mode choices, which is supported by many previous researchers. Kenworthy (2008b) suggests that land use characteristics, such as residential and employment population densities, together with good quality public transport services, encourage public transport use. Cervero (1996a) finds that highly dense areas with mid-rise (3 to 6 storey) buildings are more favourable to public transport use than are other land use variables. Similarly, White (2004) find that increases in population density in cities' central areas reduces the negative impacts of urban sprawl on public transport demand. Moreover, Klinger (2013) confirm that settlement density and population size most strongly encourage the share of public transport as travel mode. Other researchers, such as Pitombo (2011) also find that travel demand is derived from the individual's need to participate in temporally and spatially diverse activities.

Studies further show that gender and age variation in a population influences public transport use. Thus, Hanson (1981) find that men in urban areas have significantly higher trip frequencies than females for all travel purposes except shopping. Based on the National Transportation Survey in Great Britain, White (2009) finds that females tend to use public transport more than males, with a similar distribution by age. White (2004) report that age elasticities are complex because they vary depending on the type of trip and/or patrons' socio-economic status.

Much research has examined the relationship between employment density and public transport use. A strong relationship is consistently reported, especially for high employment density in the CBD and inner suburbs. Hendrickson (1986) reports strong relationships in the CBD area, but not in the overall metropolitan area. Cervero (1991) finds that employment density is more influential than the degree of land use and parking supply. Relatedly, White (2004) show that high employment density in the CBD, along with good quality service provision, encourages public transport use. This is supported by Frank (1994b) who report that high employment density in the CBD, and the inner suburbs, is conductive to high public transporation use.

Other research, such as Tolley (1996), Pitombo (2011) and Curtis (2004b), highlights the important role of student population density in shaping public transport demand, since students a significant travel generator in Australia. In contrast, Bhat (2007) report that when number of households per acre (household density) is used as a density indicator, the relationship of density to car ownership is insignificant. Additionally, Greenwald (2006) finds that the effects of land use characteristics on travel mode and destination choice are relatively small compared to the more significant influences of economic diversity and mixed land use. Also, Susilo (2007) confirm that individual socio-economic factors and job market distribution are more influential on commuting trips than land use characteristics. Further, Ewing (1996) show that residential density, job accessibility, and mixed land use do not significantly influence household trip rates after socio-demographic variables are considered.

Based on such findings, Stead (2001a) like Bresson (2004) suggest including socio-economic characteristics when examining the relationship between land-use characteristics and travel patterns. One of the socio-economic factors considered in this study is average car ownership per household. Kenworthy (1989), Koushki (1988), Mokhtarian (2002), White (2004) and Pitombo (2011) suggest that public transport use has strong negative correlations with private car ownership. Paulley (2004) and Holmgren (2007) urge researchers to take into account both private car ownership and income when modelling public transport demand, since the income variable can pick up the negative effect of car ownership on public transport demand. Accordingly, income and car ownership are included in this study as socioeconmic factors.

Bresson (2004) shows the variation of income elasticities in different regions. Income elasticities are negative in England where public transport is seen as inferior good, but it is not so in France. More interestingly, White (2004) find that the sign (positive or negative) and magnitude of car ownership, and the income elasticities for public transport demand, can vary depending on income levels. They also report variation in income elasticities for different public transport modes and periods (short or long term). Accordingly, different income groups are also taken into account in this research, rather than using average income as observed variable. Additionally, Sipe (2006) highlight the important role of housing expenditure on transport mode choice. According to ABS, housing expenditure is significantly increasing, putting pressure on household expenditures that can influence transport demand. Therefore, average weekly rent is also considered in this study.

Regarding road length as one of the urban form factors, Mogridge (1990) observes that increased road capacity induces private car use and, as a result, patrons may switch from public transport to car use. This is suppored by the findings of Hansen (1993), Luk (1997), Cervero (2001a) and Cervero (2002c). Another interesting finding from Zeibots (2005) is that there was a significant shift from public transport use to private car use after a road network capacity extension was introduced in Sydney in 1992. According to Stead (2001b), the distance from city centre (i.e. distance between each home and an urban centre) has a stronger association with travel distance and transport energy consumption than with travel frequency. Therefore, total road length and distance from city centre are included as controls in the present thesis.

Webster (1982) propose considering both supply and demand factors in determining demand for public transportation. Bresson (2004) use seat km, frequency, and network densities as measures of public transport supply. They conclude that elasticities resulting from variation in public transport service frequency and network densities are significant determinants of public transport use. Barton-Aschman Associates (1981), FitzRoy (1997), Stanley (1998) and Catoe (1998), quoted in White (2004, , p.76)) similarly find that high frequency is an important factor that encourages public transport use. Additionally, Preston (1998), quoted in White (2004, , p.74) report that service elasticities for bus use can vary based on time of day. White (2004) also state that public transportation network density reflects the accessibility of services. According to Cervero’s (2002a) finding, residents within one-half mile of a train station tend to choose public transport over alternative modes. Moreover, Murray (2001) highlights the importance of evaluating the trade-offs between access coverage and stop placement efficiency in evaluating public transport policy, and the use of monitoring to increase service usage. Consequently, this study uses service frequency and average stops per km^2 as measures of public transport service provision.

Some researchers apply data mining techniques, using smart card data, to gain a better understanding of patrons’ travel patterns and behaviours. Pelletier (2011) mention that smart card data have been analysed for decision support at three levels of management, namely strategic, tactical and operational public transport planning. Notably, Morency (2007) measure the regularity of public transport use by aggregating the frequencies for the most frequently used bus stops. Some of other interesting research conducted with smart card data involves designing targeted public transport marketing campaigns, Bagchi (2005), transfer-pattern analysis to enhance timetable design Chapleau (2008), developing public transportation use in terms of the spatial and temporal variability of different modes of public transportation, Lim (2008), analysing linked-trips and turnover rates for scheduling Bagchi (2005), examining the maximum number of boarding points and return runs, which is useful for schedule coordination among different public transport methods (Chu and Chapleau, 2008), generating a comprehensive Origin-Destination matrix that optimizes service planning Munizaga (2012), examining the frequency, consistency and composition of public transportation usage by different modes of travel Utsunomiya (2006), developing a supply-dependent Integrated Intervening Opportunities Model (IIOM) Nazem (2013), and performing user-behaviour analyses of the temporal distribution of public transport use and load-profiles Trépanier (2007).

Next, the research gaps in the current literature on public transport demand and its determinants are identified. A detailed explanation of the research objectives, as derived from the research questions, is provided in this chapter.

The complexity in relationships between land use characteristics, urban forms, socio-economic factors, service provision, and public transport use arise from the interrelationships among the observed variables. Generalising about public transport use by considering only a few factors is quite risky. For example, the population elasticity for public transport use can be different in various age groups. Therefore, it is necessary to examine the collective influence of land use characteristics, urban forms, socio-economic factors, and service provisions on public transport use at a fine-grained geographical level, which can be done using smart card travel data. The main research question is as follows:

> “***What are the primary determinants which explain spatial and temporal variations in public transportation use in Perth for the year 2009?”***

A detailed explanation of the study’s research methodology is provided in two chapters. Chapter 3 presents an overview of theoretical considerations and data requirements. First, it provides reasons for applying a quantitative research approach and explains how the research is designed to contribute theoretically to public transport demand modelling. A conceptual theoretical framework is also provided to illustrate the general analytical features of the study. Next, the spatial and temporal dimensions of the research design are described to clarify its context and nature. In addition, the statistical techniques used to validate the regression models developed for the study are reviewed. Further, the chapter provides a detailed account of the origins of multiple data sources, and the meticulous processes undertaken to standardise the data formats, aggregate (or disaggregate in some cases) the data to be on the same fine-grained geographical scale, and integrate all of these datasets into a data warehouse. Handy (2005) recommends to consider the comprehensive combinations of observed variables and constructing detailed geographical. These recommendations are taken into consideration when extracting each variable from its original database and constructing the required datasets. Finally, the chapter discusses issues encountered in creating the data warehouse and solutions for them, along with the study’s limitations.

Chapter 4 describes in detail the statistical techniques and model development process used in this research. The use of factor analysis and multiple- regression is discussed and justified, as are detailed explanations of their assumptions and methods of confirming them. In addition, the chapter describes cross-validation techniques such as stepwise regression and robust regression modelling as well as the statistical software tools employed in the study.

Chapter 5 presents the empirical results of different analyses—descriptive analysis, factor analysis, and multiple regression analysis. The first part of the chapter focuses on a spatial and temporal analysis of public transport use, as well as composition of use by different types of patrons. Based on the descriptive analysis, factor analysis is performed, yielding two public transport service provision factors and three land use characteristics and socio-economic factors. The reliability and validity of the latent constructs is discussed in terms of co-linearity among observed variables, and the factor loadings are explained. Tests for normal distributions are performed, and data is transformed when necessary to satisfy the criteria of regression modelling. Next, six regression models are created using various combinations of observed variables, with a view to specifying their influence on public transport use when other variables are controlled. The models are developed using a stepwise regression method, as well as a robust regression method to cross-validate the derived model. The results of the stepwise regression are then reported, and the combination of variables with the greatest explanatory power is identified In addition, it is reported that there is relatively low (and statistically insignificant) variation in each predictor-coefficient, based on a comparison of the results of the models using the standard ‘forced enter’ regression method along with robust regression. Therefore, it confirms that the regression model is stable and best fitted to the data.

Finally, in Chapter 6, the thesis makes conclusions regarding the original research questions on the basis of findings from chapter 5, and discusses their theoretical implications. The main finding of this research is that the bus/ferry service provision density factor, along with income (i.e. a below- $2000 weekly earner dominant income group factor) and land use characteristics (i.e. students and mid-aged dominant resident population density factor) are the primary determinants of temporal and spatial variations of public transport use in Perth’s metropolitan suburbs. Recommendations for future research are made and the limitations of the study are discussed. Finally, the chapter discusses policy implications, such as enhancing the bus/ferry service provision, particularly in those areas where mostly mid-aged population~~s~~ resides and where the average weekly income is below $2000. Additionally, it is also recommended that policymakers integrate more frequent feeder-bus services with train services, and that the average stops per km^2 be increased for better accessibility.

The outlines of chapters in this dissertation can be summarised as:

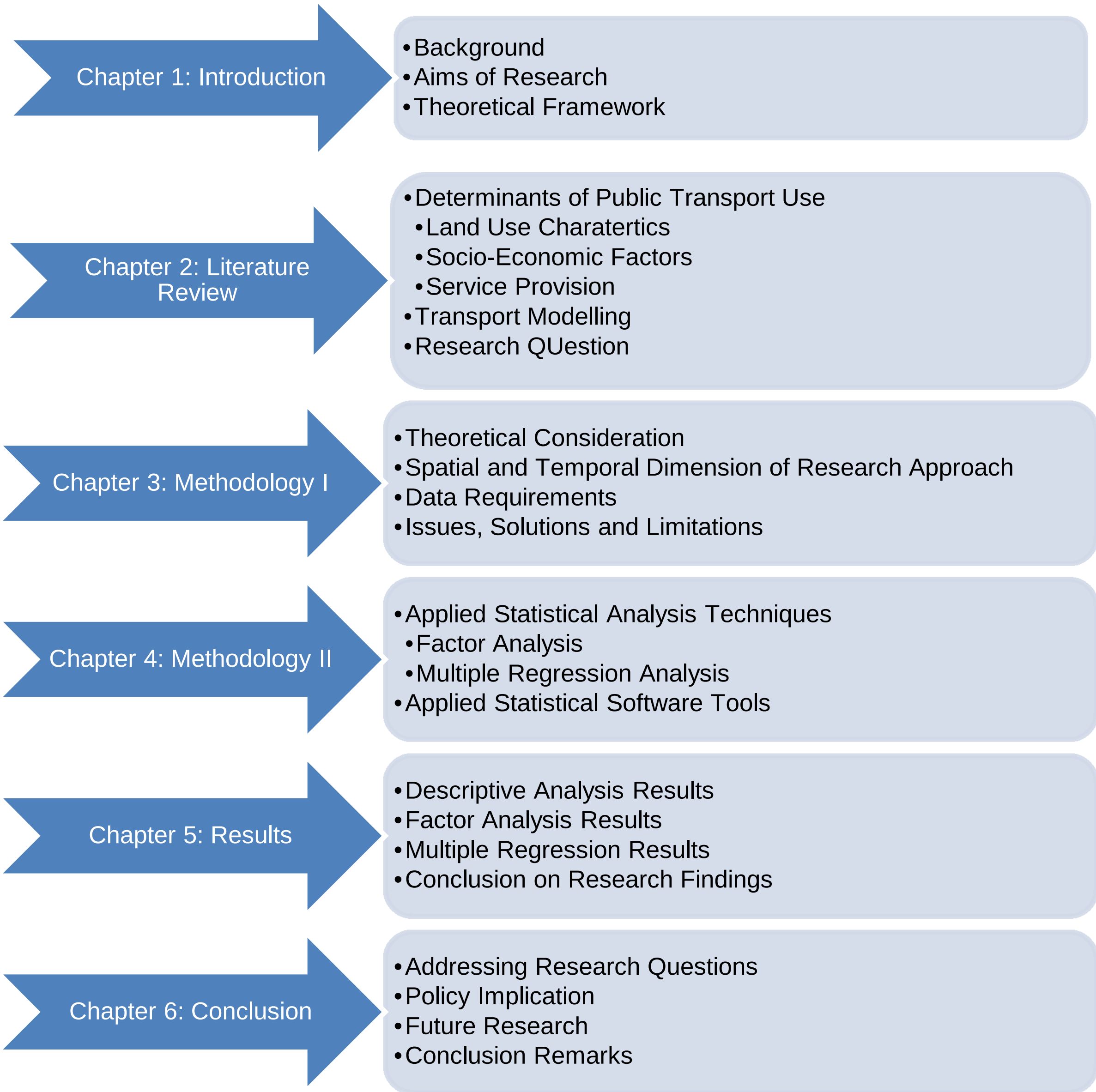


Figure 7: Outlines of chapters

2 Determinants of Public Transport Use

In the literature regarding public transportation, many studies have been conducted to explore the relationship between land use characteristics, socio-economic factors, public transport service provision, and travel behaviour (e.g. Stead and Marshall, 2001; Curtis and Perkins, 2006). As Stead (2001a)states, explanations of relationships between land use characteristics and travel behaviour are confounded with the associations between land use characteristics and socio-economic factors, as well as the impact of those factors on travel patterns. Therefore, this research considers both socio-economic factors and land use characteristics in developing a public transport usage model. Moreover, public transport service provision factors are also included in this study because they are important supply factors affecting public transport usage.

According to a review by Gwilliam (2008), the major issues examined and studied in public transport economics over the last 50 years include the analysis of cost and demand determinants through the application of sophisticated mathematical modelling and analytical techniques to issues of ownership and competitiveness among alternative modes of transportation, the relationship between public transport demand and urban development, among others. He states that public transport plays an important role in achieving a wide range of economic, social and environmental objectives. Therefore, he urges that public transport planning to be integrated with the strategic planning for the rest of urban development policy.

In their literature review of factors influencing public transport ridership, Taylor (2003) state that these factors can be categorised as internal and external. Internal factors are the ones which public transport planners have control over, such as service provision and quality. External factors, on the other hand, are those that planners do not directly control, such as socio-economic and spatial patterns involving land use characteristics and urban forms. They conclude that socio-economic factors have relatively less substantial effect on public transport use than spatial factors, which include residential and employment population densities, and that service provision is more influential than fare in determining public transport use Taylor (2003).

Similarly, based on current literature review of the demand determinants for urban public transport services, Polat (2012) provides a summary of key findings about demand forecast modelling, operational decision making, and demand management. Like Taylor and Fink, he proposes that the determinants of public transport demand can be placed into two categories: structural factors and external factors. The author states that public transport demand is very dynamic, contributed to by alternative transport modes, different types of passengers, their respective travel purposes, and varying travel volumes at different time periods. He also points out that public transport demand is time dependent, as it varies with peak and off-peak time periods, and days of the week (i.e. weekdays or weekends). He also mentions that expectations on public transport can vary based on different travel times and purposes.

As the aforementioned authors recommend, this research will examine the relationships among exogenous variables and how they collectively influence demand for public transport. Accordingly, the literature review begins with the effects of land use characteristics, socio-economic factors, urban form factors, and public transport service provision on travel behaviour. Next, exiting models of public transportation, both general and specifically for Perth (Western Australia) are discussed. Finally, there follows a detailed explanation on how these research questions and objectives address a knowledge gap in the existing literature.

2.1 Land Use Characteristics

Cervero (1989) highlights the importance of balancing land use characteristics (jobs and housing) to improve regional mobility. Much research has explored the relationship between land use characteristics and travel behaviour. Handy (1996) states that these studies can be classified into three categories:

1. **Traditional Transportation Models** used to predict differences in total travel between different neighbourhoods,
2. **Aggregate level data analysis** used to compare average travel characteristics in neighbourhoods of different design and/or cities of different densities, and
3. **Disaggregated data** used to examine differences in individuals' travel choices between neighbourhoods, and the relative influence of various land use characteristics on those choices.

Kenworthy (1989) conducted the first international-scale study of the relationship between population density and travel behaviour, comparing 32 metropolitan cities across the world. They conclude that density is most strongly correlated with automobile use, as measured in terms of petrol consumption. This is supported by the findings from Frank (1994a). Focusing on the impact of land use characteristics, such as resident population density, employment density, and land use mix, on the frequency with which modes of public transportation are chosen, they find that these relationships are nonlinear. They also point out that at the census tract scale, the relationship between land use characteristics and public transport usage is relatively weak. They suggest analysing these relationships at a smaller geographic unit of analysis.

Cervero (1997) examine how density, diversity and design of built environment influence travel demand. They use the San Francisco Bay Area Travel Survey (BATS), a detailed single- and multiple-weekday travel diary for 1990-1991, population densities data from the US census, employment data from the Census Transportation Planning Package (CTPP), and land use data from Association of Bay Area Governments (ABAG). Applying factor analysis, they find that there are three latent built environment factors: density, diversity, and design. The variables included in the **density** factor are population density, employment density, and accessibility of jobs, which measures the relative closeness and compactness of land use practices. The **diversity** factor is comprised of the proportion of different land uses within a zone of hectare grid cells; decomposed land use categories; the proportion of commercial/retail areas within a zone; the intensity of land use in different categories, including residential, commercial, office, industrial, institutional, and recreational, activity centre mixture; and proximity to commercial/retail uses. Finally, the **design** factor includes the proportion of commercial/retail areas, along with parking availabilities, pedestrian and cycling provisions, and street layouts. The authors used multiple regressions to predict miles of travel in personal vehicles, and a binomial logit model to forecast the probability of a person traveling by non-personal-vehicle (or non-single-occupant vehicle (SOV) mode) for non-work trips. These control variables are taken into account for developing the base model. Then, statistically significant built environment variables were included in built environment models. Cervero (1997) found that population density, land-use diversity, and pedestrian-oriented design generally contribute to decreases in trip rates and increases in non-private car travel. Additionally, they claimed that there were strong and positive associations between compact development and non-personal vehicle mode choices for personal business trips and non-work trips. They also find that having accessible retail activities within neighbourhood is closely associated with transport mode choice for work trips. Moreover, the elasticities between each built environment dimension and travel demand were statistically moderate, and the intensity factor (a combination of retail store density, activity centre density, retail intensity, walking accessibility, park intensity, and population density) had a fairly marginal impact on travel demand in the San Francisco Bay Area. This intensity factor has positive association with all non-work trips, personal business trips and work trips by non-personal vehicle travel but stronger magnitudes for personal business trips and work trip at 1.34 and 1.87 respectively Cervero (1997).

Using the 1991 Household Travel Survey in Boston and 1992 Travel Characteristics Survey in Hong Kong, Zhang (2004) analyses and compares how land use attributes determine travel mode choices for work and non-work trips in Boston and Hong Kong. Travel costs for bus, train, walking, biking, and car driving, along with population composition by gender, age, household composition, and annual household income are included as socio-economic variables, while population density, employment densities, public parking space, connectivity of road networks, and entropy of land use are included as land use variables. His findings suggest that including land use attributes substantially improves the explanatory power of models estimating travel mode choices. Zhang (2004) concludes that the effects of land use attributes vary depending on the trip purpose (i.e. work related or non-work related) and whether public transport use is measured at travel origins or destinations. For example, there is a statistically significant relationship between train usage and population densities at both trip origins and destinations, whereas demand for public transportation for non-work related trips in Boston is not significantly correlated with population density at the trip origin, or employment density at destinations. Nevertheless, he points out that higher employment densities at trip destination areas can encourage choosing non-driving modes of transportation. More importantly, he emphasises that it is critical to consider the composite effect of changes in land use characteristics on mode of transportation choices, and argues that, "*land use is necessary but not sufficient to influence travel*" Zhang (2004: p.357).

Furthermore, Perkins (2006) review the literature on factors influencing individual travel behaviour published between 2001 and 2006. They conclude that residents living in the areas with low density and single-land use tended to have longer travel journeys and relied more on private cars as their principal means of transport. On the other hand, public transport usage was high for the areas with higher density, mixed land use, and greater access to public transportation. In contrast, Bhat (2007) report that when number of households per acre (household density) is used as a density indicator, the relationship of density and car ownership is insignificant.

White (2004) analyse how land use policies, which include industrial and commercial zoning locations, the location of new residential developments, mixed land use developments, and transit oriented development influence transportation use. Their findings are supported by Kenworthy's (2008) comparative analysis of sustainable transport cities. He compares public transport usage, and the factors contributing to it, in different cities using data that includes twenty-six sustainable transport variables for 84 cities around the world. The cities were classified into five clusters, from least to most sustainable transportation, by applying a ranking and clustering technique. Kenworthy (2008a) reports that the most significant factors affecting transport use in these cities are urban density (measured as persons per hectare), proportion of jobs in Central Business District (CBD), less parking space per 100 CBD jobs and total public transport seat kilometres per capita. This suggests that land use characteristics, such as residential and employment population densities, together with good quality public transport services, encourage public transport use.

Additionally, Greenwald (2006) studied the relationship between land use, destination selection, and travel mode choice. Modelling intra-zone trips with multi-nominal logit and binary logistic methods, he finds that due to intra-zone trip characteristics, the mode and destination choices for these trips were most likely to be influenced by elements of land use characteristics, such as street design and housing concentration, but that the magnitude of this influence is relatively small compared to significant influence of economic diversity/mixed use to alter travel behaviour. Moreover, the study emphasizes the importance of proximity of origin and/or destination in discrete choice and/or activity based trip distribution models.

Susilo (2007) use data from the Dutch National Travel Survey to conduct an exploratory analysis of the effects of built environment on commuting behaviour, departure time, commuting time and distance, and the transport modal split. Using regression analysis and choice models, this study shows that land use characteristics variables consistently influenced the parameters of commuting journeys. Nevertheless, individual socio-economic factors and others (such as job market distribution) were more critical determinants of commuters' travel behaviour. Relatedly, by way of an individual-activity based analysis, Shiftan (2008b) developed an activity and trip generation model to generate complete and optimised daily activity schedules, which include activity types and sequence; locations and means of transportation; activity start times and durations; and household structure and its members' activities calendars.

More recently researchers have used advanced statistical techniques to model the relationship between land use characteristics and transport use. For example, Souche (2010) applies robust econometric methods such as double least squares (2SLS), seemingly unrelated regression (SUR) and the three least square method (3SLS) to measure structural determinants of urban transport demand and private travel demand. Here, she uses data on demography, urban structure, the economy, and transportation data from the IUTP (**U**nion **I**nternationale de **T**ramways/Internationaler **P**ermanenter Strassenbahn-Verein, International Association of Public Transport) collected in 1995. The findings from an initial ordinary least squares regression show that increases in the average cost of private car use, urban density, and public transport supply encourage increases in public transport use. Then, the double least squares method (2SLS) is used to isolate the specific effects of the observed variables with heteroskedasticity, thereby avoiding the problem of auto-correction. The results show a reduction in the influence of urban density on demand for public transport, as compared with the results from the OLS model, nonetheless, urban density remains one of the two most salient variables among the structural determinants in her study.

Another interesting approach is taken by Pitombo (2011) to examine the relationships between the travel patterns of the residents in São Paulo Metropolitan Area and their socio-economic characteristics, activity participation, and land use. They apply Cluster Analysis to group and characterise the Traffic Zones, and Decision Tree method to identify unknown relations between the socio-economic characteristics, land use attributes and destination choices. By way of Exploratory Multivariate Data Analysis, they investigate which pervasive travel patterns are associated with a given group of individuals, and how the patrons' trip sequences relate to the land use patterns of their residence, their activities, and their socio-economic characteristics. Industry activity level (employees in industry per resident in traffic zone), services activity level (employees in the service sector per resident in traffic zone), commerce activity level (employees in the commerce sector per resident in traffic zone), the number of total enrolment in elementary schools, the total number of enrolment in high schools and universities, and population density were used in a two-step cluster analysis to calculate general activity means, which in turn were used to categorise 389 traffic zones into 7 clusters. Next, the decision tree method was applied to analyse the travel choice patterns. They found that car ownership was the most important variable for explaining the travel patterns in their study. Additionally, they concluded that residents living in traffic zones with high levels of activity were making short trips (distances shorter than 5 km) while residents living in traffic zones with low activity levels were travelling more than 15 km for work, leisure, shopping, etc.

Another recent study applying factor and cluster analysis to questions of supply and usage of urban transportation systems is Klinger et al.'s (2013) research on German cities. Both objective (urban form and socio-economic) and subjective (individual preferences and attitudes toward mobility) criteria are incorporated into their models, along with transport demand indicators, transport infrastructure, and supply. They show that the cities cluster into six groups across 23 variables. Among these, two groups of public transport oriented cities are "density and public transport orientation" and "metropolitan character". Settlement density and population size are among the highest and positively contributing factors to modal share of public transport.

In addition to aggregate analyses, researchers have used disaggregation to study how land use characteristics influence transport use. Thus, Aditjandra (2012) examine how changes in neighbourhood characteristics affect transport choice—especially public transportation—by applying longitudinal structural equation modelling. They use the quasi- longitudinal survey sample of 219 respondents who reported residential relocations in the metropolitan areas of Tyne and Wear, UK. Taking into account self-selection (individual preference in choosing a sustainable mode of public transportation), their findings show that changes in shopping accessibility accompany changes in public transport accessibility. This can discourage driving and it follows that mixed land use development, combined with integrated public transport accessibility, can encourage the use of more sustainable modes of transport Aditjandra (2012).

Urban development patterns also have an impact on choice of travel mode, and by channelling growth into centres designed for public transport, short drives, or walking, automobile travel can be reduced by 30% more than it would be with more conventional development strategies Douglas (1991). Furthermore, land use attributes, such as population density, the availability of a railway station and travel time, affect mode choice for medium and long distance travel, though the size of these effects varied depending on trip purposes Limtanakool (2006).

Pitombo (2011) argue that travel demand is derived from the individual's need to participate in temporally and spatially diverse activities, such as employment, study, shopping, leisure, etc. Using data collected from an origin-destination home interview survey conducted by METRO-SP São Paulo Metropolitan Area (SPMA) in 1997 which contains 98,780 respondents from 389 traffic zones, they applied two multivariate data analysis techniques to examine the relations between land use patterns, activity participation, socio-economic characteristics, and travel demand. They identify the predominant travel pattern associated with different groups of individuals (classified traffic zones) and examine how trip sequences relate to land use patterns and socio-economic characteristics of the various zones. Cluster analysis is applied to group the traffic zones based on the industry activity level, services activity level, commerce activity level, the total number of enrolments in elementary schools, the total number of enrolments in high schools and universities, and population density. Then, the decision tree technique is used to determine the relationships among land use attributes (clustered traffic zones) and socio-economic characteristics of origin traffic zones (family composition, level of education, gender, age, family income, car ownership, family size and use of transportation credit), and travel patterns. Their findings confirm that land use variables, such as resident population, population employed in industry, the service sector, commerce sector, and student enrolments in elementary, high school, and universities are the most prominent factors determining travel demand, patterns and sequences.

Relatedly, Stead (2001b) provide a review and evaluation of the international literature on the relationships between land use characteristics and travel patterns over 20 years. Their study focused on identifying the more concentrated research area, pointing out gaps in previous research, particularly with regard to data accuracy, reliability, and quality, as well as the applicability of research methods and the interpretation of findings. Their review includes previous research that uses the following nine aspects of land-use characteristics, from the regional strategic planning scale to neighbourhood scale:

a) Distance of residence from the urban centre
b) Settlement size
c) Mixing of land uses
d) Provision of local facilities
e) Density of development
f) Proximity of transport networks
g) Availability of residential parking
h) Road network type
i) Neighbourhood type.

The authors argue that two shortcomings of previous studies are 1) difficulty determining causal relationships from cross-sectional data, and 2) problems replicating the outcomes for comparison in different areas with different socio-economic factors. Moreover, they reported that socio-economic factors such as income, car ownership, household size and type, attitudes, personality type, driver’s license, gender, age, education, employment type, and work status also interact with each other when influencing travel patterns. Additionally, they conclude that not only the socio-economic factors, but also land use characteristics such as population size, local facilities, population density, employment density, job ratio, distance to urban centre, availability of parking, public transport accessibility, pedestrian network, road network, and neighbourhood type were interconnected, making analysis of how each one affects travel patterns difficult. Furthermore, they suggest that the strength of evidence and geographical scale of analysis (i.e. regional level, city level or neighbourhood level) should be taken into account before drawing any policy conclusions.

Following the aforementioned recommendations, the present research includes resident population (by age and gender) density, employment population densities in various industries and student population density as land use characteristics that potentially influence public transport use. Research regarding the effects of these observed land use characteristics on public transport use is discussed in the following section.

2.1.1 Resident Population Density

Hanson (1981) examine individuals' daily travel patterns, and their relationship to their characteristics by using the disaggregated Household Travel Survey data collected in Uppsala, Sweden. Applying principal components factor analysis, they examine the interrelationships among fifty one measures of travel activity patterns, including number of trips by different travel modes, number of stops, time spent in public transport for different activities, distances travelled, distance from home to city centre, among others. They identify five travel pattern factors, including frequency of travel; dispersion of destinations visited; shopping, variety and multi-stop travel; travel to work; and social travel and travel to recreation, and correlate them with individual and household characteristics (i.e. occupation, education and employment status of household head, household income and composition by gender and age, and car ownership). Using five-week travel/activity diaries, their study stresses the significance of gender differences in the travel activities of urban residents. They find that men in urban areas have significantly higher trip frequencies than females for all travel purposes except shopping. The findings also confirm that women with higher social status, education, and access to cars make more trips than their counterparts.

Building on these findings, Cervero (1996a) uses the 1985 American Housing Survey to examine how land use characteristics influence public transport use with respect to choice of travel mode. In this study, single family detached, single family detached/low-rise multi-family buildings, mid-rise multi-family buildings, and high-rise multi-family buildings within 300 feet of unit, and the presence of retail services in the neighbourhood are included as variables, along with number of cars owned per household, distance from home to work, residence within central city, and public transport service adequacy. They find that highly dense areas with mid-rise (3 to 6 storey) buildings are more favourable to public transport use than are other land use variables. Notably, public transport service adequacy and residence in the city centre have stronger positive influence on public transport use in high-density areas, while car ownership has stronger negative one. Moreover, in terms of the probability of using public transport, density has stronger impact than the presence of mixed land use

(Cervero, 1996). Similarly, Frank (1994a) also find that the relationship between resident population density and public transport use is nonlinear, with the population densities at trip origins and destinations only strongly affecting use of public transport for shopping trips.

Dunphy and Fisher (1996), quoted in Badoe (2000, , p.249), use the 1991 Federal Highway Administration (FHWA) highway statistics to investigate the relationship between residential density and public transport usage. They argue that higher densities substantially increase use per capita at the zone-level, but not at the level of total use. The results from a comparative study by Schimek (1996), also quoted in Badoe (2000, , p.238), show that not only the density, but also the interactions among other variables, need to be considered as determinants of public transport use. He develops time-series models of public transport use in Toronto and Boston to compare the factors by which it is influenced in each city. The study highlights how greater residential population density, higher employment density in CBD or inner suburbs, better public transportation services, and lower income levels collectively have a greater effect on public transport use in Toronto than in Boston. He also stresses that when income is considered, households in areas with higher population densities tend to own fewer cars, leading to more public transport usage. This is also supported by the findings from Messenger (1996).

Similarly, Balcombe et al. (2004) find that increases in population density in central cities reduce the negative impacts of urban sprawl on the public transport demand based on data from a 1999 census conducted in France. To compare the model developed from this 1999 census in France, they use operator data derived from ticketing systems and travel activity diary survey collected from the British National Travel Survey to model the elasticities of public transport demand in UK. They state that public transport demand tends to be greater in areas with high population densities and settlement sizes where there is more access to required services and facilities. Mixed land use, combined with good access to everyday facilities for shopping, school, recreation, etc., tends to lessen the journey lengths and car dependence; it does not, however, always leads to increases in public transport demand. They also mention that local population density can be used to formulate the function for forecasting the new public transport services such as new train station or new service line.

Using more detailed data on population composition, White (2009) finds variations in public transport use by age and gender in his analysis of data from Great Britain's National Transport Survey. Accordingly, the present study includes resident population by age and gender among its explanatory variables. The following section reviews the effects of these variables on public transport use.

2.1.1.1 Gender

Based on the National Transportation Survey in Great Britain, White (2009) states that females tend to use public transport more than males, with a similar distribution by age category. In their *Demand for Public Transport: A Practical Guide*, Balcombe et al. (2004) mention that there are two previous empirical studies showing that males have more access to private cars, and are therefore more sensitive to fare changes than females. Meanwhile, Mackett (1990) develops a micro-simulation model by considering the progression of socio-economic and demographic attributes. He reports that the public transport elasticity for males is, in the long run, fairly higher (-0.59) than the one for females (-0.39). This finding also suggests that males tend to respond more to fare changes than female patrons. This is supported by findings from research conducted by Wardman (2000) which examines how public transport elasticises vary by income, gender, age and journey purpose. His findings indicate that in a patron group whose income level is £5000 or less per annum, males are more sensitive to fare changes than females, and that females with higher incomes are relatively less responsive than those with low income-levels.

2.1.1.2 Age

Goodwin (1987), quoted in Balcombe et al. (2004, p.60), point out that bus fare elasticity declines with age, with figures of -0.87 for the youngest versus -0.25 for the oldest age group. Findings from a study of fare elasticities conducted by Preston (1998), also quoted in Balcombe et al. (2004, p. 60), indicates that the elasticity for children is higher than adults, the elderly, and the disabled, but not when used for schooling. Therefore, Balcombe et al. (2004) explain that age elasticities are complex because they vary depending on the type of the trip and/or patrons' socio-economic status. As most of trips generated by the elderly are non-compulsory, age elasticity could be high. Nevertheless, many elderly people have low incomes, low rates of car ownership, and difficulty with mobility, all of which lower their elasticity, as public transport may be the only option they have (Balcombe et al. , 2004).

White (2009) also finds that there are variations among public transport use by age and travel mode. Working age groups from 21 to 59 had the highest trip rates and highest share in train usage in 2006, though this group contributed only 4 percent to bus and coach usage. Age groups 17-20 and 70 and over contributed the most to overall bus and coach use, at 15 percent and 12 percent, respectively. By contrast, there was not much variation in train usage across age groups.

According to the findings from a modal choice model developed by Soltani (2006), females in four suburbs of metropolitan Adelaide are less likely to use public transportation for their journeys because they are, presumably, more likely to be car passengers. These models also indicate that school-aged residents have higher tendency to be car passengers than to use public transport.

2.1.2 Employment Population Density

Hendrickson (1986) examines changes in public transport usage in 25 large metropolitan areas of the United States between 1960 and 1980. Significant relationships between public transport usage and CBD employment are reported. The author also concludes that public transport usage was strongly related to CBD employment, but not to overall metropolitan area size.

Cervero (1988) states that mixed-use development can improve suburban mobility. His study was conducted in 57 large suburban employment centres in the United States, including Bishop Ranch and Hacienda Business Park east of San Francisco, Warner Centre and South Coast Metro near Los Angeles, Tyson's Corner outside of Washington, and downtown Stamford near New York City. To determine the relationship between mixed-use development and suburban mobility, the percentage of floor space in office use and retail use were used as measures of land-use composition, while number of restaurants, banks, shopping and retail centres, their square footage, and employment totals were included as indicators of consumer services. Stepwise regression was applied to examine which factors influenced the percentage of work trips in the drive-alone mode. The author found that the percentage of solo-commute trips increases as suburban workplaces become more office oriented. Additionally, mixed-use work environments encouraged reductions in auto dependency.

In a subsequent study, Cervero (1991) examines how land use characteristics influence the percentage of work trips made with public transportation. In his study, the number of available parking spaces, degree of mixed uses (i.e. mixed office and retail uses or office related uses only), the number of stories in office buildings, and car occupancy levels are considered as explanatory variables. His stepwise regression results show that an area with 10 story office buildings generates approximately four percent more public transport use than a comparable area with 1 story office buildings, and it can be inferred that density also influences public transport use. In addition, his findings indicate that the degree of mixed land use has a stronger relationship with public transport use than private car use. His stepwise regression models suggest that the degree of mixed land use (office and retail uses) has positive relationship with public transport use at 3.236 and negative relationship with private car use at -3.207. More importantly, his computations of elasticities between land use characteristics and public transport use suggest that density has a stronger association with public transport use than degree of land use and/or parking supply. These findings, in turn, indicate that mixed land use encourages non-auto travel choices.

Numerous studies have been conducted to examine how mixed land-use environments impact transport demand and mode choices. Handy (1992) compares the shopping trips generated by residents from four neighbourhoods in the San Francisco Bay Area, finding that mixed-use neighbourhoods encourage reductions in private car use for shopping trips. Meanwhile, from the comparison between work and non-work travel patterns by residents from six communities in Palm Beach County, Florida, Ewing (1994) find that mixed land-use environments with recreation and shopping facilities significantly lower the number of vehicle hours travelled per capita.

Schimek (1996), quoted in Badoe (2000, : p. 238), asserts that high employment densities in CBD and inner suburbs are conductive to higher public transport usage. This is supported by some findings from Frank (1994a). To evaluate the impact of mixed land use and density on transport mode choices, the latter authors develop a regression model from data on travel behaviour, demographic variables, public transport service levels (from Puget Sound Transportation Panel), population density (from the US Census Bureau), employment density (from Washington State Department of Economic Security), urban form variables (from Puget Sound Regional Council), and mixed land use (from King County). They also consider family size and mixed-use developments (for retail, services, office, entertainment, institutional, industrial and manufacturing) to calculate the level of land use mix. Their findings indicate that the relationship of public transport usage for work trips to employment density is stronger (at 0.59) than to either population density (at 0.19) or land use mix at (0.15). Similarly, employment density has the strongest relationship with public transport usage for shopping trips (at 0.44), followed by population density (at 0.16). Additionally, their regression results indicate that the percentage of public transport usage is highly influenced by employment density at trip origins and trip destinations, with a beta value of 0.65 for work trips and 0.32 for shopping trips, confirming that the relationship between employment density and public transport use is nonlinear. Notably, they also identify the employment density thresholds that encourage people to shift from using cars to public transport/walking. Specifically, 20-50 employees per acre produces a moderate shift, while more than 75 employees per acre produces a significant one.

In another study, Cervero (1996a) uses data from the American Housing Survey (conducted in 1985) to examine how the presence of retail activities influences residents' choice of transportation mode. Based on a discrete binomial choice model, his findings show that mixed land-use environments and the presence of retail stores—and other non-residential activities—increased demand for public transportation. The results from this study reveal that the adequacy of public transportation services, car ownership per household and residence within city centre areas have a stronger influence on public transport demand than do land-use variables. Moreover, it also discloses that density is more influential than land-use variables in determining public transport demand. The author concludes that while mixed land-use environments can influence public transport demand, density has stronger influence. He also highlights how the presence of retail services in a neighbourhood (between 300 feet and 1 mile) can reduce public transport use because people prefer to link work trips with grocery or consumer shopping by a car. In a later study, Cervero (2002b), again using a binomial choice model, reports that diversity in land-use (employment density and population density relative to county ratio) at both origins and destinations can significantly increase public transport usage. His analysis suggests that density and diversity factors encourage public transport usage, with an increases in gross densities at origins and destinations having a greater effect than car driving or ride-sharing; further, a higher degree of mixing of residential and employment populations also favours public transportation use.

Sohn (2005) examines whether commuting patterns are a good indicator of urban spatial structure. His comparison between the spatial structures evinced by commuting patterns and the actual urban density distribution shows that the former reflect the employment distribution, but not the distribution of employed residents. Relatedly, Balcombe et al. (2004) show that a greater degree of centralisation for employment and required facilities fosters demand for public transportation.

2.1.3 Student Population Density

Tolley (1996) brings attention to the role of universities in public transport systems as large trip generators. His study shows that commuting trips were responsible for 97.8 percent of CO_2 emissions, and commuting trips by students accounted for three quarters of total commuting at the University of Northumbria in 1994. Similarly, the total commuting trips by staff and students (33 m km per annum) contributed to more than 99 percent of the total CO_2 emissions in the University of Central Lancashire. This pioneering study indicates that student population density is an important variable in accounting for the generation of trips within the public transportation system. Similarly, in their study of active commuting behaviour among university staff and students at the University of Western Australia (UWA), Shannon (2006) confirmed that the universities are major trip generators. Their results show that both staff and students generate a significant number of trips, accounting for 1.61 million staff trips and 2.91 million student trips per academic year. The findings from Pitombo (2011) also confirm that individuals' participation in study activities influences their travel demand and trip sequences.

Similarly, Curtis (2004a) observe that universities are significant travel generators in Australia. With student and staff populations in the thousands, Australia's university campuses generate a significant number of trips. They note that parking demands are growing considerably on the universities campuses located in suburban areas, and highlight the complexity of travel patterns generated by university campuses as a consequence of diversity in ages, life styles, and occupations. Further, they emphasise the importance of public transport service provision in these areas, along with other travel demand management programs Curtis (2004b).

2.1.4 Urban Form Factors

Two urban form factors taken into account in the present research are road network density and distance from city centre. Findings from previous studies relating to these factors are discussed below.

2.1.4.1 Road Network Density

Mogridge (1990) states that increased road capacity can attract more car use and, as a result, intensify shifts from public transport use to car use. This can then lead to loss of revenue for public transport services and pressure to reduce service provision (Mogridge, 1990). Hansen (1993) produces a report on the impact of highway capacity expansion on trip generation and land use change in urban counties of San Diego, San Francisco, and Los Angeles for the period of 1973-1990. They apply econometric analysis to examine the relationship between road capacity extension and land use development on a panel consisting of eight corridors where freeway capacity had been expanded. They analyse time series data on travel mode choice and land use variables such as single-family housing, multifamily housing, office development and industrial development. Their research shows that there was a remarkable increase in single housing development after the completion of the freeway capacity extension. There was also an increase in multi-family housing development, though not of the same magnitude as single housing development. Their findings also show that commercial development starts increasing after the freeway capacity extension was completed, and continued to accelerate over several years. However, the immediate effect on industrial development from this freeway capacity extension was negligible (Hansen et al., 1993).

In an appraisal of induced demand and road investment in Australia, Luk (1997) review empirical results from the US and UK, followed by an analysis of induced demand in an Australian case study. They use traffic data from the South Eastern Arterial and its adjacent arterial roads in Melbourne, along with train patronage data from the Dandenong and Glen Waverley train lines in the south-eastern corridor. Their review of previous empirical studies finds that the inducement level depends on various other factors, such as population density and existing congestion level. For example, the addition of road capacity in London can be detrimental to public transport use, while travel times to city centre by car or public transportation remain the same. Results from the Melbourne case study show that there was a shift in modes, from train to car, when the new South Eastern Arterial was introduced in 1988 (Luk and Chung, 1997).

In a study of the effects of land use characteristics on public transportation use, Kenworthy (1999a) investigated patterns of car dependence and public transport use in forty-six cities and their relationships to urban wealth and land use practices. The findings demonstrate that: 1) *the wealth of cities alone does not provide reliable or consistent evidence to explain the degree of automobile dependence in different cities (pg. 718);* 2) the development which encouraged car dependence in cities would not bring benefits for economic performance due to its creation of higher road construction and maintenance costs; 3) higher levels of rail service encourage more public transportation use and reduce car dependence; and 4) higher urban density has a strong correlation with lower automobile dependence, as well as higher public transport use, which leads to lower total costs of operating urban transportation systems (including both private and public transport costs for infrastructure and investments). These findings highlight the importance of policies to strategically reshape urban land use, emphasise non-auto infrastructure investment, and develop physical planning strategies aimed at reducing car dependence.

Cervero (2001a) reviews research employing model forecasts, as well as area studies using proxy elasticities and partial elasticities that disaggregate the effects of mode choice on transportation demand. Based on his meta-analysis, he provides a very interesting discussion of induced travel demand. He explains that expansions of road network supply that are intended to ease traffic congestion actually account for increases in miles travelled by private cars, which in turn decreases public transport use, Cervero (2002c). He also recommends financial resource allocation be reconsidered, specifically that resources be transferred from road network expansion to the provision of public transport.

In a later study, Cervero (2002a) use simultaneous equation analysis to clarify how the expansion of road network capacity impacts travel demand. They apply this analysis to both pooled time-series and cross sectional data on road supplies and travel demand, along with various control variables, for the period of 1976 to 1997 in the state of California. The control variables considered in their research are vehicle meters travelled (VMT) as a measure of travel demand, lane-miles as supply, fuel price, residential and employment population densities, income, weather conditions, air quality, and politics. The findings indicate a two-way relationship between road supply and demand, and show the significance of induced demand and induced investment demand. They report that the impact of lane-miles (supply) on vehicle miles travelled (demand) seems to be stronger than it is the other way around (Cervero and Hansen, 2002). Subsequently, Cervero (2003) carried out a path-analysis of the relationships among road expansion, urban growth and induced demand, which he tested with data from a panel of California freeways. He explains that short-term models can indicate the instant effect on travel levels through changes in road supply over a year or so; longer term models on the other hand, can show how road expansion induces urban activities, which can then spur travel demand. The findings from his near-term path model show that induced travel effects are positive and significant, with an elasticity of 0.24. His longer-term path model considers not only the operating speed of cars on a roadway but also building activities, employment density and population density, black proportion and Hispanic proportion. Empirical findings from the longer-term path model indicate that "*an estimated 80% of California's freeway capacity additions were absorbed by traffic spurred by faster speeds and land use shift*" Cervero (2002a: p.158). A further suggestion is to consider how the impact of road network capacity extension on land use development activity in surrounding areas affects travel behaviour in the longer term (5-6 years).

Another interesting analysis of the relationship between road network capacity and public transportation use, especially train use, is provided by Zeibots (2005). Their case study examines how travel modes shifted from Western Sydney Rail Lines to private car use in May 1992 after Sydney introduced the M4 Motorway sections from

Mays Hill to Prospect. Using time series regression analysis, they quantify the extent to which travel modes shifted during 117 four-week accounting periods from July 1988 to May 1997. Their negative and significant regression coefficients indicate that the road network capacity extension was associated with a notable decline in train use. Further, they show that the loss of train patronage for the Western Sydney Rail Line during the accounting period instantly following the opening of the M4 Motor section amounted to approximately 200,000 journeys; at the same time, there were more than 6,000 additional car trips per day, with an average vehicle occupancy rate of 1.2.

2.1.4.2 Distance from City Centre

According to an international literature review conducted by Stead (2001b), the distance from city centre (i.e. distance between each home and an urban centre) has a stronger association with travel distance and transport energy consumption than with travel frequency. The study by Sohn (2005) uses data from surveys of commuting collected in 1987, 1990 and 1995 from the Seoul Metropolitan Region—which is made up of Metropolitan Seoul, Metropolitan Inchon, and Kyungki Province—to investigate the association of commuting patterns and urban spatial structures. To standardise the comparison, all administrative zones were aggregated into 57 zones, with locational variables used to measure the distance between the centroid of these zones and the city centre, as well as the distance between the centroid of the origin and destination zones. The findings show that locational variables have a significant influence on commuting patterns, and that the dispersion of employment was less than that of employed residents over the studied year. This confirms Weber (2003) finding that distance has a mixed impact on urban spatial structure, as trip distance minimisation is not the only factor determining workplaces, residential locations, and commuting routes.

Some researchers include not only the land use characteristics, but also socio-economic factors, in their studies, with the aim of examining their combined effects on transport use. Cervero (2002b) uses a logit model to examine the influences of three core dimensions of built environments—density, diversity and design, and socio-economic attributes of travellers—on travel mode choices (drive alone, group ride automobile, public transport) in Montgomery County, Maryland. Total travel time differentials (average time travelled via three travel mode options) are considered in comparing modal attributes, and socio-economic attributes include car ownership per household and driver's license ownership. Additionally, population density, employment density, job accessibility, labour-force accessibility, land use diversity, and ratio of sidewalk miles-to-road miles are included as land-use attributes. Based on logit modelling and elasticity estimates, the author argues that land-use variables and socio-economic factors have significant effects on travel modes by Montgomery County residents.

In a study relating land-use patterns at the neighbourhood level and non-work trip generation, Boarnet (1998) develop an ordered probit model of a travel demand function. They include socio-demographic variables (gender, race, education, household income, car ownership) and four land-use variables such as % GRID (the percentage of the street grid characterised by four way intersections within a quarter mile radius of residential areas), population density, retail jobs density, and service jobs density. And they compare results at the neighbourhood and post code level. Their findings suggest that residence location choice and different levels of geographical details are important in studying the relationship between land-use and travel behaviour. The results also show that the land-use characteristics, as measured in their model, are statistically insignificant at neighbourhood level. In a later Boarnet and Crane's (2001) study, evidence from Orange County in Los Angeles revealed the important role of measurements in different geographical scales because its findings indicate a relationship between land use and private car trip generation only when the land use characteristics are measured in postal code areas. The authors draw the attention on the importance of the complexity of travel behaviour when examining correlations between urban design and travel behaviour, particularly when concluding that urban design changes can *cause* changes in individual travel.

Boarnet (2001) argue that several specification and estimation issues involved in estimating the relationship between land use characteristics and travel behaviour could be overcome by systematically isolating the respective effects of urban design characteristics on travel and then analysing individual level data. Their analysis, informed by microeconomic theory, included the following information in the travel behaviour model:

- Price of travel and income level of the individual or household;
- Several socio-demographic "taste" variables such as gender, education levels, age, and number of persons in the household; and
- Measures of land use and urban design characteristics near the residences of individual travel diary respondents.

Relatedly, Riekko (2005) examines the relationship between land use characteristics and public transport, using data on individual trip attributes for 136,000 households representing 374,000 individuals from The Transportation Tomorrow Survey, along with socio-demographic variables data from the Census of Canada and land use characteristics and land use variables from DMTI Spatial which is a Canadian geographical information systems company providing location based business intelligence software and Canada's most accurately and comprehensively address database integrated from 7,300 data sources. Combining descriptive and inferential analysis, he suggests that grid or mixed street patterns should be emphasised in transit oriented development to encourage the public transport use because grids street patterns can create more direct routes to destinations and transit stops, also allowing public transport operation to increase its efficiency. Riekko (2005) also recommends that transit oriented development should incorporate mixed land uses such as recreation and commercial and employment intensive light industry use. This study reveals that residential density and population density have stronger association with travel mode choices than urban form variables and land use diversity and they were positively associated with public transport usage. In addition, the author finds that socio-demographic variables explain a large proportion of variance in public transport mode choice more than the land use characteristics do.

In light of the preceding review, socio-economic factors are included among the explanatory variables in this research. The effects of socio-economic factors on transport use are discussed in the next section.

2.2 Socioeconomic Factors

Many previous studies investigate the effects of social and economic factors on public transport use. For example, Hanson (1981) highlight the extreme complexity of travel behaviour patterns and the influence of individual or household socio-economic characteristics on them. This is strongly supported by the findings of Ewing (1996) that residential density, job accessibility and mixed land use do not significantly influence household trip rates after socio-demographic variables are considered.

In a later study, this was also confirmed by Stead (2001a), who examined the relationship between socio-economic factors, land use factors, and travel patterns by using two types of datasets: National Travel Surveys collected in 1978/79, 1985/86, 1989/91 and 1991/93; and local travel surveys conducted in Kent and Leicestershire. Based on his findings that socio-economic characteristics such as family composition by age and gender, income level and car ownership have significant relationships with travel patterns especially travel distance, he urges researchers to consider how socio-economic and land use factors influence travel patterns.

Subsequently, Cervero (2002b) developed utility-based models of mode choice based on density, diversity and design. Using the Trip records of Montgomery County residents extracted from the 1994 Household Travel Survey (yielding 5167 observations), he employs maximum likelihood estimation to the problem of choice. He finds that the model's predictive power is enhanced when land use variables are considered along with the socio-economic factors. The number of households with no car ownership strongly influences the choice of travel mode, while female population with driver's licenses are less likely to choose public transportation.

More recently, Pope (2005) shows that the relationship between one explanatory variable and transport demand can be measured by elasticity. Income and price elasticities are the most commonly used. He identifies the economic determinants of passenger transport demand as individuals' income and wealth, the price and availability of competing modes, and individuals' preferences and family structure. He also explains that different types of travel, such as business, meetings, luxury-leisure and non-luxury-leisure, have different income elasticities. The income elasticity of transport demand can be defined as follows:

$$\frac{\%\ change\ in\ the\ quantity\ demanded\ of\ transport\ mode}{\%\ change\ in\ net\ income}$$

According to Balcombe et al. (2004), increasing car and license ownership, along with income growth and the declining cost of car ownership, may influence personal travel patterns in four key ways:

1. Increases in income could lead to either increase in car ownership or increase in public transport use,
2. Increases in car ownership will lead to decreases in public transport use,
3. Income level is the main factor for determining the sign and magnitude of usage elasticity for public transport, as against car availability, and
4. Increases in average trip length can be expected when income growth becomes higher.

These relationships are supported by findings from a study conducted by Paulley (2006), which concluded that income and the negative effects of car ownership growth are fundamental to determine public transport use, while fare elasticity tends to change over time (these results are based on several data sources, including operator ticket sales and information from the British National Travel Survey).

Bresson (2004) state that substitution between public transport use and car use is represented by the cross-elasticity with respect to the price of fuel (a determinant of car use). They conclude that increasing car ownership has a negative impact on public transport usage and point out the influence of variables such as age, gender, zone of residence inside the public transportation area, and number of cars owned by the household differed across population categories. Further, they find that public transport usage elasticities changed over the period of their study, and that there was considerable variation in these elasticities among different geographical areas.

In addition, socio-economic characteristics such as the size of households, number of adults participating in the labour market, presence of young children, age, gender, education level, annual household income, and particularly car availability have a strong influence on travel mode choice for every trip purpose, Limtanakool (2006). The importance of socio-economic factors as a determinant of travel mode choices is also endorsed by Cervero (1996b, , p.134) who state that: "*the type of neighbourhood exerts a significant influence on mode choice for shopping and other non-work trips*"; this conclusion is based on a study of work trips and non-work trips by residents in two neighbourhoods in the San Francisco Bay area.

Further, Lin (2008) present an interesting approach to identifying the neighbourhoods and neighbourhood types that uses the log-likelihood clustering technique on socio-economic, demographic, and land use characteristics, as derived from the 2000 Census Transportation Planning Package data. Within this framework, they examined five household travel measures (i.e. number of trips per household, mode share, average travel distance and time per trip, and vehicle miles of travel (VMT)) and compared them across ten neighbourhood types. They find that the households' daily trip generation was greatly influenced by their size and life cycle status. Further, public transport availability at the residential location tends to increase the share of that public transport mode regardless of household car ownership and income level, and job-housing trade-offs were evident when mobility was not of concern. The study also highlighted the significant influence of racial/ethnic residential preference.

Other socio-economic factors such as fuel price, interest rate, inflation rate, and mortgage cost should be included in studying public transport usage, since these factors have a strong impact on household travel expenditures. According to the Reserve Bank of Australia (2009), the new policy relationship driven by rising fuel prices, interest rates, and mortgage costs has significant implications for the socio-economic and financial circumstances of households in Australian cities. The Reserve Bank of Australia (2009) also expected that growth in household spending will remain subdued over the Dec 2008 – Dec 2011 period as a result of a deteriorating labour market and weaker outlook for inflation, caused in turn by a higher exchange rate, higher oil prices, and the impact of a global economic recession. Financial constraints and modest-income households seeking home ownership in fringe areas can result in 'locational disadvantage', because many outer estates lack high quality access to employment and social and community services, which include public transportation Gleeson (2006).

Holmgren (2007) conducted a meta-analysis of demand elasticities from previous studies of demand for public transportation. Using a meta-regression method, he includes 81 estimated price-elasticities, 58 observations of elasticities regarding vehicle-kilometres supplied (i.e. public transport service provision), 22 observations of income elasticities, and 8 observations of car ownership elasticities. He concludes that increases in income encourage car ownership, resulting in declining patronage for public transport. Thus, public transport use is highly sensitive to rates of car ownership. Two main suggestions from this study are:

1. The effects of car ownership and income should be taken into account when analysing public transport demand; and,
2. Public transport demand per capita should be used if the population is excluded from the explanatory variables.

Nijkamp (1998) conducted also a meta-analysis using rough set analysis to evaluate the factors influencing public transport demand in Norway, Finland, the Netherlands, and the United Kingdom. They conclude that country-specific factors play a major role in producing variations in price sensitivity due to differences in culture and public transport service provision and quality.

Currie (2010) proposes a method for measuring the relative quality of public transport supply and its spatial distributions with regards to patron needs in Melbourne. He uses a combined measure of access distance to each stop/station and total number of service arrivals per week to quantify public transport supply. Meanwhile, a transportation-needs index was calculated based on the number of adults without cars, accessibility (straight line distance to Melbourne CBD), number of persons aged over 60 years, persons on a disability pension, low income households, adults not in the labour force, students, and persons 5-9 years of age. Then, the Australian Bureau of Statistics Index of Relative Socio-Economic Advantage/Disadvantage, developed by Adhikari (2006), was used to identify need-gaps. The authors conclude that only 10% of outer Melbourne has public transport walk catchment coverage, as compared to 90% of inner Melbourne; and that these outer areas have 'very high' need for public transport.

Accordingly, socio-economic factors are included in the present research on Perth's public transportation use patterns.

2.2.1 Average Car Ownership per Household

McFadden (1974) uses data collected before and after the introduction of the Bay Area Rapid Transit (BART) transport system in the San Francisco Bay Area to propose new approaches in the behaviour theory of travel demand. This survey samples 213 households residing and working in BART areas. Observed variables in the study include trip purpose, frequency, origins, destinations, and different travel modes (private car, bus and rail); as well as car ownership, resident locations, and end-of-trip activities. Using maximum likelihood methods, the author finds that the "*most effective way to increase bus patronage is to increase the auto costs"* McFadden (1974: p. 324).

Subsequently, Kenworthy (1989) examined land use, car dependence and transportation patterns in the world's major cities and the correlations between transportation and land use characteristics resulting from these data. They suggest that public transport use has strong negative correlations with gasoline and private vehicle use, as well as significant positive correlations with urban population and job densities, the latter producing increases in public transport passenger kilometres per person, a higher proportion of total passenger kilometres on public transport, greater public transport service provision per person, more annual trips per person, and a higher proportion of workers using public transport. Later, based on studies of the USA, European cities, and Australian cities, Kenworthy (1999c) introduced traffic calming (street environments which force traffic to travel at a slower speed), light rails, and urban villages—three integrated factors for solving the multiple problems of car-dependency and increasing public transport service and use.

Badoe (2000) affirm a fairly consistent finding from the literature on transportation and land-use, which is that households in highly dense areas tend to own fewer cars and use public transport more often, resulting in fewer vehicle miles travelled (VMT). They point out that the role of car ownership decisions has been underestimated in the models of integrated land use and transportation interaction, and recommend an integrated research approach for any study of this topic.

Relatedly, on the basis of an attitudinal survey of 389 university students in Hong Kong, Cullinane (2002) concluded that male university students want to own a car more than female students, who are content to rely on the good public transportation services. In a later study based on a survey of 401 car owners examining the reason why people own cars in Hong Kong despite the presence of a good public transport service, Cullinane (2003) report that car ownership is perceived as necessary for reasons of lifestyle, which results in longer travel distances and more trips.

Mokhtarian (2002) explore how residential neighbourhood types influence travel behaviour, taking into account a range of socio-demographic factors, residential location, attitudes, and life style characteristics. They use micro-scale site survey data on road networks, public transport services, land use, and travel diaries collected in San Francisco Bay Area neighbourhoods. Based on structural equation models, they find that the total negative effect of car ownership on public transport usage is greater than that of any other factors analysed. This is in line with findings from previous studies (e.g. Bagley and Mokhtarian, 2002; Koushki, 1988). For example, Koushki (1988), who collected a systematic random survey sample of 2259 households, compared the impact of socio-economic characteristics on transport mode choice in Saudi Arabia and the USA. His results confirm that high income and car ownership contribute significantly to lowering public transport use in Saudi Arabia, which is less than half that of the USA (Koushki, 1988).

Paulley (2004) state that the relationships between income, car ownership and public transport have been well documented. From the National Travel Survey, they find that increases in car ownership per household tend to decrease demand for public transportation, especially for buses (whereas train demand is less elastic). They also stress that car ownership should be included with income when modelling public transport demand, as the income variable can pick up the negative effect of car ownership on public transport demand.

Bresson (2004) analyse panel data, including an annual time series from 1975 to 1995, for 62 urban areas in France. They develop a conventional fixed-effects model, using a Bayesian approach to compare the elasticities of public transport demand over a twenty-year period. Their results show that elasticities can vary over time, and depend on whether only economic determinants, or both economic and structural determinants, are included in the model. They assert that public transportation appears to be an “inferior good” when only three economic determinants (vehicle km – public transport supply, income and price) are included. When other structural determinants, such as population composition by age and gender, household location, and car ownership are introduced, the results show that the “income effect” has a significant impact on the “motorization effect”. That is, increasing income encourages more car ownership, which negatively impacts public transport use. Their results also show that car ownership is not only influenced by changes in income, but also by variations in population composition by age and gender, as well as how it is spread out geographically. For example, the proportion of women living in single-car households has a significant impact on public transportation use.

Balcombe et al. (2004) mention that the public transport elasticity of those who have an alternative mode of transportation is more sensitive to fare changes. Their findings highlight the joint effects of car ownership and income on public transport demand. They confirm that car ownership has negative influence on public transport demand. Moreover, they emphasise the important role of car ownership as an explanatory variable for models of income elasticity. If a car ownership variable is not considered, the elasticities of income would capture the negative influence of car ownership on public transport demand, which could make public transport seem like an inferior good (Balcombe et al., 2004). Additionally, the authors point out the difficulty of separately interpreting the effects of income and car ownership due to their strong correlation. Another interesting fact that they highlight is the variation in extent of car ownership elasticities for different modes of public transport, where car ownership elasticities have a more negative impact on train demand than bus demand; further, there are large discrepancies among market segments and across distance zones.

The significant influence of car ownership on travel mode choice can also be found in research conducted by Greenwald (2006). To examine the relationship between land use and travel mode choice, he used the data collected from the 1994 Household Activity and Travel Diary Survey in the Portland Metro area, and developed multinomial logit and binary logistic models. The relative risk ratios in these models indicate threshold effects of socio-economic and mixed land use on transport mode choices. The most prominent results reported are that the tendency to choose public transportation can be reduced by 99% relative to car use as a consequence of increases in the rate of car ownership, and that growth in the number of license holders can lead to 83% reduction in the propensity to use public transportation over cars.

Bhat (2007) use the 2000 San Francisco Bay Area Travel Survey, especially the section for Alameda County which consists of 233 transport analysis zones, to conduct a comprehensive analysis of how built environment factors, along with transportation, demographic, and network characteristics influence residential choice and car ownership. They find that employment density has marginally a significant influence on car ownership, as does the resident population living in multifamily housing units and low household income groups. Moreover, there is a significant influence of public transport availability on car ownership.

Bohnet (2008) use travel survey data from Hanover Region to develop a predictive model of joint household car ownership and transportation mode choice. Their results reveal that car ownership is influenced not only by socio-economic factors, but also geographic characteristics, and that the need for car ownership can be reduced by integrated urban and transport planning, including high activity density areas with good quality public transport service. The findings from Pitombo (2011) also support the strong relationship between car ownership and travel mode choices. However, as Olszewski (2007) emphasises, car ownership can only be limited via fiscal measures up to a certain point. This lesson is taken from the experiences of Singapore over a 30-year period, where the income elasticity of car ownership turned out to be stronger than price elasticity. Some countries, like Singapore, are implementing transport policies that restrain car ownership in order to minimise the negative effects of pollution and urban traffic congestion. However, the author warns that these fiscal measures will ultimately not be able to restrain car ownership, and urges that transportation planners recognize that it rise along with rapidly growing incomes despite increasing car prices.

2.2.2 Income

Bresson et al. (2003) find that different locations or regions exhibit distinctive dynamics in terms of their income elasticity of public transport use. Their study shows that both short-run and long-run income elasticities in England are negative and significantly high on public transport use while this is not the case for France. The income elasticity of public transport use in France being zero indicates that public transport is seen as an inferior good in England, but not in France. Thus, attitudes towards public transport use could vary by region. Soltani (2006) conduct another comparative analysis on the travel patterns of four suburbs in metropolitan Adelaide. Using multinomial logit models, their findings indicate that people tend to choose travelling by car over public transport when their income increases. Pitombo (2011) also found that individuals with higher incomes tend to have longer trips and use both transport modes (public and private) more often.

Dargay (2002), analyzing bus use in the UK, confirm that the income elasticity of bus use is negative, indicating its positive effect on car ownership and its negative impact on bus usage. Notably, they report that income elasticity in metropolitan areas is more responsive to changes, both in the short and long term, than in rural areas (Dargay and Hanly, 2002).

The findings from the public transport demand analysis conducted by Balcombe et al. (2004) show that generally, income has fairly strong positive effect on train demand, and that income elasticities for both bus and train demand can be expected to increase over time. They also explain that changes in income level have various effects on public transport demand, including different modes and trip lengths. Price elasticities for patrons with high income-levels are high for both buses and trains. Moreover, it could be expected that patrons with low incomes have higher elasticities for short trips, and that those with high incomes have higher elasticities for longer strips. They highlight the interrelationships among income, car ownership, and public transport use. Therefore, they also suggest including car ownership as a control variable when measuring the income elasticity of public transport demand, since the model would pick up the negative effects of car ownership on public transport demand.

Balcombe et al. (2004) report four key findings regarding the effects of income on transportation demand:

- Depending on income levels, income growth can lead to higher car ownership or to an increase in public transport use,

- Where other variables (situations) remain constant, an increase in car ownership (also availability) can result in a decrease in public transport demand,
- The sign (positive or negative) and magnitude of car ownership and income elasticities for public transport demand can vary depending on the income level,
- A consequence of income growth can be an increase in average trip length.

The authors report that income elasticities, when considering car ownership, have a negative effect on bus demand, ranging from -0.5 to -1.0 in the long run (even though this magnitude tends to be smaller in the short run, indicating a substantial decline in bus demand over time). Nevertheless, these income elasticities can change over time when car ownership approaches saturation levels. Then, negative income elasticities can diminish with a corresponding rise in income elasticities for trains.

Research by Thompson (2012) highlights the significant effects of income over other land use variables on public transport use. Using data on bus transit work trips (including origin and destination traffic analysis zones) in Broward County, they examine a range of variables, including population density, population composition by race, employment density, income, walkability, walking and waiting time for public transport, parking fees, vehicle travel times, and whether living downtown or not are considered. Their results show that income has more (negative) sensitivity to public transport use than do land use characteristics. In addition, the positive impact of population density on public transport use is quite significant; however, employment population densities are less influential (Thompson et al, 2012).

Holmgren (2013) uses annual public transport use data for 26 Swedish counties during the period from 1986 to 2001 to develop estimates of how income, fare, and car ownership affect public transport demand. In particular, he considers the effects of income on car ownership and the joint capacity and quality of public transport services. There is high variation in public transport demand (aggregated number of trips and average number of trips per capita), percentage increase in demand, and public transport service levels (average number of vehicle kilometres supplied per km^2). Stockholm stands out among these counties as the region with by far the highest public transport demand and service levels. Nevertheless, the variations in income and car ownership are relatively insignificant among these counties. His findings indicate that car ownership has the highest negative elasticity for public transport demand, as compared to fare, vehicle-kilometres, and income. He also concludes that changes in income have effects on public transport not only directly, but also indirectly, through changes in car ownership. "*Direct effect is positive while the indirect effect is negative*" (Holmgren, 2013: p.106).

2.2.3 Housing Expenditure

Another socio-economic factor considered in the present research is average weekly rent[5], included as a measure of housing expenditure. According to the Australian Bureau of Statistics (15 February 2006, (6 September 2011)), transportation and housing are the highest household expenditures. According to these reports, housing expenditures have increased from $144 per week (16% of total household expenditure) in 2003-04 to $223 per week (18% of total household expenditure) in 2009-10. Similarly, transport expenditure has gone from $139 per week in 2003-04 to $193 per week in 2009-10, along with a nearly 30% increase in average fuel prices, from 90.6 cents per liter in December 2003 to 117.1 cents per liter in December 2009 Department of Commerce (2011).

Sipe (2006) highlight the importance of housing tenure in suburban transportation. They emphasize the vulnerability of low income groups, who only can afford lower housing expenditures at outer suburbs, to the effects of current economic changes, such as industrial restructuring, rising interest rates, and increasing unemployment; and also the need for improvements in access to public transportation for these groups. Further, based on their index of vulnerability assessment for mortgage, petrol and inflation risks for Perth, Dodson (2007) state that it is difficult to interpret the impact of public transport on the level of mortgage and oil vulnerability in Perth because its impact is not even along the train networks in the city. Martinez (2008) suggests that changes in land rents should be considered to be a measure of transport benefits when examining the extent to which transportation projects attract urban development. Location, land use and rents are also included in their land use model, which is based on classical utility maximization assumptions (Martinez, 2008).

In a recent study, Liao (2014) examined the housing expenditures that people are willing to make by way of discrete choice experiments. The explanatory variables included in the model include distance to work, distance to public transport, street design, rent/home prices compared to what participants are currently paying, distance from home to shops, restaurants, public libraries, and schools, housing type, and parking availability. They apply the latent class analysis of preference heterogeneity and find two groups of residential choice behavior. One of them shows compact

[5] Either average weekly rent or mortgage is used to measure housing expenditures for individual households. Before developing a multiple regression model, correlation tests were conducted to satisfy all assumptions. The correlation between average weekly rent and mortgage was found to be very high and their use is nearly identical. Therefore, it is necessary to select only one of these two variables for the regression model. In the present study, average weekly rent is chosen as a measure of housing expenditure because this dataset is available annually, and better suited to use for further modeling and/or model extension and validation with datasets later than 2009.

development is preferred by low income and renter occupied households, they favor use of public transport and are sensitive to price changes. 76.8% of the respondents from the other group are living in single-family houses characterised by a pattern of spatial dispersion and their higher car ownership is contributing to lower public transport usage.

In all cases, the groups of factors that influence travel patterns are socio-economic factors, land use characteristics, and the quality of public transport. A range of models already exists, including several for Perth. However, they all exhibit a number of differences in the way they have been developed. Thus, the present study has the potential to contribute to knowledge of travel patterns through an improved modelling process.

2.3 Public Transport Service Provisions

Other factors reported as having an impact on public transport use are quality of service indicators such as service interval, distance and duration of journey, and accessibility. Webster (1982) propose considering both the supply and demand factors to determine demand for public transportation. They explain that the most common measure of public transport service quality is vehicle-km operated, which reflects service frequency and route coverage; also, public transport service elasticities are more sensitive than fare elasticities. Polat (2012) agrees that service frequency and waiting time determine the quality of public transport services. FitzRoy (1998) emphasize that public transport service quality is a function of the quantity supplied. Their empirical results show that a greater supply in terms of vehicle kilometres (more frequent service on a given network) can result in reduced waiting times and increased public transport demand. Bresson (2003) agree with this emphasis, and also add that the service quality of a specific mode of public transport not only depends on the supply of that mode, but also the availabilities of other modes. Similarly, Polat (2012) finds that service frequency determines passengers' preferences and mode choice with more frequent public transport services being provided in highly dense areas.

Holtzclaw (1994) examines the relationship between household automobile use (vehicle-miles-travelled; odometer data collected during biennial auto emissions inspections) and land use characteristics variables such as density, a public transport accessibility index (reflecting hourly access to public transport), a neighbourhood shopping index (fraction of households within a quarter mile of five key local commercial establishments), and a pedestrian accessibility index (reflecting continuous grids, street slopes, sidewalks, building entrances and traffic control). The findings suggest that the major statistically significant relationship is between automobile use, on the one hand, and density and public transport accessibility, on the other.

According to Perl (1995), the significant increase in car ownership during the period 1950 to 1990, along with decreasing population density in Canada's inner cities during 1970s and 1980s, has a negative impact on public transport usage in Canada (public transport usage per capita dropped from 246 in 1950 to 97 in 1980). They state that Canadian public transport policy failed to respond to the socio-demographic changes in the 1980s and recommended a three phase strategy in urban transport policy for Canada to accommodate the changes in urban mobility patterns. They also highlight the importance of public transport accessibility in urban development planning to end car-dependent land use in the last step of their three phrase strategy.

Bresson et al. (2003) conduct a comparison analysis on the main determinants of demand for public transport in England and France, applying a random-coefficients approach and Bayesian shrinkage estimators. They use journey per capita as their public transport use variable (dependant measure), and vehicle km per capita as a measure of public transport supply. Their empirical results show that the elasticity of public transport service provision is three times higher in England than in France. This suggests that public transport service provision levels are more important for public transport patronage in England. As already reported, they conclude that public transportation is perceived as an inferior good in the UK more than in France, due to the latter's better public transport service quality and a higher income elasticity for car ownership in the UK. In a related analysis, Bresson et al. (2004) use seat km, frequency, and network densities as measures of public transport supply. Their results show that seat km is the most significant determinant of public transportation use, as compared to frequency and network densities. They also explain that variations (increases or decreases) in these two later variables actually reflect changes in seat km. Therefore, they conclude that elasticities resulting from variation in public transport service frequency and network densities are significant determinants of public transport use.

Balcombe et al. (2004) observe that public transport service quality can be defined by a wide range of attributes, such as access and egress time, in-vehicle travel-time, service intervals, and accessibility of services. Other attributes such as comfort, safety, reliability and comfort are seldom used because their quantitative measurement is difficult.

Bresson et al. (2004) explain how seats per kilometre can measure public transport service quality, while frequency and density of network can measure service quantity. By using the annual time series data from 1975 to 1995 in 62 urban areas of France, they examine the economic and structural determinants of public transport demand They develop a conventional fixed-effect model with Bayesian shrinkage estimators by taking into account economic factors (income and price) and structural factors (population by age and gender, car ownership and house locations). Based on their results from this model, they conclude that the impact of service quantity (seats km) is more influential than service quality (frequency and density of network).

Using operator data for 46 counties for the financial years from 1987/88 to 1996/97 (from the Department of the Environment, Transport and the Regions), Dargay (2002) estimate the short term and long term fare elasticities of bus usage per capita in England, especially English counties. They consider income per capita, bus fare (average revenue per journey), private driving costs, demographic factors, and service levels (bus vehicle kilometres). Their findings indicate that the magnitude of service elasticity is higher than fare elasticities for bus use in both metropolitan and rural areas, both for the short term and the long term. They also report that bus use in rural areas is more sensitive to changes in service levels than those in metropolitan areas (in both the short and long terms) because bus services in rural areas are poorer, such that when they are improved, it has a larger impact on bus use. This is further supported by their finding that the English shire counties and Wales areas have the lowest levels of bus use, along with lowest levels of services; whereas English metropolitan areas have the second highest level of bus use, even though their bus services are just mid-level in terms of bus vehicle kilometres per capita.

According to findings by Preston (1998), quoted in Balcombe et al. (2004, p. 74), the service elasticities for bus use can vary based on the time of day. In the short run, bus service provision on Sunday is most sensitive (at 1.05), followed by the one on Saturday (at 0.52). Nevertheless, in the long run, the two highest service elasticities are in the evenings and on Sunday (at 1.95 and 1.61, respectively) and the service elasticities in early morning/peak and on Saturday respond quite equally (the former at 0.58 and the later at 0.67). It is remarkable that the service elasticity in inter-peak periods is not significant in either short run at (0.17) or long run (at 0.3) (Preston (1998) quoted in Balcombe et al., 2004: 74).

The relative risk ratios in multinomial logit and binary logistic models developed by Greenwald (2006) indicate the strong influence of public transport availability on choosing public transportation over car use. This notable result suggests that public transport availability can produce a 56% increment in the inclination to choose public transport modes of travel over car use (Greenwald, 2006).

Moreover, Buehler (2012) performed a detailed comparative analysis of public transport demand in Germany and the USA using national travel surveys conducted in 2001/2002 and 2008/2009 for both countries. They argue that it is essential to have a coordinated package of mutually supportive policies, including more frequent and better quality public transport services and land-use policies that support compact and mixed land-use developments, along with attractive fares, full modal integration, and restrictions on car ownership and use. In addition, Bar-Gera (2003) explain that one of the main challenges of transportation modelling is the interdependency among travel patterns, level of services, and associated congestion. Therefore, public transport service provision (frequencies and average stops per squared km) will be considered in this demand modelling analysis.

2.3.1 Service Provisions (Frequencies)

Hauer (1971) states that service frequency can determine the quality of public transport services, and can encourage or discourage use depending on the availability of alternative modes of transportation. This relationship has been the subject of much subsequent research.

Barton-Aschman Associates (1981) prepared an interim handbook on how people respond to transportation systems changes for the U.S Department of Transportation. Findings from this book regarding peoples' response to changes in public transport frequency, scheduling, and new service provision indicate that scheduling and frequency are the stronger determinants of public transport service quality. They report that scheduling changes to provide more frequent services enables people to follow schedules more easily—for example, service availability every 15 minutes, rather than at particular scheduled time—and this can attract more public transport users. FitzRoy (1997) confirm this result, finding that high frequency is an important factor to maintain market share and the demand for train systems in European countries.

Webster (1982) review elasticities associated with fare changes and service quality, and suggest that supply and demand factors should be jointly considered in transport planning. Their findings regarding service provision (frequency or route changes) elasticities show its positive impact on public transport use, ranging from 0.2 to 1.2 (from time-series analysis) and from 0.6 to 1.4 (from aggregated cross sectional analysis) while 'before and after' studies find positive elasticities between 0.2 and 0.5. Catoe (1998), quoted in Balcombe et al. (2004, p.75) lends support to these findings, reporting positive elasticities of 0.82 (Big Blue Bus Systems) and 0.97 (Lincoln Blvd Route) in Santa Monica, CA, after their service frequencies had been improved.

Stanley (1998) and Catoe (1998), quoted in (Balcombe et al. (2004, p.76)) find substantial positive effects from enhancement to service frequency and hours in Santa Clarita and Santa Monica, California, reporting service elasticities of 1.14 and 0.82, respectively. Additionally, based on a comparative analysis of service elasticities for buses in a number of European cities (by population), the European Commission (1997) finds that service elasticity is higher in large cities (higher populations) than in small cities due to the range of other public transport service modes available. Further, Kilcoyne (1998a and 1998b), also quoted in (Balcombe et al. (2004, p.75)), reports that service frequency elasticity was positive (at +1.14) when headway is improved from 60 to 30 minutes, along with enhanced service hours in Santa Clarita CA. This is supported by Stanley (1998), who explains that introducing new weekend service and improving current frequency in Foothill Transit, LA, has had a positive impact on public transport use at elasticity (at 1.03).

Mees (2000b) observes that Australia's public transportation systems are facing persistent difficulties from operating below potential capacity, and that mode integration (using buses to feed the train system) needs improvement. Dodson (2007) support this modal integration proposal by suggesting circumferential links between the radial train lines, which should encourage more public transport use. He also describes how in Australia, higher income groups are mainly concentrated in inner suburbs with high public transport accessibility, whereas lower income groups living in middle or outer suburb areas face disadvantages as a result of poor public transport service provision.

White (2004) analyse some of the main findings and issues identified in the report "The Demand for Public Transport: a Practical Guide" published by Balcombe et al. (2004). They point out that public transport service intervals can be measured as total vehicle kilometres, frequency, service intervals (headway), wait times, and schedule delays.

The service interval elasticity of bus demand is more sensitive in the long run (at 0.7) as compared to short run (at 0.3). Additionally, they show that the service interval elasticity for bus demand is significantly greater on Sundays and in the evenings, where service provision is relatively low, compared to peak times on weekdays. Finally, service interval elasticities in rural areas are higher than in the metropolitan areas with populations of over 50,000.

Balcombe et al. (2004) also finds that service is valued more highly in areas with higher income levels, and that service interval elasticities differ depending on the size of the cities. They emphasize the negative impact of waiting time on public transport demand is due to such consequences as passengers' discomfort and the inability to use the time for other purposes. Also, public transport services with high frequency can provide patrons with the flexibility of arriving independently of rigidly scheduled services (Balcombe et al., 2004).

Curtis (2004b) conducted a survey to examine travel demand management status at Australian universities based on such aspects as public transport accessibility and service provisions, university support for public and/or private transportation infrastructure use, and any methods used to influence mode choice on university campuses. They recommend that public transport services be provided frequently in the suburbs, where university campuses encourage its use by staff and students. Further, they find that one quarter of universities in Australia do not have public transport services with a frequency of 15 minutes or less—a problem, since frequent public transport service is one of the most important aspects of any travel demand management plan for influencing mode choice (Curtis and Holling, 2004a, 2004b).

The availability of public transportation at residential locations has a great influence on public transport use regardless of household car ownership and income level Lin (2008). But there could also be different service sensitivities in various time periods. In the Australian Transport Council Guidelines quoted in John Taplin et al. (2014, : p.48) state that the short run frequency for service elasticities is 0.25 for peak times, and 0.50 for off-peak times.

Curtis (2011b) assess the impact of increasing public transport service provisions—by introducing a new train line from Perth to Mandurah—on the non-residential land use in the context of three emerging transit oriented development (TOD) precincts. They report growth in number of businesses at new development sites in Cockburn central, along with less significant growth in Wellard and Bull Creek. Further, their findings regarding transport mode shares indicate that the majority of employees in these three TOD precincts are traveling to work by car (Curtis and Mellor, 2011). Another interesting finding from their survey is:

"*In Bull Creek and Cockburn Central, the dominant mode of travel for customers is thought to be the car, although in Cockburn Central public transport use is evident. Only in Wellard are, a higher number of customers perceived to use the public transport. In this case, this may be a reflection of the lower socio-demographic resident profile of the precinct rather than enhanced accessibility for these modes" (*Curtis and Mellor, 2011: 159)

The authors also point out that it is necessary to clearly specify the meaning of "high trip generating development"—specifically, whether the number of trips is used as simple measurement or land use types are also considered for their office and retails uses, as well as health and education.

Based on the preceding research, the present study includes service provision (frequencies) in 3 hours periods of weekdays, Saturday, and Sunday, along with other land use characteristics, urban form, and socio-economic factors.

2.3.2 Average Stops per km^2

The present research uses average stops per km^2 as a measurement of the accessibility of public transportation services, particularly with regard to the stops located within a given suburb. Accordingly, the more stops there are in an area (stop density), the more accessible is public transport for patrons. Balcombe et al. (2004) also observe that public transportation network density reflects the accessibility to the services. The proximity of train stations is also relevant to mode choice, as per Cervero's (2002a) finding that residents within one-half mile of a train station tend to choose public transport over alternative modes.

Asensio (2000) uses monthly data collected between January 1991 and December 1995 in the eleven Spanish urban areas where RENFE (the publicly owned Spanish National Railway Network) operates. He models the short-run and long-run elasticities of train service in Spain, considering suburbanisation (ratio of peripheral to city central population), price, population, service quality (average number of places where train service is provided divided by the total length of train networks in each city), petrol, and other dummy variables to capture the seasonal variation in use over different months of the year. He finds that in terms of service quality elasticity for train use, quality is a more important determinant than changes in price for maintaining and expanding public transportation's market share, especially in large cities. The service quality elasticity in large cities is reported to be 0.78, whereas it is only 0.39 in small Spanish cities. He concludes that better service quality and a higher average number of places where train services are accessible are more important for commuters travelling longer distances in large cities.

Polat (2012, : p.1217) observes that the access time and distance that a person faces in getting to a public transport service stop are significant factors in determining public transport demand. He also defines public transport service accessibility as "*the degree to which a system is usable by as many people as possible"*. Further, he also states that accessibility is important at both points of origin and destination. Cervero (2001c) similarly finds that an increase in accessibility (time or distance) can lead to a decrease in public transport demand. Krygsman et al. (2004) support this finding, reporting the influence of accessibility on the public transport use, as people tend to choose alternative modes of travel if accessibility exceeds the maximum acceptable threshold.

According to Polat (2012), the access coverage of public transport services is another determinant of demand. Murray (2001) highlights the importance of evaluating the trade-offs between access coverage and stop placement efficiency in public transport policy evaluation, and of monitoring, to increase the use of services. He also recommends that variations in travel time and optimum coverage should be considered in assessing the efficiency of stop placement, as this will increase competitiveness with private transportation; further, the standard distance required to access public transportation should be region-specific.

Based on Granger causality test results, Holmgren (2013) states that the relationship between public transport patronage and service provision (vehicle-kilometres) are in a two-way relationship, which should be considered in public transport demand model development. Another fact that he puts forward is that capacity and quality are joint products. Growth in capacity will result in increased public transport service quality, because more frequent services can lead to reductions in waiting times.

Bass et al. (2011) use experimental survey data collected from Catholic University employees in a longitudinal study of the city of Santiago before and after the introduction of a new public transport system ("Transantiago"). The first round of this survey was conducted before the change; the other three rounds were conducted 3 months, 10 months and 20 months afterward. Information was collected on the socio-economic characteristics of the participants and travel details such as origin, destination, mode of travel, and level of service availability. Using a frequency compliance index (ICPH) to measure the quality of service, the authors developed a nested logit model to determine the probability of migration from public transport to other modes. They find that age over 49, income, transport mode availability, and quality of service are significant variables predictors of migration. Among transportation mode characteristics included in the model, newly available modes of private travel (car ownership) and bus/train frequencies have significant effects on mode choice.

Dodson (2004), Buchanan (2005), Sipe (2006) observe that the accessibility of public transportation services can be used as a representation of socio-economic status among Australian cities. Affluent residents with high socio-economic status are located in the areas which have better quality public transport services, whereas those with lower socio-economic status face social and travel disadvantages due to getting much poorer services in the fringe areas where they reside (Dodson et al., 2006; Mees, 2000b; and Morris et al., 2002).

In light of prior studies, the present investigation includes both the frequency of service provision and average stops per km^2 among service provision factors affecting public transport use.

2.4 History of Transport Demand Modelling

According to Bates (2008) the fundamentals of transportation modelling were established in the 1950s by the Detroit and Chicago Transportation Studies. Their techniques were later adopted in the UK for the London conurbation study in the early 1960s. Over the next 20 years, mainstream techniques evolved and alternative ‘paradigms’ were developed. These alternative approaches led to a unifying framework, underpinned by economic theory~~,~~ in the 1970’s. Bates (2008) also states that the remarkable growth in computing power over the last few decades significantly contributed to transportation demand modelling research, expanding the scale and granularity of the problems addressed and providing for various complex mathematical modelling techniques. The fundamentals of these models are informed by economic theory—i.e. demand and supply are used to model transportation (Bates, 2008).

Meyer (1997) observe that demand for transportation is commonly described as derived demand because it is not generally required for its own sake—rather, it reflects need for other activities, such as buying goods or services, going to work or school, seeking leisure activities, and so on. They explain that in the early stages, transportation demand estimates were determined by price, income and other variables—that is, “neoclassic” variables. Over time, land-use or spatial-location characteristics came to offer a more accurate representation of the underlying structure of travellers’ behaviour.

Additionally, the authors propose that the main determinant of transportation demand is location choice, and that all demand forecasts should be based on the geographic distribution of economic activities (Meyer and Straszheim, 1997). They explain the first step of four steps travel demand modelling is that passenger trips and traffic generation are based on patterns of land use and population distribution, and forecasts of passenger trip generation are normally made at the individual or household levels, which provide the estimates for trip origin and destinations demands. These generated trips are then assigned or distributed across various zonal networks. The flows between any two regions can be estimated by one or more activity parameters, such as population or income level, or employment, etc. The next step is that these assigned trips are associated with service costs and price characteristics for each available mode to predict modal choice. Time saving and comfort are also considered in passenger transport mode choice models to minimise total costs and to maximise the utility. The last step is route assignment by mode. They also describe that estimating residential and work-place locations and the relationship between them is the primary emphasis of public transport studies. Route assignment includes specifying patterns of traffic flow for each transportation mode, and describing the considerable details of all relevant linkages. Minimum path algorithms are applied to identify and select the shortest route with least time, distance, and cost for each trip.

Disaggregated travel demand models have also been developed. McFadden (1997) observes that disaggregated models for forecasting transportation demand are based on the assumption that all travel demand is generated by individual choice behaviour, guided by maximization of utility, or preferences. They are flexible in dealing with various problems and allowing transportation planners to address particular questions: for example, predicting the demand response from patrons to a new mode of transportation, or to changes in timetables for public transport service provisions (McFadden, 1997). He explains that disaggregated models can be developed from individual utility functions, which are probability distributions of unobserved variables that determine the utilities associated with each alternative transportation mode. The most commonly used techniques to standardise disaggregated behavioural multinomial logit models are maximum likelihood estimation, non-linear least squares, and the Berkson-Theil method for the data that can be grouped easily. He also suggests that aggregated transport demand models can represent physical regularities in aggregate flows, while disaggregated transport demand models can express regularity in individual choice behaviour among various market segmentation.

2.5 Existing Transport Usage Models

Many mathematical models for transportation demand analysis have been constructed, using different sets of explanatory variables (Paez (2012), Badoe (2000), White (2004), Bass (2011), Bellacicco (1987), Bhat (2005), Bowman (2001), Chiang (2011), Enrique Fernández L (2008), Horn (2002), Ichikawa (1982), Mattsson (2008), McFadden (1974), McNally (1997), Nurlaela (2012), Willumsen (2011), Paez (2012), Prevedouros (1990), Waddell (2002), Western Australian Planning Commission (April 2005), Zhang (2005), Lin (2008)). According to Chang (2006), there are four main techniques for modelling the relationship between land use and transportation demand: spatial interaction models, mathematical programming models, random utility models, and bid-rent models. He explains the pros and cons of each type as follows:

- Spatial interaction models are very simple and comprehensive, but they do not capture the unique characteristics of location and individual preferences due to being aggregated,
- Mathematical programming models also rely on aggregated data, but they can be applied to explicitly examine the partial interactions of land use and public transport demand,
- Random utility models can represent the unique characteristics of locations, along with individual behaviour, to determine utility maximizing conduct, but cannot establish clearly the interaction of land use and public transport demand; and,
- Bid-rent models can capture the characteristics of location and individual preferences, and define the decision-making process through the bidding mechanism and utility maximisation; they cannot, however, determine the outcome of this interaction (Chang, 2006).

Handy (1996) states that several choice models have been developed in the literature to demonstrate the importance effects of land use characteristics factors—given the influence of socio-economic factors—and to predict travel mode choices based on variables such as travel distance, destination choice, combinations of choices (mode and destination), and travel frequency. Among the several urban land-use transport modelling frameworks currently in existence, six are widely used:

1. **Integrated transportation and land use package:** consists of a Disaggregate Residential Allocation Model and Employment Allocation Model that uses the Lowry-derivative form Lowry (1964)(Lowry, 1964). Models that include trip assignment in this package use employment, population, trips, activity rates, household types, trip generation, and distribution rates data. These models test different improvements in linkages within transportation networks, with support from geographic information system (GIS) databases Hunt (2005).
2. **MEPLAN:** an aggregate model where quantities of households and economic activities are allocated to zones in the city, and flows of transportation are generated from interactions

among these factors in different zones Clay (2006). For example, this framework has been used to predict transport use generated by different land use scenarios, such as minimal construction, extensive highway construction, light rail transit (LRT) construction, and transit oriented development Rodier (2002).

3. **NYMTC-LUM:** a framework that simultaneously models the interactions between residential housing, commercial floor space, labour, and non-work travel markets by finding prices and wages that cause the usage and supply in public transport markets (Hunt et al., 2005).

4. **TRANUS:** this package is one of the most general urban models that can deal with land use, transportation, and environment, both temporally and spatially, by using a more restricted set of functional forms and modelling options than does MEPLAN (Hunt et al., 2005).

5. **UrbanSim:** consists of many models which predict patterns of accessibility based on car ownership, changes in households and jobs, the movement of households or jobs within an urban region, the location choices of households and jobs, and new development and redevelopment and the price of land at each location Waddell (2002).

6. **MUSSA:** provides an equilibrated forecast of travel use and land use by adjusting the amount of building stock supplied to consumers' expectations for their housing; it uses traffic analysis zones as its spatial unit of analysis (Hunt et al., 2005).

Gaudry (1975) uses monthly time-series data from the Montreal metropolitan area to formulate a public transportation demand estimation model, based on an aggregate time-series analysis. He explains that public transport demand can be a function of prices for alternative modes of transport, time and comfort factors, income, and activity levels. The factors included in the study are travel distance, changes in fares and waiting time, weather related comfort variables, and other location- specific variables. His findings show that public transport is not an inferior good in the Montreal metropolitan area, and that elasticities for waiting and travel times are more sensitive than price elasticities.

In a later study, Bellacicco (1987) developed a traditional, four step transport model of dynamic transportation, taking into account urban structure characteristics such as industrial employment in particular zones for different income groups, total employment, total population, cost of trip by travel mode, average value of travel time, average value of excess time, capacity, flows of network links, set of links which form the best path from origin to destination, by different travel modes. Their **trip generation** functions for origin and destination only consider land use activities and exponentially weighted average generalised costs per trip. They used the number of trips generated (by origin as well as destination) along with distribution model parameters and generalized trip costs, to develop a trip distribution function. Then, a **modal choice** function was developed based on the modal costs and sensitivity of each person-group to model cost differences. The travel assignment model was then calculated as a function of the flows, capacities, and shortest possible path of the network links. Additionally, they developed a generalised inter-zonal cost figure as a sum of the money cost for the trip, average value of travel time for different travel modes, average value of excess time for travel modes, and the origin and destination 'terminal' times for travel modes.

Bowman (2001) develop an activity-based discrete choice travel demand model system. They forecast demand based on an individual's activity and travel schedule. Using the Boston Travel Survey (conducted in 1991) and transportation service-level data, they analyse the activity patterns, including the day's activity and types of trips based upon the number, purpose and sequence of activity stops, along with travel patterns which consist of travel times, destinations and transport modes influenced by these activity patterns. The authors conclude that, "*the choice of activity pattern is influenced by the expected maximum utility derived from the available trip alternatives"* (Bowman and Ben-Akiva, 2001:p.1) and their findings confirm the fundamental precept of activity-based travel theory: transport demand is only derived from demand for activities when the net utility of activity and travel exceeds the utility achievable from activities without any travel.

Hensher (2001) introduces the fundamental principles and practices for forecasting transport choice and demand. He states very simple and robust principles based on the phenomenon of choice, and of demand derived from the preferences of individuals, their budget constraints, and the quality, availability and costs of alternative transport modes. He describes two main ways to develop transport demand models. First, there are those based on changes in attributes that determine travel choice or demand. Second, there are models based on indicators of people's behavioural responses to changes in attributes (called elasticities[6]). He also states that both single cross-section and time-series data are used in transport demand modelling. Cross-sectional data allow to compare behavioural differences among various groups and to gain a better understanding of factors influencing transport demand. On the other hand, time series data enables researchers to analyse changes over time and to do projection analyses at the aggregate level. Further, he notes that linear regression models have been developed to analyse the relationship between socio-economic characteristics, land use (activity) attributes, transport service provision, and the availability and price of other modes of transportation. Discrete choice and multinomial logit models have also been applied to analyse how individuals make the travel-related decisions, choosing from among a set of available alternative modes to achieve desired outcomes. These models enhance understanding why one alternative is chosen over others, and how observed variables influence these choices (Hensher, 2001).

Pitombo (2011) use exploratory multivariate data analysis and cluster analysis to analyse land use characteristics in the Sõao Paulo Metropolitan Area. They divide the area into 389 transit

[6] Hensher (2001) defines elasticity as a measure of the relationship between percentage changes in levels of attributes and corresponding changes in transport demand.

zones, and then use a two-step cluster algorithm and log-likelihood similarity measures to explore the following variables:

- Industry employment per transit zone inhabitant
- Service sector employment per transit zone inhabitant
- Commerce sector employment per transit zone inhabitant
- Elementary school enrolment per transit zone inhabitant
- High school enrolment per transit zone inhabitant
- Population density.

These variables are then used to characterize transit zones based on their activity levels and total number of residents based on the known relationship between activity distribution and destination choices. Travel patterns are defined as trip linking and associated with a series of attributes based on activity sequences, travel modes, and destinations/travel sequence distances. The independent variables chosen by the classification and regression tree algorithm are:

a) Socioeconomic variables: Position of individuals in their family, level of education, gender, age, family income, car ownership, family size, and use of transportation “credits” for transit tariff.
b) Activities participation: work, study
c) Land use characteristics: total number of household, population, school enrolments, jobs, automobiles, population density, employment density, and trip origin clusters: these trip origin clusters are identified from grouping and classifying the traffic zones in the studied areas based on total number employees in the industry, in the service sector, in the commerce sector, total number of student enrolment in elementary, high schools and universities and resident population density.
d) It is important to mention that for decision tree (DT) application, individual land use and cluster variables were used. The intention was to let the DT choose the variables most important to data segmentation.

Through data segmentation, one can observe the influence of different independent variables on travel choice patterns. In their study, Pitombo et al. (2011) include car ownership, use of transportation “credit” for transit tariff, study and family income, study and position of the individual in the family and origin location clusters. Overall, their findings show that socio-economic variables are significant determinants of transportation mode sequences, and that participation in such activities as study and work is also an influential factor in transport demand. Moreover, the activity levels of the places where people reside determine their travel distances.

Finally, Enrique Fernández L (2008) propose a new, state of the art mathematical method for demand responsive public transit system design, with the aim of developing a new transportation structure in the city of Santiago, Chile. They use the current operational information on Santiago's public transit system and network infrastructure, and consider two levels of design (physical and operational) in formulating transportation network topologies, including the frequencies and capacities of each itinerary. They compare the optimized solution, based on the current system, with the integrated system they propose. They show that the integrated system, which includes a centralised fare collection system, modernised systems to control and manage public transportation fleets, the construction of new stops and intermodal terminals, and segregated corridors can reduce the social cost of restructured system and operating costs (Enrique Fernández L et al., 2008).

John Taplin et al. (2014) reviewed and summarised the strategic transport models used in the major cities of Australia. These models are compared based on their structure criteria as well as integration of different transport networks. They identified EMME and Cube Voyager as the software packages used for transport modelling in Australian cities – EMME is for strategic transport evaluation models (STEM) and Cube Voyager for the regional transportation models (ROM/ROM2) of Perth's Greater Metropolitan areas. In these models, public transport network, 7 STEM and 11 ROM/ROM2 trip purposes and commercial vehicle trips are included to plan for transportation by applying a 4-step procedure. One of the drawbacks in these models is that they do not include land use modelling.

In the following section, existing transport models, specifically for public transport modes are discussed.

2.6 Existing Public Transport Models

Balcombe et al. (2004) review different methods for modelling the elasticities of public transport demand: 'revealed preference' methods and 'stated preference' methods. The key points are as follows:

- Revealed preference methods use observational data from ticketing systems and/or ticket sales from public transportation operators, as well as trip activity diaries and household surveys. These methods have been used in different ways, depending on data availability and research purposes:
    - *Revealed Preference Aggregate Cross-Section Models*
      Public transport demand is forecast or estimated as a function of effects of exogenous factors like population density, employment density, and car ownership.
    - *Revealed Preference Aggregate Time Series Models*
      Changes in demand are measured as a function of changes in exogenous variables such as fares, service provision, income levels, and car ownership.
    - *Revealed Preference Aggregate 'Before and After' Models*

This method is used to examine the impact of a specific change(s) on public transport demand.

They point out that aggregated revealed preference models have limitations because it cannot reveal whether changes in public transport use are caused by changes in its use by existing users or by new users, or specify changes in individuals' choices based on their socio-economic circumstances.

- Stated preference methods, by contrast, are used to examine how individuals respond to proposed changes in public transport systems—such as new fares or new rail systems—based on their preferences.

Mattsson (2008) state that there are three main types of models for determining demand for alternative modes of public transportation:

a) Simple elasticity models, which are derived as a function of price and travel time
b) Public transportation assignment models, which are generated by distributing the demand for each public transportation service and mode on the basis of travel time and price, also taking into account departure time; and,
c) Multinomial logit models of individual choice among modes of transportation, which are based on price, travel time, taste, and unobserved attributes (measurement error).

Doti (1991) developed a multiple linear regression model based on the total number of passengers using public transportation services over quarterly periods from 1974-1988 in Orange County, California. Variables included total wages and salaries, employment, ratio of public transport vehicle service miles to the total population, and the ratio of average fares to gasoline prices. Some dummy variables were also included to take account of seasonality and external shocks. The Cobb-Douglas functional form and Cochrane-Orcutt iterative procedure were applied to eliminate the multicollinearity, and the model was used to forecast ridership during the 1989-1993 period. In this study, public transport usage was defined as a function of the potential number of patrons, the level of relative public transport service, the relative price of public transport services, and seasonality. External shocks included work stoppages and oil shortages. The authors report that when other variables are held constant, a 1% increase in employment could lead to a 1.74% increase in public transport usage in a given annual quarter. The second most influential variable was the ratio of public transport service vehicle service miles to the total population. Here, at 1% increase can lead to and increased of 0.37% in public transport use. The ratio of average fares to gasoline prices was also a significant explanatory variable, with a 1% increase relative fares resulted in a 0.31% decrease in public transport usage. Finally, seasonality was found to be significant; nonetheless, though, employment and service provisions were the most influential predictors of public transport use.

Relatedly, Bain (2011) develop a demand-modelling framework for a regional transportation and land use model system (R-Tresis) in New South Wales, Australia). They used the national visitor survey (NVS), conducted annually by Tourism Research Australia, applying a Bayesian imputation multinomial logit (BI-MNL) method combining elements of Bayes' theorem and multinomial logit choice models. Predictive models were developed for different modes of travel (car, bus, train and plane) based on population, travelled distance, and mean household size. Their results indicate that fare levels have the most significant negative effect on bus use, while waiting time most strongly dissuaded train use, and followed by changes in fare. On the other hand, train frequency at origin is strongly encourages train use (Bain et al., 2011).

Much research has sought to develop public transport demand forecasting models that apply to the specific issues or needs of a city or region. Chiang (2011), for example, developed a model to forecast public transportation use in Metropolitan Tulsa. They consider the number of individuals receiving food stamps from the Food Research and Action Centre, the annual budget for Tulsa Transit's operation and capital funds, changes in fixed route fares, and seasonality (for August and September). Various techniques, including regression analysis, neural networks and autoregressive integrated moving-average (ARIMA) time series analysis models are combined with forecasts. They conclude that operation funds (as a measure of public transport service supply) have the most significant positive influence on use, while the number of food stamps (reflecting the unemployed population) has a negative impact. Further, after comparing forecast results from different methods, they argue that a combined forecasts method can provide more accurate projections (Chiang et al., 2011).

Recently, researchers have begun to look at how to maintain public transport demand stability. Bass (2011) propose a model of demand stability that is based on a nested logit structure that considers people's travel mode choices among alternative modalities, their time dependence, and their socio-economic characteristics. The model also takes into account external factors at two levels of aggregation—e.g. the elementary alternative level and public-transport-private level. The authors claim that their model can be used to identify the underlying reasons why people switch from public to private transportation, the probability of this happening, and remedial measures to prevent or correct this. Using an experimental survey administered before and after a public transport change in Santiago, they found that a reduction in weekly working hours can significantly reduce public transport use; the second most impactful factor being increases in waiting time, which reflect poor service quality. Additional factors predicting a switch to private transportation are: being over 49 years old, activity levels for those under 35, new car availability, and availability of parking. Therefore, they suggest implementing more sensitive public transportation services (Bass et al., 2011).

2.7 Existing Transport Models Used in Perth

According to a review from the Western Australian Planning Commission (April 2005), there are three main transport models currently used in Perth:

1. DPI's **Strategic Transport Evaluation Model (STEM)**: a high level strategic multi-modal model for land use and transport policy assessment which illustrates traffic flows from different scenarios and measures the performance of the metropolitan transport system in terms of economic efficiency, social impact, and broad environmental impact.
2. Main Roads Western Australia's (MRWA) **Regional Operational Model (ROM)**: a medium level strategic/operational road-based model for assessing the impact of road infrastructure projects on traffic flow and traffic volume on the road traffic system, such as at interchanges and intersections.
3. The city of Perth's detailed **City Centre Operational Model (CPM)** analyses traffic flows and car parking impacts in the CBD and outputs the detailed traffic flows, including turning movement flows, travel times and delays, and traffic system performance measures such as degrees of saturation at intersections.

To be able to improve public transport service provisions, and thereby increase usage, Porta (June 2008) developed a GIS-based tool that evaluates the centrality and connectivity of the public transport spatial network in Perth (and Melbourne)—which are important aspects of its quality. The criteria that they use to assess the connectivity and efficiency of the system are the impediment value of route segments between two nodes, the degree centrality of each node, the closeness centrality of each node, betweenness centrality for route segments, and the connectivity index for each node. The model evaluates integrated train and bus networks in Perth (both before and after opening the Perth-to-Mandurah railway) and provides a comprehensive, comparative analysis of geographical coverage, network connectivity and centrality.

Curtis (2009) enhance this spatial network analysis for multimodal urban transport systems (SNAMUTS) by defining more activity nodes in Perth's Network City, Western Australia Planning Commission (2004) to examine proposed scenarios of land use-integration. They produce simulations of future urban development in metropolitan Perth until 2031 on the basis of seven measures of centrality, connectivity and performance indicators. The latter indicators are *degree centrality by transfer ("average minimum number of transfers between origin and destination"), closeness centrality by impediment ("average minimum cumulative travel time divided by service frequency per segment"), efficiency change ("before and after ratio of inverse cumulative impediment for all network paths"), contour catchment ("the number of residents and jobs in activity nodes accessible within 30 min"), between centrality ("weighted* *by cumulative impediment and combined activity node size") and network connectivity ("the propensity of each activity node to act as a transfer hub") (*Curtis and Scheurer, 2009: 76).

Subsequently, Curtis (2010) state that the SNAMUTS model has evolved to provide more comprehensive land use-public transport integration analysis. The model was enhanced to analyse how changes in public transport service provision affect levels of centrality in a particular activity node and/or in the network as a whole, and the effect of this, in turn, on the spatial distribution of residents and employment. They also report that the SNAMUTS model has been used to conduct before and after analyses of the impact of some public transport network reconfigurations. This allows for evaluation of various proposed scenarios for land use-transport integration and the development of benchmarks to compare public transport network accessibility in different cities.

Meanwhile, Curtis (2011a) reports that SNAMUTS can be used to evaluate the outcomes of public transport network connectivity and centrality for various proposed scenarios in land use-transport integration projects. Considering population growth forecasts, along with outer metropolitan growth; the intensification of activities in the central city and middle suburbs; activity corridor projects; policymakers from state planning; main roads; and public transportation departments in Western Australia, the model determines which activity centres should be focused on. Changes in service intensity and average catchment for different proposed scenarios are compared, and the author discusses how the model facilitates more collaborative and deliberative discussion on planning choices, and how it can be calibrated based on feedback form the practitioners (Curtis, 2011).

In their review of transportation modelling, John Taplin et al. (2014) make recommendations for determining the appropriate application of macroscopic, meso-scopic, and microscopic models; identifying the required land use variables and their limitations in terms of data collection; and compiling and developing an integrated transport model. They mention that Main Road Western Australia is deploying a meso-scopic modelling trial to develop a congestion management strategy, with a focus on Canning Highway and its interchange with the Kwinana Freeway. They propose an integrated transportation-modelling suite called PLATINUM for Perth, because it can integrate existing models: Perth Strategic Transport Model, Perth Road Transport Model, Perth External Travel Model, and Perth Freight Transport Model. The proposed system can also be used to generate required reports by integrating the currently used transport matrices, such as the road passenger/vehicle Origin-Destination metrics by time of the day and planning years; delay information on network travel time; and freight vehicle and external vehicle trip matrices. Additionally, they highlight the need for continuous panel data to measure changes in travel behaviours and preferences in conjunction with SmartRider, Perth and Regions Travel Survey (PARTS) and census data.

2.8 Smart Card Data Use in Public Transport Research

Up until now, the rich data resources inherent in public transport smart card systems seem not to have been widely exploited for academic research. A detailed review of these studies is provided in this section. Some researchers apply data mining techniques, using smart card data to gain a better understanding of patrons' travel patterns and behaviours. One example to this is the use of smart card data to explain non-habitual overcrowding on public transport in Singapore Pereira (2015).

Morency (2007) discuss how data mining techniques can provide a better evaluation of the spatial and temporal variability of public transport use by various population segments. The data they use were collected in 2005 through the Smart Card system used by Société de Transport de l'Outatouais (STO) in Gatineau, Quebec. Notably, they measure the regularity of public transport use by aggregating the frequencies for the most frequently used bus stops. Their results shows that commuter-type patrons (adult-interzone and adult-express) display routine-behaviour, using a smaller number of stops and contributing a higher proportion of the boarding as a consequence of travel between home and the workplace. On the other hand, student and senior patrons show more variation in use of bus stops (Morency et al., 2007). From their temporal variability analysis, they show that commuter-type patrons have more zero-boarding days on weekends, while significant fluctuations exist among senior patrons, who have many zero-boarding days on weekdays but low zero-boarding days on weekends. They also find that 50% of adult-interzone patrons show the highest overall temporal regularity (at 93%).

Further, Pelletier (2011) conducted a literature review of smart card data use in public transport research. They concluded that smart data have been analysed for decision support at three levels of management in various local and national contexts:

a) Strategic public transport planning:
   - Identifying temporal variation in public transport use by user-types to better understand user behaviour (Morency et al., 2007),
   - Computing trip rates, linked trips (bus-to-bus interchange), and turnover analysis to determine the consistency of patrons' travel behaviour patterns over time, which can assist in designing the targeted marketing campaigns Bagchi (2005),
   - Doing transfer-pattern analysis by using an algorithm to detect transfer coincidence for itinerary reconstruction, along with load profile analysis, which can assist planners in understanding the detailed spatial-temporal progression and passenger-load-variability along travel routes and enhance timetable design Chapleau (2008) and,

- Developing public transportation use profiles based on the volume of transfers, boarding, and the spatial and temporal variability of different modes of public transportation Lim (2008).

b) Tactical public transport planning:

- Analysing linked-trips and turnover rates to suggest adjustments in public transport offerings and schedules (Bagchi and White, 2005),
- Examining the transfer activities and the linked and non-linked itinerary to identify the maximum number of boarding points and return runs, which are useful for schedule coordination among different public transport methods (Chu and Chapleau, 2008),
- Developing public transportation network design problem algorithms (via multilevel programming structure) based on the itineraries, frequencies, and capacities of a set of public transportation services to propose optimised service routes, frequencies and capacities (operational design) (Enrique Fernández L et al., 2008),
- Deriving an algorithm to estimate potential destinations based on smart card and GPS data, which can be used to generate a comprehensive origin-destination matrix that optimizes service planning Munizaga (2012),
- Conducting longitudinal analysis of spatial and temporal variability in public transport use, classifying smart cards based on boarding patterns to gain a better understanding of user behaviour (Morency et al., 2007),
- Examining the frequency, consistency and composition of public transportation usage by different modes of travel, along with access distance to services, to facilitate provision adjustments Utsunomiya (2006),
- Developing methods for identifying transfers and complete journeys, and linking journey-stages to complete journeys, which are used to identify direct links or reroutes and minimise transfers Seaborn (2009),
- Developing a supply-dependent Integrated Intervening Opportunities Model (IIOM) for public transportation trips based on employment population, resident population, school enrolment, origin-destination matrices from surveys, and public transport supply characteristics (such as spatial density of public transport stops, frequency, and trip durations) to provide a tool that forecasts changes in use patterns resulting from modifications to public transportation networks, or from changes in socio-demographic and socio-economic characteristics Nazem (2013).

c) Operational public transport planning:

- Proposing ways to detect and analyse errors, inconsistencies, and anomalies in transactional data, such as boarding date, time, and location, and how to improve data integrity Chapleau (2008),

- Performing user-behaviour analyses of the temporal distribution of public transport use and load-profiles for a particular route, and then developing a trip destination estimation model to estimate the alighting location for each boarding, for use in service provision planning Trépanier (2007),
- Identifying and predicting the non-habitual large crowds called as overcrowding hotspots by investigating the public transport usage smartcard data from Singapore and five social network and event websites such as Facebook, google hits, eventful.com and last.fm by applying Bayesian hierarchical additive model, Pereira (2015).

Along with smart card data use, geographic information systems have been used in public transport demand analysis and modelling. Kenneth J. Dueker: Zhong-ren (2008) state that the geographic information system (GIS) has been used in transport modelling for its four key functions: digital mapping, data management, data analysis, and data presentation. Further, three GIS database queries—spatial queries, attribute queries, and combined spatial and attribute queries—can be used to merge data (linking point locations from a land map to transportation network map), data districting (aggregating smaller zones into larger transportation analysis zones), data measurement (measuring the distance between the points or lengths of particular routes, or areas of transport analysis zones), data buffering (illustrating the buffers to reflect the specified size of geographic features, e.g. line buffer can be used to illustrate a bus route catchment area of 400 m around its route), and data overlaying (a public transportation stops map and a suburb boundary map can be overlaid to extract the number of stops in each suburb).

As the present literature review indicates, many researchers have examined various combinations of land use characteristics and transport facilities, including public transport service provision, to propose more sustainable approaches to transportation and land use development. Paez (2011) propose a very compelling geo-demographic framework to examine how the socio-economic characteristics of patrons that are stored in smart cards can be used to identify transportation zones with commercial potential, which in turn can increase public transport use. Travel household survey data, collected from the Greater Montreal area in 2003, identifies business locations and types, the number of employees, and annual sales, (Paez et al., 2011). By using data processing, analysis, and visualization functions from GIS, the authors' geo-demographic findings locate commercially viable zones based on patrons' use patterns. They also identify the types of businesses that should be integrated into public transport zones, given patrons' travel behaviour and socio-economic characteristics (Paez et al., 2011).

One of the advantages that smart card travel data provides is ability to conduct comprehensive analyses of spatial and temporal patterns in daily public transport use, like the one carried out by Nishiuchi (2013). They use smart card travel data collected from the bus and train networks serving Kochi city, Japan, for the period of June 2010. Their cumulative curve of frequency (number of passengers) shows that 40% of passengers are less frequently using the public transport. Their usages are accounted for 1 or 2 days during the studied 30 days period. Further, their spatial-temporal analysis of individual daily trip patterns shows that there is no significant variation among the number of trips by day of the week. Notably, they find that adult patrons who registered smart cards have relatively low levels of spatial and temporal routine. This probably indicates a weak routine-commitment to using public transportation for employment and education. On the other hand, children have comparatively high use routines, reflecting their regular journeys to school or college (Nishiuchi et al., 2013).

Based on smart card data from the Beijing bus and train systems for 5th-9th July 2010, Ma (2013) analyse individual travel patterns to identify the regularity clusters. The attributes considered in their cluster analysis are "*number of travel days, number of similar first boarding times, number of similar route sequences and number of similar stop ID sequences"* (Ma et al., 2013: p.6). They use the K-Means++ algorithm to identify five regularity clusters in public transport use: very low, low, medium, high, and very high regularity. These findings allow public transport authorities to identify types of patrons, travel locations and times, and service availability in these areas, thereby helping to improve public transport and attract more users. Moreover, they carried out comparison analysis on different data mining algorithms for clustering, using their cluster analysis results from the K-Means++ algorithm as baseline index. Further, they recommend using the rough set-based algorithm over alternatives like the Naïve Bayes Classifier, C4.5 Decision Tree, K-Nearest Neighbour, and the Three-hidden-layers Neural Network algorithm when analysing datasets for clustering (Ma et al., 2013).

In conclusion, much research has made use of public transport smart card data to examine temporal and spatial variations in public transport use, as well as the diverse travel patterns of different types of patrons. By applying more advanced modelling approaches, the travel regularities of different types of patrons have been identified, yielding a better understanding of travel behaviour that can, in turn, assist in determining why alternative modes of travel may be preferred and in developing patron-retention and expansion programs.

2.9 Research Gaps

Gim (2012) states that findings regarding the relationship between land-use characteristics and travel behaviour have not resulted in a consensus. He conducts a meta-analysis on the magnitude of density-travel relationships by synthesising empirical findings from studies that were published between 1970 and 2008. His findings indicate that the accuracy of research design and application of statistical techniques produce variation in the magnitude of the relationship between density and travel behaviour. He also finds that the relationship between density and travel behaviour was significant in both the United States and Europe, and that density had been the most frequently analysed element of land-use characteristics. He notes that there have been three approaches to specifying the relationship between the density and travel behaviour:

a) Metropolitan-level studies explaining the propensity on an international scale
b) Replicating and validating the outcomes of area-specific studies, and
c) Syntheses of quantitative and qualitative outcomes from all of these studies.

Based on their review of previous studies, Badoe (2000) also observe that the results of transportation-land-use interactions vary a lot because of data limitations and methodological weaknesses. Many empirical findings indicate that land use characteristics are important determinants of transportation choice. On the other hand, some researchers, such as Ewing et al. (1996) find that residential density, job accessibility, and mixed land use do not significantly influence household trip rates once socio-demographic variables are considered. This is supported by Boarnet (1998) who, in examining the relationship between non-work travel and land use characteristics in Southern California, conclude that land use characteristics do not explain non-work travel behaviour at the neighbourhood level.

Relatedly, Badoe and Miller (2000) recommend more comprehensive examinations of the interactions between land-use, socio-economic and neighbourhood characteristics, and public transport supply, as well as analysing transport mode choices at more granular spatial level, since aggregated zone-units can comprise different levels of variability. Moreover, proper measurements of density are also important in explaining its influence on travel mode choices.

In a recent meta-analysis, Gim (2012) makes additional recommendations for measuring land use characteristics. In examining the differences in the size of density-travel behaviour relationships reported in research on the United States and Europe, he finds that the relationship is significantly stronger in Europe. The author uses two main variables, research design rigorousness and technical rigorousness, in his sensitivity test of previously reported relationships. Research design rigorousness was coded as having the lowest score where metropolitan-scale data is used; a modest score for small-scale data; and the highest score for studies using longitudinal (temporal precedence) data. Technical rigorousness was coded according to whether studies employed bivariate or multi-level regression, or structural equation modelling. Gim (2012) states that both research design and technical rigorousness are important for finding the true magnitude of the relationship between density and travel behaviour. Some of his recommendations mirror those previously proposed by Handy (2005), who suggests that empirical research using a low research design with broader geographical scale data should apply more rigorous techniques to compensate for limitations in design, data availability, accuracy, and the granularity of geographical scales.

As stated in the earlier sections of this literature review chapter, the relationships between land-use characteristics, urban form, socio-economic factors, and public transportation service provision and use are very complex due to the interrelationships among the observed variables. It is very difficult and risky to generalise about public transport use when only considering a few factors. For example, increases in public transport service provision often encouraging increased usage, but there are scenarios where this does not hold. Therefore, there is a need to study the synergistic influences of land use characteristics, urban form factors, socio-economic factors, and service provision on public transport use at more granular geographical scales, and with revealed preference usage data that can be extracted from smart card databases.

2.10 Research Objectives and Questions

As stated in the previous section, there is a need to study the synergistic influence of land use characteristics, urban form factors, socio-economic factors, and service provisions on public transport use at more granular geographical scales and with revealed preference usage data. This section states the research objectives and questions made relevant by gaps in the existing literature.

2.10.1 Research Objectives

The main objective of this thesis is to develop a comprehensive understanding of the determinants of public transport use in the suburbs of Western Australia's main metropolitan area – its capital city Perth. This study seeks to analyse the whole picture with an original dataset which addresses aspects of public transport demand that other studies have only partially examined. The study also aims to construct a meticulous, granular database comprising the most detailed information on physical planning infrastructure, socio-economic characteristics, public transport service provision, and other variables, ever brought together on a fine-grain geographical scale in any city. Further, it aims to feature a comprehensive, detailed analysis of the processes driving temporal and spatial patterns of public transport use in Perth, including information on types of patrons, the origin suburbs of their journeys, and the day of travel (weekdays and/or weekends) by particular time segments, along with data on revealed preferences derived from smart cards.

The extent of the relationships between public transport use density and other factors vary depending on whether they are measured with bivariate or multivariate correlations. Therefore, unlike previous studies, all land use characteristics, socio-economic, urban form factors and public transport service provisions factors are taken into account in developing a predictive model. Further, a theoretical contribution of this research is that it analyses the relationship of land use characteristics factors and public transport use before examining how the strength of this relationship can be changed when moderating factors (i.e. socio-economic and urban form factors) are taken into consideration.

There is always a concern about problems of multicollinearity when a large number of variables are included in a regression model. Accordingly, factor analysis is used to derive latent land user characteristics and socio-economic factors and service provision factors, thereby preserving the richness of the observed data. With fewer latent variables and less distortion from multicollinearity, a more robust regression model can be developed, with implications for public transport policy. Moreover, at a very granular spatial level the study aims to produce mathematical equations to predict how much changes in land-use characteristics and socio-economic factors, supply factors, or urban form variables can impact public transport use density at the suburb-level. This will help facilitate operational, tactical and strategic public transport planning.

In summary, the study's two aims are:

1. To provide detailed descriptions of public transport usage patterns in Perth based on types of patrons and various temporal and spatial factors, and
2. To conduct a comprehensive and rigorous analysis of use-determinants and develop a robust predictive model which can inform policymaking.

2.10.2 Research Questions

The major research question is as follows:

What are the primary determinants that explain spatial and temporal variations in public transportation use in Perth for the year 2009?

The following research sub-questions are necessary to investigate this overarching question:

1. What is the most appropriate way to construct a model to predict public transport usage based on factors, such as land use characteristics, urban form, socio-economic conditions and public transport availability? and
2. What explanatory power does such a model provide and how can it inform policy making?

Specific tasks that the construction of such a model require are:

- To provide detailed analysis and spatial and temporal characterisation of transport usage patterns across the suburbs of Perth.

- To examine land use characteristics, such as resident population density by age and gender, student (including university student) population densities, and whether they account for explaining the variances in public transport usages across the suburbs in Perth. If so, which mixed land use characteristics factors has a greater potency in accounting for these variances,

- To analyse whether land use characteristics factors, such as the employment densities in various industries, significantly relate with discrepancies in public transport usages across suburbs in Perth,

- To investigate whether urban form factors, such as the distance from city centre and total road length per square kilometre, influence public transport use across suburbs in Perth. If so, to what extent can they determine this public transport use,

- To examine whether socio-economic conditions, such as income, average rent as main household expenditure and average car ownership per household, can be used as an explanatory factor in understanding the differences in public transport usages across the suburbs in Perth. If so, which socio-economic conditions are more analytically salient and empirically cogent in explaining the variations,

- To investigate whether it is possible to use the availability of public transport services to explain differential usages of public transport across the suburbs in Perth. If so, how would it induce differential patterns of public transport usage in Perth,

- To analyse how sensitively changes in these observed variables impact on variations in public transport use,

- To determine which factor has the highest explanatory power when land use characteristics, urban form, socio-economic and public transport service factors are considered at once and

- To develop the most appropriate way to construct the model to predict the public transport usage based on all these factors.

2.11 Significance of the Research

It is well understood that relationships among socio-economic traits, land use characteristics, and travel behaviour are complex. Therefore, it is important to interpret what a study finds within the context of its limitations and implications for future research. The majority of previous studies were based on survey/census data, which were derived from stated preference methods for hypothetical travel behaviour analysis. Louviere et al. (2000) suggest that stated preference and revealed preference approaches can be comparable for the valuation of attributes, but that stated preferences may yield biased results. Accordingly, the present study examines causal relationships among socio-economic traits, land use characteristics, and travel behaviour using data derived from a revealed preference approach that reflects actual travel behaviour.

The policy-relevance of the model developed in this research is that it enhances the understanding of the factors shaping public transport use patterns in Perth. More specifically, the ***expected benefits from policy implications of the model*** are as follows:

a) The model can assist in predicting changes in public transport use due to changes in socio-economic factors such as average car ownership, income, and housing expenditures.
b) It will help to accommodate predicted usage and inform policymaking related to increasing public transport service provision in terms of frequency and service density (average stops per square km). It will also assist in predicting supply requirements on the basis of a range of socio-economic factors and the land use characteristics of individual suburbs.
c) The model will be able to inform suggesting required policy interventions for reducing car dependence and fuel consumption. In particular, the model can specify show how land use characteristics, public transport accessibility, and quality of service can be transformed to increase public transport use in the context of the socio-economic realities in each suburb of Perth.

Therefore, the outcomes of this research are:

- A better understanding of the factors affecting public transport use based on an actual revealed preference approach and using a unique set of data for Perth, thereby contributing to new knowledge in the field, and
- A practical contribution in helping to predict the requirements of public transport services based on land use characteristics, socio-economic factors, and public transport service provision in each suburb, which can in turn help to increase public transport usage.

3 Methodology I: Theoretical Considerations and Data Requirements

This chapter sets out the methodology followed to investigate the research questions posed in Chapter 1. The chapter explains the purpose of the research, addresses theoretical considerations, and reviews data requirements, and is organized as follows:

3.1 Application of quantitative social science approach taken in this research, and why it was chosen over other methods. A justification is then provided for the use of non-experimental methods,
3.2 Explanation of the theoretical basis for the research approach,
3.3 Detailed explanation of the theoretical framework,
3.4 Explanation on geographical scope selected for this research,
3.5 & 3.6 Discussion of the critical nature of spatial and temporal dimensions,
3.7 Validation of the research approach,
3.8 Explanation of the data requirements for a predictive model of public transport usage,
3.9 Detailed descriptions of the origins of the data and the manipulation procedures used to aggregate, or compartmentalise, the variables so that they are on the same spatial level and
3.10 Discussion of technical issues encountered and adopted solutions and
3.11 Limitations of this research

There have been many quantitative studies McNally (1997); Kenworthy (1996); Boarnet (1998); Buehler (2012); Curtis (2008); Kenworthy (1991); Cervero (1991); Cervero (2002b); Lin (2009); Pitombo (2011) Stead (2001a); White (2009); White (2004); Paulley (2006); Bellacicco (1987), using a variety of methodologies and examining the relationship between travel behaviour, on the one hand, and land use characteristics and socio-economic indictors on the other. Handy (1996) states that the methodologies used to examine the relationship between land use characteristics and travel behaviour, a key concern in this research, fall on either end of a spectrum of data granularity. At one end of the spectrum are studies using aggregate data based on the characteristics of cities or zones; at the other are studies applying choice models and activity-based analyses of individuals or households. The techniques used in these approaches also vary, ranging from simple correlation tests to complicated behaviour estimation methods. Handy (1996) identifies five main research methodologies deployed in travel behaviour research based on land use characteristics, namely simulation studies, activity based analyses, aggregate analysis, disaggregate analysis, and discrete choice models. She also suggests the following aspects be considered to improve transportation research methodologies: (1) address which aspects of land use characteristics have significant influence on travel choices, (2) conduct a sensitivity analysis on the impact of a change in land use characteristics on travel behaviour, (3) analyse the extent to which public transport can replace the automobiles, (4) explore appropriate ways of measuring the observed variables for better and more meaningful interpretation and (5) take into account geographic location, which could lead to variations in transport usage. The design of this research considers all of Handy's (1996) suggestions and is described in this chapter.

3.1 Application of Quantitative Social Science

According to Neuman (2006), there are three approaches in social science: positivism, interpretive social science and critical social science. Bryman (2004) defines positivism as "*an epistemological position that advocates the application of the methods of the natural sciences to the study of social reality and beyond".* Painter (2008) explain that an interpretive approach can be applied when the research explores the subjective reasons, internal reality and meanings which trigger people's social actions, typically by collecting qualitative data through interviewing or observation. Finally, Schwandt (2007) explains that critical social science is derived from critical theory, and has the aim of integrating theory and practice to gain more awareness of, and challenge the contradictions and distortions in, various belief systems and social practices, specifically by developing an imminent critique of them. It is more practical and normative, and encourages self-knowledge and self-reflection through inquiry to achieve individual or social transformation, rather than just description or explanation. The positivist approach is the one used in this study, as the aim is to gather facts and produce objective scientific findings about the topic. The interpretive method is not appropriate because the aim is not to explore people's subjective reasons for using public transportation. Similarly, the critical method is not applicable because this research is intended neither to raise awareness of the contradictions in social practices by using immanent critique, nor to transform individual or social practices.

This research examines the main determinants of public transport usage in the Perth metropolitan suburbs, and how these variables individually and collectively influence such usage. According to Cozby (2005), there are two main approaches to the study of relationships among variables, (1) the non-experimental and (2) the experimental method. In this research, the non-experimental method is applied by measuring and correlating variables of interest to study the relationships between public transport use, on the one hand, and service provision factors, socio-economic factors and land use characteristics, on the other. The correlations between observed variables are investigated without direct manipulation and control of variables. A major drawback of this non-experimental method is that it is difficult to determine the directions of cause and effect among the variables, another reason why, together with the cross-sectional design, inferences about causality are not attempted.

In this research, land use characteristics variables, such as employment population densities, estimated resident population densities, student population densities, socio-economic factors, and public transport service provision factors are aggregated at the suburb level to analyse whether and how they influence public transport usage in Perth. One of the weaknesses of aggregate analysis is that the findings cannot reflect the characteristics of public transportation users. Therefore, the level of detail in the observed variables is carefully constructed to encapsulate and maintain the overall characteristics of the place. Aggregate analysis does not use the average values for many variables because they do not represent how the variables are actually distributed. For example, average weighted income does not reflect the proportions of people in different income groups within an observed spatial location. Factor analysis is applied to capture these variations by analysing the correlations among sub-categorised variables and constructing latent variables among them.

3.2 Research Approach

The first approach applied in interpreting the findings is descriptive research. Neuman (2006) defines descriptive research as presenting a picture of the specific details of a situation, social setting or relationship. Using data from Perth's Smart Rider ticketing system collected in 2009, the research begins with a descriptive analysis of a well-defined set of public transport service provision variables, including socio-economic factors, land use characteristics indicators, and public transport usage.

Next, in accordance with the main purpose of this research, a predictive model of public transport usage in metropolitan Western Australia is developed using the aforementioned variables. The model helps to achieve a better understanding of the relationship between these diverse factors in determining public transport usage variations in each suburb. The theoretical framework developed in section 1.4 is used to gain a better understanding of which aspects of land use characteristics significantly influence choices about whether to use public transport in metropolitan suburbs of Perth. This framework is employed as the guide to interpreting the strength of relationships between predictor and outcome variables, as well as how the strength of each relationship changes when all of predictors are considered together.

This research is designed to inform operational and tactical decision making, for example by predicting the percentage increase in public transport usage when there is a corresponding increase in observed variables in a given suburb. Furthermore, the model is designed to inform strategic decision making about issues such as how mixed land-use, or land use characteristics factors, should be changed to achieve public transportation goals.

In the following section, a detailed explanation of the theoretical and empirical contribution to knowledge about public transport demand modelling is provided.

3.3 Theoretical Framework

In this section, the theoretical framework for this research is introduced to outline the general analytical feature of the study. Specifically, the focus is on the linkages between three main analytical pillars: land use characteristics, socio-economic characteristics and public transport service provision, so as to explicate their potential contribution in explaining public transport usage in the case study of Perth metropolitan.

Fawcett (1992) points out that a conceptual model can contribute to theory development by focusing on particular things, and also can act as a guide for the development of new theories by drawing attention to certain concepts and their relationships in a distinctive context. Batey (1971) also advises that even though two researchers may observe the same situation or event, their perceptions of why it occurs, their conceptual approach to the problem, the knowledge base on which they focus, and their analytical examination of that problem may all differ. Therefore, the theoretical framework for this research needs to be conceptualised at its initial phase. According to Raykov (2008), a theoretical framework should have the following conceptual features:

1. A clear definition of phenomena to be studied,
2. A comprehensible explanation of the peculiar nature of the problem to be studied,
3. Clear identification of the data to be collected,
4. Directions dealing with the research design, instruments, and methods to be deployed,
5. Guidance on the methods to be employed for reducing and analysing the data, and
6. A coherent explanation of the nature of the contributions that this research makes toward the advancement of knowledge.

All of these considerations inform the conceptual model developed for the scope and objectives of this research. According to McNally (2008), the initial development of transport demand modelling in the early 1950s was based on trip generation, distribution and diversion. The first comprehensive application of such a model was in the Chicago Area Transportation Study, which considered land use characteristics and economic evaluation. This was a Four-Step Model designed to forecast trip generation, trip distribution, mode choice, and route choice. Vuchic (2005) states that public transport demand models have since developed based on this four-step methodology, with most models involving variations on these steps. He also explains that numerous phases of transportation planning are based on theoretical models representing relationships between observed factors and trip generation. This model calibration process is applied to develop transportation-planning models on the basis of observed variables, as counted or measured at the time of planning. The derived planning models are then used to project future travel demand based on the assumption that the relationships among individual behaviour, conditions, and travel characteristics will remain the same in the future. If a change in these relationships is indicated, the travel forecasting models can be adjusted accordingly.

The most commonly applied sequence of steps in transport modelling is illustrated by Fertal (1966) as follows (see figure 8):

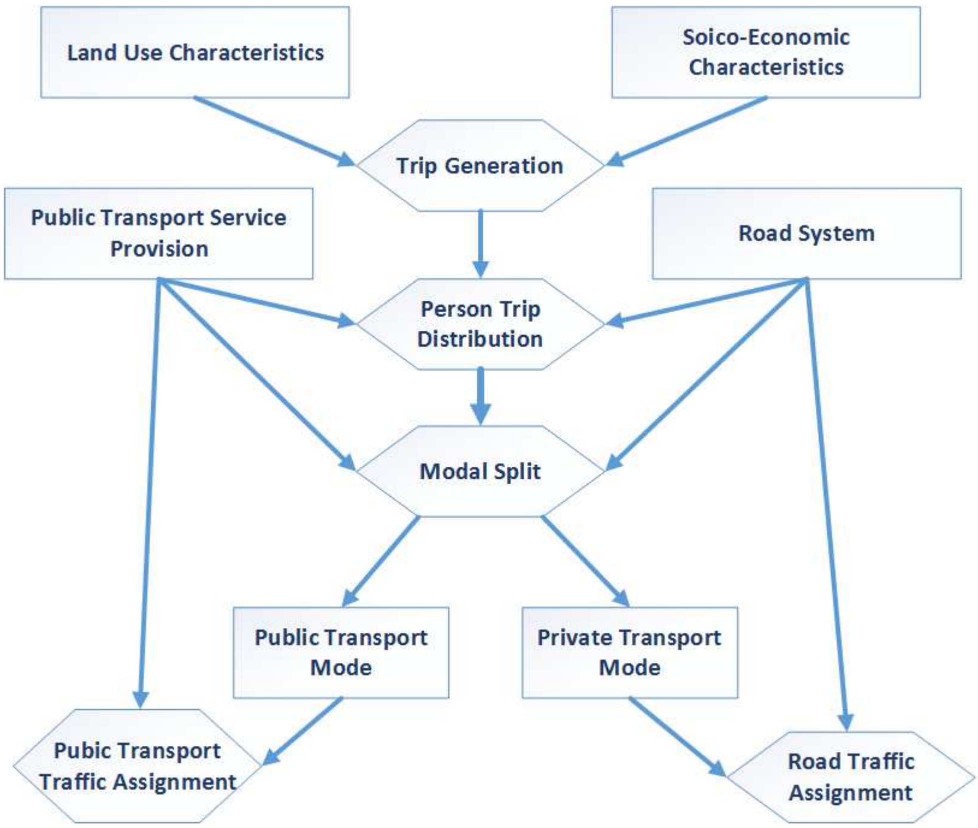


Figure 8 : Sequence of Steps in Travel Volume Forecasting Fertal et al. (1966)

According to Bates (2008), the four steps transport demand modelling has been widely used in the transportation research due to its logical application. These four steps include trip generation, trip distribution, modal split and traffic assignment. Oppenheim (1995) explains that the initial step this model (trip generation) is to estimate numbers of trips in a particular origin and/or destination based on their land use characteristics and socio-economic factors. Two approaches can be applied in trip generation estimation process such as aggregate level of origins or destinations and disaggregate level at household. Linear regression or quasi-linear regression methods are applied to develop the trip generation demand models. In addition, various statistical techniques such as analysis of variance, factor and cluster analysis, contingency tables and discriminant analysis are used to classify the patrons. He also describes that after estimating trip generation is completed, trip distribution can be performed by using the 'synthetic' models or 'gravity' model based on the attributes of travel destinations. As third step in transport demand modelling, total number of origin-destination trips can be split into various alternative travel modes based on the attributes of travel modes such as cost, level of service, socio-economic factors of travellers and their car ownership. He explicates that after the modal split is conducted, the trip assignment can be performed based on the network links and routes between origin and destination. Logic of this four steps modelling is in sequential approach and there is a knowledge gap in transport demand modelling research area to consider the magnitude of correlations among all factors influencing in each step of this traditional transport demand modelling approach. For example, in most of the areas, more public transport service could be provided in highly dense area and this more frequent public transport service provision may encourage more trip generation, not only at the modal split step.

Therefore, this research is aimed to develop a public transport usage predicting model by considering the influence from all observed variables illustrated in Figure 5, such as land use characteristics, socio-economic characteristics, public transport service provision and road systems (measured as road length (in km) per km^2 of suburb area and magnitudes of correlations among these observed variables. This research also takes into account variations in the sequence of steps and correlations among the observed variables. The conceptual model outlining the scope and main purpose of this research is discussed in the following section.

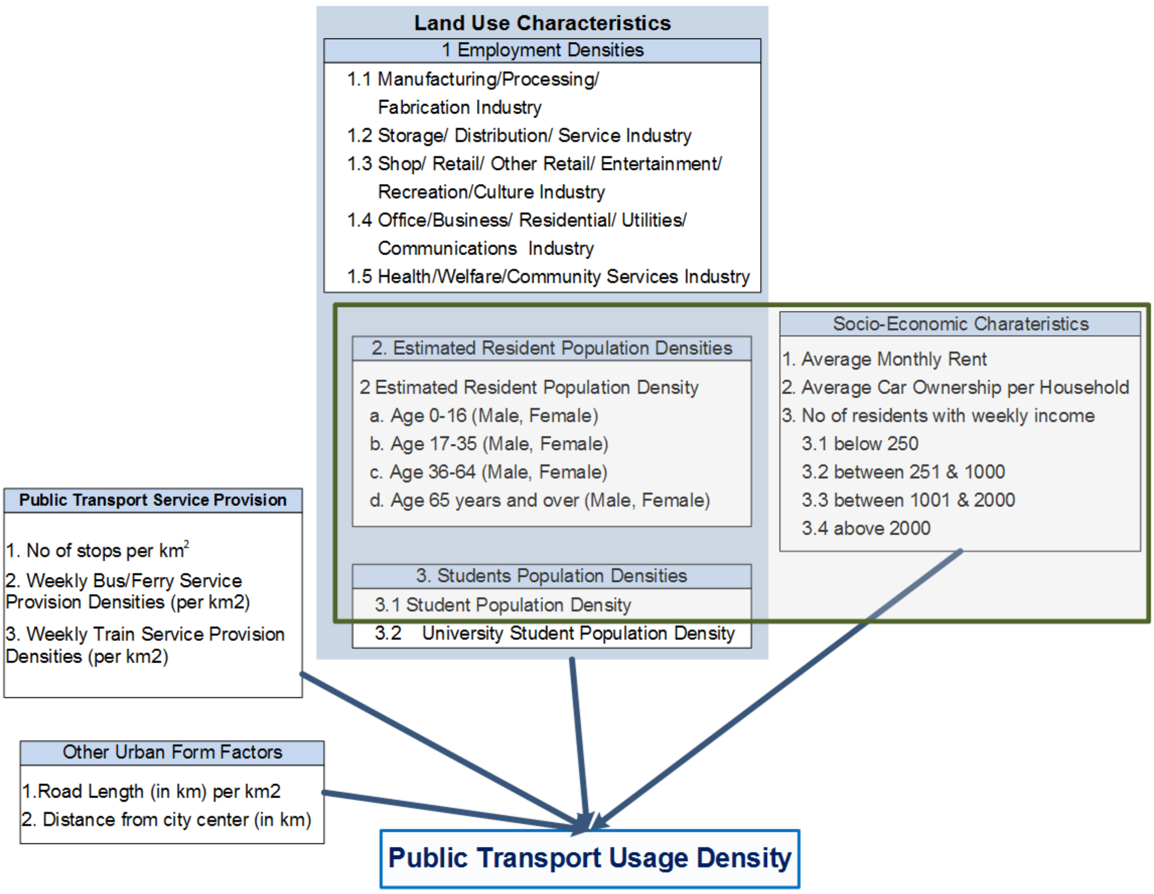


Figure 9: Conceptualised Theoretical Framework

The following changes in sequence of steps are applied in developing the predictive model:

a) The correlations among one of the land use characteristics (student & estimated resident population densities) and their socio-economic characteristics are analysed, and latent variables are extracted from their correlations;
b) The relationships of all observed variables to public transport usage density are examined together to explore the magnitude of their influence.
c) Latent variables are derived from public transport service provision densities during different time segments to examine how service provision for each time segment influences public transport use;
d) In addition to the highways, the freeways, roads, streets, and spaces that encourage private transportation are also considered as urban form factors; finally,
e) The distance from the city's centre is used to capture the nature or location of the areas studied, since Perth is widely dispersed city.

Following Raykov (2008), the following theoretical framework is conceptualised as a guideline contributing to the theoretical and practical understanding of demand planning:

1. Clear definition of the phenomena to be studied:
    a. This research investigates the driving forces behind the spatial variations of public transport usage in Perth.
2. Explanation of the peculiar nature of the problem to be studied and the aims to be fulfilled by this research:
    a. Socio-economic factors such as income, primary household expenditure (rent), average car ownership per household as well as different types of population densities (resident population densities according to different age and gender groups, university student population densities, student densities) and employment densities for various industries will be analysed as driving factors contributing to public transport use. More specifically, it is assumed that university students have a usage rate that differs from other age and gender-specific population groups, such as population under 16 years old. In addition, different estimated resident population densities from four age groups (0-16, 17-35, 36-64 and 65 and over) by gender are taken in account in this research to gain better understanding on the magnitude of contribution by each group towards public transport usage.
    b. Based on a synthesis of previous studies, the public transport service provision factors, such as average number of stops per km^2 of urbanised suburb area, bus/ferry service provision, and train service provision by time of weekdays and weekends will be employed as supply factors which could moderate public transport use. It can be argued that having high population densities would not result in high public transport usage when public transport service provision is not sufficient or attractive enough to influence the high demand converting into actual usage. In addition, the availability of different modes of public transportation (such as bus, ferry and train) in different suburbs could contribute to variations in use rates. Therefore, the provision of service by bus/ferry and trains (by different time of weekdays and weekends in each suburb of Perth will be considered as supply factors.
    c. In addition, land use characteristics factors such as total road length (in km) per km^2 and distance from the city centre will also be included as intervening geographical factors which could moderate the use of the public transport.
3. Identification of the data to be collected, the subjects who are to provide it, and the settings in which it will be gathered:
    a. The main dataset needed for this research tracks public transport usage in Perth. It is also necessary to collect data regarding bus, ferry, and train stop locations and their timetables to aggregate the average stops per km^2 and

amount of service by time of day (weekday and weekend). Transperth, the public transport systems servicing Perth (operated by the Perth Transport Authority) provided these data for the purpose of this study.

b. It is also required to collect data on the socio-economic conditions of the suburbs, as this data is important in examining the linkage between usage-driven factors and public transport usage. Average rent data was acquired via the Real Estate Institute of Western Australia (REIWA). The information on average car ownership per household, estimated resident population by different age groups, and number of residents by different income groups is available through the Australian Bureau of Statistics. Both student population data, which is regularly collected by the Department of Education in Western Australia, and university student population, was accessed via the Business Intelligence Data-warehouse, hosted by the Office of Strategy and Planning at Curtin University. Moreover, the information on employment densities for eleven different industries was obtained through the Department of Planning and Infrastructure and the Western Australian Planning Commission.

c. One of the land use characteristics factors, namely total road length (in km) per km^2 of urbanised suburb area, is generated from the road centreline vector dataset, available through Landgate, the Department of Land Information Authority, in Western Australia. To compute the distance of each suburb from the city centre, Google Earth software was utilised.

d. A detailed explanation of how each of the datasets was collected, and either aggregated or decomposed to apply at the suburb-level is provided in the Data Requirements section of this chapter.

4. Directions for dealing with the research design, instruments and methods to be deployed in the investigation:

   According to Figure 9, a cross-sectional research design is used to examine how socio-economic, land use characteristics, and service provision factors contributed to public transport usage in 2009. The framework also provides a basic schematic template for understanding the relationships among land use characteristics, socio-economic factors, supply factors, and geographical factors, so that an appropriate regression model can be constructed.

5. Guidance on the methods to be employed in reducing and analysing the data:

   There are seventy variables (including sub categories) in the (fifteen) land use characteristics, (six) socio-economic and (two) urban form factors and (forty six) public transport service provision densities variables used in this research. Due to the large number of variables, it is necessary to examine the correlations among them before investigating their influence on public transport usage to minimize multi-collinearity, or bias that can result from their interdependencies. Therefore, factor analysis is done for all correlated variables before putting them into a regression model. The use of factor analysis and multiple regression analysis in this research is addressed and justified in next chapter.

6. Explanation of the nature of the contributions that this research will make to the advancement of knowledge:
    a. This research brings together an elaborate combination of physical planning infrastructure, land use characteristics, socio-economic, public transport service provision, and other variables on a fine-grained geographical scale for a city in order to develop a predictive model of public transport use and explain public transport usage patterns. This has not been done before whereas previous studies have addressed parts of this modelling, the present one covers the whole picture using a rigorously constructed set of original data.
    b. This research is the first investigation of the main determinants of public transport use in Western Australia. It is conducted at a very granular spatial level by performing scrupulous and thorough data extraction, verification, aggregation or disaggregation (as necessary), transformation into common data formats, and building of a comprehensive big data warehouse such most elaborate integration of city transportation infrastructure, land use characteristics, socio-economic factors, urban form attributes and public transport service provision is conducted as first time on very granular geographical scale for any city.
    c. The research advances knowledge of public transport demand on both theoretical and practical levels. Theoretically, the findings from this research expand our understanding of the main determinants of public transport use. As the results will show, the influence of each observed variable on transport use is modified when all others are taken into account. This, in turn, reveals which variables have the largest influence on the outcome, controlling for the others. Further, by applying factor analysis, the research model reduces the number of variables in the data while retaining the richness of each variable's contribution. On the practical side, the predictive model developed here can be used for operational and tactical decision-making and also has implications for transportation policy.

The scope of this study is identified based on the dimensional framework proposed by Florian (1988). As they explain, the transport planning management perspective can be categorised in terms of three levels:

- Strategic level: an aggregate and overall analysis that informs policy decision making and involves long lasing impacts, investments, or improvements, for example to government initiatives (e.g. transportation policy, choosing to support particular mode of travel method, regulation, etc.);
- Tactical level: a narrower perspective focusing on particular issues, resource allocation, and prioritization directed at improving the efficiency or productivity of public transport operations;
- Operational level: the most disaggregated and narrowest perspective, this level addresses short term issues to improve day to day activities, such as the impact of weekly and daily variations in scheduling of public transportation use.

The authors also suggest including the activity location, demand, performance, supply actions, cost minimisation, production and transport system levels as further dimensions in transport planning research (see Figure 10: Two-dimensional Conceptual Framework, Florian et al. (1988).

	Planning Management Perspectives		
	Operational	Tactical	Strategic
Activity Location	✓	✓	✓
Demand	✓	✓	✓
Performance	✓	✓	✓
Supply actions	✓	✓	✓
Cost minimisation			
Production			
Transport system levels			

Figure 10: Two-dimensional Conceptual Framework, Florian et al. (1988)

The areas attended in this study are highlighted in grey in Figure 10. Activity location (at suburban level), demand from various patronage groups, performance of different public modes of transportation, and supply actions (i.e. variations in public transport service provision over numerous time segments) are considered in this research to gain a better understanding of the main determinants of public transport usage in Perth metropolitan suburbs.

From a strategic perspective, by identifying the main determinants of public transport use, the research can indicate the amount of public transportation required to reach Perth’s sustainable transportation target. The findings will also assist decision marking about integrated land use and sustainable transportation planning, especially for public transport services. Further, based on the projected economic and population growth in the Perth metropolitan areas, the demand model developed here can forecast future changes in the need for transportation services. This, in turn, is useful for calculating the benefits of changing usage rates that could be brought by different modes of public transportation.

Referring to Figure 10, the predictive model of public transport usage developed here is based on activity location, demand, performance and supply action. Based on the findings from this research, cost minimisation, production and transport system level should also be included in future research to show how the supply of public transport such as service provision could be improved to meet the rise/fall of demand from changes in land use and the socio-economic characteristics of activity locations.

By taking into account detailed, comprehensive datasets, the predictive model will also aid in tactical decision making for medium-term plans. In estimating the urgent need for public transport service provision, the challenges resulting from changes in the medium-term, such as increases in employment population densities in particular industries, or fluctuation in estimated resident population densities of particular age groups, can be anticipated and addressed.

The research is conducted by constructing a meticulous and granular data-warehouse and performing comprehensive and thorough regression modelling along with applying factor analysis to encapsulate the weights of contributions from all categories within an observed variables. For example, factor analysis is conducted to derive the latent variables among public transport service provision by different travel modes in different time segments, instead of aggregating the total service provision by all travel modes. The factor loadings and scores from these detailed service provision by different travel modes in different time segments are taken into account for predicting the changes in public transport usage. This facilitates the capability for operational decision makings such as how service provision should be increased for bus or ferry or train at a particular time segment (6-9 am on a weekday or 9am -12 noon on a weekday or 3-6pm on Saturday, etc.).

The developed theoretical framework for the study provides the foundations for particular modelling and data handling techniques to be applied. They are explained further in the remaining of the chapter.

3.4 Geographical Scope

Perth is the state capital of Western Australia, Kennewell (2008). This research is conducted to analyse the public transport usage in the suburbs of metropolitan Perth. According to Australian Bureau of Statistics, metropolitan Perth is defined as Perth statistical division consisting of 31 local government areas, namely Peppermint Grove, Nedlands, Cottesloe, Cambridge, Claremont, Subiaco, Mosman Park, Perth, East Fremantle, South Perth, Vincent, Melville, Joondalup, Serpentine, Jarrahdale, Mundaring, Kalamunda, Cockburn, Fremantle, Canning, Stirling, Victoria Park, Bayswater, Swan, Wanneroo, Rockingham, Gosnells, Bassendean, Armadale, Belmont, Kwinana and Mandurah Department of Local Government and Communities-Government of Western Australia (n.d)_ENREF_66. The metropolitan Perth area stretches from Two Rocks in the north to Singleton in the south and expanded in east to the Lakes. This study examines 292 out of 318 suburbs (91.82%) in the Perth Metropolitan Area where Transperth provides the public transport services. In this thesis, Perth refers to the state capital of Western Australia and Perth Central Business District (CBD) refers to the area consist of four suburbs such as West Perth, Perth suburb, Northbridge and East Perth, Landgate (2013)_ENREF_124.

3.5 Spatial Dimensions of Research Approach

The spatial dimensions of this study are quite important. Aggregate measures along with correlations and regression analyses are used to compare relatively large spatial areas such as cities, neighbourhoods or zones. These aggregated analyses enable us to examine the possible effectiveness of land-use policies in reducing automobile dependence. Nevertheless, such analyses compromise the richness of data and the ability to explore underlying factors and rationales, which together influence individual travel decisions. More micro-level research tries to address this by using variance, correlations, or regression analysis to examine relationships between transportation use and the socio-economic, land use, and travel characteristics of individuals or the household. However, such disaggregated analyses generally only account for a small proportion of the observed total variance in travel behaviour. The present research overcomes the shortcomings of both designs by carefully selecting the degree of spatial aggregation so as to perform an aggregated analysis while minimising data loss or reduction.

Datasets for socio-economic variables and some of the land use characteristics are collected from the Australian Bureau of Statistics. The 'Census Collection District' (CCD) is the most granular spatial unit in the Australian Standard Geographical Classification used for the Census. These CCDs are amalgamated to form "Suburbs", see Figure 11: Census Geographic Areas Structure Charts,

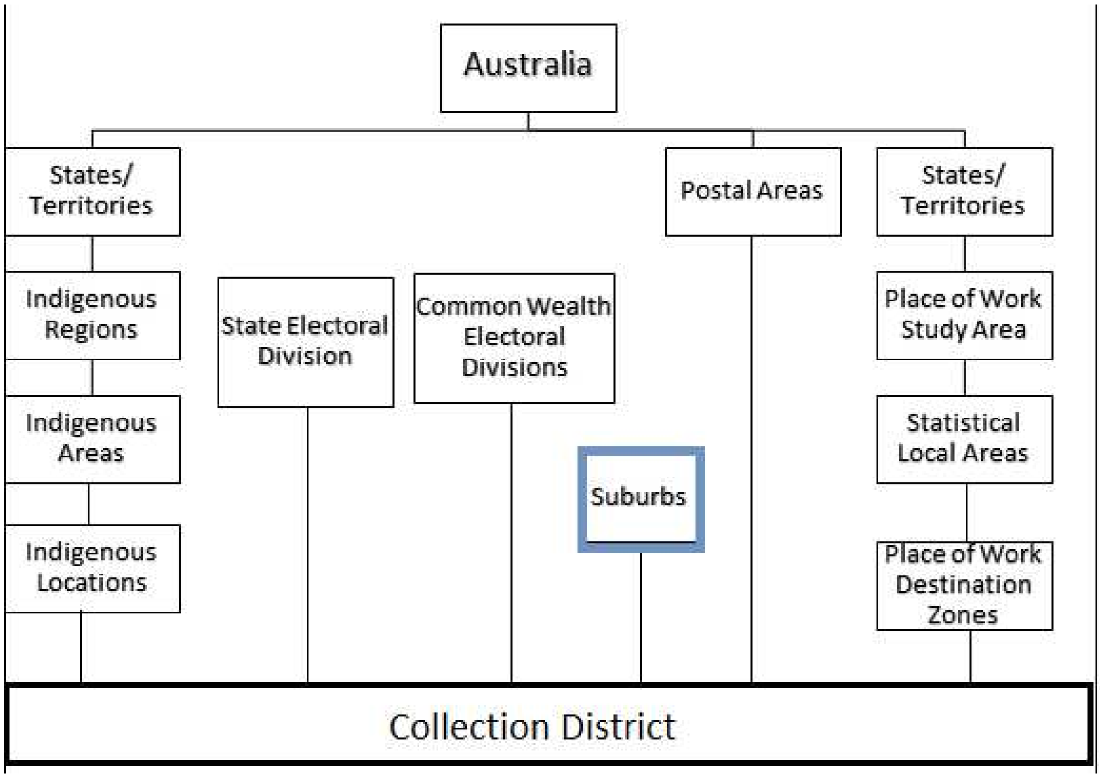


Figure 11: Census Geographic Areas Structure Charts, Australian Bureau of Statistics (2007)

As shown in Figure 11, these suburbs are the third most granular spatial unit after Census Collection District (CCD), and strike the best balance for ABS data between aggregation and richness. In this study, therefore, suburbs are used as the spatial unit of analysis. Census Collection Districts are the second smallest geographic area defined in Australian Standard Geographical Classification (ASGC). According to ABS, the Collection Districts (CD) have been used in the Census of Population and Housing as the smallest unit for data collection and processing. The suburb-level is chosen for the following additional reasons:

1. Even though some data, such as public transport usage, service provision, employment densities, and land use characteristics can be analysed at the CCD level, the most granular level for other data, including estimated resident population density by age and gender, number of residents in different income groups, average rent and average car ownership per household, is the suburb level.
2. Suburb level analyses will most effectively facilitate tactical and strategic decision-making for policy makers. This level of spatial analysis at suburb level has been used in public transport research in Australia. One recent study that deploys suburb level spatial analysis can be found in the development of the public transport demand model in South East Queensland using a Melbourne and public transport user survey. Public Transport Victoria (August 2012) developed a public transport demand model at the suburb level in Melbourne by considering the role of population, employment and

education growth in an elasticity-based demand forecast model. Additionally, TransLink (n.d.) also conducted a public transport user survey at the suburb level in South East Queensland over a few weeks in May 2010 to gain a better understanding of the main purposes for public transport usage, passenger type profiles by location/route and mode splits.

Overall, spatial analysis at the suburb level provides the most meaningful and rich information both statistically and practically. Therefore, all suburbs to which Transperth provides public transport services will be included in this research. These suburbs and their corresponding public transport zoning are illustrated in Figure 12.

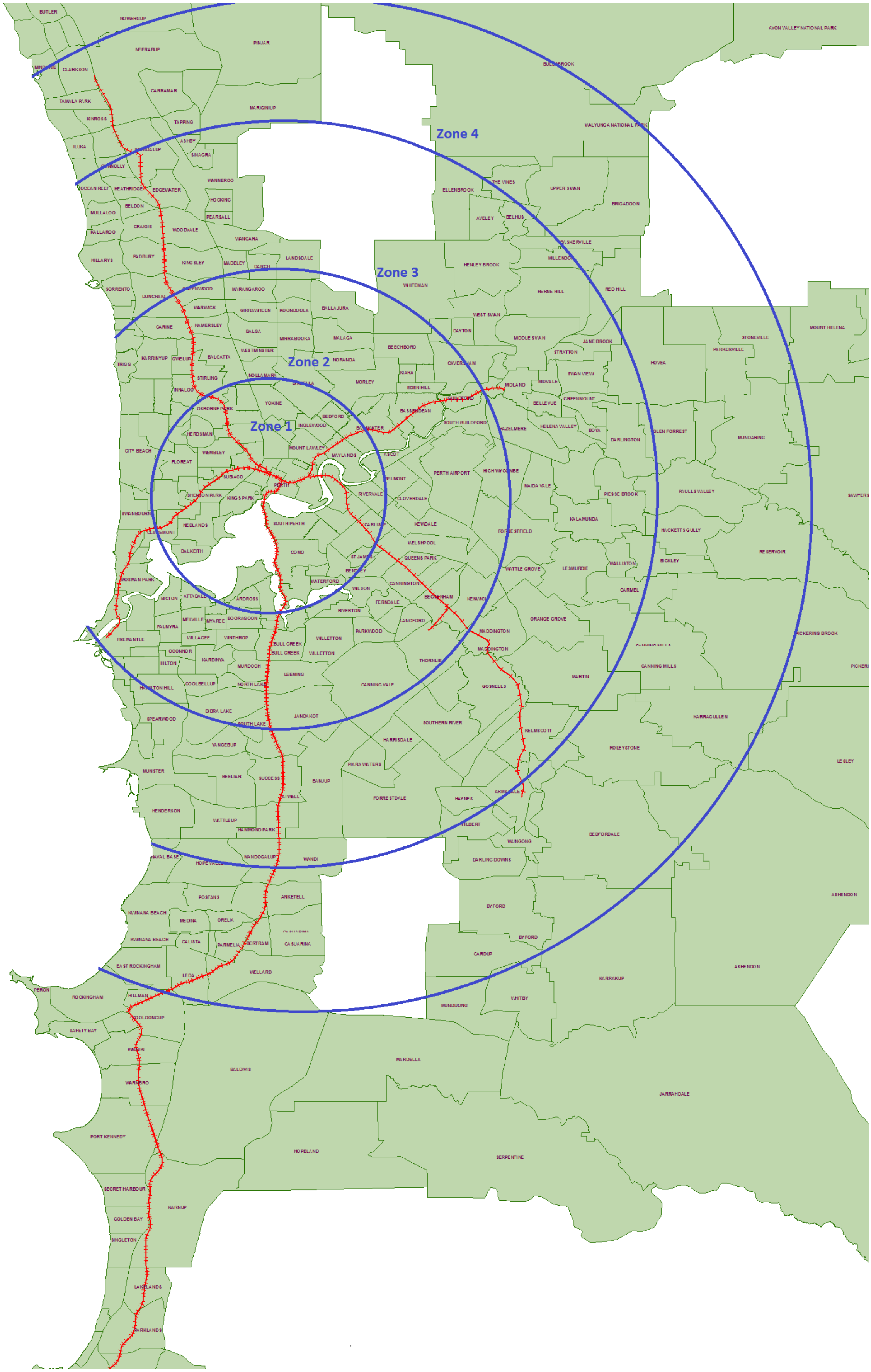


Figure 12: Transperth Zone Map, TransPerth (n.d.)

3.6 Temporal Dimensions of Research Approach

The temporal dimension for any research is very important. This study is based on cross-sectional analyses. According to Kelsey (1986), cross-sectional research uses observations at a single point in time, or over short time period across different parameters. Bryman (2004) points out that the cross-sectional research approach is mainly employed when the researcher is interested in the variation of a particular phenomenon, as opposed to focusing in detail on one or a few particular cases. Cross-sectional analysis is applied in this research to investigate the variation of aggregated public transport usage among Perth metropolitan suburbs where Transperth provided the services in 2009, in contrast with examining the disaggregated individual travel mode choices as a result of changes in socio-economic characteristics or public transport service provision over a time period. In line with this, the study examines 292 out of 318 suburbs (91.82%) in the Perth Metropolitan Area where Transperth provides public transport services and uses them to analyse variations in public transport usage. Bryman (2004) also suggests that cross-sectional design should only be used to examine relationships between variables, not to observe the direction of causal influence. In this study, there is no causal analysis. As previously described, the primary aim is to describe the relationships between public transport usage and land use characteristics, socio-economic, urban form factors and public transport service provision factors.

Hence, this study presents a cross-sectional snapshot of public transport usage in Perth in 2009. The year 2009 was chosen as the focus of this thesis as it coincided with the start of the PhD research. The Global Financial Crisis, the most precarious economic crisis since the Great Depression of 1930s, started in 2008. There had also been many significant bailouts of global financial institutions in 2009 (Havemann, n.d) and other severe economic repercussions across the globe. This however was not the case for Western Australia. According to the Department of Treasury and Finance (n.d), the annual change in real gross domestic product at the state level during 2008-2009 remained quite stable for Western Australia as shown in Figure 13. Hence, 2009 was a typical year for Perth and Western Australia without any unusual or extreme events. In other words, the year 2009 represents a typical year for public transport use in the Perth metropolitan areas.

.

Figure 13: Real Gross State/Domestic Product (Annual Change: 1996-97 to 2009-10)

Source of Data from Department of Treasury and Finance (n.d, , pg.1)

3.7 Validation of Research Method

Holmgren (2007) argues that it is important not to exclude relevant explanatory variables from any model because this could result in biased estimates when there is correlation between the excluded explanatory variables and predictors. He also urges researchers to determine whether there is a change in the statistical significance of models if a particular variable is included or excluded. Therefore, to gain a better understanding of which determinants are most important in explaining public transport usage, a number of land use characteristics, socio-economic, urban form factors and service provisions factors are taken into account in this research to develop a usage predictive model by applying standard regression after conducting data verification and performing data transformation to satisfy the assumptions of the standard regression method. To validate the robustness of this developed predictive model, the following other regression methods are used:

- A stepwise method is applied to examine which explanatory variables can explain most of the variance in public transportation use, and whether adding more explanatory variables to the model improves its predictive power.
- A robust method is used to develop another predictive model by placing the weight on each case depending on their individual leverage (the measurement of how much the values of an independent variable deviate from its mean). The predictive model derived from the standard regression method (after performing the most suitable data transformation) is simpler to use for policy development. However, the model derived from

the robust regression method is used to validate the magnitudes of influences from individual observed variables on public transport usage in Perth metropolitan suburbs.

3.8 Introduction to Data Requirements

This study brings together a wide-ranging set of data to address the research questions put forward in Chapter 1. As shown in the literature review (Chapter 2), there have been a very large number of studies over the years from a wide range of disciplines that have investigated relationships between public transport use and other factors. However, this study uses much more detailed data on many different possible explanatory factors for public transport use, which are brought together in a systematic way to investigate variations in use rates.

This section gives a broad overview of the data collected and analysed for this research. The individual variables are classified under specific headings. One variable is used to measure public transport use, two variable groups characterise public transport service provision, four different types of density variables are considered under the heading of land use characteristics, three variable groups are used to measure socio-economic factors and road length and distance from the city centre are used as measurement for urban form factors. Subsequent sections in this chapter provide specific details about the origin and nature of these data.

The **Exploratory variable** (outcome variable) used in this research is “public transport usage density”. The author gained access to a unique set of data held by the Public Transport Authority in Perth, collected through its SmartRider fare collection system. The specific measure for public transport usage adopted in this research was obtained by aggregating the total number of journeys from their origin suburbs. A detailed explanation of how SmartRider transactional data was aggregated is provided in section 3.9 (Origins of Multiple Data Sources Integrated in Datawarehouse).

The explanatory variables used in this study are public transport service provision factors, land use characteristics, and socio-economic factors.

For public transport service provision, two key variable groups were developed:

1. Average stops per km in each suburb – Asensio (2000), Bass (2011), Holmgren (2013), Polat (2012) and
2. The total frequency of service during eight 3 hour periods on weekdays and weekends (covering total service provided in a whole week for each suburb in Perth) resulting in forty six public service provision variables – White (2004), Barton-Aschman Associates (1981),

Dodson (2007), Curtis (2011b), Holtzclaw (1994), Mees (2000a), Polat (2012), Webster (1982).

For land use characteristics, the following variables have been used in previous public transport research and will also be included in this study:

3. Estimated resident population density by age and gender in urbanised suburb area (per km^2) – Badoe (2000), White (2004), Cervero (1997), Cervero (1991), Cervero (1997), Perkins (2006), Mackett (1990), Souche (2010), Susilo (2007), White (2009);
4. School student (up to year 12) population density gender in urbanised suburb area (per km^2) – Pitombo (2011), Tolley (1996);
5. University student population density gender in urbanised suburb area (per km^2) – Curtis (2004a), Curtis (2004b), Shannon (2006);
6. Employment density gender in urbanised suburb area (per km^2) – Cervero (1988), Cervero (1991), Cervero (2002b), Ewing (1994), Hendrickson (1986), Schimek (1996).

A wide range of data were collected from a variety of sources for public transport users' socio-economic attributes based on the literature review as:

7. Number of residents whose weekly income falls into four different groups in each suburb – Bresson (2003), Pitombo (2011), Dargay (2002), White (2004), Thompson (2012), Holmgren (2013);
8. Average monthly rent of each suburb – Sipe (2006), Dodson (2007), Liao (2014), Martinez (2008);
9. Average car ownership per household in each suburb – McFadden (1974), Kenworthy (1989), Kenworthy (1999c), Mokhtarian (2002), Cullinane (2003), Paulley (2006), Bresson (2004), White (2004), Greenwald (2006).

Referring to the literature review, the following variables are collected as urban form factors:

10. Average street length (in km) gender in urbanised suburb area (per km^2) – Mogridge (1990), Hansen (1993), Kenworthy (1999a), Cervero (2001c), Cervero (2002a), Cervero (2003), Zeibots (2005) and
11. Distance from the city centre – Stead (2001b), Sohn (2005), Weber (2003), Cervero (2002b), Boarnet (2001; Boarnet (1998; Riekko (2005).

Multiple datasets were collected from different sources including Transperth, the public transport operating agency, the Australian Bureau Statistics, the Western Australian Department of Planning and Infrastructure, Landgate, and other organisations. The details of each dataset are described in the following sections.

3.9 Origins of Multiple Data Sources Integrated in Datawarehouse

In addition to the aforementioned data, the urbanised area for each suburb data was used to calculate density variables. Details of how these datasets were collected and calculated are now explained.

3.9.1 Suburban and Urbanised Areas

To calculate the urbanised area of each suburb in metropolitan Perth, two datasets were collected: (1) the area of the administrative boundaries and (2) the non-urbanised area of each suburb. For these calculations, the areas considered non-urbanised were agricultural lands, meadows, pastures, regional scale parks, urban forest, forest, wasteland, and water surfaces. The administrative boundary area was calculated using the dataset provided by Landgate (the Western Australia Land Information Authority). These areas were linked with the States' cadatstral and tenure data boundaries in a database called "Spatial Cadatastral Database", which is maintained by Landgate and carries the same point and line accuracies with the suburb boundaries used by Australia Bureau Statistics. The data file for these boundary areas of each suburb in metropolitan Perth was provided by Landgate in ESPRI (one of the most commonly used Geographical Information Systems) shape file. The administrative boundary area of each suburb, as measured in square kilometres, was computed with the ArcGIS application.

Second, to calculate non-urbanized areas, the ESPRI shape file was converted into kml file format using the "ogr2ogr" software. As ESPRI shape file cannot be imported into Google Earth application, this "ogr2ogr" application is used to convert this ESPRI shape file into kml file. The main reason for using Google Earth is that it was the only available source for accessing historical imagery of urbanised areas in 2009. The administrative boundary area of suburbs map file was then loaded into Google Earth to map the administrative area of each suburb first. After loading the administrative boundary area map into Google Earth, the non-urbanised area within each suburb was mapped by using the 2009 historical land-use imagery and the total non-urbanised area in each suburb (within administrative boundary area) was calculated.

Third, the non-urbanised areas identified in Google Earth were validated against the non-urbanised areas defined in the Perth Metropolitan Region Scheme (MRS) of the Department of Planning (Western Australia) (see http://www.planning.wa.gov.au/5223.asp). This publicly available Perth Metropolitan Region Scheme map, downloaded on 11 Feb 2011, identifies the urbanised and non-urbanised areas amended up until 15 January 2010. This validation process was conducted to ensure that all non-urbanised areas identified by the Department of Planning by 15 January 2010 were included in the non-urbanised area map.

Finally, the urbanised areas of the Perth metropolitan suburbs were as calculated by subtracting the non-urbanised area from the total administrative boundary area. The resulting urbanised area (in m^2) computations were then used to calculate the density variables.

3.9.2 Public Transport Usage

Transperth is the operational and brand name used for the public transport system provided in Perth metropolitan areas by the Public Transport Authority. It is comprised of a bus network, train system (five train lines) and a ferry service across the Swan River from South Perth to Perth CBD. Since 2000/2001, the growth of patronages who use SmartRider or cash ticket for initial boarding has been continued by 1.6 percent and total boarding by 2.2 percent, according to Pubic Transport Authority (n.d-b). It is also stated in the Public Transport Authority 2009/2010 annual report that the overall customer satisfaction with Transperth for all modes reached 85.8 percent in 2010 with the steady growth from 85.4 percent in 2009.

SmartRider is the electronic ticketing systems on Transperth services, Transperth (n.d-b). This is the type of ticketing that regular patron uses. A cash payment is also available from any station or bus driver. However, the cash ticket is used less. According to the Public Transport Authority 2009/2010 annual report, SmartRider covers 70% of all public transport use in 2009. Transperth (n.d-b) states that the types of SmartRider are:

- Standard SmartRider: available for any passenger,
- Concession SmartRider: available for passengers who are holding any concession entitlements from Centerlink and Interstate pensioners,
- Seniors SmartRider: available for Western Australia seniors,
- Pensioner SmartRider: available for passengers who are receiving aged pension or disability support pension,
- Veterans SmartRider: available for passengers who are a veteran, a war widow or war widower (not receiving income support supplement ISS) or a DP (Disability pension 70-100%), Extreme Disablement Adjustment (EDA), International Students who are fully funded by specified Australia Government Scholarship (INT), Pensions of War (POW), Totally and Permanently Incapacitated (TPI) or Totally and Temporarily Incapacitated (TTI) pensioner

- Student SmartRider: available for primary and secondary school students (up to year 12)
- Tertiary SmartRider: available for passengers who are currently enrolled as full time students at a Western Australia university or TAFE college or interstate primary or secondary school student or a secondary school students who are 19 or older.

This study is mainly based on passenger trip records from the SmartRider System, provided by Transperth for Jan-Dec 2009. The nature of these SmartRider data is important to understand. The SmartRider card is a form of pre-purchased, stored value ticketing whereby public transport users “tag on” and “tag off” when they enter and leave any public transport vehicle, with the system automatically determining how much money is due for the trip and deducting it from the balance on the card.

The public transport usage data files were extracted from the SmartRider tag on and tag off tables. All of these data files were provided to the author in csv (comma-separated values) file format and consisted of data fields that are explained later in this section. It is necessary to understand these data, as the information is critical to the way the public transport data for Perth are extracted and processed. The unique nature of the dataset used in this research makes the work undertaken in the thesis an original contribution to the field. In total, the database contained 141.3 million trip segments on which there were nineteen useable trip characteristics, giving a matrix of 2.7 billion data entries. The data therefore represent millions of public transport trips across the Perth metropolitan region for an entire year, accompanied by additional socio-economic, land use and public transport service timetable data for research investigations. It is estimated that the data from the SmartRider public transport usage represent approximately 66.1% of the total (2008/2009) annual usage, Pubic Transport Authority (n.d-a) and 69.9% of the 2009/2010 total annual usage, Pubic Transport Authority (n.d-b) of public transportation in Perth, and therefore constitute an extremely valuable resource for research that has not been exploited before.

According to Perth’s Public Transport Authority’s 2009/2010 annual report, SmartRider covers 70% of all public transport use in 2009. Not all the SmartRider data however was usable because of users or the system not recording the complete information about the start (tag-on) and end (tag-off) of the journey. The SmartRider transactions which were missing tag-on or tag-off details are excluded from this analysis Therefore, the SmartRider transactions used in this study account for 66.1% of the total (2008/20098) annual usage of public transport in Perth.

Even though only 2009 travel data were used to develop a predictive model of public transport use, the travel data (Jul-Dec 2008) were also imported to verify the conversion process by which trip segment transactions were transformed into whole trip records. Transperth uses the financial year calendar as its reporting calendar, and this begins in July instead of January. Therefore, the total trips generated during July 2008 and June 2009 were aggregated and checked against the total number of trips reported by Transperth for this 2008/2009 financial year[7].

This study examines an extensive amount of information on all suburbs in Perth for one period of time (2009) for the following reasons:

- To analyse the temporal and spatial variations in the public transport usage patterns in metropolitan Perth, Western Australia;
- To examine the relationship between public transport usage and land use characteristics, as well as their correlations with socio-economic attributes and the influence of moderating factors, namely public transport service provision;
- To generate a public transport usage function for Perth based on these factors; and
- To develop a comprehensive and rigorous regression model to predict the changes in public transport usage on the basis of changes in its determinants.

Holmgren (2007) conducted a meta-analysis on elasticity estimates from previous public transport demand studies. He suggests that if population is not included among the exploratory variables, then variables used should be in per capita form. In this research, public transport usage per capita is not used as a measured variable because all employment, estimated resident population, student (up to year 12) and university student population densities are included as predictor variables in the regression model and all these variables are presented in relative terms.

Before explaining the data aggregation process, the definitions of the measurements used for public transport usage in this research are presented first:

- Travel segment is an individual travel transaction which initialises when a patronage gets on board on any public transport mode by tagging on his/her SmartRider card and completes when that patronage alights from it by tagging off his/her smart card.
- A journey can be defined as a movement between origin and destination. Journey is a whole travel record which comprised of one or more than one travel segment

[7] The financial year in Australia is from 1 July to 30 June. Hence the 2009 calendar data are across two financial years. While the Public Transport Authority reports public transport usage in financial years, this study uses the public transport usage in 2009 calendar year to be temporally compatible with other datasets such as socio-economic and land use characteristics datasets from the Australian Bureau of Statistics.

by taking into account all transfers involved within the same trip to move between origin and destination[8].

- Transfer is the act of changing the different public transport modes or the different services of the same mode.

According to White (2009), the various ways of measuring the use of public transport are

- the absolute number of trips or boarding collected from ticketing systems,
- distance travelled (passenger-km) calculated by multiplying the number of trips by an average length estimates collected from travel surveys, either on-vehicle or household questionnaire or travel diary,
- user expenditure on public transport per kilometre to measure the high expenditure per kilometre for high quality public transport services,
- trip rates per head of population as an indicator to compare the usage per capita among different areas and
- market share to compare the usages among different public/private transport mode.

In this research, the absolute number of journeys comprised of one or more than one boarding (trip segment) is used to measure the public transport usage in Perth metropolitan suburbs. Data importing, cleaning and aggregating process for public transport usage is explained in the following section.

3.9.2.1 Data Importing, Cleaning and Aggregating Process for Public Transport Usage

The total number of public transport journeys originating in each suburb was aggregated first, after which the aggregated total journeys per km^2 were used to compute a "public transport usage density" outcome variable. Before aggregating the total number of journeys originating in each suburb, all trips (travel segments) were converted into journeys. Each journey consists of one or more travel segments. For example, when a Curtin University student travels from Curtin University to Perth City centre,

- He/she may be taking the 72/75 Bus directly to the city as one journey consisting of one travel segment, or

[8] Trip and travel segment are used as interchangeable terms in this thesis.

- He/she may be taking the 100/101 Bus to Canning Bridge train station (one travel segment) and then the train to the city (another travel segment), which would be one journey comprising of two travel segments.

Each travel transaction (travel segment or trip) was converted into a whole journey based on the reference numbers of the one or more of the travel segments that contribute/s to the whole journey. After transforming travel segments into whole journey data, total monthly journeys (based on the origin and annual usage data for 2009) were aggregated for each suburb.

The data warehouse for this research was built using Microsoft SQL server, 2008[9]. There were many issues this data warehouse, such as the maximum file size for data imports and the complexity of the data manipulation tasks. The maximum number of rows (655356 rows) that could be imported at a time was 655,356. Another issue encountered in data loading process into datawarehouse was data file format incompatibility. Public transport usage data was provided in Unicode data file format by Public Transport Authority. And SQL was incompatible with Unicode data files, which made the importation process tedious.

The following steps were taken during the data import process:

- Converting the data from Unicode to ANSII files by using the Unix application in Windows (Cygwin)
- Splitting the data files by using Cygwin, so that each file contains a maximum number of rows less than 655,000. The importing could then be performed successfully using SQL's BULK INSERT function.

After preparing and importing the files, the following trip characteristics were imported in the data warehouse:

a) CardId: An internal database of reference IDs identifies individual cards/users.
b) OnTran: This field is used to identify the validity of the transaction and consists of two values: 40 = normal tag on, and 42 = synthetic tag on. "Synthetic tag on" occurs either at controlled stations (when someone gets caught by inspection for not tagging on when he/she boards on public transport service), or when someone tries to tag off without tagging on; the software usually identifies the error. Therefore, all transactions in this column with a value of 42 represent

[9] Microsoft SQL server is the relational database server management application developed by using Structured Query Language (SQL programming language).

corrupt/invalid transactions, and these were discarded before converting the valid transactions into trips.

c) Jrny: A unique ID is assigned to this field for each trip made by individual patron on a particular day. As mentioned before, each trip can be made of one or more segments or legs. Therefore, the following onRef field is used to distinguish each segment or trip during the same journey.

d) OnRef: This transaction reference for boarding is used to identify each travel segment in the whole journey. The total number of segments during the same journey can be aggregated by using the Jrny and OnRef fields together. Also, based on this transaction reference, the segments in each trip can be linked together in their sequential order to convert the transaction segments into trip records.

e) OnDate: This field records the date and time of a patron's "tag on" when he/she boards any public transport service.

f) OffDate: This field records the date and time of a patron's "tag off" when he/she exits from any public transport vehicle.

   After linking all travel segment transactions from the same trip, the above two fields (OnDate and OffDate) can be used to identify the first boarding and last exiting date and time for the whole trip. Among all travel segment transactions for the same journey, the earliest boarding time (OnDate) with the smallest OnRef number is identified as the first boarding of the whole journey. The latest exit time (OffDate) with the largest OnRef number is used to identify the last exit for the whole journey.

g) OnType: This field is comprised of two valid values: 0 = initial boarding and 1 = transfer, and it is useful for reconstructing whole trips, rather than just individual legs.

   The Jrny, OnRef, OnDate, OffDate and OnType fields are mainly used to establish the sequence of segments/legs and to transform these segment transactions into journey records.

h) OnMode: This field indicates the travel method as 0 = Bus or 1 = Train or 2 = Ferry.

i) OnZone: This field identifies the zone in which the user boarded[10].

j) OnLocation: When a patron tags on for train stops, the stop number (ID) of train station is recorded in this field.

k) OnLandmark: When a patron tags on for bus or ferry stops, the stop number (ID) of bus or ferry stop is captured in this field.

[10] There are 8 zones which are circular bands with eight or ten width, from Perth city center. Fares are charged based on the number of zones travelled and time limit. As time limit, two hours is allowed for a journey of one to four zones and three hours is allowed for a journey of five or more zones. Transperth (n.d-a)

l) OffTran: The field indicates the validity of the patron's exiting with two valid values: 41 = normal tag off, and 43 = synthetic tag off. "Synthetic tag off": occurs when a patron leaves the public transport vehicle without tagging off, and it usually identifies this as an error. Therefore, all transactions in this column with a value of 43 are corrupt/invalid and were deleted before transforming the valid segment transactions into journey records and loading them into the data-warehouse.
m) OffZone: This field identifies the zone where the passenger exits.
n) OffLocation: This field identifies the number of the train station where the passenger alights.
o) OffLandmark: This field identifies the 5-digit bus or ferry stop number where the passenger alights.
p) FareType: There are two fare types available: 0 = normal, 1 = default fare.
   Default fare is charged if a patron does not tag on at the start of his/her trip but tags off at the end, or does tag on when boarding but does not tag off when exiting. The default fare for buses is the cash fare of the longest trip a patron could have taken on that service. But for trains, it is the cash fare for a seven-zone trip. The default fare for ferries is the amount to be charged for the trip between Barrack Street and Mends Street (TransPerth, n.d). The journeys with default fare are included in the public transport usage aggregation because the public transport usage in this research is measuared by aggregating the total boardings from origin suburbs and these journeys have valid boarding stop numbers.
q) Origin: OnLocation or OnLandmark of the first boarding in a journey. This field is useful in determining the true start and destination points for a multi-leg journey and verifying that the origins of the converted journeys are correct.
r) TransferFrom: This is the mode of travel for a previous leg where -1 = initial boarding, 0 = bus, 1 = train, 2 = ferry.
s) Token: The concession types for the fare. This field is used to identify the type of passenger, whether they are a standard user, senior concession, healthcare, Public Transport Authority (PTA) concession, PTA free pass, PTA rail pass, student travel permit (Tertiary), student travel permit (up to Year 12), veterans concession, or veterans free.

In order to generate spatially resolved data and arrive at an accurate picture of public transportation use in Perth, the researcher carefully processed this data file. The following procedure was used to generate aggregated usage data for each suburb:

1. All travel data files were imported to a "Travel Temporary" table. Altogether, 141,302,390 travel segment transactions were imported.
2. All bus stop location data files were imported. Details of the bus stop data files are explained in section 3.9.3.1 (Average Stops per km^2).These data files include all bus stop status details necessary to identify a stop's status (active or

discontinued). The status field is used to verify that all of the bus/ferry stops and train stations considered in this research were active in 2009.

3. All imported travel segment transactions were transferred to the respective monthly segment table for 2009. The whole imported dataset was split into monthly tables, making data verification easier and removing the invalid/error transactions from the dataset.
4. All invalid travel segment transactions—those with synthetic tag on and tag off values in the OnTran and Off Tran fields—were discarded during the data cleaning process.
5. Based on the Jrny and OnRef values for each travel leg transaction, all travel leg transactions for each trip were joined together and converted into single trip records containing origin and destination points. The 141,302,390 travel segment transactions then converted into 103,141,779 trip records (suggesting an average transfer rate of 1.37).
6. Origin and destination suburbs were added to each trip transaction using the suburb locations of the stop numbers in the bus/train stops table.
7. Public transport usage was aggregated to the suburb-level based on the origin suburb (i.e. where the patrons began their journeys).
8. Public transport usage density (public transport journeys per km^2) was then calculated based on the urbanised area of each suburb.
9. The Public Transport Authority (PTA) stated that Transperth fee-paying boardings in 2008/09 (recall that the PTA uses financial years, not calendar years, for these total data) reached 76.467 million (Public Transport Authority, 2009) and that SmartRider boarding accounted for an average of 66.1 per cent, or 50,544,687, of the fare-paying boarding for the system as a whole (Public Transport Authority, 2009). This figure can be used to verify that the conversion process from travel leg to travel trip record was correct. After the trip segment transactions were converted into trip records, the data warehouse contained a total of 49,281,446 travel trip records for the period from July 2008 to June 2009. Comparing the variation in July 2008-June 2009 dataset namely 49.28 million journeys with the reported total trips by PTA, namely 50.54 million journeys, there was only 2%, which is insignificant for this large dataset. This comparison verifies that the data conversion process was valid.

3.9.2.2 Errors in Public Transport Usage Data Collection

Like any other information system, errors must be identified and corrected (if possible) before proceeding further with aggregation. The main causes of possible errors in the system are as follows:

- Computer and database-related errors (rare);
- Errors due to mishandling of the Automated Fare Collection (AFC) system by drivers and operators;
- Card-reading and –validating errors, due to on-board GPS reader equipment problems;
- Errors due to the desynchronizing of planned service data and operating service data as a result of emergency detour on a route.

The following data verification processes were conducted for each travel transaction to check any potential system errors:

1. Checking that each passenger's alighting tag off date and time is later than their boarding tag on date and time. The researcher identified some system errors in public transport usage data for March-April 2008. For many passengers, the alighting tag off date and time were earlier than the boarding tag on date and time. Hence, the OnRef number for these transactions was not recorded properly, and converting the travel segments/trips into journey could not be performed correctly. Therefore, the complete datasets for 2008 were discarded and the whole data collection was conducted again for 2009. Only data for the second half of 2008 is used to validate the data processing for converting trips into journeys. The 2009 dataset didn't have such serious problems.
2. Discarding corrupted or invalid transactions with a synthetic tag on, which either occurs at inspectors' stations or when someone tries to tag off without tagging on.
3. Checking the boarding travel method against OnLocation and OnLandmark values. If a passenger's initial travel method is train, the boarding location must be recorded as "train station" in the OnLandmark field. If a passenger's on-boarding travel method is bus, the on-boarding location must be recorded as "bus stop" in the OnLcation column.
4. Checking the on-boarding travel method against the alighting travel method to identify errors caused by desynchronizing planned service data and operation service data, which could result from an emergency detour on a route.
5. Discarding corrupted or invalid transactions with a synthetic tag off, which occurs when a passenger alights the public transport service without tagging off. These travel transactions are invalid and cannot be used in converting travel segments into journeys, as it is impossible to determine whether there are any subsequent travel segments.

Following all these data checks and integration, the 2009 public transport usage information was cleaned and organised in the data warehouse prepared to be used for the development of the model.

3.9.3 Public Transport Service Provisions

The PTA provided a variety of files from which it was possible to calculate the number of different types of public transport stops (bus, ferry and train) and also the number of departures made by different modes of public transport departing from each stop. The “public transport service timetable” data file contains extensive data identifying the stopping date/time of bus/ferry/train vehicles at every single public transport stop in Perth metropolitan areas. The researcher extracted this timetable dataset on 22 June 2009. The data fields in the timetable are as follows:

a. STOPNUMBER: A unique 5-digit number identifying bus or ferry stops or train stations.
b. ROUTE NO: This field identifies the public transport service routes. It is an important field in the data file because many different buses or trains stop at a particular bus-stop or train-station. Hence, this field is used to distinguish multiple routes at a given stopping point.
c. ROUTE TYPE: This attribute differentiates whether the particular route is standard (everyday regular route), or used by school or seasonal service for occasions such as ANZAC day.
d. ROUTE STATUS: This determines whether a route is currently active or discontinued.
e. DATE1: This is the date a particular route begins actively running.
f. DATE2: This is the date a particular route is discontinued.
g. STOPPING TIME: The time when a particular public transport vehicle stops at a given stop or station.

This data file was imported into the data warehouse as **Timetable table.** The DATE1, DATE2 and ROUTE STATUS fields can be used to identify all public transport service routes that were actively running on a particular day. The details of how these fields are used to calculate the total public transport service provision are explained in section3.9.3.2 (Total Frequencies as Public Transport Service Provision Densities).

The three main data files needed to calculate public transport service provision are bus/ferry stop numbers and locations file, train stop numbers and locations file and timetable data file.

<u>Bus/ferry stop numbers and locations data file</u>: This contains geographic coordinates from which the exact location of the stop can be found and allocated to a suburb. There are only two ferry stops in Perth, but thousands of bus stops. The files contain the following information:

a. STOPNUMBER: A 5-digit stop number linked to SmartRider data. This file does not include train stations.
b. ROAD: The road name assigned to stop.
c. SUFFIX: The road name's suffix, e.g. "street", "avenue", etc.
d. STOPNAME: The name assigned to the stop, usually the nearest intersection.
e. SUBURB: The suburb where the bus stop is located
f. STATUS: This field identifies stops as "Active" or "Discontinued" if they are no longer in use.
g. POSITIONX_MGA. The horizontal MGA (Map Grid of Australia) co-ordinate of the stop.
h. POSITIONY_MGA: The vertical MGA (Map Grid of Australia) co-ordinate of the stop.

The MGA coordinates x and y were used to verify the suburb location of each public transport stop. By default in the files provided by PTA, some stops were assigned to suburbs that are reserved lakes, parks or forests, such as Herdsman, Whiteman, and Lake Monger. As there is no socio-economic data associated with them. The statistical software discarded any trip originating from these stops by coding it as missing data. Nevertheless, a meticulous process of data validation was undertaken to update these stops and assign them to corresponding suburbs adjacent to the reserved areas, because these trips were generated by land-use activities involving residents, employees, and student populations residing nearby. These MGA coordinates data were used to locate the exact positions of these stops so it could be determined whether the name of the adjacent suburb should be assigned to them. As a result of this verification process, the trips originating from the stops assigned to reserved areas were not discarded carelessly; rather, they were considered legitimate trips generated from land-use activities in the corresponding suburbs.

<u>The train station numbers and locations data file</u> contains:

a. STOPNUMBER: A 5-digit train station number linked to SmartRider data.
b. LOCATION: The name of the road/street where the train station is located
c. NAME: The name of the train station.
d. SUBURB: The suburb where the train station is located.

The data from these two files was imported into the **public transport stops table** in the data warehouse.

3.9.3.1 Average Stops per Km^2

Using the public transport stops table, the total number of bus/ferry stops and train stations in each suburb were counted first. Then, the total number of stops and stations that were available in 2009 were divided by the urbanised area of the corresponding suburbs in Perth's metropolitan areas to calculate the average stops per km^2 in each suburb.

3.9.3.2 Total Frequencies as Public Transport Service Provision Densities

Total frequencies of all public transport services available within an average week were used as service provision variables. The bus/ferry or train service provision schedules[11] were categorised into 23 time segments for descriptive analysis:

Weekday	12am-3am
	3am-6am
	6am-9am
	9am-12noon
	12noon-3pm
	3pm-6pm
	6pm-9pm
	9pm-12midnight
Saturday	12am-3am
	3am-6am
	6am-9am
	9am-12noon
	12noon-3pm
	3pm-6pm
	6pm-9pm
	9pm-12midnight
Sunday	3am-6am
	6am-9am
	9am-12noon
	12noon-3pm
	3pm-6pm
	6pm-9pm
	9pm-12midnight

[11] Bus/ferry or train service provision on public holidays is the same as provision on Sunday.

The timetable dataset was used to compute public transport service provision variables by taking the following steps:

- Aggregating the weekly total frequencies of bus/ferry services and train services for each suburb in 2009 based on the above 23 time segments and
- Dividing these forty six weekly public transport service provisions in each suburb by its urbanised areas to derive public transport service provision densities.

These service provision variables were subsequently used in a factor analysis to analyse of how they correlate with each other, and how much each service provision variable contributes to the latent variables produced by the factor analysis.

3.9.4 Land use characteristics

The land use characteristics variables examined in this study include estimated resident population densities, students up to year 12 population density, university student population density, employment population densities, road length per km^2, and distance from the city centre. In this section, the data collection and aggregation processes for these variables are described in detail.

3.9.4.1 Estimated Resident Population Density

The “Estimated Resident Population Density in 2009” dataset was provided by the Australian Bureau of Statistics as customised data from its Information Consultancy Services. The estimated residential population for state suburbs within Perth’s metropolitan areas was collected according to four age groups: 0-16, 17-35, 36-64, and 65 years and over, and partitioned by gender. These estimates correspond with Estimated Resident Populations by Statistical Local Area, released on 30 March 2010 (in Regional Population Growth, Australia, 2008-09: cat.no.3218.0). Estimates based on Collection Districts are customised data available for purchase from an information consultancy.

According to the Australian Bureau of Statistics Population Concepts (Australia, 2008: Catalogue No.3107.0.55.006), the estimated resident population (ERP) is the official measure of Australia’s population. The resident population is estimated by counting all residents living in Australia for 12 or more months, regardless of nationality or citizenship with the exception of foreign diplomatic personnel and their families. The estimates used in this research are based on state suburbs aggregated from collection districts (CD), which in turn are based on estimated resident populations. Resident population estimates for collection districts (CD) are prepared by starting with the usual population counts, by age and sex, from the most recent census, then updating each CD total by estimating the growth in post-census years implied by changes in electoral roll counts and prorating the updated CD totals to the ERP totals at suburb level. An iterative proportional fit process is then used to produce the CD totals and the State suburbs based estimated resident population estimates.

Instead of using the estimated total resident population, this research uses estimated resident populations categorized by age and gender to gain a better understanding of the differential contribution of each of these groups to public transport use in Perth. Then these estimated resident populations by age and gender in each suburb are divided by its urbanised area to calculate the estimated resident population density (by age and gender).

3.9.4.2 Students Up to Year 12 Population Density

Data on students up to year 12 were extracted from Curtin University business intelligent data warehouse, which is available to all staff and research students. At http://planning.curtin.edu.au/bitools/, the warehouse makes available many educational datasets, including the WA State Government Education Department dataset (previously DETWA dataset).

This dataset includes detailed information on student enrolments in particular schools, along with school addresses, for two terms in 2009. There was low variation between the student enrolments in these two terms, and the average of the student enrolments in two terms was calculated for each school first. These student enrolments were then agglomerated based on the suburb in the school address. Then, student enrolments for each suburb were divided by its urbanised area to calculate the population density for students up to year 12.

The student enrolments in particular school were validated against the student enrolments data available at http://www.det.wa.edu.au/schoolsonline/home.do (published by Department of Education, Western Australia). There were 69 suburbs without schools in 2009.

3.9.4.3 University Student Population Density

A similar process was applied to compute the university student population data from the Curtin University business intelligent data-warehouse, and to calculate the average student enrolments in two semesters. Subsequently, the researcher divided the university student populations by the urbanised areas of their corresponding suburbs to generate the university student population density. The six suburbs with university student populations include Crawley, Bentley, Murdoch, Mount Lawley, Fremantle and Joodalup.

The university student enrolments retrieved from the Curtin University business intelligent data-warehouse were verified against the student enrolments reported in the Annual Reports from the University of Western Australia, Curtin University, Murdoch University, Edith Cowan University, and University of Notre Dame.

3.9.4.4 Employment Population Density

The Department of Planning conducts the Perth Employment Survey every five years. The last survey was done between March 2007 and June 2009 and captures employment, floor space and land-use data in areas specified as commercial, industrial, public purposes, and recreation/open space (Department of Planning, December 2009). Approximately 113,000 activities in the Perth Metropolitan region and the Mandurah and Murray local government areas were measured. The results of this survey were published in the report “The Evolving City: An atlas of change in the City of Perth 1990-2007”. The survey of land use focused on the number and type of establishments, the floor-space occupied, and the number of employed persons to examine significant patterns, themes and emerging trends in Perth. These three main variables were coded according to the Western Australian Standard Land Use Classification (WASLUC) coding system and the Planning Land Use Categories (PLUC).

For this research, the employment population data was provided by the Western Australia Department of Planning in .map file format containing the land use activities as well as the shape and locations of complexes. Four types of complexes are included in this study: commercial, industrial, public purpose and recreation and open space. Both full time and part time employment are included in this analysis. There are eleven Planning Land Use Categories identified as types of industries in this research: Primary/Rural, Manufacturing/Processing/Fabrication, Storage/Distribution, Service, Shop/Retail, Other Retail, Office/ Business, Health/ Welfare/ Community Services, Entertainment/ Recreation/ Culture, Residential, and Utilities/ Communications.

In the employment map file, the shape and location of all complexes in the studied areas were defined by associating the Local Government Areas (LGA) with one another. These Local Government Areas are higher-level geographical area representations than suburbs, and each LGA contains multiple suburbs. Therefore, ArcGIS software was used to overlay the employment map file and Western Australia suburbs map file to determine the list of complexes within each suburb in the Perth

Metropolitan area. Then, employment details of each complex were extracted from the map files to aggregate the employment population at suburb level. There were two scenarios in this data aggregation process:

1. Some complexes were not split into more than one suburb i.e. the complex was wholly within an identified suburb. In this straightforward scenario, the employment population data for each of these complexes were simply aggregated.
2. Some complexes, however, fell into more than one suburb i.e. they straddled the administrative boundaries of these suburbs. In this scenario, the researcher determined the floor space areas of each suburb by overlaying the location and shape of a complex in the administrative boundary area and shape of associated suburbs. Then, the employment population was split among the corresponding suburbs using the percentage of floor space associated with the part of the complex in each suburb where it was located. This meticulous process was replicated for all complexes located across more than one suburb.

Subsequently, the employment population densities in each industry were calculated by dividing the employment population by the urban area of each suburb.

3.9.5 Other Urban Form Variables

Two urban form variables taken into account in this research are road length (in km) per km^2 and distance from city center. This section will explain how these datasets are collected and processed.

3.9.5.1 Road Length (in km) per km^2

The roadcentreline dataset was provided by Landgate (Western Australia Land Information Authority) as a GIS .map file. This dataset was used to calculate the total road length in each suburb based on the Locality (Suburb) data field. All road types, such as minor road, closed road, road, street, highway, freeway, bus lane, and roundabout, are included; the lane counts, however, are not considered in calculating the total road length in km. Then, the researcher divided each suburb's total road length by its corresponding urbanised area to generate a measure of road length (in km) per km^2.

Total road lengths (in km) of ten sampled suburbs, including Bentley, Osborne Park, Perth, Mirrabooka, Success, Fremantle, Middle Swan, Success, Bull Creek, and Joondalup were measured manually in Google Earth to verify the road lengths retrieved from roadcenterline dataset. In all cases, the results were satisfactory.

3.9.5.2 Distance from City Center

The suburb administrative boundaries dataset was provided by Landgate (Western Australia Land Information Authority) as a GIS .map file. ArcGIS application is used to calculate the centroids of all polygons representing the suburb boundaries in GIS .map file. The distance to city center is defined as the direct distances between centroid of Perth suburb polygon and centroids of other suburb polygons because Perth suburb is the city center of Perth metropolitan area.

3.9.6 Socio-Economic Variables

The socio-economic variables used in this research are:

- Number of residents in different income groups
- Average weekly rent[12]
- Average car ownership per household.

There were eight suburbs for which socio-economic factors were insignificant: Welshpool, Karrakatta, Perth Airport, Malaga, Neerabup, Malaga, Carabooda, and Kwinana Beach. This is because of their very low estimated resident population densities. However, employment densities in these suburbs were significant, as they are industrial areas. In this research, not only the estimated resident population densities, but also employment densities and student population densities, are considered as determinants of public transport usage. Therefore, these eight suburbs are still included in this research even though they are having insignificant socio-economic factors as a result of very low estimated resident population densities.

3.9.6.1 Number of Residents in Different Income Groups

According to the Australian Bureau of Statistics (15 May, 2014), a regular Population Census has been taken since 1911. The majority of datasets used in this research are from 2009, and the two censuses closest to 2009 are the 2006 Census and 2011

---

[12] Average weekly mortgage was included in the initial data analysis stage. However the correlation between average weekly rent and average weekly mortgage is significantly high and these two variables are almost identical. To reduce multicollinearity between the regressors in multiple regression models, it was necessary to exclude one of these variables. Average weekly rent is selected because this dataset is available annually and will allow for validation and verification with other studies in the future.

Census. Of these two census years, data collected in 2011 are used in this research because it is closer to 2009 than is 2006.

Data on the number of residents whose weekly income falls into four different income groups were obtained from the 2011 Census. The TableBuilder Pro tool was used to extract this information from the Census database. It was a very useful tool for constructing income data cubes for different income groups at a suburb-level. The four weekly income groups used in this research are as follows: (1) weekly income below $250, (2) between $250 and $1000, (3) between $1000 & $2000, and (4) above $2000. The number of residents in the four different weekly income groups was used instead of average weekly income in each suburb to determine each group's differential contribution to public transport usage. A major problem with calculating average weekly income for a suburb is that one would need to assume an average or median figure for each income range reported in the Census and the highest bracket, this is problematic because there is no upper limit defined.

3.9.6.2 Car Ownership per Household

Other socio-economic data collected from the 2011 Census is car ownership per household. According to the Australian Bureau of Statistics (20 May 2011), the census counted the number of registered cars which were either owned or used by the members of a household, and which were parked in garages or near the occupied household on census night. Vans and company vehicles kept at home were included in this count. However, motorbikes or scooters were excluded. The number of household is counted is counted based on the number of motor vehicles parked at its premises at census night. The number of motor vehicles ranges from 1 to 30. Using VEHD (Number of motor vehicles) from the 2011 census, the researcher aggregated the total number of registered motor vehicles and the total number of dwellings in each suburb. Then the average car ownership per household is computed by dividing the total number of registered motor vehicles by the number of dwellings.

3.9.6.3 Average Monthly Rent

The rental data for metropolitan Perth was retrieved from the historical rent dataset, which is publicly published by REIWA (Real Estate Institute of Western Australia). This dataset can be accessed at http://reiwa.com.au/Research/Pages/Perth-Rental-Data-Search.aspx?reg=perth. Rental data were collected based on the rental properties listed and leased in 2009. All types of dwellings, such as townhouses, units, duplexes and villas, were included in this data collection. The rental data for each suburb was manually retrieved for each quarter in 2009. Then, average monthly rent in each suburb was enumerated. Accordingly, only residential rentals were taken into account, while commercial rentals were excluded.

3.10 Issues and Solutions

The issues encountered and the solutions applied in the course of the data manipulation and verification processes for each dataset are described in the previous sections. The technical issues faced in constructing the data-warehouse and integrating the datasets extracted from differently formatted data sources are as follows:

- Based on considerations about financial affordability, Microsoft SQL server 2008 was selected to build the data-warehouse for this research in 2009. Due to limitations on the maximum number of rows (1,048,576 rows) in excel 2007 (Microsoft, n.d), the program Mircorsoft SQL server used to upload data, there were issues importing information from the public transport usage data files, as the latter contained one million rows for all travel transactions. Therefore, it was necessary to split the large files into smaller ones with fewer rows to successfully import them into the data-warehouse. The cygwin application was used to perform the data file manipulation. To split the files at the end of a row, it was necessary to find "Next Line" tag in the data files. The public transport usage data files are encoded as Unicode, and it was necessary to decode them into ASCII formats. After decoding files into ASCII formats, it was possible to count the number of rows by finding "Next Line" tags in the data file. Then, large data files were split into smaller ones consisting of 1,000,000 rows. After this scrupulous data file manipulation process, all public transport usage data were imported into the data-warehouse. All of these travel segment transactions (legs) were then converted into journeys, which were verified using the total number of journeys reported by Transperth in their annual report.
- Incompatibility among geo-code data sources, such as urbanised areas generated in Google Earth, the road centreline data file (in shape format), and the employment survey data file (in .map file) was another challenge that arose during this research. Various geographic information systems (GIS) applications such as Google Earth, MapInfo, and ArcGIS were used to overlay the maps, change to common format, and integrate the data at the same geographical level (the suburb-level).
- The most significant and tedious challenge for the researcher was determining the most granular geographical scale possible. As described in previous sections, very detailed data were collected from numerous sources with different detailed geographical zoning levels. For example, employment population data was collected based on building complexes; public transport usage and service provision data were collected at the service stop level; and some observed variables were collected at the suburb or collection district level. Various GIS applications were used to overlay these maps with different geographical boundaries and to aggregate/disaggregate the data at the suburb level.
- Some statistical issues in this analysis include outliers in the data and high correlations among some of the observed variables. A very comprehensive data verification

process was conducted to make sure that these outliers were not artefacts of data import errors. These outliers were found to reflect real differences in rates of public transport usage in Perth's metropolitan suburbs. In particular, there were significantly high usage rates in Perth, Bentley, Crawley, Fremantle, Murdoch and Joondalup, which is due to their high employment densities and the presence of universities. Some outliers also existed in the observed variables, such as employment population densities in various industries and estimated resident population densities by gender and age group. This resulted from city planning for industrial areas and residential areas. Statistically, these outliers can distort the regression model. Therefore, rigorous data transformation was conducted to manage these outliers. Further, high correlations among observed variables can violate the assumptions of regression analysis. Even though aggregating across various categories for each variable can solve this problem, factor analysis was conducted to derive the latent variables. The factor loadings and scores were used to develop a predictive model.

In summary, there are many technical issues encountered in this research regarding the constructing the data-warehouse especially in integrating the datasets extracted from differently formatted data sources and aggregating or decomposing them to be on the same geographical level which is at suburb level. And these issues are dealt with meticulous data manipulation and verification processes. Another statistical issue with having important outliers and significant correlations among observed variables are also handled by conducting rigorous data transformation processes to satisfy all assumptions and requirements for developing regression model.

3.11 Limitations

The limitations of this research are as follows:

- Data availability is another limitation. Census data on socio-economic characteristics was not available for 2009. Therefore, 2011 census data was used for socio-economic variables such as income and average car ownership per household.
- Only patrons paying SmartRider fares are counted as public transport users for this analysis. Those who paid with cash, however, are not included in the data. SmartRider usage accounted 70% of the public transport usage in 2009. SmartRider cards are used by the majority of frequent or regular patrons.
- One limitation of the smart card data is that it is impossible to determine what activities people are engaging in between travel segments. Therefore, estimating the time thresholds between travel segments is important to identify the sequential travel segments which can be considered as part of the same journey. In this research, the reference number of travel segments which have been identified in the original travel data file provided by Transperth is used to convert the travel segments into whole journey records for further aggregation.

Based on the abovementioned limitations, the future researches are proposed as follow:

- To collect the data during the same time period for public transport usage and all observed variables to validate or enhance the predicting model developed in this research,
- To suggest Transperth to issue the cash sales ticket with barcode so that patrons can tag on and tag off by scanning these barcodes which can then be used to identify the origin of their journeys to include in aggregating the public transport usage density and
- To identify the stops with estimated journey activities such as going to school or going shopping or going for leisure activities by searching whether these stops are next to school, university, shopping centres or recreational area, to analyse the public transport usage patterns based on journey activities[13],

[13] Even though it was possible to do so, it was unable to conduct the analysis on public transport usage patterns based on journey activities in this research due to having time and resource constraints.

4 Methodology II: Statistical analysis techniques and model development

This chapter presents the research approach and statistical techniques used in the study. A detailed description is provided of the applied factor analysis methods, which compute the correlations among the observed variables and extract latent factors from them, thereby reducing the total number of variables while preserving their meanings. An explanation is given for the application of multiple-regression to develop a predictive model of public transport usage based on public transport service provision factors, socio-economic factors, and urban form factors.

The variables used in this research to examine the main determinants of public transport use in Perth based on 2009 are as follow:

- **Forty seven public transport service provision variables** (forty six public transport frequency variables and average public transport stops per km$^{2)}$,
- **Fifteen land use characteristics** (including five employment density variables, university student population density and student (up to year 12) population density variables and eight resident population density (by age and gender) variables,
- **Two urban form variables** such as road length in km^2 and distance from city center variables) and
- **Six social economic variables** (average rent, average car ownership per household, and four weekly income variables.

4.1 Factor Analysis

Factor analysis is applied to solve the problem of multicollinearity among the public transport service provision density variables, and to derive latent public transport service provision density factors. Further on, factor analysis is also used to identify latent variables among some of the land use characteristics, such as estimated resident population densities by age/ gender groups, student population density and their associated socio-economic factors.

According to Kim (1978), this technique can be used to represent a set of variables in terms of a smaller number of hypothetical variables. Kim (ibid) suggests that there are two ways of using factor analysis:

1) Exploratory Factor Analysis: The first step of this analysis requires an examination of the interrelationships among the observed variables, which in this research are land use characteristics, socio-economic, urban form and public transport service provision variables. Then factor analysis can be applied to find out if a smaller number of latent factors can explain the observed correlations. This approach is a practical way of

ascertaining the minimum number of latent factors needed to account for the observed covariance, and also a means of exploring the data for possible data reduction.

2) Confirmatory Factor Analysis: This is used not only as an exploratory tool for discovering underlying factor structures, but also as a means of testing specific hypotheses. Depending on the researcher's knowledge and/or findings from previous research, it can be hypothesised or anticipated that there are particular underlying factors and that certain specific variables belong to these factors. Factor analysis can then be used to test the hypothesis.

Raykov (2008) also suggests that factor analysis can be used to identify latent factors that supposedly underlie contribution of each case (suburb) on a given set of observed variables, as well as their interrelationships. These latent factors can be derived from the pattern of interrelations among observed variables. As a consequence, latent factors can be used not only to explain the high correlations among observed variables, but also to interpret them substantively. The principal component scores can be used as an estimate of the proportion of each subjects' contribution to the factor score in subsequent statistical analyses.

Kim (1978) states that the main assumption in exploratory factor analysis is that the covariance among observed variables is a result of their sharing of common factors, not a consequence of their causal interaction with each other. He notes that both the hypothetical and observed variables are standardized to have unit variance called factor loadings. When only one common factor is involved, these factor loadings are the same as correlations between latent factors and observed variables. Also, when multiple common latent factors are orthogonal to each other, these factor loadings are still equivalent to the correlations between factors and observed variables. When the number of variables is reduced, the contributions of each variable are expressed as factor loadings. Therefore, the covariance among socio-economic and public transport service provision factors is examined first, followed by exploratory factor analysis extract latent factors underlying the data and determine their influence on public transport usage in Perth.

Tabachnick (2000, : p.582) state that factor analysis is applied

"…to a single set of variables when the researcher is interested in discovering which variables in the set form coherent subsets that are relatively independent of one another. Variables that are correlated with one another but largely independent of other subsets of variables are combined into factors."

According to Gorsuch (1983), factor analysis is used to summarise the interrelationships among the variables in a succinct but precise manner as a means of conceptualisation. This technique also assists in minimizing the number of variables for further research while also maximising the amount of information in the analysis. The smaller set of variables reduced from the original set of variables can be used as operational representatives of the constructs fundamental to the complete set of variables. This analysis can also be deployed to search data for possible qualitative and quantitative distinctions especially when the total amount of available data exceeds comprehensibility. Factor analysis thus represents the summation of the observed variables by factors and the degree of generalisation between each observed variable and each factor. This degree of generalisation between variables and factors is presented by the factor loadings and can be used to compare the generalisation of the same variable on several factors.

According to Raykov (2008), the purpose of factor analysis can be outlined as follows:

- to examine whether a smaller set of latent factors can be derived based on the high correlations among the observed variables by applying the principle component method as a data reduction method;
- to compute the number of latent factors which can be substantively interpretable in subject-matter terms; and
- to evaluate the studied subjects based on the principal component scores and use them as factor scores in the subsequent statistical analyses.

Raykov (2008) states that a factor analysis model is based on a set of equations derived from correlations of observed variables and associated distributional assumptions regarding these observed variables. Therefore, a factor analysis model can be mathematically defined as follows:

$$x_1 = \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1m} f_m + \varepsilon_1$$

$$x_2 = \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2m} f_m + \varepsilon_2$$

$$\vdots$$

$$x_p = \lambda_{p1} f_1 + \lambda_{p2} f_2 + \cdots + \lambda_{pm} f_m + \varepsilon_p$$

where:

$f_1 + \cdots + f_m$ represent latent factors

λ_{ij} represents factor loadings (ιth observed measure on the jth factor)

$\varepsilon_1 + \cdots + \varepsilon_p$ represent error or residual terms

The Factor analysis conducted for this research follows the steps below:

1. A correlation matrix (R-Matrix) is constructed. Field (2013) explains that a correlation matrix (R-Matrix) is based on the correlation coefficients between the variables. When there are clusters of coefficients between subsets of observed variables, it indicates that they measure aspects of some underlying dimensions called latent factors. According to Raykov (2008), the latent factors should be substantively interpretable and should explain why certain sets of observed variables are highly correlated. He states that the spectral decomposition of a covariance or correlation matrix can be denoted $\sum$ as:

$$\Sigma = \upsilon_1 e_1 e'_1 + \upsilon_2 e_2 e'_2 + \upsilon_p e_p e'_{p'}$$

   Where υ_i = eigenvalue and e_i = eigenvectors

   Therefore, correlation matrices are constructed for the observed public transport service provision, land use characteristics and socio-economic variables before conducting factor analysis in order to verify the existence of clusters of coefficients between these variables.

2. Principal Component Analysis is applied to construct the initial latent factors. Raykov (2008) points out that principal component analysis can be used to reduce the complexity of the interrelationships among a large number of observed variables to a comparatively smaller number of factors that represent linear combinations of these variables. With this method, the latent factors are generated based on the correlations and covariance among the observed variables. The factors derived from this principal component analysis can be interpreted based on the degree of association between the factors and each variable. Even though the number of derived factors is relatively small compared to the number of observed variables, they still encapsulate the variability of the original data set while representing purer measures of the underlying latent dimensions. The weighted combination of correlations and covariance among the observed variables constitute factor loadings for each factor. These factor loadings represent the maximum possible variances in the initial dataset. According to Kim (1978), factor loadings are coefficient values in the factor pattern or structure matrix, which is a matrix of coefficients where the columns represent common factors and the rows represent the observed variables. Kim (ibid, pg. 77) states that, "*elements of the matrix represent regression weights for the common factors where an observed variable is assumed to be a linear combination of the factor; for an orthogonal solution, the pattern matrix is equivalent to correlations between factors and variables".* The factor loadings of land use characteristics, socio-economic factors and public transport service provision factors are used along with urban form factors to develop the regression model to predict public transport usage in Perth.

3. Varimax factor rotation is applied to develop a final solution for the factor pattern. Kaiser (1958) states that the ultimate principle of a rotational procedure is factorial invariance. He defines the Varimax criterion as:

$$v = \sum_s ([n \sum_i \left(\frac{a_{is}^2}{h_i^2}\right)^2 - \sum_i (a_{is}^2/h_i^2)]^2]/n^2)$$

Where: v = varimax

n = number of test cases

a_{is}^2= factor loading of i^{th} test case of s^{th} factor

h_i^2= communality of i^{th} test case

Since this research applies the principal component method of factor analysis, the Varimax orthogonal rotation method must be used to determine that the latent variables are not related to one another by construction. Field (2013) recommends the use of Varimax orthogonal rotation because this method encourages maximum dispersion of loadings within factors, loading a smaller number of variables onto each factor.

After applying the above extraction and rotation methods for exploratory factor analysis using SPSS, the following criteria help to analyse the results:

a) Firstly, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy test (KMO) is used to validate the adequacy of sampling in the factor analysis. Field (2005) recommends using the KMO to evaluate the appropriateness of factor analysis because it calculates the ratio of the squared correlation between the variables to the squared partial correlation between variables which assesses the quality of sampling. It is recommended that the KMO test result should be at least 0.5.
b) Secondly, Field (2005) also recommends Bartlett's test of Sphericity to detect high level of correlation among the variables. He points out that the test results should be significant with a value of less than 0.05 before concluding that there are correlations among the observed variables and that it is appropriate to use factor analysis.
c) The communality test is used to determine the reliability of the factor analysis. Extraction values in the communality table represent the common variances and can be used to verify if eigenvalues greater than 1 are applicable, depending on the number of variables and test cases (Field 2005). Kim (1978, pg. 75) defines communality as "…*the variance of an observed variable accounted for by the common factors; in an orthogonal factor model, it is equivalent to the sum of the squared factor loadings*". In addition, Field (2005) explains that communality is the proportion of common variance within a variable. The principal component analysis used in this research assumes that all variance associated with a variable is common. The extraction values in the communalities table are verified whether

they reflect this common variance. Field (2005) states that the extraction values in the communality table should be greater than 0.7 for an exploratory factor analysis using less than thirty variables. When more than 250 test cases are used, Gray (2012) recommends that the extraction values be greater than 0.6. In this research, 293 test cases (representing suburbs in the Perth Metropolitan Area) are used. Therefore, only extraction values in the communality tests greater than 0.6 are used for extracting latent public transport service provision factors and land use characteristic and socio-economic factors.

d) Total variances explained with initial eigenvalues are used to determine how many latent factors exist and how much of the variance in the original datasets is explained by these factors. Kaiser's eigenvalue (1) is used as the minimum criterion for retaining the latent factors Raykov (2008).

e) Scree plots are also used to confirm factor retention. Kim (1978) points out that a scree test can be used as a rule-of-thumb criterion to verify the number of significant factors to retain, based on the graphical illustration of eigenvalues (roots), and this is considered appropriate for addressing the disturbances of minor factors.

f) A component matrix and rotated component matrix are also used in this analysis to identify the loading of each variable onto each factor and to verify how these factor loadings change, providing for more meaningful interpretation after the Varimax rotation is applied.

4.2 Multiple Regression Analysis

Multiple regression analysis is widely used to investigate the relationship between one dependent variable and several independent variables. This analysis can also be used to assess the correlation of independent variables with one dependent variable. Further, it is used to develop the equation for predicting a dependent variable from several continuous (or dichotomous) independent variables.

The commonly used regression techniques are standard multiple regression, sequential (hierarchical) regression, and statistical (stepwise) regression and robust regression, Tabachnick (2000). Sirkin (2006) states that **stepwise multiple regression** allows statistical software applications such as SPSS to determine which independent variables to enter and when to enter them. He also explains that the first step in this process is finding the “best” variable among a menu of variables and generating a simple regression for predicting the dependent variable from the selected independent variable. Then SPSS finds the “second best”, and “third best” variables, and so on. Statistical (stepwise) regression is applied in this analysis to determine how public transport usage can be predicted on the basis of public transport service provision, land use characteristics, socio-economic factors, and the urban form of each suburb in the Perth metropolitan area.

There are a few cases (i.e. suburbs) that have a significantly high public transport use density. Statistically, these test cases are influential outliers and can distort the regression line of predictive model. However, they are still included in this study because they are very important and significant suburbs in metropolitan Perth. As a consequence of having these influential outliers, the normal distributions of the public transport usage data and observed variables have high kurtosis (peakness) values. Therefore, rigorous data transformation is conducted in this study to satisfy all assumptions required for multiple regression analysis. Furthermore, the robust regression method is also applied for model validation. Yaffee (n.d) suggests that the **robust regression method** can be used to detect any influential outliers in the dataset, and also as a substitution for the least squares regression method for datasets with influential outliers. He explains that in the robust regression, absolute residuals are calculated and scaled by weighting each test case to reduce the influences of outliers. A case- weighted regression is then rerun.

In this research, various multiple regression methods are applied to develop a predictive model of public transport use. These methods are:

1. **Standard multiple regression (Enter method)** is used to develop the model by considering all observed land use characteristics, socio-economic, urban form and public transport service provision factors.

2. Then the **stepwise method** is used to gain a better understanding of which observed variables are the most dominant determinants of public transport use in Perth, and to confirm that all observed variables used in the standard multiple regression make significant contributions to the robustness of the model.
3. The **robust regression** method is also applied to validate the model and verify its overall fitness and the relative importance of the observed variables.

When interpreting the results from the different methods, not only the multiple correlation coefficient but also the beta coefficient of each observed variable towards public transport use are taken in account. The multiple correlation coefficient measures the correlation between a dependent variable and the combined effect of other designated variables in the system. The coefficient of multiple determination measures the proportion of variation in the dependent variable accounted for by those other variables and the beta coefficient measures the relative importance of observed variables (Sirkin, 2006). The Beta coefficient or beta weight (β) and the standardised partial regression slope are used in developing the predictive model for metropolitan Perth.

4.3 Assumptions for Multiple Regression Analysis

Richardson (2011) explains the five assumptions for multiple regressions as follows:

a) **Assumption of normality**:

Richardson (2011) states that in multiple regression analysis, it is assumed that the values of any of the independent and dependent variables are normally distributed, resulting in a normal distribution of error terms. Tabachnick (2000) also explain that errors (standardised residuals) should be normally distributed around each and every predicted value of the dependent variable. According to Berenson (2013), the normality assumption in the errors can be evaluated by constructing a normal probability plot of the residuals along with a histogram based on the results of the residuals tally in the frequency distribution.

Field (2013) points out that the dependent variable should have a linear relationship with any predictor. Therefore curve estimates are conducted to investigate whether the relationships between the dependent variable and its predictors take linear or other (curvilinear) forms. If the relationship is found to be curvilinear, then data transformation is performed to make it linear and normalize the distribution.

Field (2013) also recommends two methods of data transformation to overcome any issues with positive skewness, positive kurtosis, unequal variances and lack of linearity, namely log transformation and square root transformation. Log transformation compresses the right tail of the distribution, which reduces positive skewness and creates a more normally distributed variable. This method can also enable a curvilinear relationship to be transformed into a linear one. A constant is added to make sure that the minimum value of variables that need to be transformed is zero or a positive number. Likewise, square root transformation brings any large scores closer to the center to normalize the distribution.

In the present research, all observed variables (both dependent and independent) are analysed to see whether they are normally distributed based on their skewness and kurtosis (pointiness). Then, any variable with high skewness, kurtosis, and curvilinear pattern is transformed by using the log or the square root method. Then, descriptive analysis is conducted on the transformed variables and the method (log or square root) which can produce better skewness and kurtosis is chosen for each variable. Detailed data transformation is discussed in section 5.3.3: Data Transformation.

b) **<u>Assumption of homoscedasticity</u>**:

Berenson (1999) state that a regression model violates the assumption of homoscedasticity if there is any systematic pattern or clustering of the residuals. They also suggest that a plot of regression-standardised predicted values against regression- standardised residuals can be used to examine whether the derived model satisfies the assumption of homoscedasticity. If there appears to be a fanning effect where the variability of the residuals increases as the regression-standardised predicted values increase on the x-axis, this indicates the lack of homogeneity, and hence a violation of homoscedasticity occurs (i.e. heteroscedasticity). According to Field (2013), this problem can be overcome by using weighted least squares regression. This research uses a scatter plot of predicted values against regression-standardised residuals to verify whether the derived model satisfies the assumption of homoscedasticity.

c) **<u>Assumption of linearity</u>**:

Tabachnick (2000) explain that the relationship between the predicted values of the dependent variable and errors of prediction should be in a linear form. When there is a non-linear relationship between the predicted values and errors (residuals) of these predictions, the problem can be solved by transforming the independent or dependent variables. While violating the assumption of linearity of residuals weakens the regression model, it does not invalidate it.

In this study, curve estimation is used to examine whether the relationships between the dependent variable and each predictor are linear or curvilinear. Based on the type of curvilinear relationship (cubic, logarithmic, power, or exponential), appropriate methods of data transformation are selected. The Detailed curve estimate results are discussed in Section 5.3.2: Curve Estimates between Dependent Variables and Predictors.

d) **<u>Assumption of non-multicollinearity</u>**:

According to Curwin (1997), multicollinearity can be defined as the interrelatedness of the independent variables. Multivariate outliers have an unusual combination of scores on two or more variables which can lead to unstable results, unreliable prediction, and inconsistent conclusions. Love (1991) recommends that any research applying multivariate analysis should test for multicollinearity, as it is one of the major potential problems associated with multivariate regression analysis.

Following Berry's (1985) recommendation, the correlation matrix is used to do an initial test for multicollinearity between predictors and outcomes. Even though the commonly used cut-off value for correlations among all independent variables is 0.8, Berry (1985) explains that the correlation between a pair of observed variables of 0.7 could indicates problems of estimation for a very small test case; whereas in larger samples, a correlation of 0.85 alerts the existence of multicollinearity. In addition, he recommends examining the standard errors of estimates. By comparison, Field (2013) suggests that multicollinearity exists when correlations between predictors are greater than 0.9.

Considering the recommendations of these authors, the present research uses 0.85 as a cut-off for multicollinearity. Thus, the aim of the tests is to determine whether correlations between public transport usage (dependent variable) and public transport service provision factors, land use characteristics, socio-economic and urban form factors (independent variables) are greater than 0.85.

Finally, the correlations between the dependent and independent variables are also analysed. If multicollinearity exists in a model, then the coefficients of the independent variables may be unstable and the model is said to suffer from autocorrelation. Therefore the Durbin-Watson test is also conducted to test for autocorrelation.

e) **<u>Assumption of independence</u>**:

Field (2013) explains that the assumption of independence means that the errors in the regression model are not related to each other. The equation used to estimate the standard error is valid only when the observations are independent. It they are not independent (or this assumption of independence is violated) then the confidence intervals and significance tests used in developing the regression model become invalid.

Berenson (2013) suggests that the assumption of independence can be examined with a scatter plot of residuals. He also recommends using the Durbin-Watson statistic, which can be generated as part of the regression model summary output in SPSS to verify whether there is any autocorrelation in the model. This study adopts these recommended tests.

4.3.1 Cross-validation of the model

The public transport usage prediction model is cross-validated or assessed for accuracy with the two main methods suggested by Field (2013): Adjusted R^2 and stepwise multiple regression.

a) Adjusted R^2 (Coefficient of Determination)

According to Field (2013), the R^2 value represents how much variance in the dependent variable (Y) is accounted for by the regression model, based on the estimation sample; while the adjusted R^2 represents the variance in Y accounted for by the regression model for the population. Thus, adjusted R^2 indicates the loss of predictive power. Consequently, R^2 and adjusted R^2 can be compared to validate the predictive model.

Aiken (2002) state that $r\,(\text{sample correlation coefficient})$ represents the measurement of association between two variables, while $R\,(\text{population correlation coefficient})$ is used to measure the association between a dependent variable and multiple independent variables. Achen (1982) suggests that a better criterion for measuring the goodness of fit is minimising the errors (residuals), which involves examining each residual individually and identifying the patterns with large errors. However it could be tedious, and not always possible, to examine every residual in a large number of test cases. Therefore, he recommends using the standard error of estimates to examine the true standard deviation prediction errors, as it has the advantage of not relying on the variance of the independent variables. The standard error of estimates is calculated as follows:

$$\sigma^2 = [\, var(y_i) - var(y_i)]n/(n - k - 1)$$

where σ^2 = standard error of regression, n = number of test cases, and k = number of independent variables (not counting the intercept).

Achen (1982) suggests using R^2 as only one method for measuring a regression model's goodness of fit. He recommends the use of the standard error of estimates to evaluate the fitness of the developed regression model. The R^2 is calculated as follows:

$$R^2 = \frac{\hat{\beta}^2 \text{x } var(x)}{\hat{\beta}^2 \text{x } var(x) + \sigma^2}$$

where $\hat{\beta}^2$ = coefficient β, σ^2 = variance of the residuals and can be interpreted as :

$$R^2 = \frac{(causal\ strength)^2 \text{x } var(x)}{(causal\ strength)^2 \text{x } var(x) + goodness\ of\ fit}$$

Therefore, not only the R^2, but also the standard error of estimates, is examined to evaluate the goodness of fit of the regression model for this research.

b) Stepwise method in multiple regression

Vogt (2005, pg:312-313) defines stepwise regression as:

"*A technique for calculating a regression equation that instructs a computer to find the "best" equation by entering independent variables in various combinations and others. Stepwise regression combines the methods of backward elimination and forward section. The variables are in turn subject first to the inclusion criteria of forward selection and then to the exclusion procedures of backward elimination".*

He also explains that stepwise regression is different from hierarchical regression analysis where the researcher identifies the order of the variables in developing the regression equation. Vogt (2005) suggests that stepwise multiple regression is useful for examining which combinations of explanatory variables can best predict the outcome variables based on the strength and significance of the correlations between them. Another advantage of using stepwise regression is that it allows to find out whether adding more predictors can enhance the ability of the derived regression model to predict the outcome.

Therefore, the stepwise regression method is used to examine which determinants can best explain the public transport usage in the Perth metropolitan suburbs, and to validate the addition of explanatory variables to the model.

c) Robust regression

Andersen (2008) states that it is important to detect and handle the outliers properly because unusual observations can distort the estimates from the regression. He also recommends using the robust regression method as diagnostic tool to identify the potentially problematic cases, and to deal with outliers that cannot be fixed with data transformation. Further on, he explains that robust regression handles the heavy-tailed error distributions and outliers by assigning the weights to each unusual case, thereby providing more efficient estimates and limiting the influence of unusual cases on their values. In this research, a robust regression is performed with the statistical software package STATA.

4.3.2 The Process of Fitting a Regression Model

Field (2013) suggests that it is necessary to check whether there is any bias or unusual cases in the observed data that could violate regression assumptions. Therefore, descriptive analyses of the dependent variable (public transport usage) and each predictor are conducted first to obtain a better understanding of their distribution and determine whether there are outliers.

Due to the nature of public transport usage in the suburbs comprising the Perth metropolitan area, the suburb called “Perth” is an unusual case among the observed suburbs. It is the centre of the city and includes the CBD where many public transport trips originate and terminate due to the concentration of jobs in the area. Field (2013) defines an outlier as an unusual case which differs considerably from the main trend of the data and which affects the estimates of the regression coefficients. Barnett (1994) recommends using the robust regression method to accommodate the outliers, with weights assigned to the extreme values in the estimation. He recommends the use of robust regression as an alternative to least squares regression to test the validity of the model when there are outliers or influential observations in the datasets. It is therefore important to detect influential observations. In order to deal with these matters, the derived public transport usage prediction model using the standard regression method in SPSS, is validated against the model derived using the robust regression method in STATA.

After performing descriptive analyses for the dependent and independent variables, a multiple regression model is estimated based on the process recommended by Field (2013) for fitting a regression model. This is depicted pictorially in Figure 14: The Process of Fitting a Regression Model.

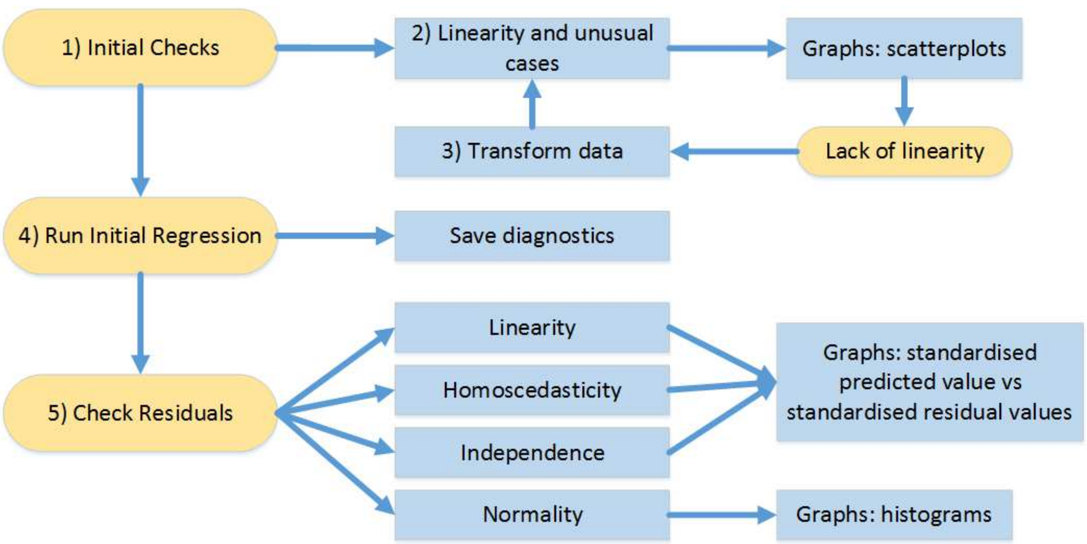


Figure 14: The Process of Fitting a Regression Model, Source: Field (2013: p.316)

As shown in Figure 14:

1) Initial checks are conducted which are explained further in 5.3.1: Initial Checks for Normal Distribution.
2) Curve estimates are used to evaluate linearity and examine the unusual cases and results, as discussed in 5.3.2: Curve Estimates between Dependent Variables and Predictors.
3) Due to lack of linearity between the dependent variable and predictors, data transformation is conducted to make the distribution more linear, as explained in section 5.3.3: Data Transformation.
4) Initial regression including correlation analysis is carried for the observed variables to ensure that there is no multicollinearity. This is discussed in detail in section 5.3.4: Initial Checks for Multicollinearity.
5) After the regression model is derived with the standard method, its residuals are evaluated to confirm that it satisfies all assumptions for multiple-regression, as explained in section 5.3.4.4: Multiple Regression Model Validity.
6) Finally, the stepwise multiple regression method and robust regression are applied to validate the derived predictive model, as explained in section 5.3.5: Cross-Validity of Derived Multiple Regression Model.

4.4 Statistical Software Tools

Two software tools are used in this research to conduct the statistical analysis.

1. SPSS version 22 is used to conduct the factor analysis and multiple regression analysis (both standard and stepwise methods).
2. STATA version 12 is used to develop the predictive model with the robust regression method.

.

5 Results

This chapter reports the results from the quantitative analyses of data that were gathered from the following sources:

- Transperth: public transport usage and service provisions data,
- Department of Planning (Government of Western Australia): employment population survey data,
- Australian Bureau of Statistic: estimated resident population density and their socio-economic data,
- Department of Education (Government of Western Australia): student population data,
- Curtin University Office of Strategy and Planning : university student population data,
- Landgate (Western Australia Land Information Authority): road centreline and administrative boundary data,
- Google Earth: distance from city center.

The results are presented in three sections:

1. Descriptive analysis of public transport usage, public transport service provision, land use characteristics, urban form variables, and social economic variables.
2. Factor analysis results for public transport service provision factors, land use characteristics and socio-demographic factors.
3. Public transport usage regression model.

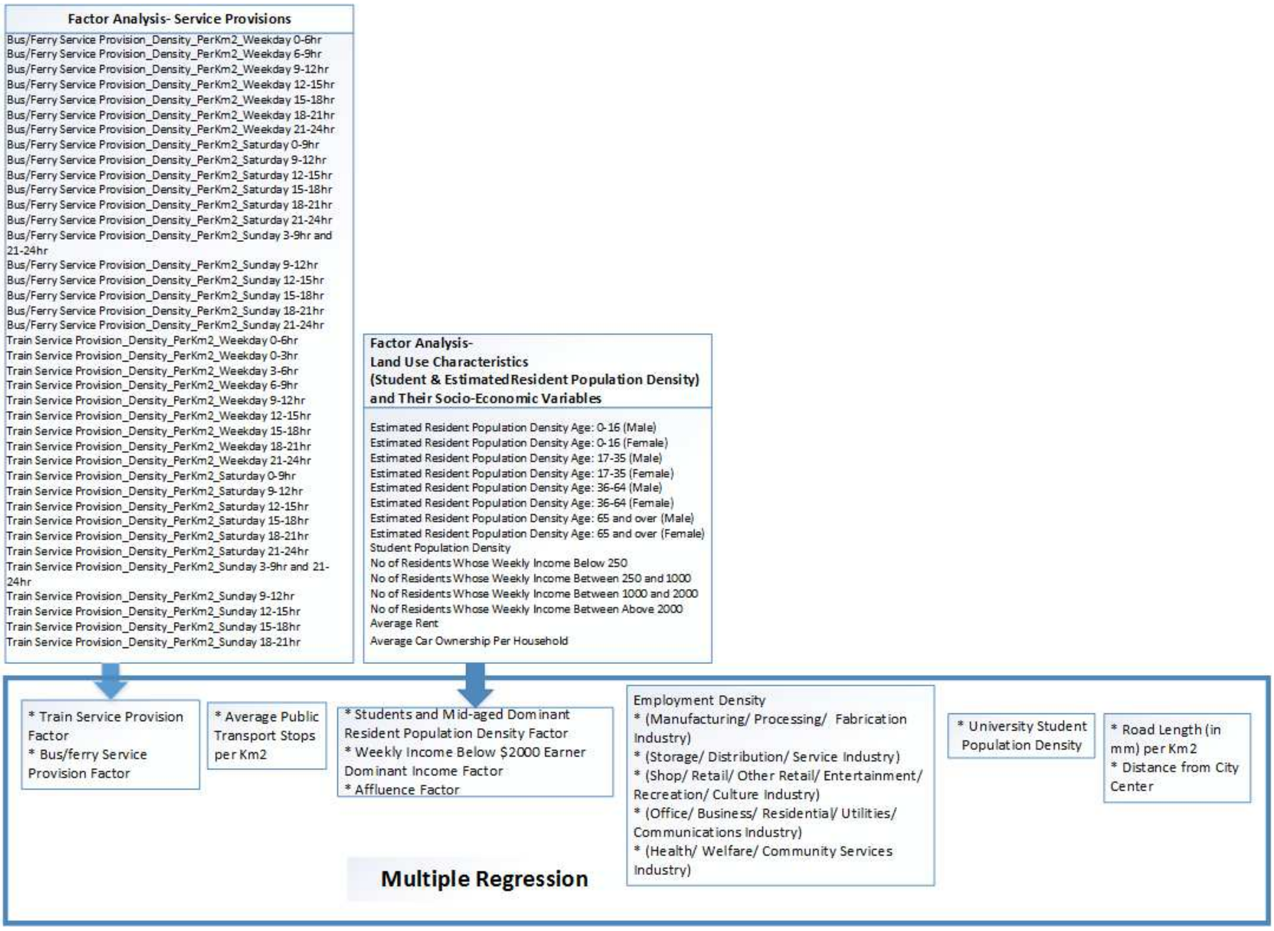


Figure 15: Statistical Analysis Process Summary

Two factor analyses were conducted to identify the latent variables among the highly correlated observed variables. In order to satisfy the assumption of multicollinearity, these latent variables are used in the subsequent multiple regression analysis to develop the predictive model shown in Figure 15: Statistical Analysis Process Summary. The first factor analysis of public transport service provision variables disclosed two latent factors: (1) train service provision and (2) bus/ferry service provision. A second factor analysis of land use characteristics (resident population density by age and gender) variables and socio-economic variables was then performed. It revealed three latent factors: (1) student and mid-aged dominant resident population density, (2) weekly income below \$2000 earner dominant income, and (3) affluence. All latent factors were then combined with other land use characteristics variables, including employment densities in various industries, university student population density, road length per km^2, and distance from city center, in a standard multiple regression analysis to develop a model that predicts public transport usage. The findings from this analysis are verified against the findings from the stepwise regressions and robust regressions to gain a better understanding of which determinants are significant in explaining public transport usage in Perth, as well as to validate the robustness of the derived predicting model.

5.1 Descriptive Analysis

In this section, a systematic and detailed description of the independent and dependent variables is provided. The following measurements are used in the descriptive analysis:

1. Mean: Vogt (2005) defines the arithmetic mean as the average by adding up all values for each case as a total value and dividing this total value by the number of cases. It is used to report the central tendencies of observed variables by identifying the tendency of data in the observed variables clustering around some central values.
2. Standard Deviation: Hair (2010) define this as "*an estimate of the average variability-spread of a set of data measured in the same unit of measurement as the original data".* It is used to describe the typical amount by which scores in a given dataset vary from the mean.
3. Skewness describes the extent to which one of the tails of a variable-distribution is pulled away from its center (i.e. the degree of asymmetry in the distribution), Field (2013).
4. Kurtosis indicates the amount of variation due to outliers, with higher values suggesting greater influence of extreme values, Field (2013).

5.1.1 Spatial-Temporal Analysis of Total Public Transport Usage

The section begins with a spatial analysis of the variation in public transport usage by suburb. This is followed by a description of the variation in public transport usage by month for the year 2009. In the first spatial analysis, both the aggregated total journeys originating in the suburbs of Perth and transport-usage per capita are used to compare high public transport usage (stemming from high population densities) with high public transport usage generated by individual patrons. Then, data on public transport usage by different types of patrons is examined to find out which of these contributes most to overall use-patterns in Perth.

5.1.1.1 Variation in Public Transport Usage by Suburb

For each suburb, the information on total public transport usage is aggregated from the SmartRider dataset, which is based on the origin of each journey. All tag-on and tag-off transactional data is used to transform the segment details into journey details. A journey can be comprised of more than one segment, as people can use more than one mode of public transport for a journey. All statistics described in this section are calculated on the basis of 293 suburbs in metropolitan Perth. Some suburbs, such as Kings Park, Whiteman Park, Tamala Park, Burns Beach and Medora Bay are only used as recreational venues or environmental reserves. Therefore, these suburbs are excluded from this analysis due to unavailability of social-economic data. Eleven outer suburbs, including Banksia Grove, Barragup, Baskerville, Beechina, Belhus, Carramar, Darling Downs, Falcon, Furnissdale, and Henderson are also excluded due to lack of employment data.

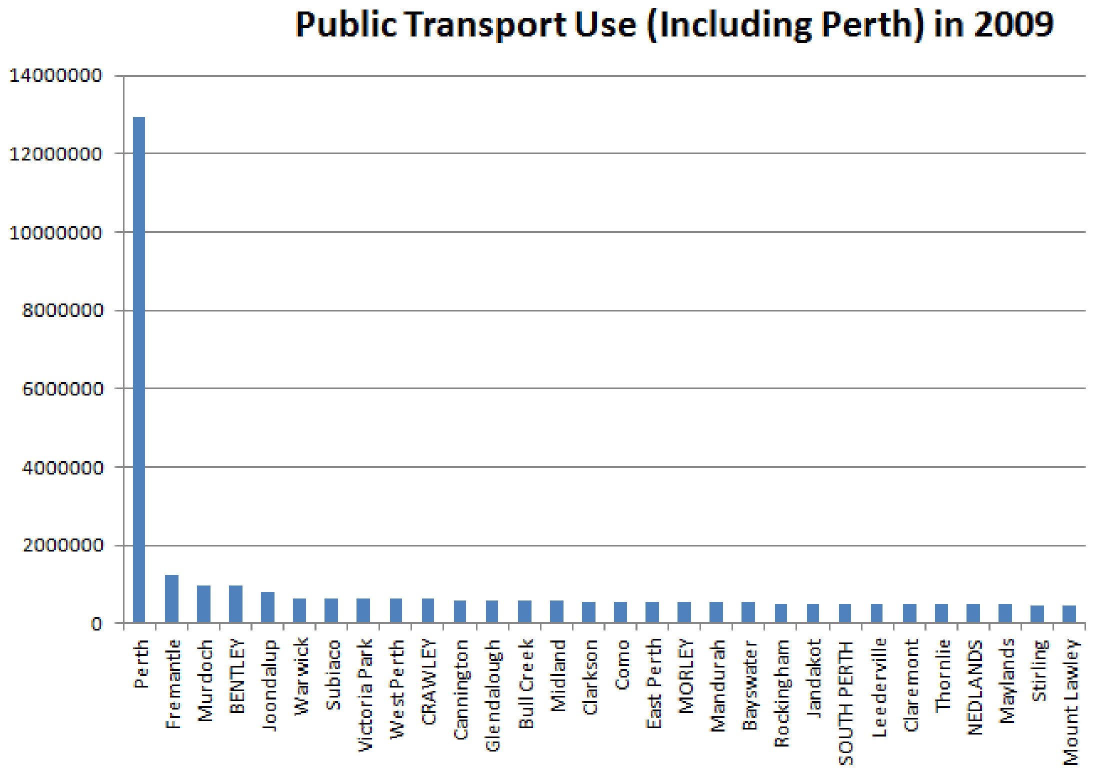


Figure 16: Top 30 Suburbs for Public Transport Usage Including Perth[14] suburb in 2009

Figure 16 shows the top thirty suburbs where the number of trips originated in Perth suburb is significantly higher than the rest of the suburbs in Western Australia during the year 2009. This is mainly driven by its significantly high employment. Statistically, Perth suburb can be considered as an outlier in the data analysis. However because of its significance in land use characteristics and socio-economic factors, the Perth suburb cannot be eliminated. Therefore, the log transformation method is applied to normalise the public transport usage data and Perth is included in all data analyse in this research.

[14] Perth here is referred to Perth suburb which is one of four suburbs constituting as Central Business District.

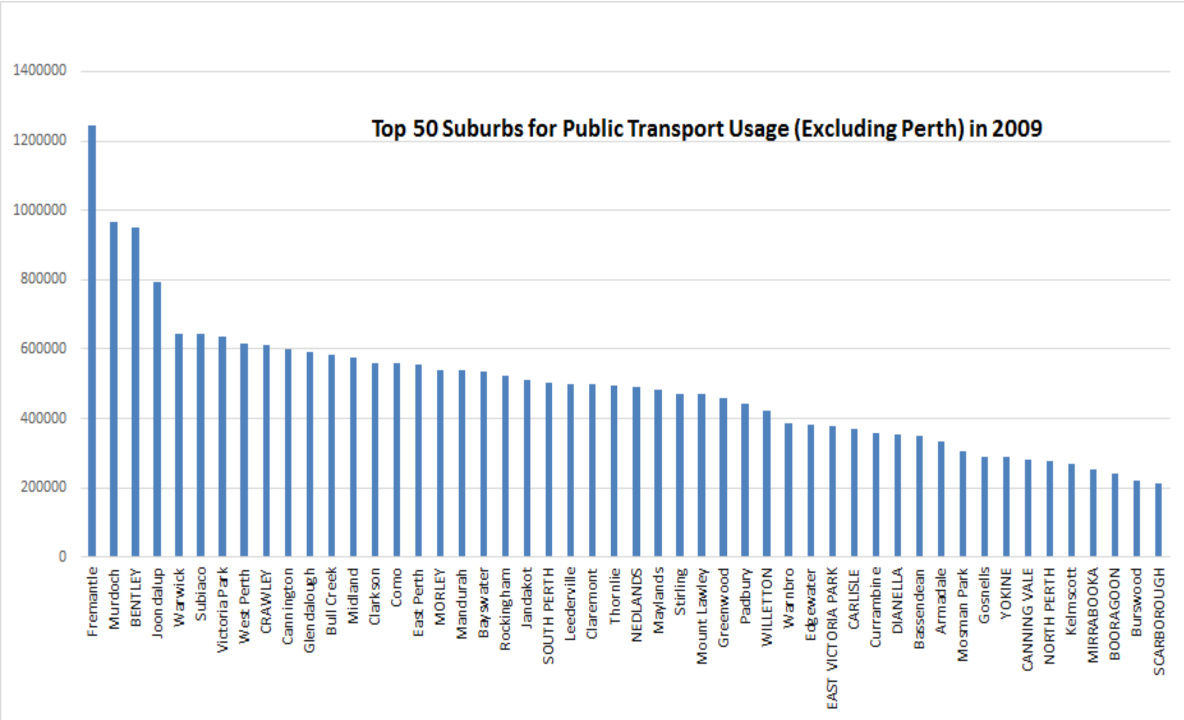


Figure 17: Top 50 Suburbs for Public Transport Usage Excluding Perth Suburb in 2009

Figure 17 shows the top 50 suburbs for public transport usage excluding Perth suburb in 2009 in order to gain better understanding and comparison of the variation among their usage. In Figure 16: Top 30 Suburbs for Public Transport Usage Including Perth suburb in 2009, this was not clearly identified due to having extremely high usage in Perth. This figure illustrates that the suburbs with high university student populations, such as Fremantle, Murdoch, Bentley, Joondalup, and Crawley are also among the top 10 suburbs with high public transport usage.

To clarify whether this high public transport usage is due to high estimated resident population, employment population, student (up to year 12) and university student population or individual high usage, public transport usage per capita is used to demonstrate the importance of distinguishing aggregate from per capita trends, particularly in regard to activity populations. **Activity Population** is an aggregate measure of estimated resident population, employement population, student (up to year 12) population and university student population.

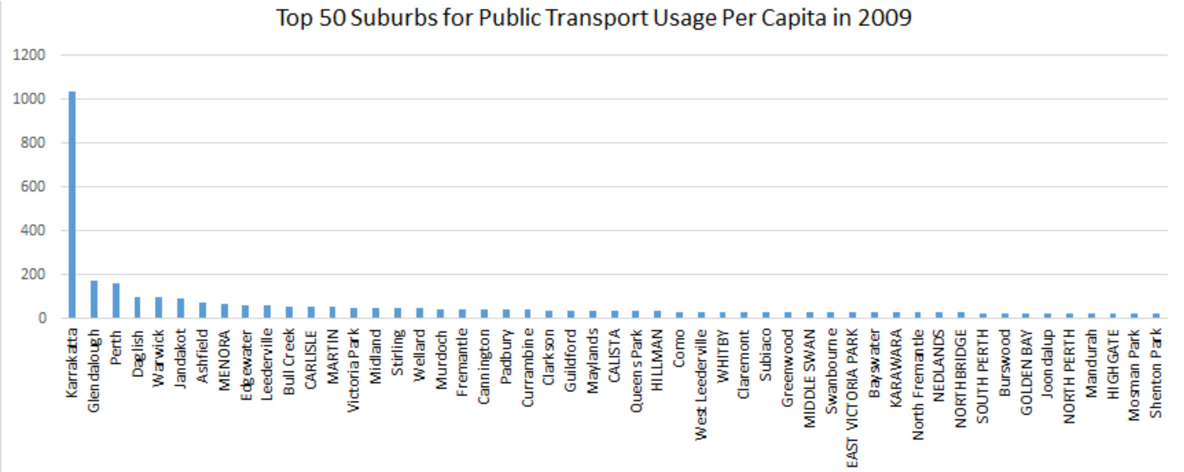


Figure 18 Top 50 Suburbs for Public Transport Usage per Capita including Perth in 2009

Figure 18 shows the 50 suburbs with the highest public transport usage per capita in 2009. Public Transport Usage per Capita is calculated by dividing the total public transport usage by activity population. All these different types of populations are taken into account in calculating activity population because the public transport journeys are generated not only by people who reside in a particular suburb, but also by people who work and study in that suburb.

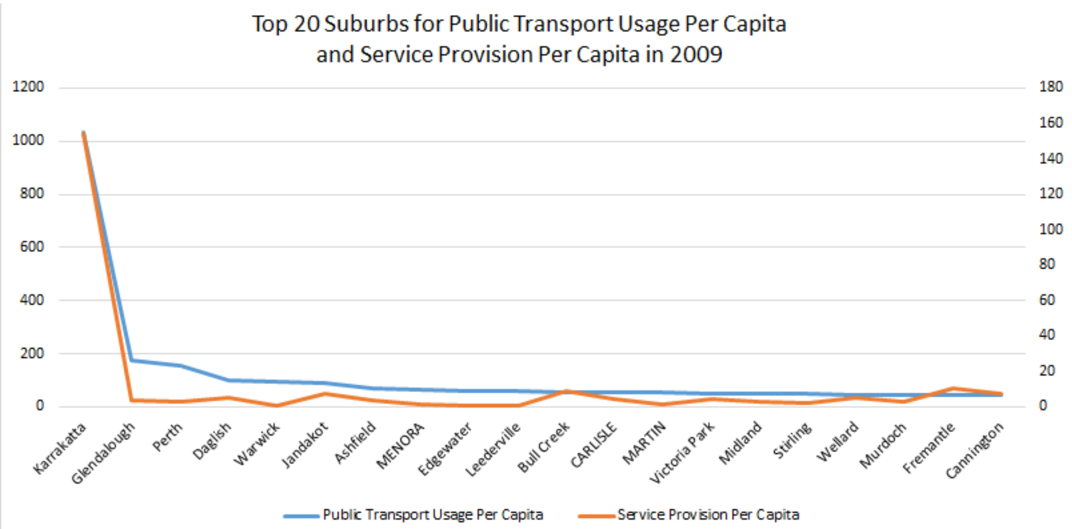


Figure 19 Top 20 Suburbs for Public Transport Usage Per Capita and Service Provision Per Capita in 2009

In the Figure 19 Top 20 Suburbs for Public Transport Usage Per Capita and Service Provision Per Capita in 2009, public transport usage per capita in Karrakatta is significantly higher than the rest of the suburbs, having total public transport usage 38218 despite a very low activity population (which is only 37). In addition, service provision per capita in Karrakatta is significantly high compared to other suburbs. It indicates that high public transport usage could be driven not only by activity population but also by service provision. Therefore, the nexus between public transport usage and service provision will be explored in a more systematic fashion in the section on factor analysis.

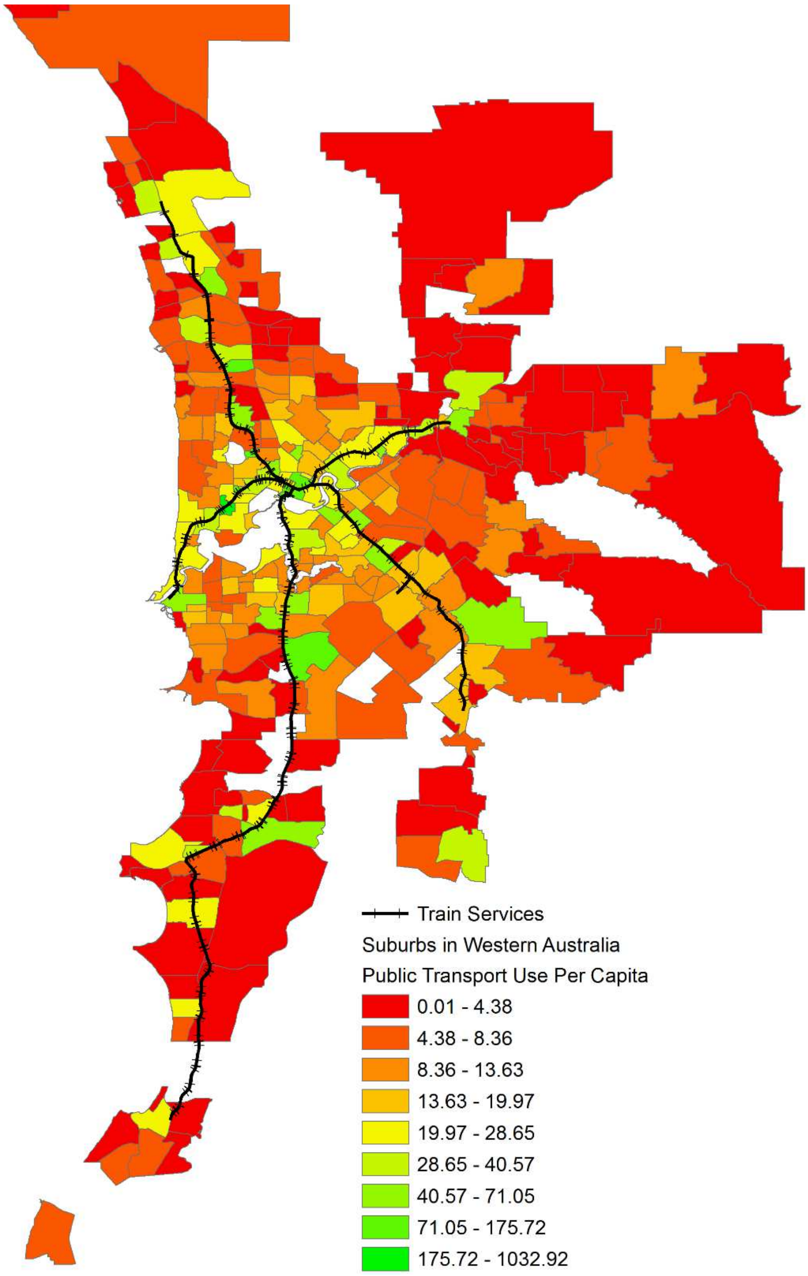


Figure 20: Public Transport Usage per Capita in Western Australia Suburbs Map

Figure 20 shows public transport usage per capita in the Perth metropolitan suburbs. It can be seen that public transport usage per capita is relatively higher in the inner suburbs, especially those which have train service and their surrounding suburbs. Moreover, public transport usage per capita is higher along the train lines when these suburbs are closer to the city center, whereas per capita usage decreases as suburbs get farther away from the central business district.

5.1.1.2 Variation in Public Transport Usage by Month (2009)

This section elucidates temporal variation analysis of how different groups of patrons used public transportation on a monthly basis in 2009.

Figure 21 shows the public transport usage by different types of patrons in each month of 2009. The standard patrons, who are the regular adult riders and not eligible for any concession, are the ones who used the public transport most in 2009, followed by university students and other students up to year 12. There is no fluctuation in the pattern of usage by standard patrons throughout the year, and we can conclude that there is no seasonal peak for this patronage group. In contrast, the public transport usages by university students and other students (up to year 12) are significantly lower in January. Their usages fluctuate throughout the year depending on school/university study periods. In addition, public transport usage by seniors, healthcare concession holders, and pensioners are relatively low compared to the other patron groups. But, their usage patterns are fairly stable throughout the year.

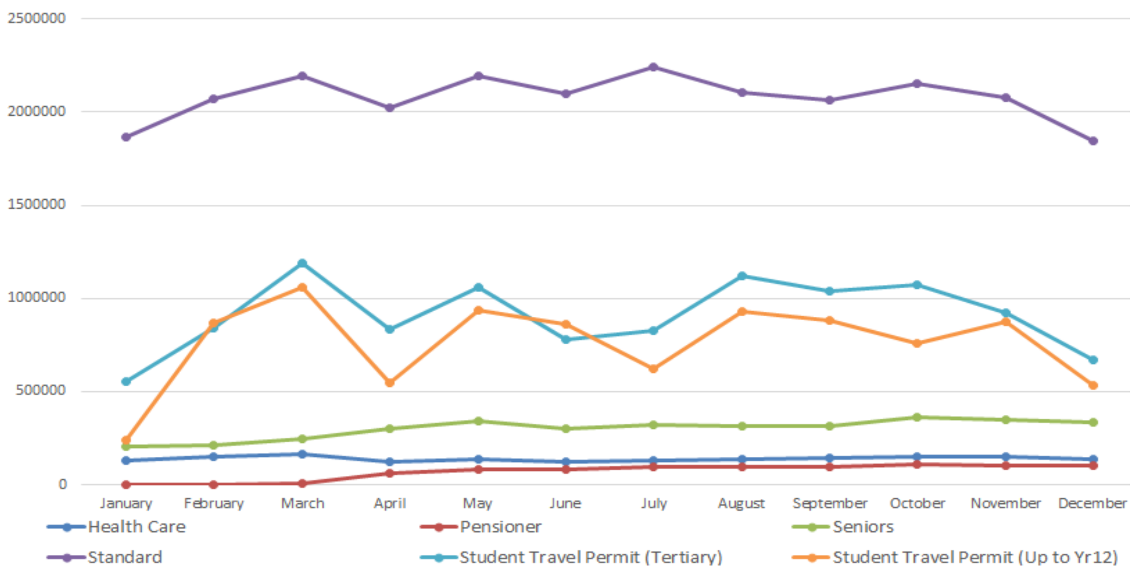


Figure 21: Monthly Public Transport Usage Patterns in 2009 by Different Types of Patrons

5.1.2 Composition of Public Transport Usage by Different Types of Patrons

In this section, the composition of public transport usage by different types of patrons is discussed to identify which group contributed the most of total usage in 2009.

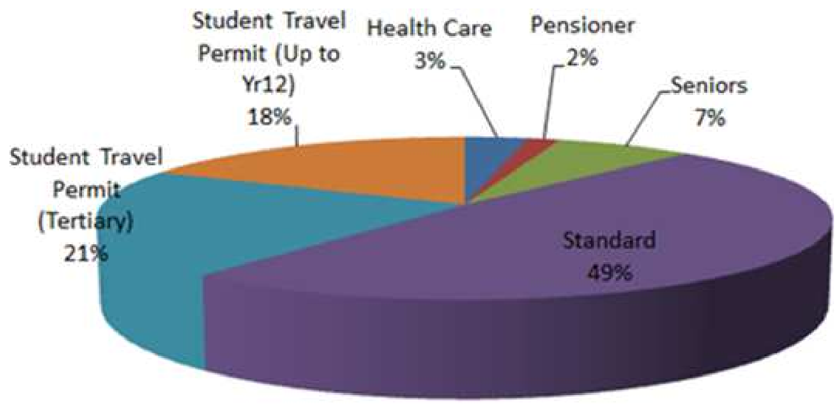


Figure 22: Percentage of Public Transport Use Percentage by Different Types of Patrons

Figure 22: Percentage of Public Transport Use Percentage by Different Types of Patrons shows that the standard patron group has the highest percentage share of total public usage and accounts for 49% of it. This is followed by university students, who as a group account for 21% of total public transport usage. The usage of other students (up to year 12) is relatively similar to that of university students, accounting for 18% of total usage. It is also noticeable that the figures for health care beneficiaries and elderly people are relatively low. Together, the total contribution from healthcare, pensioner and senior patronage groups accounts for only 12% of the overall usage,— even lower than that of students up to year 12.

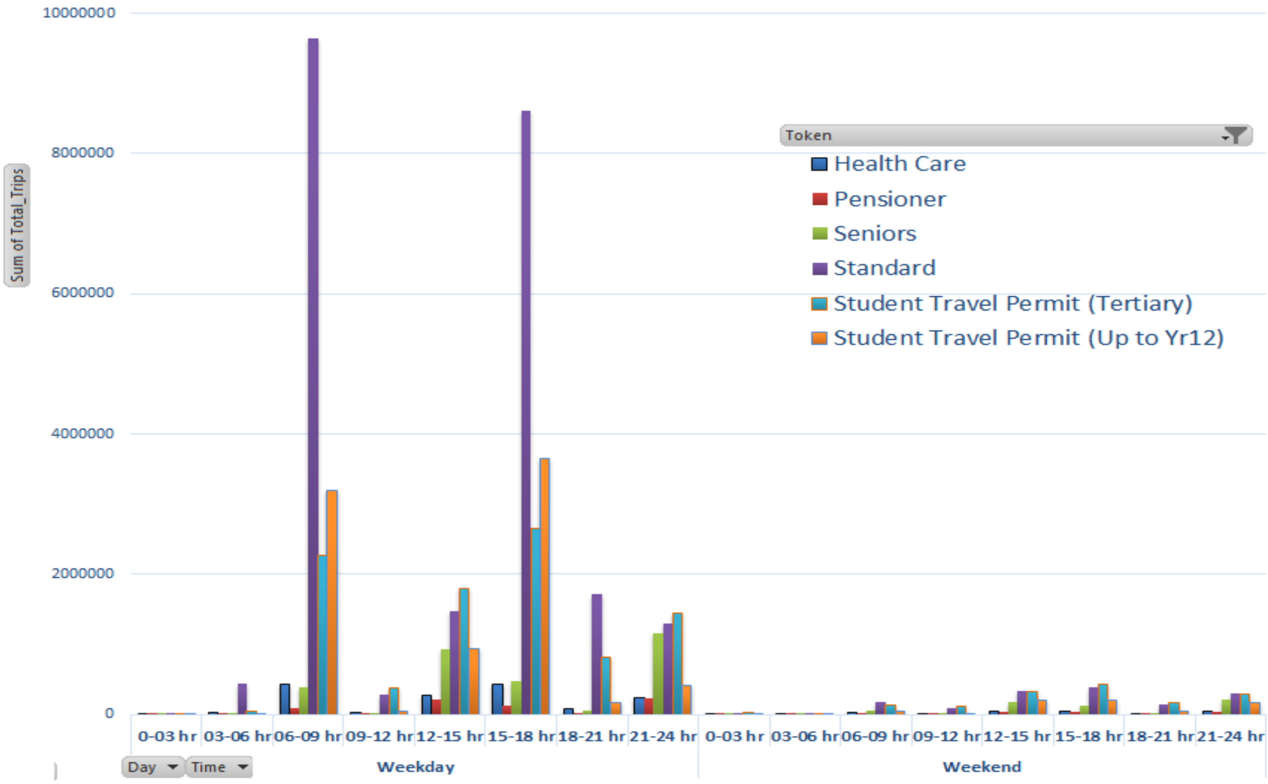


Figure 23: Number of trips by 3 hours periods of weekdays and weekends (by different patron groups)

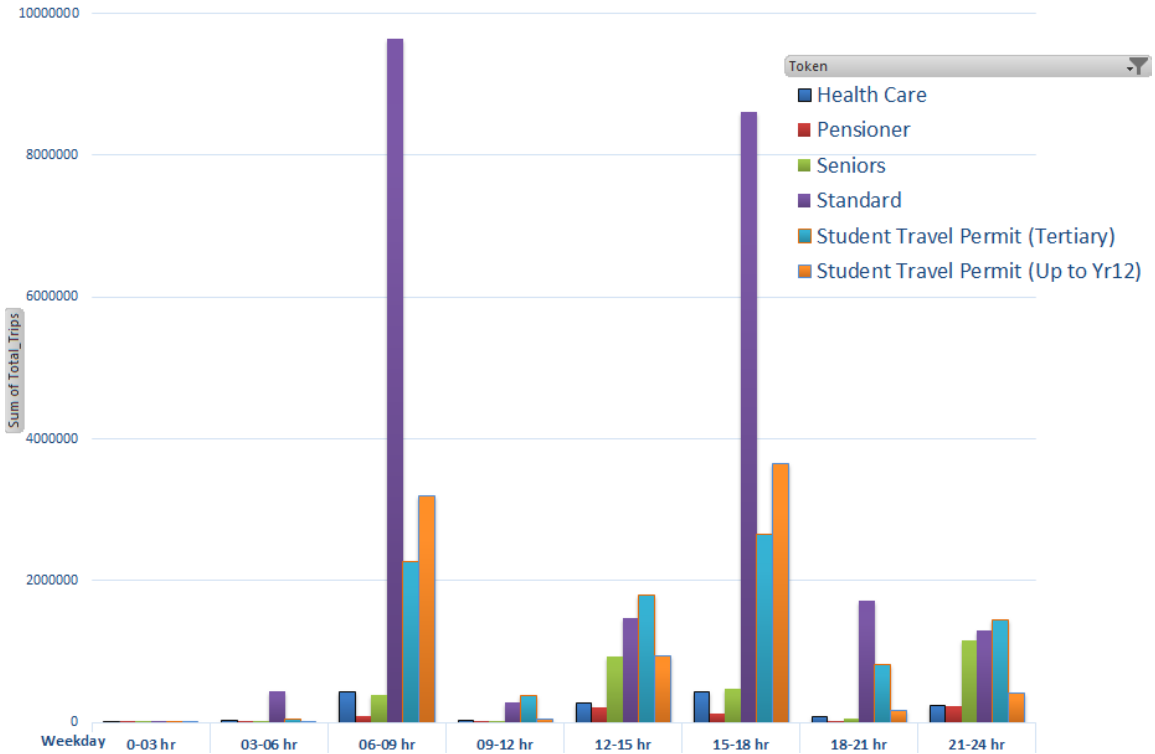


Figure 24: Number of trips by 3 hours periods of weekdays (by different types of patrons)

Figure 23 shows the comparision of number of trips by 3 hours period of weekdays and weekends by different patron groups. It is noticeable that public transport usage on weekdays is significantly different from weekends in 2009. In what follows, the weekday and weekend graphs are examined separately.

Figure 24 shows the hourly public transport usage patterns on weekdays in 2009. As expected for a metropolitan area like Perth suburb, 6-9hr and 15-18hr are the peak hours for public transport use while it is at a minimum during the 0-3 hr period, followed by 3-6 hr and 9-12 hr. The patrons using standard SmartRiders were the only transport users during 3-6 hr, one of the lowest overall usage periods. Public transport use by standard patrons during the morning peak hours (i.e. 6-9 hr) was approximately the same as the total usage during 15-21 hr. Their usage in morning and evening peak hours were remarkably higher (approximately more than twice) than the other patrons. Nevertheless, their use on weekdays dropped dramatically during 9-12 hr, after the morning peak hours were over. Finally, their usages are nearly the same during 12-15 hr and late evening periods (18-21 hr and 21-24hr).

On weekdays, the students (up to year 12) mainly used public transport during morning and evening peak hours. Public transport usage by students (up to year 12) was higher than that of univeristy students during 6-9hr and 15-18hr. The university students used public transport more than students (up to year 12) during the afternoon period 12-15 hr, as well as during evening periods (18-21hr and 21-24hr). Additionally, their usage in afternoon hours (12-15hr) was the highest among all patron groups. Moreover, their nighttime (21-24hr) public transport usage was even higher than their late evening one (18-21hr).

The other patrons' usage was relatively very low compared to that of standard and student groups. Pensioners' usage was the lowest, follwed by those using healthcare benefits. Interestingly, seniors' public transport use on weekdays was the highest in the 9pm- midnight period, followed by 12-15hr. When their usage data was drilled down in the datawarehouse which is developed for this research, it was found that Perth, Burswood and Frementle were the top 3 surburbs where they started their trips during 9pm-midnight throughout the whole year of 2009. Their usage during morning and evening peak hours were, by contrast, relatively low. Another noteworthy pattern here is that usage by seniors, standard riders, and students (up to year 12) was nearly the same during the period from 9pm till midnight.

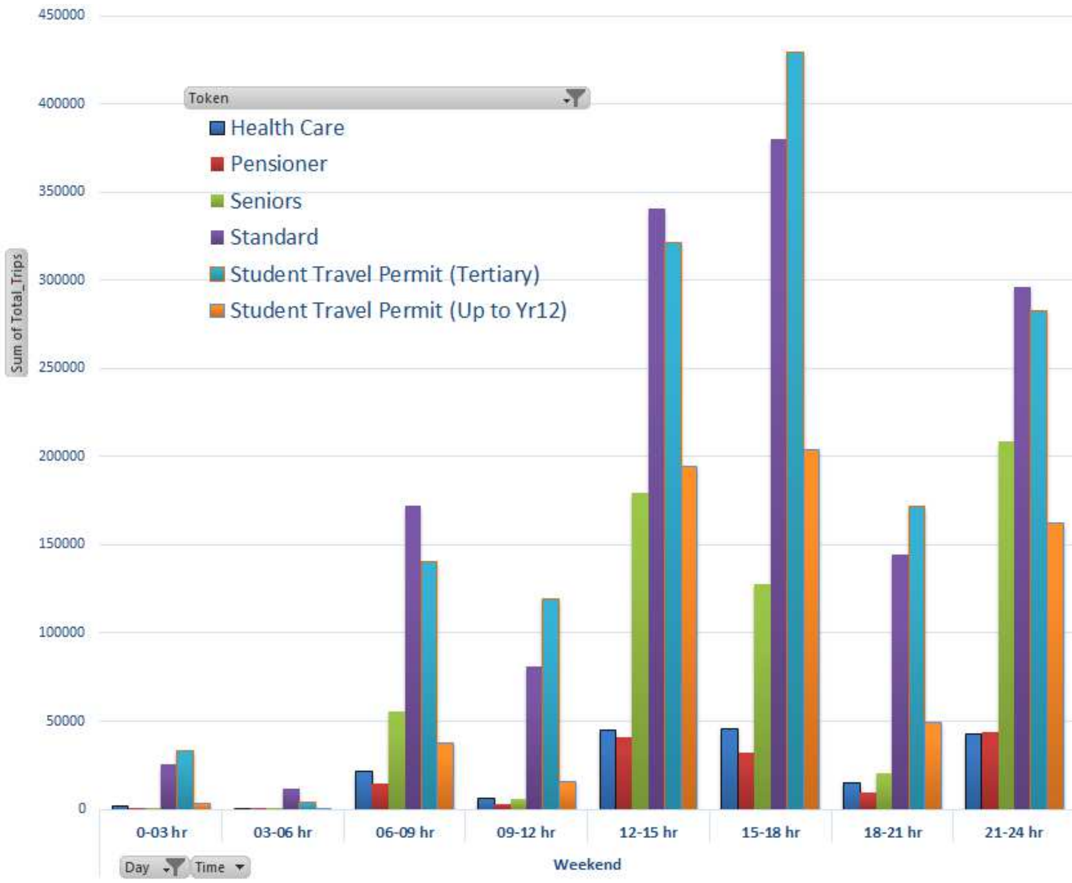


Figure 25: Number of trips by 3 hour periods on weekends (by different types of patrons)

The above Figure 25 illustrates the public transport usage patterns by different patron groups during 3 hours periods on weekends. It shows that the public transport usage patterns on weekends were quite different from the ones on weekdays. The only similarity is that the usage was the lowest during the 12am-3am and 3am-6am periods. Overall, the usage during the morning period 6am-9am was higher than 9am-12non. It then increased dramatically in all patron groups during 12noon-3pm and 3pm-6pm, before falling from 6pm-9pm. This usage then increased again in the night period (9pm-till midnight) to more than double that of the late evening period (6pm-9pm).

All patron groups followed the same public transport usage patterns on weekends. Additionally, standard patrons were the most frequent users during the morning (6am-9am), 12noon-3pm, and 9pm until midnight periods. University students were also frequent users on weekends. Usage by students (up to year 12) and seniors was relatively low compared to the first two groups.

Type of Patronages	Month	Day	Time	Suburb	Max Public Transport Usage
Standard	July	Weekday	15-18 hr	Perth	573283
Student Travel Permit (Tertiary)	March	Weekday	15-18 hr	Perth	86416
Seniors	October	Weekday	12-15 hr	Perth	27376
Student Travel Permit (Up to Yr12)	March	Weekday	15-18 hr	Perth	25649
HealthCare	March	Weekday	15-18 hr	Perth	18944
Pensioner	October	Weekday	12-15 hr	Perth	4875

Table 1: Maximum Public Transport Usage by Different Types of Patronages

Table 1 summarises the maximum public transport usage for different patrons across month, day, time and location of boarding. In 2009, standard patrons most frequently used public transport on weekdays in July from 3pm and 6pm, and boarded from Perth. The maximum public transport usage by university students, students (up to year 12) and healthcare patrons was from 3pm and 6pm on weekdays in March, and they also boarded in Perth. For pensioners and seniors, the highest public transport usage took place from 12noon to 3pm on weekdays in October. As the table indicates, Perth was the most frequent starting point for all patron groups—Perth being the suburb that provides the most public transport services.

5.1.3 Descriptive Analysis- Public Transport Usage and Its Use per Capita

Public transport use per capita is calculated as the total public transport usage in each suburb divided by the total number of residents, employees, and students—both university and up to year 12. A descriptive analysis of public transport usage and its use per capita will be provided in this section to gain a better understanding of their variations across metropolitan suburbs.

		Public Transport Usage	Public Transport Use per Capita
N	Valid	293.00	293.00
	Missing	0.00	0.00
Mean		173059.68	18
Standard. Deviation		771393.49	62.71
Variance		595047915016.44	3933.17
Skewness		15.62	14.74
Standard. Error of Skewness		0.14	0.14
Kurtosis		258.42	237.10
Standard. Error of Kurtosis		0.28	0.28
Range		12929447.00	1032.91
Minimum		12.00	0.01
Maximum		12929459.00	1032.92
Percentiles	10	4252.60	1.89
	20	12740.80	3.46
	25	17429.50	3.85
	30	24974.00	4.33
	40	37785.60	5.94
	50	53635.00	8.08
	60	73155.20	10.10
	70	112918.00	15.31
	75	132850.00	18.23
	80	194214.40	22.57
	90	471043.60	31.38

Table 2: Descriptive Analysis on Public Transport Usage

Table 2 illustrates that average public transport usage in Perth metropolitan suburbs is 173,060 trips per year. Public transport usage across suburbs ranges from 12 trips to 12,929,459 trips. Percentile values show that public transport usage for more than 75% of the metropolitan suburbs is below its mean value. Only 25% of these suburbs use the public transport more than the average of 173,060 trips per year. Additionally, public transport use per capita is 18 per annum. On average, Casuarina has the lowest public transport use with 0.01, while Karrakatta has the highest, with 1033 trips per year on average. Public transport use per capita in more than 70% of the metropolitan suburbs is below the average 18 trips per year. This implies that a quarter of the metropolitan suburbs are contributing to the bulk of total public transport usage.

The standard deviation of public transport usage is 771,393 trips, and it indicates that public transport usage in suburbs such as Perth, Fremantle, Murdoch, Bentley and Joondalup are far away from the average. Total public transport usage of Perth in 2009 is 12,929,459 trips and this contributes to the high standard deviation in the dataset. Standard deviation of public transport use per capita is also high (63) because Karrakatta, with its extreme value of 1033 trips, is pulling the distribution away from the center.

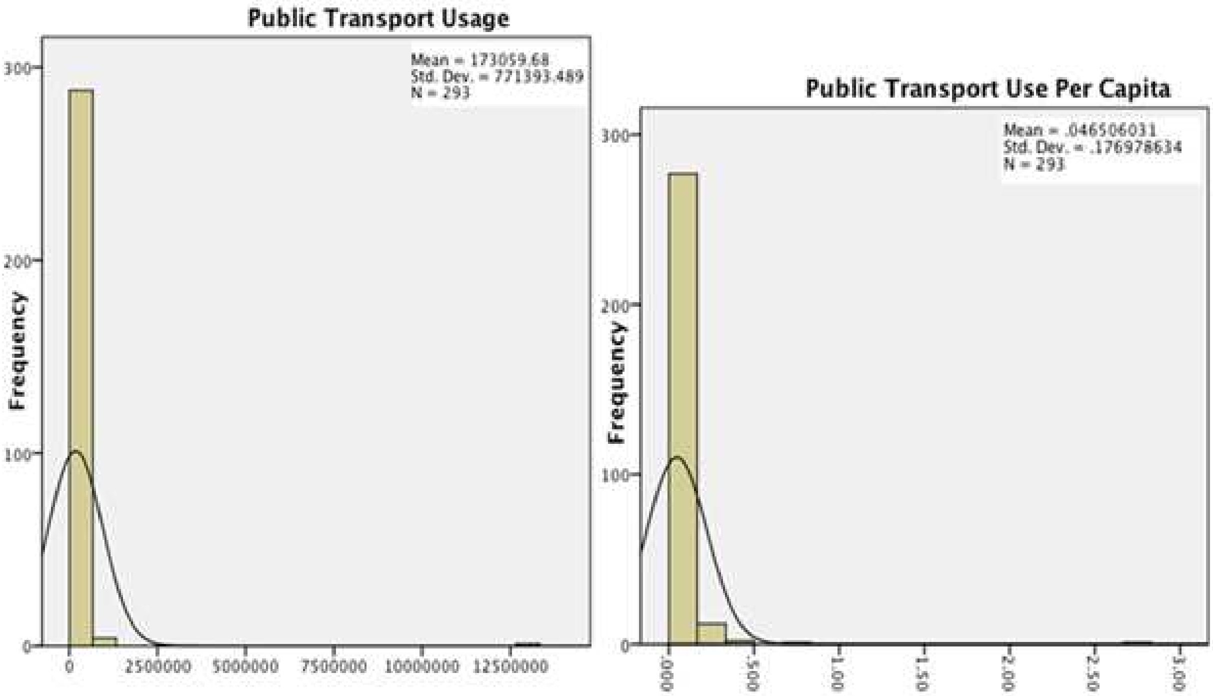


Figure 26: Histograms of Public Transport Usage and Public Transport Use per Capita

The values of skew and kurtosis for public transport usage and public transport use per capita are similar to those shown in Figure 26. They both have positive skewness (15.62 for public transport usage and 14.736 for public transport use per capita) and this indicates that their distributions are left skewed. Similarly, they both have distributions with positive kurtosis (leptokurtic distribution) so that many scores are in the tails and the curve is pointy, as illustrated in Figure 26.

5.1.4 Descriptive Analysis - Public Transport Service Provision Densities

Public transport service provision densities are calculated as total service frequencies at all stops in each metropolitan suburb divided by the urbanised area of the suburb. This section discusses how these public transport service provision densities vary across the Perth metropolitan suburbs.

Table 3 shows descriptive statistics for public service provision densities with 3 hours intervals and analysis is provided below by mode of transport.

			Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Bus/Ferry Service Provision Density	Weekday	12am-3am	0.00	10.91	1.0113	1.62740	2.231	6.646
		3am-6am	0.00	22.36	2.5810	3.23775	2.066	6.241
		6am-9am	.06	1292.58	81.2175	114.51289	5.544	47.797
		9am-12noon	0.00	1154.01	63.3005	97.80355	6.249	58.301
		12noon-3pm	0.00	1079.05	60.6611	92.24440	6.130	56.241
		3pm-6pm	0.00	1458.42	90.4215	126.59791	5.839	52.061
		6pm-9pm	0.00	454.34	44.4171	50.71418	3.043	16.716
		9pm-12midnight	0.00	220.35	16.2752	22.46994	3.752	25.070
	Saturday	12am-3am	0.00	49.98	1.5707	3.58705	9.101	115.436
		3am-6am	0.00	5.03	.2210	.73157	4.090	18.486
		6am-9am	0.00	361.20	25.8131	34.08970	4.592	35.407
		9am-12noon	0.00	645.16	41.4702	57.92652	5.381	45.379
		12noon-3pm	0.00	626.98	41.1700	57.15176	5.237	42.968
		3pm-6pm	0.00	581.55	39.8426	54.50740	4.988	38.997
		6pm-9pm	0.00	481.60	26.8972	39.80937	6.298	61.929
		9pm-12midnight	0.00	420.26	17.5766	31.60016	8.007	93.075
	Sunday	3am-6am	0.00	2.65	.0516	.28569	7.101	55.079
		6am-9am	0.00	56.09	5.8718	8.11449	2.606	9.370
		9am-12noon	0.00	308.95	25.6142	33.84936	3.858	23.026
		12noon-3pm	0.00	358.92	26.7543	36.22063	4.412	30.350
		3pm-6pm	0.00	370.28	26.2189	37.01268	4.598	32.234
		6pm-9pm	0.00	154.47	17.4485	22.72372	2.702	10.051
		9pm-12midnight	0.00	49.51	3.4220	7.47196	2.941	9.620
Train Service Provision Density	Weekday	12am-3am	0.00	13.63	.1856	.93191	11.198	152.143
		3am-6am	0.00	27.26	.3483	1.79108	12.225	177.267
		6am-9am	0.00	215.81	2.9481	14.82484	11.101	149.404
		9am-12noon	0.00	170.38	2.3461	11.65183	11.164	151.669
		12noon-3pm	0.00	165.83	2.3590	11.47421	10.860	144.789
		3pm-6pm	0.00	213.54	3.1234	15.22578	10.284	129.893
		6pm-9pm	0.00	138.57	2.0059	9.61507	10.761	142.873
		9pm-12midnight	0.00	86.32	1.1233	5.82020	11.580	160.835
	Saturday	12am-3am	0.00	43.16	.5938	2.92870	11.347	156.237
		3am-6am	0.00	13.63	.1430	.87006	13.260	200.564
		6am-9am	0.00	106.77	1.5763	7.50584	10.477	136.053
		9am-12noon	0.00	163.56	2.2355	11.18216	11.198	152.153
		12noon-3pm	0.00	163.56	2.2355	11.18216	11.198	152.153
		3pm-6pm	0.00	163.56	2.2341	11.18180	11.199	152.178
		6pm-9pm	0.00	113.58	1.5612	7.76688	11.172	151.788
		9pm-12midnight	0.00	86.32	1.1257	5.82100	11.574	160.725
	Sunday	3am-6am	0.00	9.09	.1122	.59442	12.372	180.666
		6am-9am	0.00	56.79	.7813	3.83930	11.404	158.018
		9am-12noon	0.00	156.75	2.2440	10.78127	11.000	148.120
		12noon-3pm	0.00	163.56	2.3051	11.20267	11.122	150.744
		3pm-6pm	0.00	163.56	2.3051	11.20267	11.122	150.744
		6pm-9pm	0.00	118.13	1.6317	8.01721	11.340	156.018
		9pm-12midnight	0.00	81.78	1.0981	5.59928	11.187	151.526

Table 3: Descriptive Analysis on Public Transport Service Provision Densities

Bus/Ferry Service Provision Densities

On weekdays, the highest average bus/ferry service provision density was in the 3pm to 6pm interval, with 90 service provisions per $km^{2.}$ The lowest average service provision was on Sunday from 0am 3pm, with 0.05 provisions per km^2. In weekdays, the bus/ferry services were second highest during 6-9am before gradually falling during 9am to 12 noon and 12-3pm periods. The same time-interval comparison of bus/ferry service provision densities on Saturday and Sunday shows that these services were approximately 35% less on Sunday from 9am to 9pm, but at a similar level on Saturday.

Similar to the overall average service provision densities, the highest variability in bus/ferry service provision densities are found during the weekdays from 3pm to 6pm, while the lowest is on Sundays from 12am to 3 am. The bus/ferry service provision in all time intervals has positive skewness, and the provision on Sunday 0am to 3am time interval has the largest skewness of 8.007. Additionally, they all have positive kurtosis, with the Saturday 12am to 3am time interval having the highest kurtosis value (115.436).

Train Service Provision Densities

The highest average train service is provided during weekdays from 3pm to 6pm which is 3.12, followed by weekday 6am to 9am (2.95) and weekday 12noon to 3pm (2.36). The lowest average train service is provided during Sunday 0-3am. The data shows that the train service provision within the time period for 3-6pm during the weekday has the highest level of variability. In weekends, the average train services are provided slightly more frequently on Sunday than on Saturday within the time periods of 9am to 9pm. In addition, the train service provisions for all different time intervals have positive skewness and kurtosis with the largest skewness (13.26) and highest kurtosis (200.564) during 3-6pm on Saturday.

5.1.5 Descriptive Analysis of Land use characteristics Variables

In this section, urban form variables will be discussed. The variables taken into account in this study are estimated resident population densities by age and gender; employment densities in various industries; students (up to year 12) population densities and university students population densities; road length (in km) per km2; and distance from the city center.

5.1.5.1 Estimated Resident Population Densities

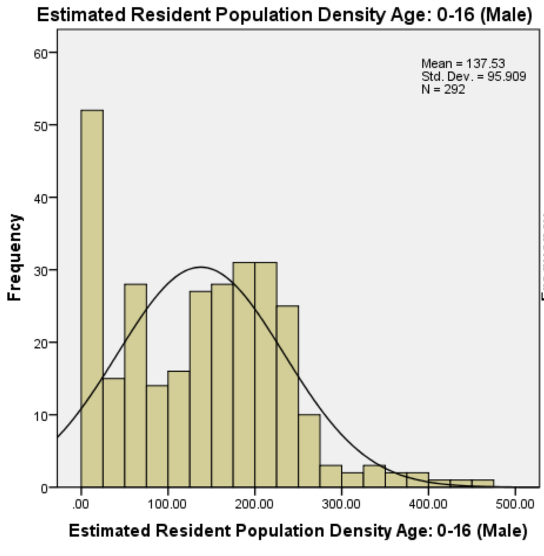
		Estimated Resident Population Density Age:0-16 (Male)	Estimated Resident Population Density Age:0-16 (Female)	Estimated Resident Population Density Age:17-35 (Male)	Estimated Resident Population Density Age:17-35 (Female)	Estimated Resident Population Density Age:36-64 (Male)	Estimated Resident Population Density Age:36-64 (Female)	Estimated Resident Population Density Age:65 and over (Male)	Estimated Resident Population Density Age:65 and over (Female)
N	Valid	292	292	292	292	292	292	292	292
	Missing	0	0	0	0	0	0	0	0
Minimum		0	0	0	0	0	0	0	0
Maximum		466.71	406.55	999.35	1023.84	969.96	825.44	296.38	481.89
Mean		137.53	129.13	204.52	190.97	247.64	248.05	75.04	93.35
Std. Error of Mean		5.61	5.26	9.94	9.19	9.85	9.76	3.31	4.59
Std. Deviation		95.91	89.89	169.85	157	168.35	166.73	56.58	78.47
Variance		9,198.60	8,080.57	28,849.73	24,650.21	28,341.17	27,800.02	3,201.59	6,157.81
Skewness		0.39	0.34	1.33	1.29	0.27	0.09	0.57	0.98
Std. Error of Skewness		0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
Kurtosis		-0.07	-0.37	3.24	3.48	-0.05	-0.81	0.11	1.49
Std. Error of Kurtosis		0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
%	10	6.07	5.9	5.95	5.43	12.43	10.14	3.33	2.96
	25	56.76	52.66	65.97	61.77	101.44	97.2	23.78	24.45
	50	145.48	135.15	198.93	184.01	258.96	256.65	69.66	80.12
	75	207.72	194.46	290.56	274.74	376.59	391.56	116.2	138.38

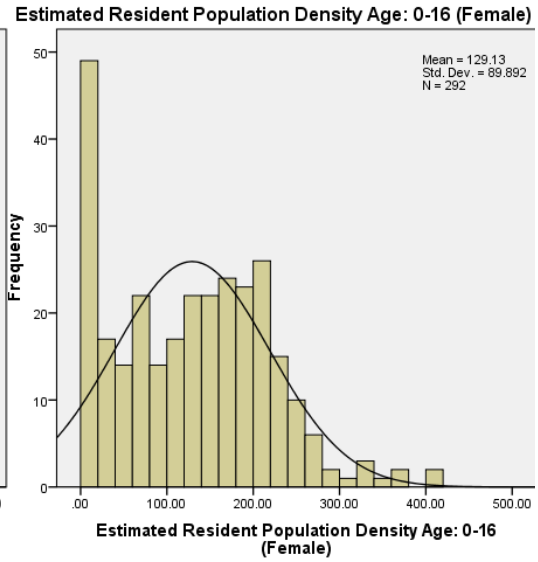
Table 4 Descriptive Analysis on Estimated Resident Population Density per Km2

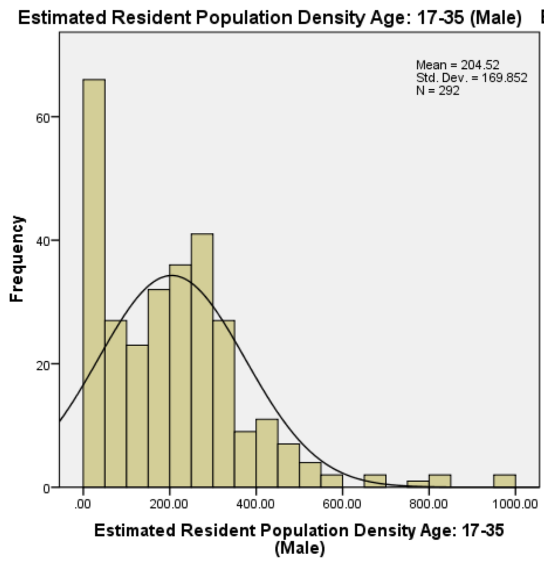
The estimated resident population density per km^2 for each age group is calculated as the number of residents in a particular suburb divided by its urbanised area. Among the eight different age/gender groups, the mean of estimated resident population is the highest at 248 per km^2 for 36-64 age female group, followed by 36-64 aged female group (mean=247.64) and 17-35 age male and female groups (mean= 204.52 and190.97 respectively). The age group with the lowest mean is residents 65 and older (75.04 for males and 93.35 for females). About 50% of the Western Australian metropolitan suburbs have approximately the average estimated resident population density for all eight age/gender groups.

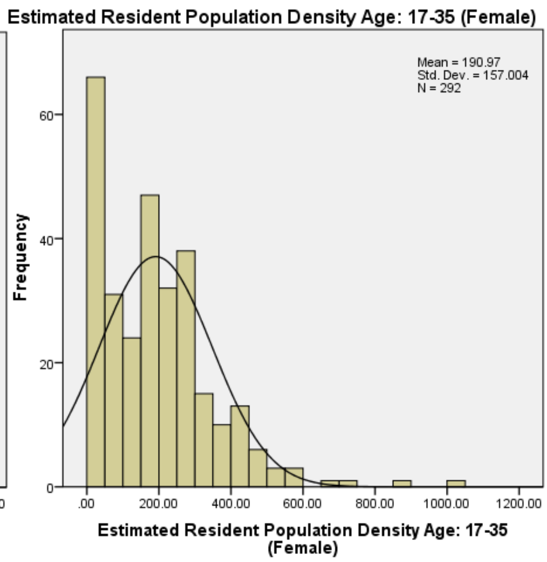
Additionally, the variation of estimated resident population density is the highest for males age 17-35 at 169.85. The values are also relatively high for the groups age 36-64, 168.35 for males and 166.73 for females. Those in the 0-16 age groups have a comparatively low standard deviation: 95.91 for males and 89.89 for females. Among these groups, males 65 and over have the lowest standard deviation of 56.58, while women 65 and over have the second lowest standard deviation of 78.47.

The skewness values for all eight groups range from -3 to 3. All of their kurtosis values are also within the -3 and 3 range. Therefore, it can be concluded that the estimated resident population density variables are normally distributed and satisfy the regression method's assumption for normality.









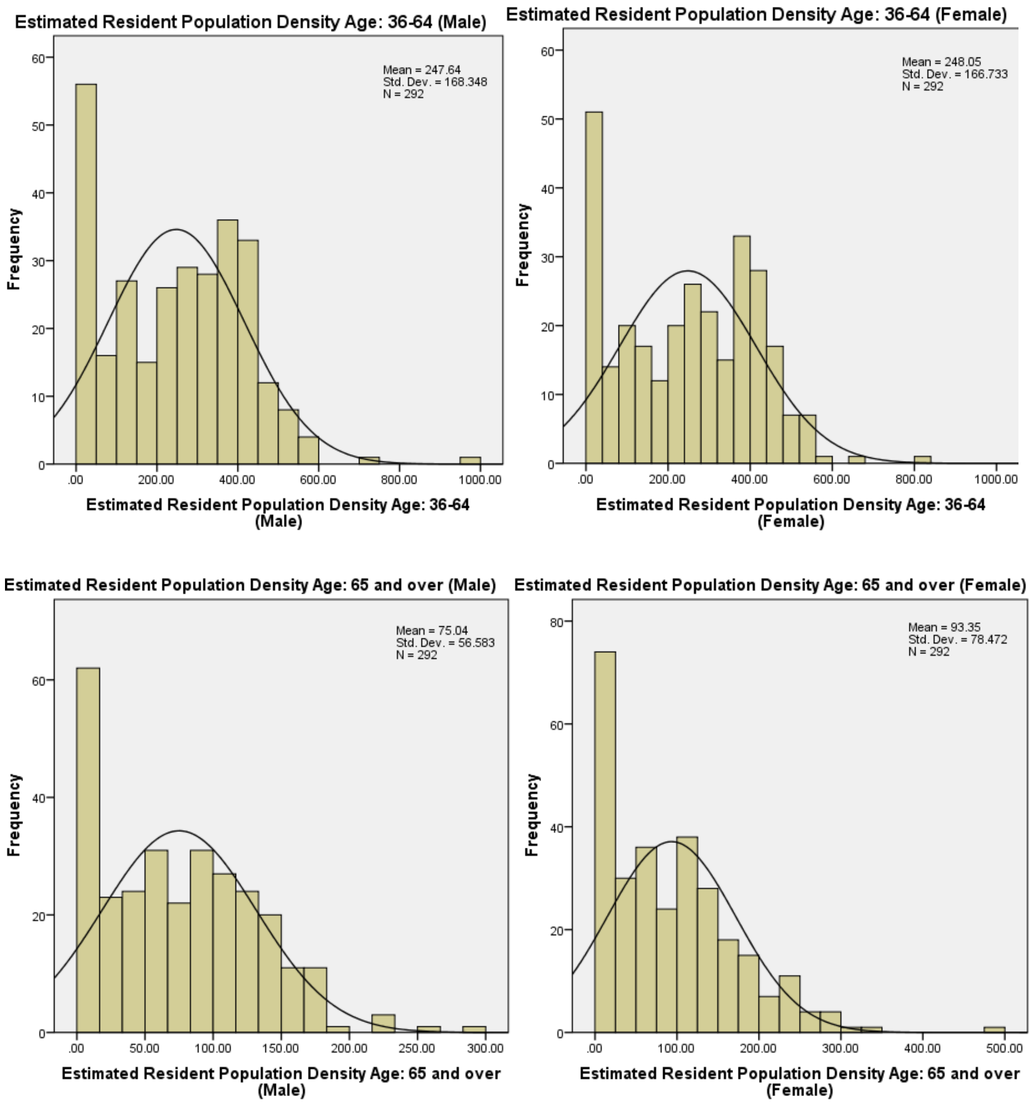


Figure 27: Histograms of Estimated Resident Population Densities in Different Age/Gender Groups

The estimated resident population densities for all age-by-gender groups have positive skewness, with males aged 17-35 having the highest skewness (1.33), followed by females aged 17-35 at 1.29. The values of kurtosis for the 17-35 and 65 and over age groups are positive, while the other two age groups have negative kurtosis. Overall, there are no significant differences in population density among the different age-by-gender groups. Therefore, it can be concluded that the sharpness of the distributions shown in Figure 27 are not significantly different from each other.

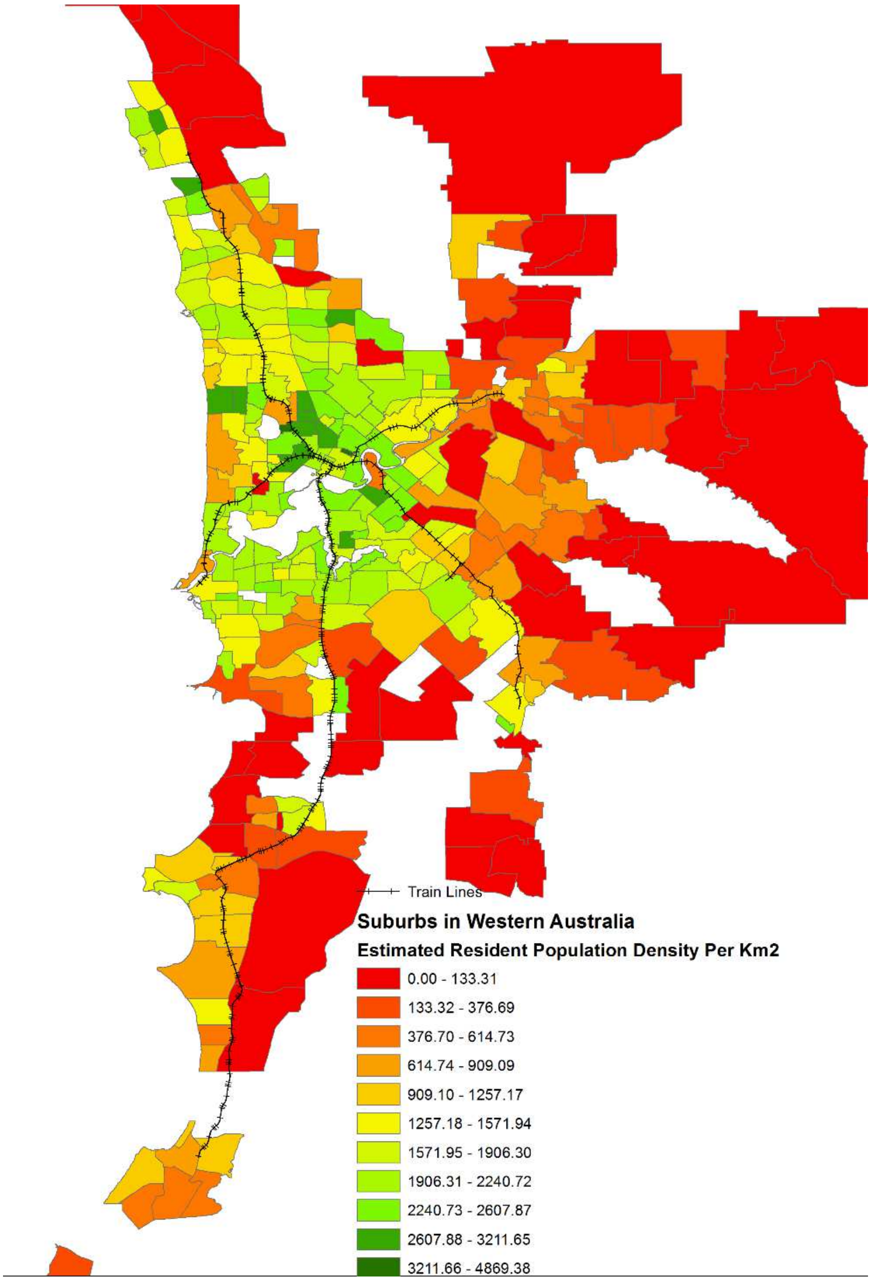


Figure 28: Estimated Resident Population Density per km^2 Map

Figure 28 illustrates that the suburbs with higher estimated resident population densities per km^2 are the ones closer to the city center. Nevertheless, there are some outer suburbs that have high estimated resident population densities. These are Ridgewood, Kinross, Alexander Heights, Scarborough, Doubleview, Tuart Hill, Yokine, Wembley, Glendalough, Subiaco, East Victoria Park, Victoria Park and Karawara. The train lines in this map suggest that the most parts of several train lines, namely the Midland, Armadale and Mundurah lines, do not serve the highly populated areas. The areas with high population densities but without train service are in the North Eastern and South Western metropolitan regions.

5.1.5.2 Employment Densities

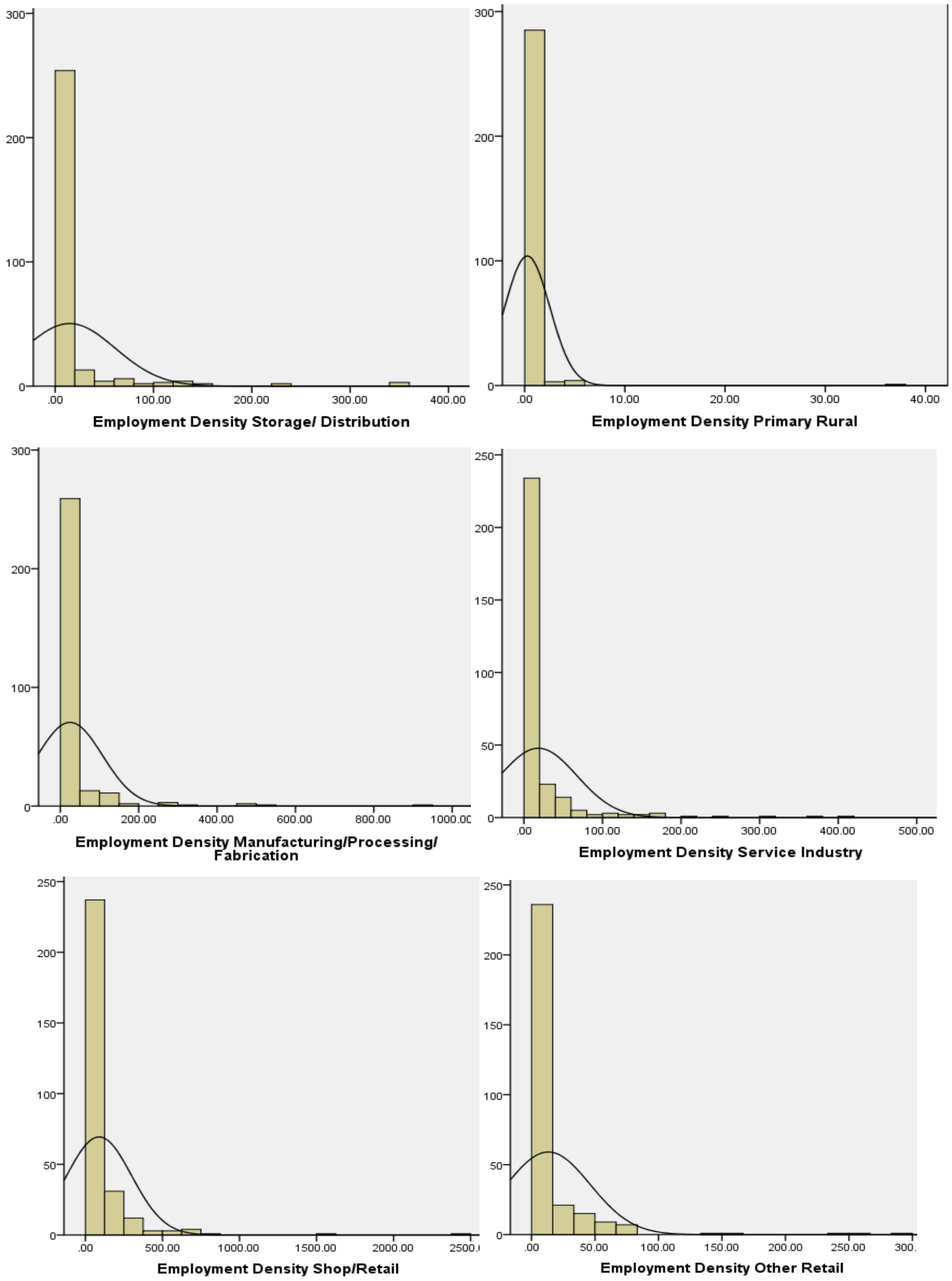
	Mean	Std. Deviation	Skewness	Kurtosis	Min	Max	Percentiles			
Employment Density							25	50	75	80
Primary/Rural Industry	0.26	2.25	15.13	245.18	0	36.99	0	0	0	0
Manufacturing/ Processing/ Fabrication	24.72	82.94	6.59	55.13	0	911.97	0	0.6	10.9	17.83
Storage/ Distribution	14.71	46.46	5.09	30.03	0	356.12	0	0	4.93	8.7
Service Industry	18.42	48.99	4.84	27.94	0	401.44	0	1.38	12.9	20.22
Shop/Retail Industry	89.2	210.79	6.72	61.98	0	2432.96	2.3	23.7	89.9	117.7
Other Retail Industry	13.27	33.03	5.31	35.14	0	286.39	0	1.85	9.94	16.5
Office/ Business Industry	215	887.64	8.28	79.9	0	10530.8	5.1	24.2	86.1	121.5
Health/ Welfare/ Community Services	42.24	145.59	9.19	99.52	0	1846.38	2.7	11.8	31.5	40.65
Entertainment/ Recreation/ Culture	18.7	78.79	11.57	155.48	0	1154.01	0.1	2.95	13.9	18.7
Residential Industry	6.49	30.15	7.63	64.25	0	293.91	0	0	0.96	2.26
Utilities/ Communications	6.47	28.11	12.05	174.23	0	427.15	0	0.3	2.72	4.08

Table 5: Employment Density per km2

Employment density per km^2 is calculated by dividing the employment population of each industry in a particular suburb by its urbanised area. Table 5 indicates that the average employment density in primary/rural industry is the lowest among all industries at 0.26. Following closely behind are employment densities in residential and utilities/communication industries at about 6.5 per km^2 each. On the other hand, the average employment density per km^2 in office/business industry is the highest at 215, followed by shop/detail and health/welfare/community industries, which are at 89 and 42, respectively.

In addition, 75% of the Western Australia metropolitan suburbs have below average employment density in shop/retail industry; 20% of these suburbs have greater-than-average employment densities in service, other retail, and entertainment, recreation and culture industries. The employment figures for primary/rural, manufacturing/processing/ fabrication, storage/distribution, office/business, health/ welfare/ community service, residential and utilities/ communication industries fall below the average in 80% of the metropolitan suburbs.

The employment density variation is the highest for office/ business industry, which has the highest standard deviation (887.64). The employment density of the shop/ retail industry also has a relatively high level of variability (210.79), followed by health/ welfare/community service at (145.59). Furthermore, the levels of variation in employment densities for manufacturing/ processing/ fabrication and entertainment/ recreation/ culture industries are relatively similar, at 82.94 and 78.79 respectively. The very close standard deviation values of employment densities in the service and storage/ distribution industries (48.99 and 46.46, respectively) suggest that the employment populations in these two industries are similarly distributed across the Western Australia metropolitan suburbs. Additionally, the other retail, residential, and utilities/ communications industries have relatively similar levels of variation in their employment densities (33.03, 30.15 and 28.11 respectively). Primary/rural industry has both the lowest average employment density (0.26) and also the lowest standard deviation (2.25).



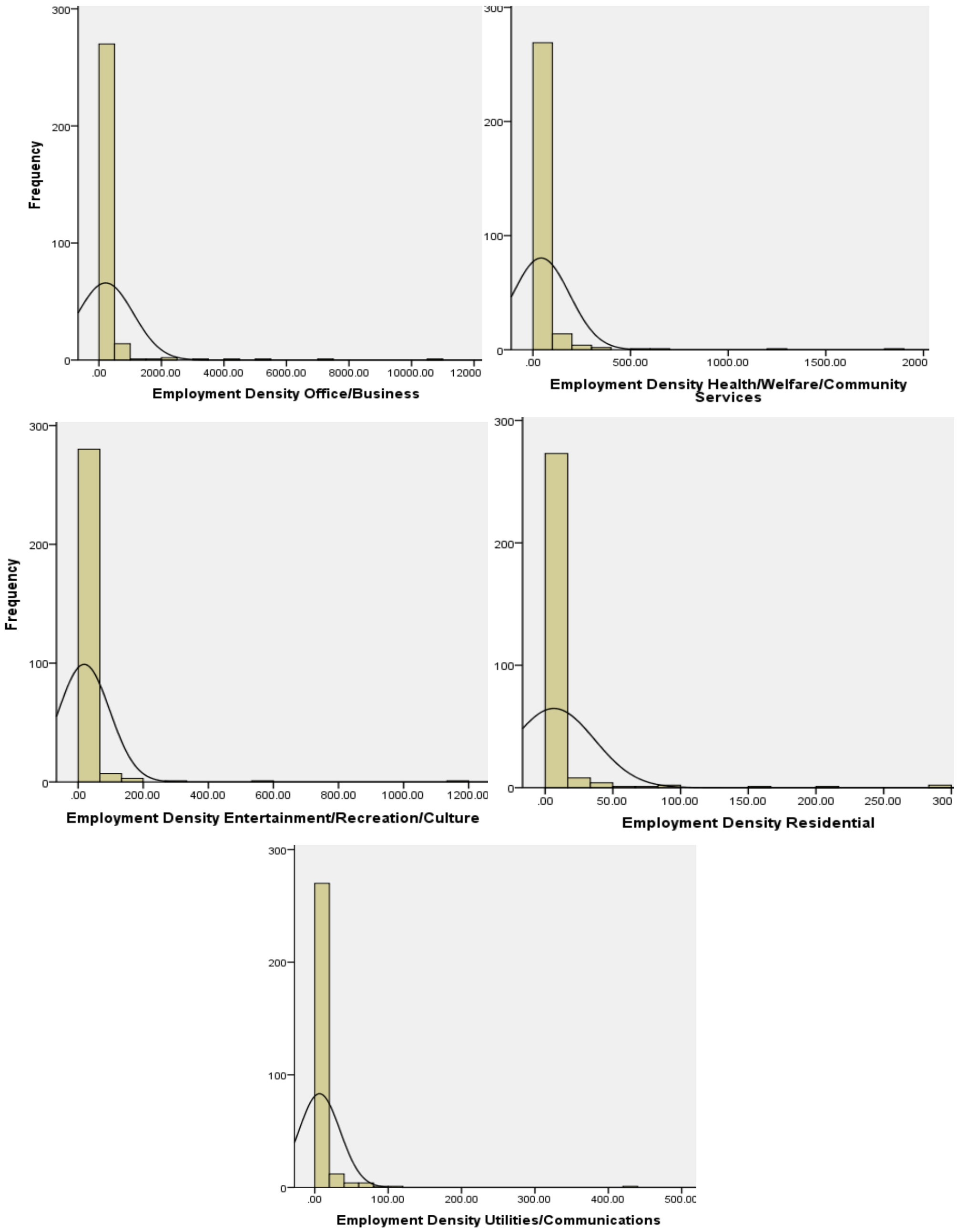


Figure 29: Histograms of Employment Densities in Various Industries

As shown in Figure 29, the employment densities in all industries have positive skewness. It is the highest in primary/ rural industry at 15.16 and lowest in storage/ distribution industry at 5.10. The kurtosis values for all of these industries are positive, the highest being in primary/rural industry (246.01) and lowest in service industry (28.04). These kurtosis values can be seen in the sharpness of the distribution curves in Figure 29 (above). Therefore, these employment densities data need to be transformed and normalized. This is done in section 5.3.3 Data Transformation.

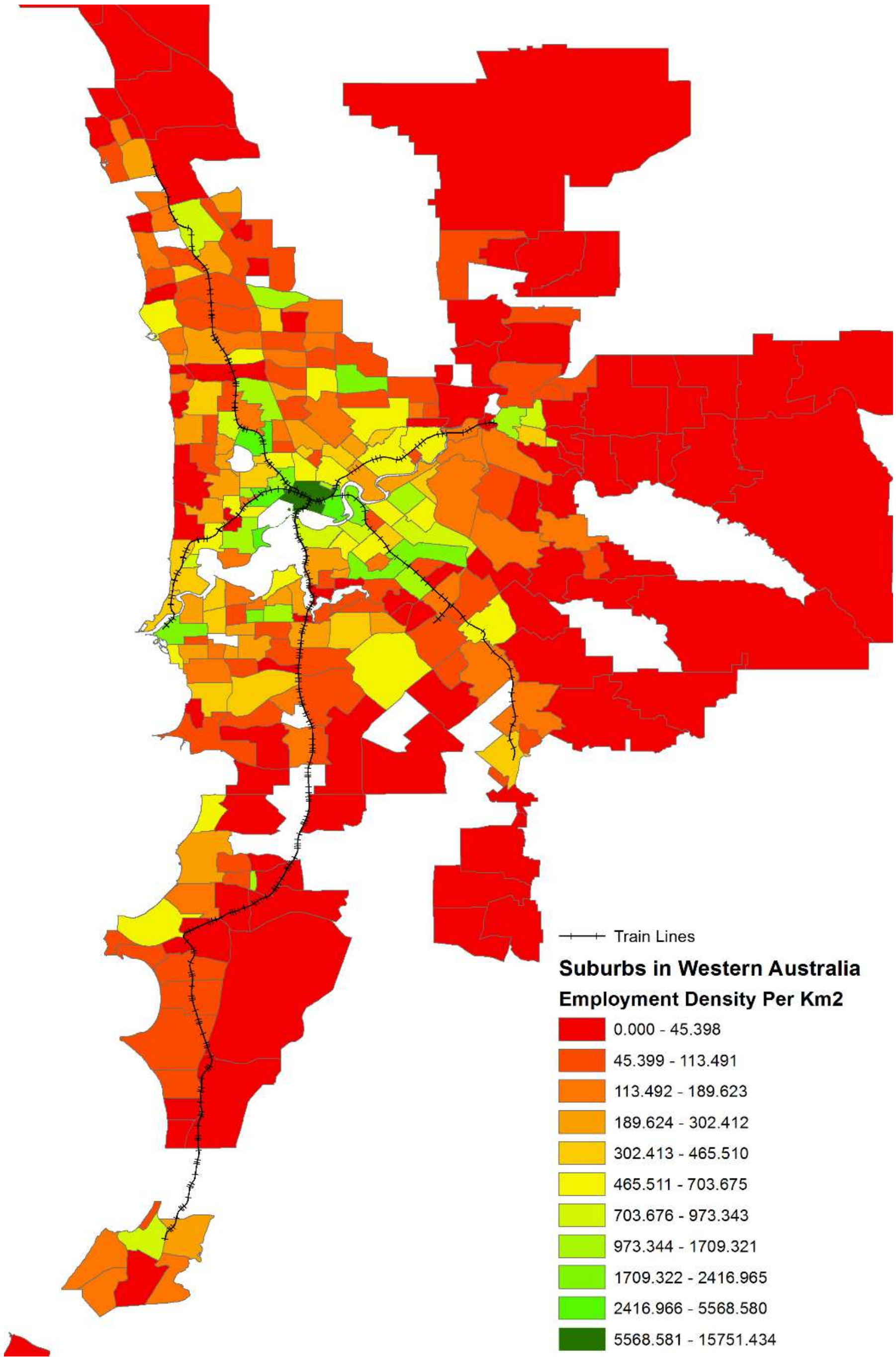


Figure 30: Employment Density per Km^2 Map

Figure 30 (which figure does this refer to?) shows that employment is highly dense in the central business district area and inner suburbs. The suburbs with the highest employment densities per km^2 (above 5569) are Perth, West Perth and Northbridge. In particular, the employment densities in the office/ business, health/ welfare/ community service, entertainment/ recreation/ culture, and utilities/ communication industries are very high in these suburbs. Employment densities per km^2 in Crawley, Bentley and Fremantle are also high, ranging between 2417 and 5569. This is due to the fact that these suburbs have universities and hospitals. The suburbs with industrial areas, such as Osborne Park, Wangara, Malaga, Leederville, West Leederville, Welshpool, O'Corner, Myaree, Midland, Kwinna Town Center and Cannington, also have employment densities above 1709 per km^2.

5.1.5.3 University Students and Students (up to year 12) Population Densities

This section examines the variations in the university student and student (up to year 12) population densities. Densities are calculated by dividing the total student population in a particular Perth metropolitan suburb by its urbanised area.

		University Student Population Density	Student (up to year 12) Population Density
N	Valid	292	292
Mean		101.8173	198.7401
Median		0.0000	113.2390
Std. Deviation		970.02039	284.76692
Variance		940939.553	81092.202
Skewness		12.611	3.212
Kurtosis		175.953	14.571
Minimum		0.00	0.00
Maximum		14534.64	2259.37
Percentiles	25	0.0000	2.5688
	50	0.0000	113.2390
	65	0.0000	190.3910
	75	0.0000	276.5248
	97	0.0000	989.2026
	98	1060.5715	1214.5793

Table 6: Descriptive Analysis of University and Student (up to year 12) Population Densities

Table 6 indicates that the average population densities for university students and students (up to year 12) are 101.82 and 198.74, respectively. The standard deviation of university student population density is significantly high at 970, with a minimum 0 and maximum 14,534. A high proportion, namely 97% of the metropolitan suburbs have the below average university student population densities because there are only seven suburbs in which universities are located. Crawley, where the University of Western Australia is located, has the highest university student population density at 14,534.64, followed by Bentley (Curtin University) at 6294.63, Murdoch (Murdoch University) at 4195.79, and Mount Lawley (Edith Cowan University) at 2379. The location of Notre Dame University in Fremantle is not that pronounced.

The standard deviation of student (up to year 12) population density is 284.77, with a minimum of 0 and maximum of 2259.37. Approximately 75% of the Perth metropolitan suburbs have below average student (up to year 12) population densities.

Churchlands, Highgate, Claremont, Peppermint Grove and Rossmoyne have the highest student (up to year 12) population densities at 2259.37, 1654.56, 1477.12, 1460.59 and 1237.65, respectively. Percentile results for student (up to year 12) population densities also show that 75% of the metropolitan suburbs have 276 students per km^2. Only 2% of these suburbs have more than 1214 students per km^2. The histograms of student (up to year 12) population densities also confirm the large variation between a very few suburbs with student (up to year 12) and the majority of the Perth metropolitan suburbs.

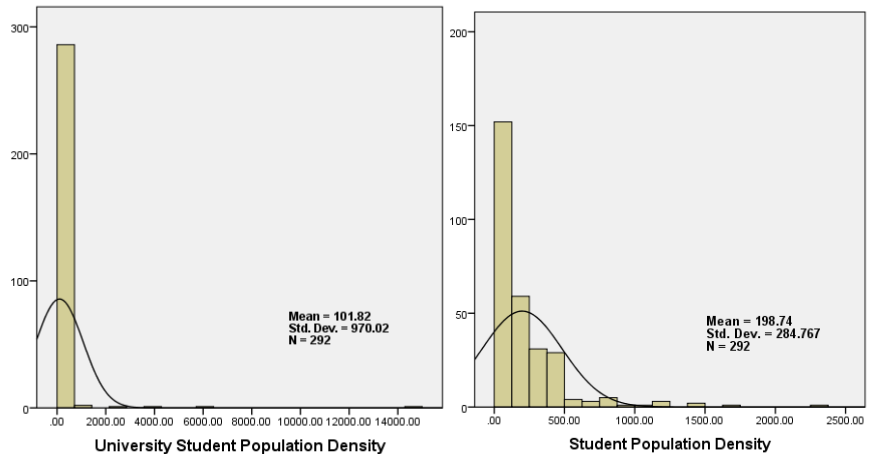


Figure 31: Histograms University and Students (up to year 12) Population Densities

5.1.5.4 Road Length (in km) per Km2 and Distance from City Center

The other urban form variables included in this research to explain the public transport usage in Perth metropolitan suburbs are road length (in km) per area (km^2) and distance from city center. The findings from descriptive analysis on these variables are follow:

		Road Length (in km) per Km2	Distance from City Center
Mean		9.71	18.95
Std. Deviation		4.55	13.87
Skewness		-0.22	1.57
Kurtosis		-0.54	2.94
Minimum		0.11	0
Maximum		22.25	79.29
Percentiles	10	2.73	5.3
	20	4.97	7.91
	30	7.48	10.59
	40	9.33	12.65
	50	10.65	15.1
	60	11.63	18.49
	70	12.53	21.73
	80	13.36	27.73
	90	14.51	36.18

Table 7: Descriptive Analysis on Road Length per Km2 and Distance from City Center

As shown in Table 7, the average road length per km^2 is 9.71km. Road length (per km^2) in West Perth is the highest at 22.25km and lowest in Chidlow at 0.11km. Road length variability is relatively low at 4.55. As can be seen in Figure 32, the road length per km2 in the Perth metropolitan suburbs has a bimodal distribution, even though the normal distribution curve seems to be symmetric.

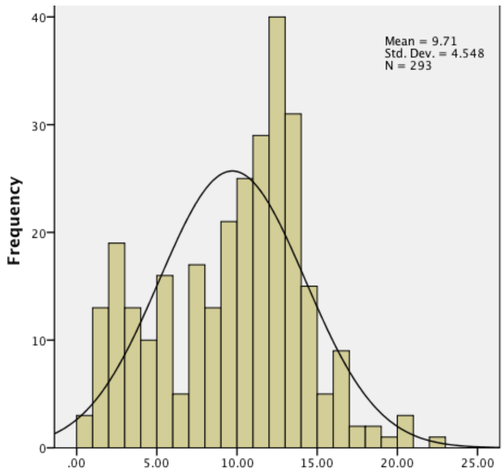


Figure 32: Histogram of Road Length (in km)per km2

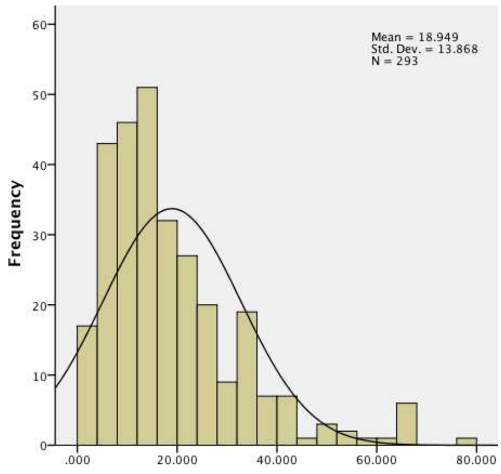


Figure 33: Histogram of Distance from City Centre

The average distance from the city center for all suburbs is 18.95 km. Dawesville is the farthest suburb, located 79.29km from the city center, while Northbridge is the closest at only 0.11 km away. In addition, 70% of the Perth metropolitan suburbs have below average distance from the city center. For 90 % of these suburbs, the longest distance is only 36.18 km. Therefore, most of the variance is accounted by only 10% of the suburbs. Figure 33 illustrates this large variation in a few suburbs. The number of suburbs that are more than 40 km away from the city center is relatively low. Distance from the city center in the case of Perth metropolitan suburbs has a positively skewed distribution and a relatively low positive kurtosis (2.94).

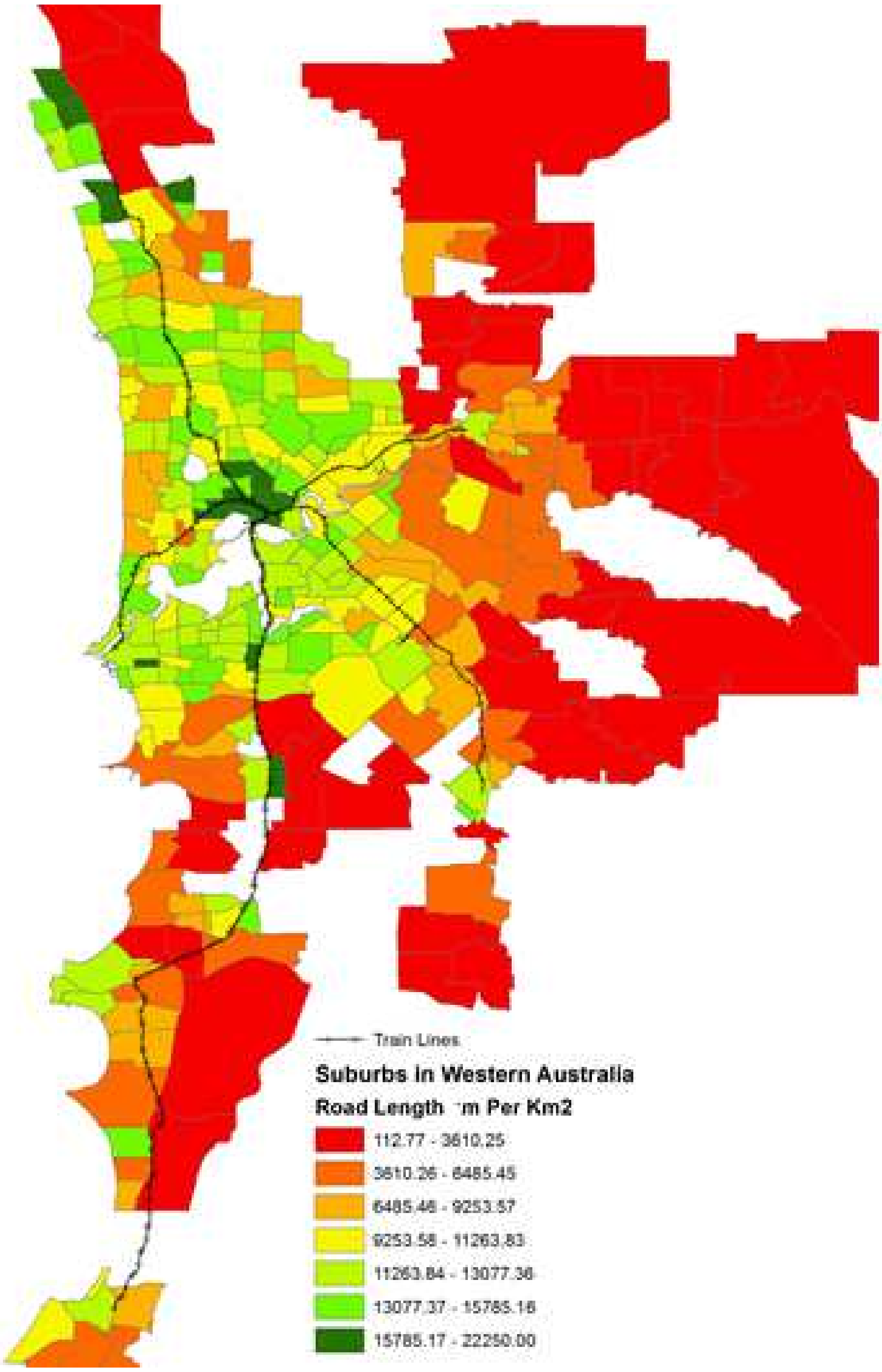


Figure 34: Road Length (m) per Km2

Figure 34 shows that more roads are built in well-developed areas such as Butler, Ridgewood, and Tapping in the northern metropolitan area; in North Perth, Mount Hawthorn, Joandanna, Subiaco, West Perth, Perth and East Perth in the central metropolitan area; and in Winthrop, Beaconsfield and Atwell in the southern region. Piara Waters, Banjup and Wandi suburbs are not well-developed with less roads even though they are relatively close to the city center.

5.1.6 Descriptive Analysis of Socio-Economic Variables

The socio-economic variables examined in this study, with the aim of developing a predictive model of public transport usage, are: number of residents in different income groups; average weekly rent; and average car ownership per household in Perth metropolitan suburbs in 2009. This section provides a descriptive analysis of these socio-economic variables.

5.1.6.1 Number of Residents in Different Income Groups

The number of residents is counted according to four different income groups, namely with weekly earnings below $250, between $250 and $999, between $1000 and $1999 and equal to or above $2000. Instead of using average weighted income as aggregated data, these four different income groups are used to find out how each of these income groups is correlating with public transport usage differently.

		No or residents whose weekly income				Average Weekly Income
		below $250	between $250 & $999	between $1000 & $1999	Equal to or above $2000	
Mean		766	1746	650	172	712
Std. Deviation		652	1474	562	207	142
Variance		425090	2173520	316238	42966	20131
Skewness		2	2	2	2	0
Std. Error of Skewness		0	0	0	0	0
Kurtosis		4	4	4	5	4
Std. Error of Kurtosis		0	0	0	0	0
Maximum		3740	8618	3605	1170	1206
	25	313	712	236	37	620
	50	626	1425	496	90	705
	60	750	1741	623	131	743
	70	936	2187	832	184	775
	75	1035	2465	945	221	799
	80	1176	2647	1076	272	815
	90	1632	3641	1358	419	885

Table 8: Descriptive Statistics on Weekly Income Groups

Table 8 shows the descriptive analysis for the number of residents in each weekly income group. The weekly income group between $250 and $999 has the highest average number of residents at 1746, while the weekly income equal to or above $2000 group has the lowest at 172. It is noticeable that 70% of the Perth metropolitan suburbs have a below average number of residents in all different income groups. The highest standard deviation, and therefore variation, is in the weekly income between $250 and $999 group (1474), while the income equal or above $2000 group has the lowest standard deviation at 207.

The skewness values for the number of residents in all different weekly income groups are positive, ranging from 2.13 to 1.62. All of these income groups have positive kurtosis, and these values are not substantially different from each other, with the highest value at 4.85 and lowest at 3.72.

Residents living in industrial areas such as Welshpool, Kwinana, Perth Airport, Karrakatta, Naval Base, Malaga, and Wangara span all income groups. People who are weekly earnings below $250 mostly live in Dianella, Thronlie, Morley, Gosnells and Gellajura. Additionally, people with weekly income between $250 and $999 live in Thornlie, Dianella, Morley, Gosnells and Canning Vale. Similarly, those with weekly income between $1000 and $1999 also mostly live in Canning Vale, Dianella, Thornlie, Willetton and Duncraig. But those with high weekly income equal or above $2000 mostly live along the Swan River or the western coastal line, in places such as Nedlands, South Perth, Subiaco, Cottesloe and City Beach.

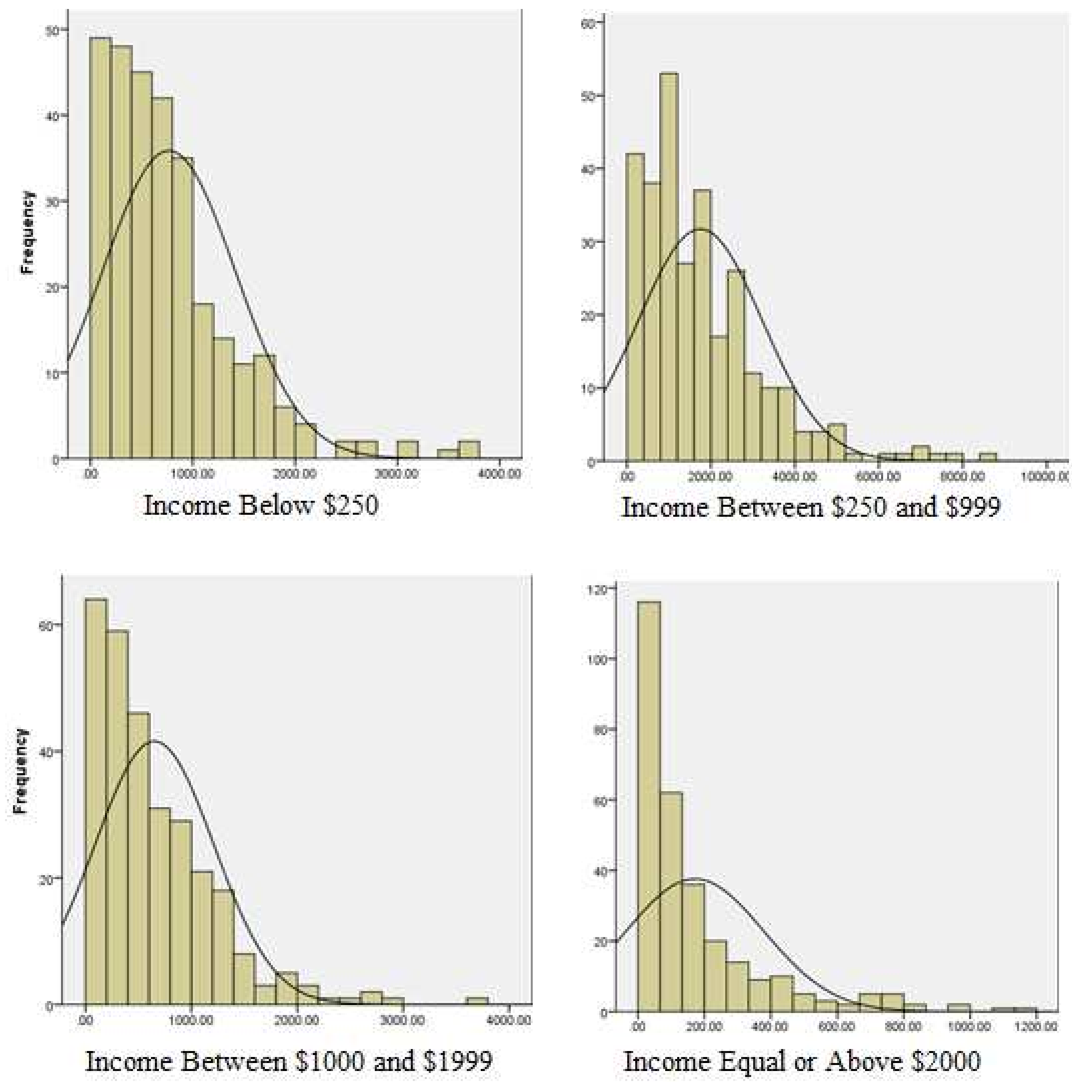


Figure 35: Histograms of Number of Residents in Different Weekly Income Groups

The resulting Figure 35 shows the comparison of frequency distributions in different income groups. All of these histograms are asymmetric and positively skewed with the higher frequency scores clustered at the lower end, and the tail pointing toward the higher scores. This indicates that larger proportions of population earn less toward the lower bound of the income ranges for all groups.

The variance among kurtosis values of all income groups is relatively small. The distributions for all income groups have similar peak levels of around 40. Therefore, it can be concluded that all normal distribution curves have similar pointiness and divergent dispersions. Since all of these income variables have similar patterns, an initial check is made to determine whether there is any co-linearity among them before using multiple regression. This is followed by a factor analysis to the calculate income factor scores without losing the richness of different income groups' data. The average weighted income is used here for comparative analysis of income dispersion in the Perth metropolitan suburbs. However, in the subsequent regression analysis, factor scores are used as variables, since they maintain the contributions from each income group.

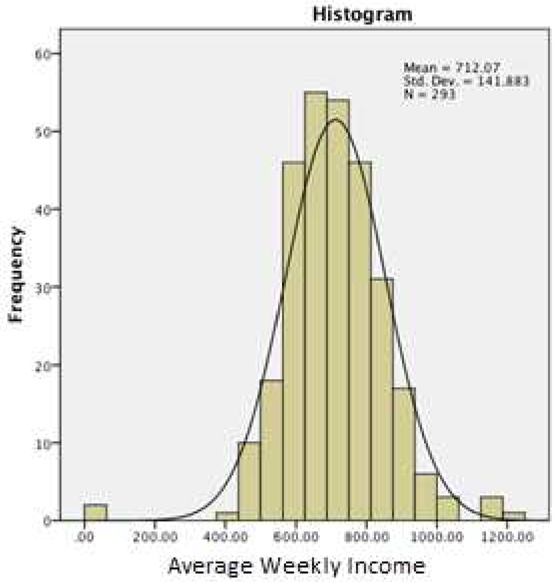


Figure 36: Histogram of Average Weekly Income

Figure 36 shows the frequency distribution of the average weighted weekly income across Perth metropolitan suburbs. The normal curve suggests that the average weekly incomes in all suburbs are distributed symmetrically, ranging from $200 to $1200, with the majority of them lying around the center. This indicates that people in the majority of suburbs are earning about $700 weekly. The frequency distribution of average weekly income is analysed to obtain the better understanding of income distribution across all studied suburbs. But the number of residents in different weekly income groups is used in this research to explore the latent factors among these income variables and the estimated resident population densities by age and gender variables.

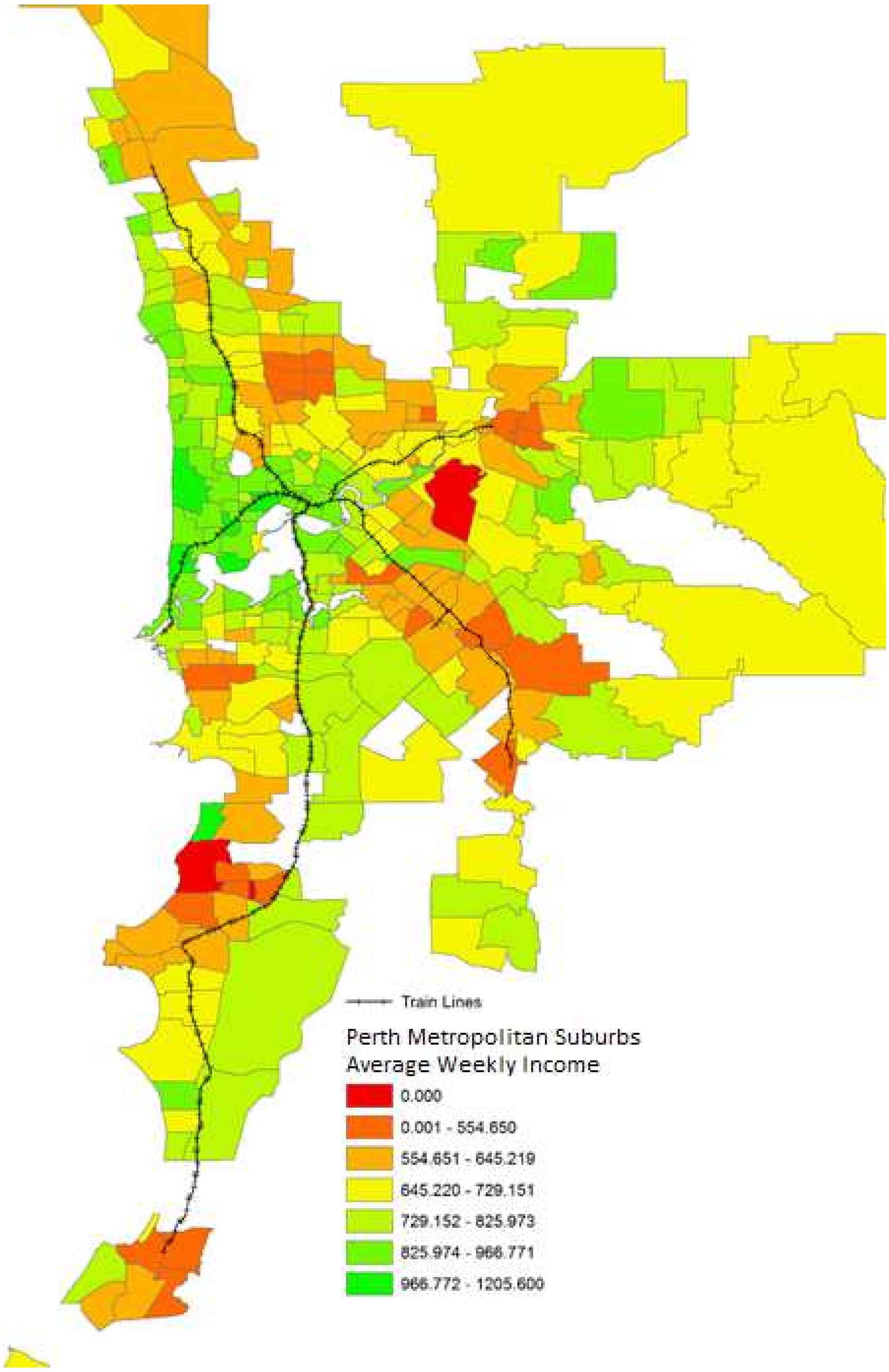


Figure 37: Average Weighted Weekly Income Map

Figure 37 shows that suburbs along the Swan River and northern coast have high average weekly income (above $967). The suburbs with relatively low average weighted weekly incomes are mostly in the North Eastern, South Eastern and Lower Western metropolitan areas. The Amardale train line serves more suburbs with a low average weighted weekly income (below $ 555). The two suburbs highlighted in red namely Perth airport and Kwinana Beach because there is no resident and the average weekly income in these suburbs are 0.

5.1.6.2 Average Weekly Rent

Another socio-economic factor considered in this research is average weekly rent. Descriptive analysis on this observed variable (average weekly rent) is discussed in this section to achieve better comprehension on its variance across the Perth metropolitan suburbs.

Average Rent		
Mean		377.87
Std. Deviation		108.62
Skewness		0.69
Kurtosis		8.59
Minimum		0
Maximum		1012.5
Percentiles	10	292.72
	20	321.11
	30	339.51
	40	350.88
	50	365.11
	60	388.59
	70	403.47
	80	440.02
	90	484.3

Table 9: Descriptive Statistics on Average Rent

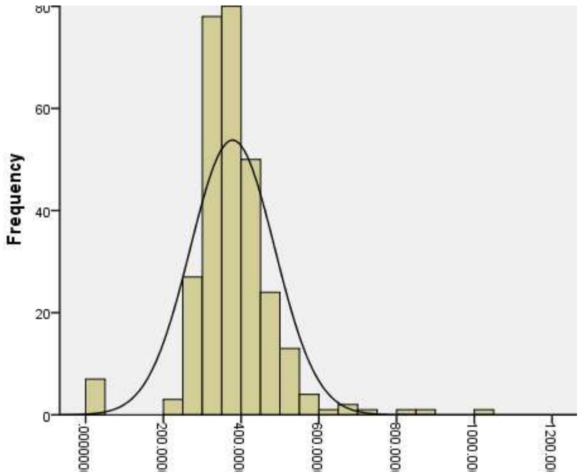


Figure 38: Histogram of Average Rent

Table 9 shows that the mean of the average rent in the Perth metropolitan suburbs is $377.87, with values below the average for 50% of them. The maximum average rent is $1012.5, and the highest average rent for 90% of these suburbs is $484.3. This indicates that there is a large variation in the average rents for the most expensive 10% of these suburbs, ranging from $484.3 to $1012.5. The normal distribution of the average rents has a high level of positive (left) skewness (0.699), and the sharp pointiness of the distribution is 80.59. Figure 38 illustrates that the average rent in majority of the Perth metropolitan suburbs lies between $200 and $600.

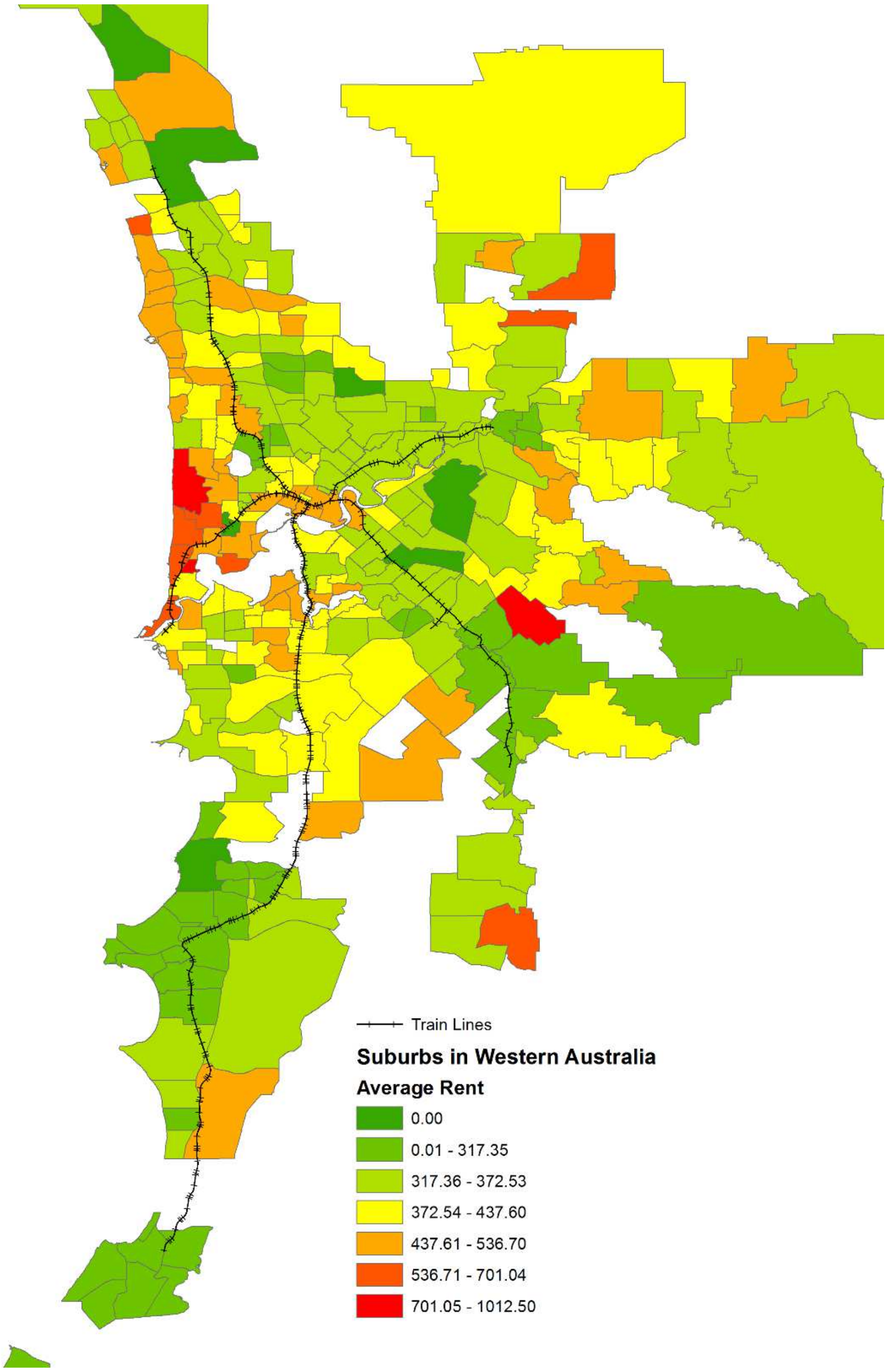


Figure 39: Average Weekly Rent Map

Figure 39 shows the average weekly rent in the Perth metropolitan suburbs. When this map is compared with Figure 37 (average weekly income map), it can be noticed that the average weekly rents in suburbs with high average weighted weekly income are high, ranging between $537-1013 per week. These suburbs are along the Western Coastal line and Swan River. The average weekly rents in the northern east, southern east, eastern and lower western metropolitan areas are low similar to the average weighted weekly income.

5.1.6.3 Average Car Ownership per Household

One of the socio-economic factors taken into account in this study is average car ownership per household. The findings from the descriptive analysis on this variable are presented in this section.

Average Car Ownership per Household		
Mean		1.80
Std. Deviation		0.44
Skewness		-0.72
Kurtosis		4.10
Minimum		0.00
Maximum		3.32
Percentiles	10	1.39
	20	1.48
	30	1.60
	40	1.71
	50	1.82
	60	1.89
	70	1.98
	80	2.11
	90	2.29

Table 10: Descriptive Statistic on Average Car Ownership per Household

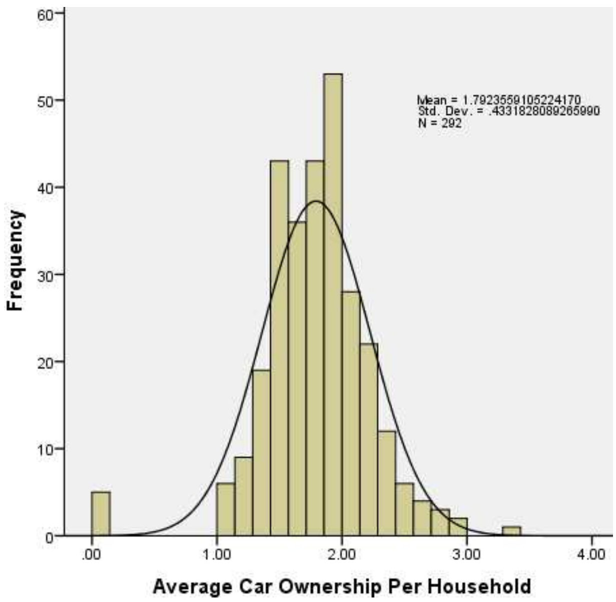


Figure 40: Histogram of Average Car Ownership per Household

As shown in Table 10, the average car ownership per household in Western Australian suburbs is 1.8. Moreover, average car ownership per household for 40% of these suburbs falls below the average value. Its low standard deviation indicates that average car ownership levels among the suburbs in this study are narrowly concentrated around the mean. Further, its negative skewness (-0.722) indicates that its normal distribution curve is skewed to the right towards higher car ownership, with relatively high kurtosis (0.435).

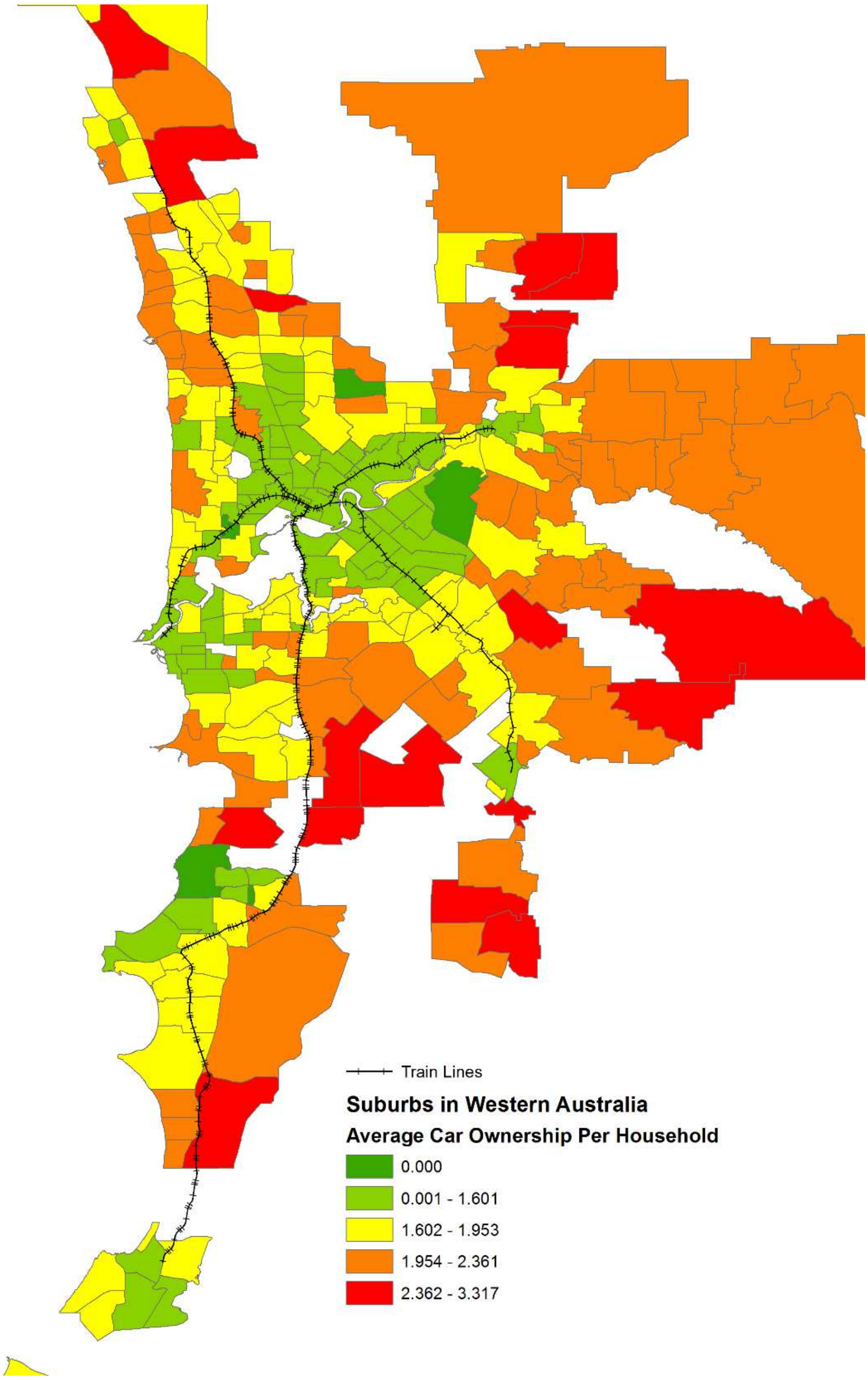


Figure 41: Average Car Ownership per Household Map

Figure 41 illustrates that there are 3 areas in which average car ownership per household are 0. These suburbs are Malaga industrial area, Perth airport, and Kwinana Beach where there is no resident population. Average car ownership per household is lower in inner suburbs, and gets higher as they get farther away from the city center. This map shows that residents who live in the outer suburbs tend to rely more cars as compared to the ones living closer to the city center and the train lines.

5.2 Factor Analysis Results

To overcome the problem of multicollinearity among independent variables, and thereby satisfy the assumption that predictor variables are not highly correlated, factor analysis was conducted to extract latent variables. Factor extraction was conducted specifically for public transport service provision density variables. It was also used to extract latent urban form factors and socio-economic factors.

5.2.1 Factor Analysis on Public Transport Service Provision Density

Gorsuch (1983, : p.2) states that factor analysis is used "*to summarize the interrelationships among the variables in a concise but accurate manner as an aid in conceptualization".* He also points out that factors can be used to minimize the number of variables while maximizing the amount of information in the analysis. The odds of a random pattern appearing significant are higher when using multiple regression analysis with many independent variables, Gorsuch (1983: 2). Factor analysis decreases this probability while the degrees of freedom are increased by transforming the number of independent variables into a smaller number of factors, which can then be used to predict the criterion variable.

Likewise, Comrey (1973) posits that factor analysis can be used to gain a better understanding of the complex interrelationships among large number of independent variables. According to Leech (2011), Principal Component Analysis (hereafter PCA) can be used to extract unobserved (or latent) variables, using as much information as possible from the observed variables.

In this research, PCA is used to extract latent variables from the public transport service provision variables, as well as the socio-economic variables. Leech (2011) states that the two main conditions for PCA are 1) to have large sample size in relation to the number of variables, so as to generate more reliable factors and 2) to have relationships among the variables. These conditions are explained below.

1) Regarding the sample size, there are 309 suburbs where the Perth Transport Authority (PTA) provides public transport services. Only 17 out of these 309 suburbs are excluded from this research: five because they are only occupied by parks and beaches namely Kings Park, Tamala Park, Burns Beach, Whiteman Park and Medora Bay, and twelve outer suburbs because and socio-economic data are not available for them. The other 292 cases, or 94% of the total population, satisfy the first condition and are included in this factor analysis.

2) Correlations among the dependent variables are analysed before conducting the factor analysis to make sure that the second factor analysis condition is met (see Table 11).

			Bus/Ferry Service Provision Density																						
			Weekday								Saturday								Sunday						
			12am-3am	3am-6am	6am-9am	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm	9pm-12midnight	12am-3am	3am-6am	6am-9am	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm	9pm-12midnight	3am-6am	6am-9am	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm	9pm-12midnight
Bus/Ferry Service Provision Density	Weekday	12am-3am	1.000	.481	.431	.428	.439	.455	.497	.540	.598	.269	.450	.406	.406	.409	.453	.466	.091	.255	.404	.404	.398	.359	.157
		3am-6am	.481	1.000	.557	.516	.530	.527	.677	.617	.354	.396	.569	.518	.528	.533	.539	.492	.046	.666	.574	.560	.550	.608	.554
		6am-9am	.431	.557	1.000	.985	.984	.986	.945	.940	.725	.152	.953	.963	.968	.966	.967	.946	.095	.592	.945	.951	.955	.889	.524
		9am-12noon	.428	.516	.985	1.000	.998	.990	.927	.916	.741	.146	.938	.958	.963	.962	.959	.934	.081	.530	.929	.939	.944	.849	.458
		12noon-3pm	.439	.530	.984	.998	1.000	.991	.930	.919	.740	.144	.937	.954	.961	.961	.959	.933	.076	.534	.929	.939	.944	.849	.462
		3pm-6pm	.455	.527	.986	.990	.991	1.000	.945	.929	.739	.110	.932	.946	.954	.954	.956	.930	.093	.557	.932	.943	.946	.862	.484
		6pm-9pm	.497	.677	.945	.927	.930	.945	1.000	.952	.649	.175	.909	.902	.915	.919	.908	.869	.139	.716	.939	.937	.933	.912	.623
		9pm-12midnight	.540	.617	.940	.916	.919	.929	.952	1.000	.696	.164	.930	.907	.914	.912	.922	.927	.128	.627	.915	.914	.913	.891	.571
	Saturday	12am-3am	.598	.354	.725	.741	.740	.739	.649	.696	1.000	.103	.739	.729	.724	.717	.797	.834	.081	.320	.668	.697	.698	.578	.218
		3am-6am	.269	.396	.152	.146	.144	.110	.175	.164	.103	1.000	.230	.207	.185	.187	.181	.144	-.028	.191	.113	.102	.108	.160	.077
		6am-9am	.450	.569	.953	.938	.937	.932	.909	.930	.739	.230	1.000	.980	.978	.975	.971	.943	.082	.585	.934	.935	.937	.895	.492
		9am-12noon	.406	.518	.963	.958	.954	.946	.902	.907	.729	.207	.980	1.000	.997	.995	.978	.942	.057	.543	.938	.945	.948	.884	.457
		12noon-3pm	.406	.528	.968	.963	.961	.954	.915	.914	.724	.185	.978	.997	1.000	.999	.979	.939	.061	.561	.951	.958	.961	.897	.481
		3pm-6pm	.409	.533	.966	.962	.961	.954	.919	.912	.717	.187	.975	.995	.999	1.000	.978	.932	.065	.573	.956	.962	.965	.904	.487
		6pm-9pm	.453	.539	.967	.959	.959	.956	.908	.922	.797	.181	.971	.978	.979	.978	1.000	.969	.088	.566	.944	.953	.956	.892	.468
		9pm-12midnight	.466	.492	.946	.934	.933	.930	.869	.927	.834	.144	.943	.942	.939	.932	.969	1.000	.090	.496	.884	.899	.903	.826	.447
	Sunday	3am-6am	.091	.046	.095	.081	.076	.093	.139	.128	.081	-.028	.082	.057	.061	.065	.088	.090	1.000	.366	.145	.133	.129	.205	.445
		6am-9am	.255	.666	.592	.530	.534	.557	.716	.627	.320	.191	.585	.543	.561	.573	.566	.496	.366	1.000	.707	.692	.678	.779	.839
		9am-12noon	.404	.574	.945	.929	.929	.932	.939	.915	.668	.113	.934	.938	.951	.956	.944	.884	.145	.707	1.000	.995	.996	.965	.628
		12noon-3pm	.404	.560	.951	.939	.939	.943	.937	.914	.697	.102	.935	.945	.958	.962	.953	.899	.133	.692	.995	1.000	.997	.953	.608
		3pm-6pm	.398	.550	.955	.944	.944	.946	.933	.913	.698	.108	.937	.948	.961	.965	.956	.903	.129	.678	.996	.997	1.000	.950	.601
		6pm-9pm	.359	.608	.889	.849	.849	.862	.912	.891	.578	.160	.895	.884	.897	.904	.892	.826	.205	.779	.965	.953	.950	1.000	.724
		9pm-12midnight	.157	.554	.524	.458	.462	.484	.623	.571	.218	.077	.492	.457	.481	.487	.468	.447	.445	.839	.628	.608	.601	.724	1.000
Train Service Provision Density	Weekday	12am-3am	.244	.172	.674	.693	.682	.666	.531	.574	.683	.080	.658	.694	.678	.662	.704	.760	-.035	.130	.555	.584	.601	.442	.067
		3am-6am	.256	.166	.679	.703	.693	.678	.535	.578	.707	.055	.655	.692	.677	.662	.711	.768	-.034	.127	.558	.591	.606	.436	.063
		6am-9am	.273	.182	.705	.722	.711	.699	.566	.611	.711	.061	.689	.720	.705	.690	.733	.789	-.032	.146	.589	.617	.634	.475	.088
		9am-12noon	.250	.171	.676	.692	.681	.667	.533	.576	.689	.070	.659	.693	.677	.661	.704	.761	-.034	.131	.558	.587	.604	.444	.070
		12noon-3pm	.249	.170	.677	.693	.681	.668	.535	.577	.684	.070	.660	.694	.678	.663	.704	.759	-.034	.132	.560	.588	.605	.448	.072
		3pm-6pm	.275	.183	.707	.722	.710	.699	.571	.616	.702	.059	.693	.722	.707	.693	.732	.787	-.033	.149	.595	.621	.638	.486	.096
		6pm-9pm	.261	.173	.685	.701	.689	.678	.546	.587	.693	.063	.667	.701	.686	.671	.712	.766	-.034	.136	.569	.597	.614	.457	.076
		9pm-12midnight	.251	.175	.679	.697	.685	.671	.534	.577	.695	.074	.659	.695	.679	.663	.708	.765	-.033	.133	.558	.588	.604	.444	.070
	Saturday	12am-3am	.247	.173	.677	.698	.686	.671	.536	.579	.687	.078	.660	.696	.680	.665	.708	.763	-.035	.134	.558	.588	.604	.443	.068
		3am-6am	.271	.159	.667	.693	.683	.672	.526	.574	.725	.029	.637	.672	.659	.643	.699	.761	-.030	.111	.540	.577	.589	.415	.046
		6am-9am	.245	.168	.672	.687	.675	.662	.532	.573	.673	.072	.656	.690	.674	.659	.698	.752	-.035	.131	.557	.584	.601	.447	.071
		9am-12noon	.245	.170	.672	.689	.677	.663	.528	.570	.684	.075	.655	.691	.674	.658	.701	.756	-.034	.130	.554	.583	.599	.440	.068
		12noon-3pm	.245	.170	.672	.689	.677	.663	.528	.570	.684	.075	.655	.691	.674	.658	.701	.756	-.034	.130	.554	.583	.599	.440	.068
		3pm-6pm	.245	.170	.672	.689	.677	.663	.528	.570	.684	.075	.655	.691	.674	.658	.701	.756	-.034	.130	.554	.583	.599	.440	.068
		6pm-9pm	.252	.172	.679	.695	.684	.671	.536	.579	.690	.070	.661	.696	.679	.664	.707	.763	-.034	.132	.560	.589	.605	.447	.070
		9pm-12midnight	.251	.175	.679	.697	.686	.671	.534	.577	.695	.074	.660	.695	.679	.663	.708	.765	-.033	.134	.558	.588	.604	.444	.071
	Sunday	3am-6am	.263	.169	.666	.691	.682	.667	.526	.570	.704	.065	.643	.679	.665	.649	.699	.757	-.034	.122	.543	.575	.589	.420	.051
		6am-9am	.267	.171	.694	.716	.706	.694	.557	.599	.707	.048	.671	.705	.692	.677	.722	.777	-.034	.137	.577	.607	.622	.459	.074
		9am-12noon	.243	.170	.672	.690	.679	.664	.531	.573	.679	.078	.657	.692	.676	.660	.702	.756	-.035	.132	.555	.584	.600	.442	.067
		12noon-3pm	.243	.170	.673	.691	.679	.665	.531	.573	.681	.078	.657	.692	.676	.660	.702	.757	-.035	.131	.555	.584	.600	.441	.067
		3pm-6pm	.243	.170	.673	.691	.679	.665	.531	.573	.681	.078	.657	.692	.676	.660	.702	.757	-.035	.131	.555	.584	.600	.441	.067
		6pm-9pm	.253	.173	.682	.700	.689	.676	.540	.583	.693	.071	.664	.699	.683	.668	.711	.767	-.034	.134	.563	.592	.608	.448	.070
		9pm-12midnight	.248	.173	.674	.691	.679	.665	.531	.575	.685	.076	.658	.692	.676	.660	.703	.759	-.033	.131	.555	.584	.600	.443	.069
			Train Service Provision Density																						
Train Service Provision Density	Weekday	12am-3am	1.000	.990	.982	.996	.995	.968	.992	.997	.999	.976	.994	.997	.997	.997	.996	.997	.988	.988	.998	.998	.998	.998	.997
		3am-6am	.990	1.000	.975	.990	.987	.955	.982	.992	.993	.981	.983	.991	.991	.991	.989	.992	.992	.984	.993	.993	.993	.992	.989
		6am-9am	.982	.975	1.000	.990	.991	.997	.995	.986	.981	.972	.989	.986	.986	.986	.991	.986	.965	.993	.985	.984	.984	.990	.987
		9am-12noon	.996	.990	.990	1.000	1.000	.978	.997	.998	.994	.977	.998	.999	.999	.999	1.000	.998	.980	.989	.998	.998	.998	.998	.999
		12noon-3pm	.995	.987	.991	1.000	1.000	.981	.998	.997	.993	.975	.999	.999	.999	.999	.999	.997	.976	.990	.997	.997	.997	.998	.998
		3pm-6pm	.968	.955	.997	.978	.981	1.000	.989	.971	.965	.953	.980	.972	.972	.972	.980	.971	.943	.984	.971	.970	.970	.978	.974
		6pm-9pm	.992	.982	.995	.997	.998	.989	1.000	.994	.989	.975	.998	.995	.995	.995	.998	.994	.973	.993	.994	.993	.993	.996	.996
		9pm-12midnight	.997	.992	.986	.998	.997	.971	.994	1.000	.996	.980	.996	.999	.999	.999	.998	1.000	.985	.989	.997	.997	.997	.998	.999
	Saturday	12am-3am	.999	.993	.981	.994	.993	.965	.989	.996	1.000	.976	.991	.995	.995	.994	.994	.996	.990	.989	.998	.998	.998	.998	.995
		3am-6am	.976	.981	.972	.977	.975	.953	.975	.980	.976	1.000	.971	.976	.976	.976	.979	.980	.981	.983	.974	.975	.975	.980	.978
		6am-9am	.994	.983	.989	.998	.999	.980	.998	.996	.991	.971	1.000	.997	.997	.997	.999	.996	.972	.989	.996	.996	.996	.997	.998
		9am-12noon	.997	.991	.986	.999	.999	.972	.995	.999	.995	.976	.997	1.000	1.000	1.000	.999	.999	.981	.986	.998	.999	.999	.998	.999
		12noon-3pm	.997	.991	.986	.999	.999	.972	.995	.999	.995	.976	.997	1.000	1.000	1.000	.999	.999	.981	.986	.998	.999	.999	.998	.999
		3pm-6pm	.997	.991	.986	.999	.999	.972	.995	.999	.994	.976	.997	1.000	1.000	1.000	.999	.999	.981	.986	.998	.999	.999	.998	.999
		6pm-9pm	.996	.989	.991	1.000	.999	.980	.998	.998	.994	.979	.999	.999	.999	.999	1.000	.998	.980	.991	.997	.998	.998	.999	.999
		9pm-12midnight	.997	.992	.986	.998	.997	.971	.994	1.000	.996	.980	.996	.999	.999	.999	.998	1.000	.985	.989	.997	.997	.997	.998	.999
	Sunday	3am-6am	.988	.992	.965	.980	.976	.943	.973	.985	.990	.981	.972	.981	.981	.981	.980	.985	1.000	.978	.984	.985	.985	.986	.981
		6am-9am	.988	.984	.993	.989	.990	.984	.993	.989	.989	.983	.989	.986	.986	.986	.991	.989	.978	1.000	.988	.988	.988	.993	.989
		9am-12noon	.998	.993	.985	.998	.997	.971	.994	.997	.998	.974	.996	.998	.998	.998	.997	.997	.984	.988	1.000	1.000	1.000	.999	.997
		12noon-3pm	.998	.993	.984	.998	.997	.970	.993	.997	.998	.975	.996	.999	.999	.999	.998	.997	.985	.988	1.000	1.000	1.000	.999	.997
		3pm-6pm	.998	.993	.984	.998	.997	.970	.993	.997	.998	.975	.996	.999	.999	.999	.998	.997	.985	.988	1.000	1.000	1.000	.999	.997
		6pm-9pm	.998	.992	.990	.998	.998	.978	.996	.998	.998	.980	.997	.998	.998	.998	.999	.998	.986	.993	.999	.999	.999	1.000	.998
		9pm-12midnight	.997	.989	.987	.999	.998	.974	.996	.999	.995	.978	.998	.999	.999	.999	.999	.999	.981	.989	.997	.997	.997	.998	1.000

Table 11: Correlation Matrix (Public Transport Service Provision Densities Variables)

Table 11 shows how public transport service provision densities are associated with each other. Correlations are highlighted in red text if they are larger than 0.5. These values indicate that there is a strong correlation between these pairs of variables. Clearly, there are correlations among bus/ferry service provision variables and also among train service provision variables. These strong correlations indicate that the second factor analysis condition is met to proceed with principal component analysis.

For factor analysis, some of these time periods need to be aggregated to get more meaningful factor analysis result for better interpretation as train service factor and bus service factor. Therefore, the number of bus/ferry service frequency variables is reduced from 23 to 18 and the number of train service frequency variables is reduced from 23 to 18. For examples, in factor analysis:

- Weekday service frequency between 0-3am and 3-6am are aggregated to 0-6am
- Saturday service frequency between 0-3am, 3-6am and 6-9am are aggregated as 0-9am

5.2.1.1 Sampling Adequacy Test for Service Provision Factor Analysis

Field (2005) suggests the use of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy to evaluate the appropriateness of factor analysis because it calculates the ratio of the squared correlation between the variables to the squared partial correlation between variables. He also explains that Bartlett's test can be used to evaluate the resemblances of the population correlation matrix to the identity matrix by detecting high levels of correlation among the variables. For the KMO and Bartlett's test, Field (ibid) recommends that KMO values should be at least 0.5, while Bartlett's test of Sphericity should be significant with a value less than 0.05.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.940
	Approx. Chi-Square	18341.081
Bartlett's Test of Sphericity	df	153
	Sig.	.000

Table 12 KMO and Bartlett's Test (Public Transport Service Provision Density Factors Analysis)

A KMO value of 0.907 means that the service provision factors analysis is appropriate, as it is greater than 0.5. Additionally, the Bartlett's Test of Sphericity's significant value 0 can be interpreted as showing that the population correlation matrix of the service provision variables is not an identity matrix.

5.2.1.2 Communality Test on Service Provision Factor Analysis

Field (2005) explains that communality is the proportion of common variance within a variable. The PCA assumes that all variance associated with a variable is common. Extraction values in the communalities table reflect this common variance. Field (2005) recommends the use of extraction values to verify that eigenvalues greater than 1 are applicable based on the number of variables and test cases. He also suggests that extraction values should be greater than 0.7 if there are less than 30 variables. Likewise, these communalities should be greater than 0.6 when there are more than 250 test cases (see also Gray, 2012).

	Initial	Extraction
Bus/Ferry Service Provision_Density_PerKm2_Weekday 0-6hr	1.000	.482
Bus/Ferry Service Provision_Density_PerKm2_Weekday 6-9hr	1.000	.966
Bus/Ferry Service Provision_Density_PerKm2_Weekday 9-12hr	1.000	.940
Bus/Ferry Service Provision_Density_PerKm2_Weekday 12-15hr	1.000	.938
Bus/Ferry Service Provision_Density_PerKm2_Weekday 15-18hr	1.000	.940
Bus/Ferry Service Provision_Density_PerKm2_Weekday 18-21hr	1.000	.942
Bus/Ferry Service Provision_Density_PerKm2_Weekday 21-24hr	1.000	.913
Bus/Ferry Service Provision_Density_PerKm2_Saturday 0-9hr	1.000	.937
Bus/Ferry Service Provision_Density_PerKm2_Saturday 9-12hr	1.000	.945
Bus/Ferry Service Provision_Density_PerKm2_Saturday 12-15hr	1.000	.954
Bus/Ferry Service Provision_Density_PerKm2_Saturday 15-18hr	1.000	.953
Bus/Ferry Service Provision_Density_PerKm2_Saturday 18-21hr	1.000	.961
Bus/Ferry Service Provision_Density_PerKm2_Saturday 21-24hr	1.000	.933
Bus/Ferry Service Provision_Density_PerKm2_Sunday 3-9hr and 21-24hr	1.000	.676
Bus/Ferry Service Provision_Density_PerKm2_Sunday 9-12hr	1.000	.963
Bus/Ferry Service Provision_Density_PerKm2_Sunday 12-15hr	1.000	.963
Bus/Ferry Service Provision_Density_PerKm2_Sunday 15-18hr	1.000	.963
Bus/Ferry Service Provision_Density_PerKm2_Sunday 18-21hr	1.000	.934
Train Service Provision_Density_PerKm2_Weekday 0-6hr	1.000	.991
Train Service Provision_Density_PerKm2_Weekday 6-9hr	1.000	.983
Train Service Provision_Density_PerKm2_Weekday 9-12hr	1.000	.996
Train Service Provision_Density_PerKm2_Weekday 12-15hr	1.000	.995
Train Service Provision_Density_PerKm2_Weekday 15-18hr	1.000	.961
Train Service Provision_Density_PerKm2_Weekday 18-21hr	1.000	.992
Train Service Provision_Density_PerKm2_Weekday 21-24hr	1.000	.995
Train Service Provision_Density_PerKm2_Saturday 0-9hr	1.000	.996
Train Service Provision_Density_PerKm2_Saturday 9-12hr	1.000	.995
Train Service Provision_Density_PerKm2_Saturday 12-15hr	1.000	.995
Train Service Provision_Density_PerKm2_Saturday 15-18hr	1.000	.995
Train Service Provision_Density_PerKm2_Saturday 18-21hr	1.000	.997
Train Service Provision_Density_PerKm2_Saturday 21-24hr	1.000	.995
Train Service Provision_Density_PerKm2_Sunday 3-9hr and 21-24hr	1.000	.996
Train Service Provision_Density_PerKm2_Sunday 9-12hr	1.000	.995
Train Service Provision_Density_PerKm2_Sunday 12-15hr	1.000	.995
Train Service Provision_Density_PerKm2_Sunday 15-18hr	1.000	.995
Train Service Provision_Density_PerKm2_Sunday 18-21hr	1.000	.997

Extraction Method: Principal Component Analysis.

Table 13 Communalities (Public Transport Service Provision Factor Analysis)

There are 36 public transport service provision variables and 292 test cases are included in this analysis. In Table 13, the extraction values for all public transport service provision density variables are greater than 0.7 with the exception of bus/ferry service provision (weekday 0-6hr, Sunday 3-9hr and 21-24hr), for which extraction values are 0.482 and 0.676 respectively. These results verify that using eigenvalues greater than 1 in this analysis is valid.

5.2.1.3 Factors Selection from Service Provision Factor Analysis

Based on their total variance explained, public transport service provision factors with significant contributions will be considered in a regression model with public transport usage eigenvalues, as well as the factor loadings described in the rotated component matrices.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	29.654	76.036	76.036	29.654	76.036	76.036	21.406	54.888	54.888
2	7.082	18.159	94.194	7.082	18.159	94.194	15.330	39.307	94.194
3	.923	2.367	96.561						
4	.598	1.533	98.094						
5	.197	.505	98.599						
6	.142	.365	98.965						
7	.101	.260	99.225						
8	.075	.193	99.417						
9	.063	.161	99.579						
10	.038	.097	99.675						
11	.029	.074	99.749						
12	.022	.055	99.804						
13	.016	.042	99.846						
14	.013	.034	99.880						
15	.009	.023	99.903						
16	.008	.021	99.924						
17	.007	.017	99.941						
18	.006	.016	99.957						
19	.005	.013	99.970						
20	.003	.007	99.977						
21	.002	.006	99.982						
22	.002	.005	99.987						
23	.001	.003	99.990						
24	.001	.003	99.993						
25	.001	.002	99.995						
26	.001	.002	99.997						
27	.000	.001	99.998						
28	.000	.001	99.999						
29	.000	.000	99.999						

Extraction Method: Principal Component Analysis.

Table 14 Total Variance Explained (Public Transport Service Provision Factors)

The Total Variance Explained table (see Table 14) lists the eigenvalues associated with each linear component before extraction, after extraction, and after rotation. It also displays the eigenvalues in terms of the percentage of variance explained. Varimax rotation is used for this factor analysis. Since it optimizes the factor structure, the relative importance of the significant factors is equalised. Table 14 shows that there are only 2 factors which explain the majority of variance in public transport service provision densities. Before rotation, the first factor accounted for considerably more variance than the second factor—76.036% as compared to 18.159%. These two factors together explain the total variance of 94.194% of all cases. Additionally, after rotation, the first factor accounts for only 54.888% of the variance and second factor explains 39.307%.

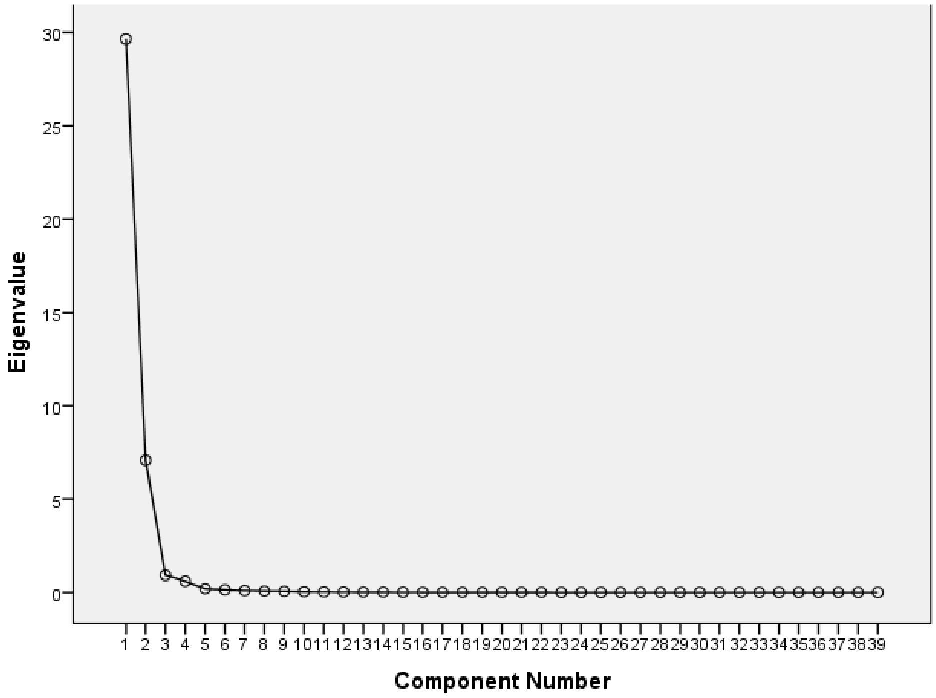


Figure 42 Scree Plot (Public Transport Service Provision Density Factors)

Figure 42 provides a graphic display of the factors and their corresponding eigenvalues. In the scree plot, also known as an eigenvalue plot, the eigenvalues of successive factors are plotted against the ordinal numbers of the factors, Gray (2012). When using PCA, factors which have eigenvalues greater than 1 are generally selected for their ability to explain the majority of common variance. In this analysis, the first two public transport service provision latent factors have high initial eigenvalues, 29.654 and 7.082, respectively. On the other hand, the initial eigenvalues of the other generated factors are less than 1. Therefore, it can be concluded that two factors should be extracted from all public transport service provision density variables. The next step is to assess their individual factor loadings to find out what each variable contributes to the extracted factors.

5.2.1.4 Public Transport Service Provision Density Factor Loadings

Rotated Component Matrix[a]

	Component	
	1	2
Train Service Provision_Density_PerKm2_Saturday 15-18hr	.956	
Train Service Provision_Density_PerKm2_Saturday 9-12hr	.956	
Train Service Provision_Density_PerKm2_Saturday 12-15hr	.956	
Train Service Provision_Density_PerKm2_Weekday 9-12hr	.955	
Train Service Provision_Density_PerKm2_Saturday 0-9hr	.955	
Train Service Provision_Density_PerKm2_Sunday 12-15hr	.955	
Train Service Provision_Density_PerKm2_Sunday 15-18hr	.955	
Train Service Provision_Density_PerKm2_Sunday 9-12hr	.955	
Train Service Provision_Density_PerKm2_Saturday 18-21hr	.954	
Train Service Provision_Density_PerKm2_Weekday 12-15hr	.954	
Train Service Provision_Density_PerKm2_Weekday 21-24hr	.953	
Train Service Provision_Density_PerKm2_Saturday 21-24hr	.953	
Train Service Provision_Density_PerKm2_Sunday 18-21hr	.953	
Train Service Provision_Density_PerKm2_Sunday 3-9hr_21-24hr	.952	
Train Service Provision_Density_PerKm2_Weekday 0-6hr	.950	
Train Service Provision_Density_PerKm2_Weekday 18-21hr	.949	
Train Service Provision_Density_PerKm2_Weekday 6-9hr	.935	
Train Service Provision_Density_PerKm2_Weekday 15-18hr	.920	
Bus/Ferry Service Provision_Density_PerKm2_Sunday 18-21hr		.946
Bus/Ferry Service Provision_Density_PerKm2_Sunday 9-12hr		.936
Bus/Ferry Service Provision_Density_PerKm2_Weekday 18-21hr		.932
Bus/Ferry Service Provision_Density_PerKm2_Sunday 12-15hr		.924
Bus/Ferry Service Provision_Density_PerKm2_Sunday 15-18hr		.917
Bus/Ferry Service Provision_Density_PerKm2_Weekday 21-24hr		.899
Bus/Ferry Service Provision_Density_PerKm2_Saturday 15-18hr		.882
Bus/Ferry Service Provision_Density_PerKm2_Weekday 6-9hr		.881
Bus/Ferry Service Provision_Density_PerKm2_Saturday 12-15hr		.873
Bus/Ferry Service Provision_Density_PerKm2_Weekday 15-18hr		.869
Bus/Ferry Service Provision_Density_PerKm2_Saturday 0-9hr		.864
Bus/Ferry Service Provision_Density_PerKm2_Saturday 18-21hr		.859
Bus/Ferry Service Provision_Density_PerKm2_Weekday 12-15hr		.859
Bus/Ferry Service Provision_Density_PerKm2_Saturday 9-12hr		.856
Bus/Ferry Service Provision_Density_PerKm2_Weekday 9-12hr		.852
Bus/Ferry Service Provision_Density_PerKm2_Saturday 21-24hr		.789
Bus/Ferry Service Provision_Density_PerKm2_Sunday 3-9hr_21-24hr		.772
Bus/Ferry Service Provision_Density_PerKm2_Weekday 0-6hr		.697

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Table 15: Rotated Component Matrix (Public Transport Service Provision Factors)

Table 15: Rotated Component Matrix (Public Transport Service Provision Factors) shows the loading of each public transport service provision variable on each component, or factor. After applying the Varimax rotation method, each factor becomes more meaningful. The main contributors to the first factor are the train service provision variables, while the primary contributors to the second are the bus/ferry service provision variables. Therefore, the first factor can be named “Train Service Provision Factor” and the second factor can be called “Bus/Ferry Service Provision Factor”. These factor scores are used in the factor regression to develop a public transport usage model.

The reproduced correlation matrix is also illustrated in the following section to analyse the correlations of variables from the model rather than the observed data. In this matrix, residual values represent the differences between the correlation values in the observed data and those of the model. In a good model, these residual values will be relatively small. Field (2005) recommends that the percentage of “non-redundant residuals with absolute values” should be less than 50%.

5.2.1.5 Validity of Public Transport Service Provision Density Factor Model

				Bus/Ferry Service Provision Density																	
				Weekday							Saturday						Sunday				
				12am-6am	6am-9am	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm	9pm-12midnight	12am-9am	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm	9pm-12midnight	3am-9am and 9pm-12pm midnight	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm
Reproduced Correlation	Bus/Ferry Service Provision Density	Weekday	12am-6am	.486[a]	.617	.598	.602	.609	.652	.629	.605	.600	.612	.618	.603	.554	.537	.654	.647	.642	.661
			6am-9am	.617	.970[a]	.958	.958	.959	.941	.937	.957	.961	.966	.966	.968	.941	.614	.955	.960	.962	.908
			9am-12noon	.598	.958	.948[a]	.947	.947	.922	.921	.946	.951	.955	.954	.958	.936	.588	.937	.944	.947	.886
			12noon-3pm	.602	.958	.947	.946[a]	.947	.924	.922	.945	.950	.954	.953	.957	.933	.595	.939	.945	.948	.890
			3pm-6pm	.609	.959	.947	.947	.947[a]	.929	.925	.946	.950	.955	.954	.957	.931	.605	.943	.948	.950	.896
			6pm-9pm	.652	.941	.922	.924	.929	.943[a]	.927	.926	.925	.935	.939	.931	.887	.680	.953	.951	.950	.928
			9pm-12midnight	.629	.937	.921	.922	.925	.927	.917[a]	.923	.924	.932	.934	.930	.893	.646	.939	.940	.940	.907
		Saturday	12am-9am	.605	.957	.946	.945	.946	.926	.923	.945[a]	.949	.953	.953	.956	.931	.600	.940	.946	.948	.893
			9am-12noon	.600	.961	.951	.950	.950	.925	.924	.949	.953[a]	.957	.956	.960	.938	.591	.940	.947	.950	.890
			12noon-3pm	.612	.966	.955	.954	.955	.935	.932	.953	.957	.962[a]	.962	.964	.939	.608	.950	.955	.957	.902
			3pm-6pm	.618	.966	.954	.953	.954	.939	.934	.953	.956	.962	.962[a]	.963	.935	.618	.952	.957	.959	.907
			6pm-9pm	.603	.968	.958	.957	.957	.931	.930	.956	.960	.964	.963	.968[a]	.946	.592	.946	.953	.956	.894
			9pm-12midnight	.554	.941	.936	.933	.931	.887	.893	.931	.938	.939	.935	.946	.935[a]	.526	.904	.914	.920	.842
		Sunday	3am-9am and 9pm-12pm midnight	.537	.614	.588	.595	.605	.680	.646	.600	.591	.608	.618	.592	.526	.619[a]	.678	.664	.656	.705
			9am-12noon	.654	.955	.937	.939	.943	.953	.939	.940	.940	.950	.952	.946	.904	.678	.963[a]	.963	.962	.936
			12noon-3pm	.647	.960	.944	.945	.948	.951	.940	.946	.947	.955	.957	.953	.914	.664	.963	.964[a]	.964	.931
			3pm-6pm	.642	.962	.947	.948	.950	.950	.940	.948	.950	.957	.959	.956	.920	.656	.962	.964	.964[a]	.927
			6pm-9pm	.661	.908	.886	.890	.896	.928	.907	.893	.890	.902	.907	.894	.842	.705	.936	.931	.927	.924[a]
	Train Service Provision Density	Weekday	12am-6am	.214	.681	.700	.689	.675	.534	.579	.680	.700	.684	.668	.710	.765	.088	.559	.589	.605	.443
			6am-9am	.240	.706	.724	.713	.700	.564	.607	.705	.724	.709	.694	.734	.786	.118	.589	.618	.634	.475
			9am-12noon	.211	.679	.699	.687	.673	.531	.576	.678	.699	.682	.667	.709	.764	.083	.557	.586	.603	.439
			12noon-3pm	.212	.680	.700	.688	.674	.532	.577	.679	.699	.683	.667	.710	.765	.085	.558	.588	.604	.441
			3pm-6pm	.246	.708	.725	.715	.701	.569	.610	.706	.725	.710	.696	.735	.785	.128	.594	.621	.637	.481
			6pm-9pm	.220	.688	.707	.696	.681	.541	.586	.687	.707	.691	.675	.717	.771	.094	.567	.596	.613	.450
			9pm-12midnight	.213	.681	.701	.689	.675	.533	.578	.680	.700	.684	.668	.710	.766	.086	.559	.588	.605	.442
		Saturday	12am-9am	.211	.679	.699	.687	.673	.531	.576	.678	.698	.682	.666	.708	.764	.083	.556	.586	.603	.439
			9am-12noon	.208	.675	.695	.684	.669	.527	.572	.675	.695	.678	.663	.705	.761	.080	.553	.582	.599	.435
			12noon-3pm	.208	.675	.695	.684	.669	.527	.572	.675	.695	.678	.663	.705	.761	.080	.553	.582	.599	.435
			3pm-6pm	.208	.675	.695	.684	.669	.527	.572	.675	.695	.678	.663	.705	.761	.080	.553	.582	.599	.435
			6pm-9pm	.213	.681	.701	.690	.675	.533	.579	.680	.701	.684	.669	.711	.766	.086	.559	.589	.606	.442
			9pm-12midnight	.214	.681	.701	.689	.675	.534	.579	.680	.700	.684	.669	.711	.766	.086	.559	.589	.605	.442
		Sunday	3am-9am and 9pm-12pm midnight	.218	.686	.706	.694	.680	.539	.584	.685	.705	.689	.674	.715	.770	.091	.565	.594	.611	.448
			9am-12noon	.209	.676	.696	.685	.670	.528	.574	.676	.696	.679	.664	.706	.762	.081	.554	.584	.600	.437
			12noon-3pm	.209	.676	.697	.685	.670	.528	.574	.676	.696	.680	.664	.706	.762	.081	.554	.584	.601	.437
			3pm-6pm	.209	.676	.697	.685	.670	.528	.574	.676	.696	.680	.664	.706	.762	.081	.554	.584	.601	.437
			6pm-9pm	.216	.684	.704	.693	.678	.537	.582	.684	.704	.687	.672	.714	.769	.089	.563	.592	.609	.446
Residual[b]	Bus/Ferry Service Provision Density	Weekday	12am-6am		-.030	-.043	-.032	-.035	.051	.045	.007	-.052	-.056	-.058	-.021	-.003	.027	-.065	-.068	-.072	-.062
			6am-9am	-.030		.027	.026	.027	.005	.003	-.005	.002	.002	.001	-.001	.005	-.034	-.010	-.009	-.007	-.020
			9am-12noon	-.043	.027		.051	.043	.004	-.005	-.007	.007	.008	.009	.001	-.001	-.074	-.009	-.005	-.002	-.037
			12noon-3pm	-.032	.026	.051		.044	.005	-.003	-.007	.005	.007	.008	.002	.000	-.077	-.010	-.006	-.003	-.041
			3pm-6pm	-.035	.027	.043	.044		.016	.004	-.013	-.004	.000	.000	-.001	-.001	-.064	-.010	-.005	-.005	-.035
			6pm-9pm	.051	.005	.004	.005	.016		.025	-.021	-.023	-.020	-.019	-.023	-.018	.017	-.014	-.015	-.017	-.016
			9pm-12midnight	.045	.003	-.005	-.003	.004	.025		.005	-.017	-.018	-.022	-.008	.034	-.022	-.024	-.026	-.027	-.016
		Saturday	12am-9am	.007	-.005	-.007	-.007	-.013	-.021	.005		.029	.022	.019	.020	.022	-.054	-.013	-.015	-.015	-.009
			9am-12noon	-.052	.002	.007	.005	-.004	-.023	-.017	.029		.040	.039	.017	.005	-.071	-.002	-.001	-.001	-.005
			12noon-3pm	-.056	.002	.008	.007	.000	-.020	-.018	.022	.040		.037	.015	.001	-.066	.002	.003	.003	-.005
			3pm-6pm	-.058	.001	.009	.008	.000	-.019	-.022	.019	.039	.037		.015	-.003	-.066	.004	.005	.006	-.003
			6pm-9pm	-.021	-.001	.001	.002	-.001	-.023	-.008	.020	.017	.015	.015		.023	-.054	-.002	.000	.000	-.002
			9pm-12midnight	-.003	.005	-.001	.000	-.001	-.018	.034	.022	.005	.001	-.003	.023		-.036	-.020	-.016	-.016	-.016
		Sunday	3am-9am and 9pm-12pm midnight	.027	-.034	-.074	-.077	-.064	.017	-.022	-.054	-.071	-.066	-.066	-.054	-.036		.017	.012	.009	.077
			9am-12noon	-.065	-.010	-.009	-.010	-.010	-.014	-.024	-.013	-.002	.002	.004	-.002	-.020	.017		.032	.034	.029
			12noon-3pm	-.068	-.009	-.005	-.006	-.005	-.015	-.026	-.015	-.001	.003	.005	.000	-.016	.012	.032		.034	.022
			3pm-6pm	-.072	-.007	-.002	-.003	-.005	-.017	-.027	-.015	-.001	.003	.006	.000	-.016	.009	.034	.034		.023
			6pm-9pm	-.062	-.020	-.037	-.041	-.035	-.016	-.016	-.009	-.005	-.005	-.003	-.002	-.016	.077	.029	.022	.023	
	Train Service Provision Density	Weekday	12am-6am	.010	-.002	.000	.001	.001	.001	-.001	-.005	-.006	-.005	-.005	-7.979E-05	.002	.013	-.001	.001	-4.359E-05	-.004
			6am-9am	.003	-.001	-.002	-.003	-.001	.002	.004	7.204E-05	-.004	-.004	-.004	-.001	.004	.003	.000	-.001	6.792E-05	.001
			9am-12noon	.014	-.003	-.006	-.007	-.005	.002	-.001	-.003	-.005	-.005	-.005	-.004	-.004	.021	.001	.001	.000	.005
			12noon-3pm	.012	-.003	-.007	-.007	-.006	.003	-.001	-.004	-.005	-.005	-.004	-.005	-.006	.021	.002	.001	.001	.007
			3pm-6pm	-.002	-.001	-.004	-.005	-.002	.003	.006	.002	-.003	-.003	-.003	-.003	.002	.000	.002	-.001	.001	.005
			6pm-9pm	.010	-.003	-.006	-.006	-.004	.004	.001	-.003	-.006	-.005	-.005	-.005	-.005	.016	.002	.001	.001	.006
			9pm-12midnight	.015	-.002	-.004	-.004	-.004	.001	-.001	-.003	-.005	-.005	-.005	-.002	-.001	.020	-.001	-.001	-.001	.002
		Saturday	12am-9am	.012	-.004	-.006	-.006	-.005	.004	.000	-.004	-.005	-.005	-.005	-.005	-.006	.021	.001	.001	8.388E-05	.006
			9am-12noon	.015	-.003	-.006	-.006	-.006	.001	-.002	-.003	-.004	-.004	-.004	-.004	-.005	.023	.001	.000	.000	.005
			12noon-3pm	.015	-.003	-.006	-.006	-.006	.001	-.002	-.003	-.004	-.004	-.004	-.004	-.005	.023	.001	.000	.000	.005
			3pm-6pm	.014	-.003	-.007	-.007	-.006	.001	-.002	-.003	-.004	-.004	-.004	-.004	-.005	.023	.001	.000	.000	.005
			6pm-9pm	.013	-.003	-.006	-.006	-.005	.002	-8.769E-06	-.003	-.005	-.005	-.005	-.004	-.004	.019	.000	1.679E-05	.000	.005
			9pm-12midnight	.016	-.002	-.004	-.004	-.004	.001	-.001	-.004	-.005	-.005	-.005	-.002	-.001	.021	-.001	-.001	-.001	.002
		Sunday	3am-9am and 9pm-12pm midnight	.011	-.002	-.003	-.002	-.001	.004	.002	-.004	-.006	-.005	-.005	-.003	-.002	.015	.000	.000	-.001	.001
			9am-12noon	.012	-.004	-.006	-.006	-.006	.003	-.001	-.003	-.004	-.004	-.004	-.004	-.006	.022	.001	2.841E-05	.000	.005
			12noon-3pm	.012	-.004	-.006	-.006	-.006	.003	-.001	-.003	-.004	-.004	-.004	-.004	-.005	.022	.001	.000	.000	.004
			3pm-6pm	.012	-.004	-.006	-.006	-.006	.003	-.001	-.003	-.004	-.004	-.004	-.004	-.005	.022	.001	.000	.000	.004
			6pm-9pm	.011	-.003	-.004	-.004	-.003	.003	.001	-.003	-.005	-.005	-.004	-.003	-.002	.017	.000	.000	-.001	.002

				Train Service Provision Density																	
				Weekday							Saturday						Sunday				
				12am-6am	6am-9am	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm	9pm-12midnight	12am-9am	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm	9pm-12midnight	3am-9am and 9pm-12pm midnight	9am-12noon	12noon-3pm	3pm-6pm	6pm-9pm
Reproduced Correlation	Bus/Ferry Service Provision Density	Weekday	12am-6am	.214	.240	.211	.212	.246	.220	.213	.211	.208	.208	.208	.213	.214	.218	.209	.209	.209	.216
			6am-9am	.681	.706	.679	.680	.708	.688	.681	.679	.675	.675	.675	.681	.681	.686	.676	.676	.676	.684
			9am-12noon	.700	.724	.699	.700	.725	.707	.701	.699	.695	.695	.695	.701	.701	.706	.696	.697	.697	.704
			12noon-3pm	.689	.713	.687	.688	.715	.696	.689	.687	.684	.684	.684	.690	.689	.694	.685	.685	.685	.693
			3pm-6pm	.675	.700	.673	.674	.701	.681	.675	.673	.669	.669	.669	.675	.675	.680	.670	.670	.670	.678
			6pm-9pm	.534	.564	.531	.532	.569	.541	.533	.531	.527	.527	.527	.533	.534	.539	.528	.528	.528	.537
			9pm-12midnight	.579	.607	.576	.577	.610	.586	.578	.576	.572	.572	.572	.579	.579	.584	.574	.574	.574	.582
		Saturday	12am-9am	.680	.705	.678	.679	.706	.687	.680	.678	.675	.675	.675	.680	.680	.685	.676	.676	.676	.684
			9am-12noon	.700	.724	.699	.699	.725	.707	.700	.698	.695	.695	.695	.701	.700	.705	.696	.696	.696	.704
			12noon-3pm	.684	.709	.682	.683	.710	.691	.684	.682	.678	.678	.678	.684	.684	.689	.679	.680	.680	.687
			3pm-6pm	.668	.694	.667	.667	.696	.675	.668	.666	.663	.663	.663	.669	.669	.674	.664	.664	.664	.672
			6pm-9pm	.710	.734	.709	.710	.735	.717	.710	.708	.705	.705	.705	.711	.711	.715	.706	.706	.706	.714
			9pm-12midnight	.765	.786	.764	.765	.785	.771	.766	.764	.761	.761	.761	.766	.766	.770	.762	.762	.762	.769
		Sunday	3am-9am and 9pm-12pm midnight	.088	.118	.083	.085	.128	.094	.086	.083	.080	.080	.080	.086	.086	.091	.081	.081	.081	.089
			9am-12noon	.559	.589	.557	.558	.594	.567	.559	.556	.553	.553	.553	.559	.559	.565	.554	.554	.554	.563
			12noon-3pm	.589	.618	.586	.588	.621	.596	.588	.586	.582	.582	.582	.589	.589	.594	.584	.584	.584	.592
			3pm-6pm	.605	.634	.603	.604	.637	.613	.605	.603	.599	.599	.599	.606	.605	.611	.600	.601	.601	.609
			6pm-9pm	.443	.475	.439	.441	.481	.450	.442	.439	.435	.435	.435	.442	.442	.448	.437	.437	.437	.446
	Train Service Provision Density	Weekday	12am-6am	.990[a]	.987	.994	.994	.975	.992	.993	.994	.993	.993	.993	.994	.993	.994	.993	.993	.993	.994
			6am-9am	.987	.985[a]	.990	.990	.974	.989	.989	.990	.989	.989	.989	.990	.989	.990	.989	.989	.989	.991
			9am-12noon	.994	.990	.998[a]	.997	.979	.996	.997	.997	.997	.997	.997	.998	.997	.997	.997	.997	.997	.998
			12noon-3pm	.994	.990	.997	.997[a]	.978	.995	.996	.997	.997	.997	.997	.997	.996	.997	.996	.996	.996	.997
			3pm-6pm	.975	.974	.979	.978	.963[a]	.978	.978	.978	.978	.978	.978	.979	.978	.979	.977	.977	.977	.979
			6pm-9pm	.992	.989	.996	.995	.978	.994[a]	.995	.995	.995	.995	.995	.996	.995	.995	.995	.995	.995	.996
			9pm-12midnight	.993	.989	.997	.996	.978	.995	.996[a]	.997	.996	.996	.996	.997	.996	.996	.996	.996	.996	.997
		Saturday	12am-9am	.994	.990	.997	.997	.978	.995	.997	.997[a]	.997	.997	.997	.998	.997	.997	.996	.996	.996	.998
			9am-12noon	.993	.989	.997	.997	.978	.995	.996	.997	.996[a]	.996	.996	.997	.996	.996	.996	.996	.996	.997
			12noon-3pm	.993	.989	.997	.997	.978	.995	.996	.997	.996	.996[a]	.996	.997	.996	.996	.996	.996	.996	.997
			3pm-6pm	.993	.989	.997	.997	.978	.995	.996	.997	.996	.996	.996[a]	.997	.996	.996	.996	.996	.996	.997
			6pm-9pm	.994	.990	.998	.997	.979	.996	.997	.998	.997	.997	.997	.998[a]	.997	.997	.997	.997	.997	.998
			9pm-12midnight	.993	.989	.997	.996	.978	.995	.996	.997	.996	.996	.996	.997	.996[a]	.996	.996	.996	.996	.997
		Sunday	3am-9am and 9pm-12pm midnight	.994	.990	.997	.997	.979	.995	.996	.997	.996	.996	.996	.997	.996	.997[a]	.996	.996	.996	.997
			9am-12noon	.993	.989	.997	.996	.977	.995	.996	.996	.996	.996	.996	.997	.996	.996	.995[a]	.995	.995	.997
			12noon-3pm	.993	.989	.997	.996	.977	.995	.996	.996	.996	.996	.996	.997	.996	.996	.995	.996[a]	.996	.997
			3pm-6pm	.993	.989	.997	.996	.977	.995	.996	.996	.996	.996	.996	.997	.996	.996	.995	.996	.996[a]	.997
			6pm-9pm	.994	.991	.998	.997	.979	.996	.997	.998	.997	.997	.997	.998	.997	.997	.997	.997	.997	.998[a]
Residual[b]	Bus/Ferry Service Provision Density	Weekday	12am-6am	.010	.003	.014	.012	-.002	.010	.015	.012	.015	.015	.014	.013	.016	.011	.012	.012	.012	.011
			6am-9am	-.002	-.001	-.003	-.003	-.001	-.003	-.002	-.004	-.003	-.003	-.003	-.003	-.002	-.002	-.004	-.004	-.004	-.003
			9am-12noon	.000	-.002	-.007	-.007	-.004	-.006	-.004	-.006	-.006	-.006	-.007	-.006	-.004	-.003	-.006	-.006	-.006	-.004
			12noon-3pm	.001	-.003	-.007	-.007	-.005	-.006	-.004	-.006	-.006	-.006	-.007	-.006	-.004	-.002	-.006	-.006	-.006	-.004
			3pm-6pm	.001	-.001	-.005	-.006	-.002	-.004	-.004	-.005	-.006	-.006	-.006	-.005	-.004	-.001	-.006	-.006	-.006	-.003
			6pm-9pm	.001	.002	.002	.003	.003	.004	.001	.004	.001	.001	.001	.002	.001	.004	.003	.003	.003	.003
			9pm-12midnight	-.001	.004	-.001	-.001	.006	.001	-.001	.000	-.002	-.002	-.002	-8.769E-06	-.001	.002	-.001	-.001	-.001	.001
		Saturday	12am-9am	-.005	7.204E-05	-.003	-.004	.002	-.003	-.003	-.004	-.003	-.003	-.003	-.003	-.004	-.004	-.003	-.003	-.003	-.003
			9am-12noon	-.006	-.004	-.005	-.005	-.003	-.006	-.005	-.005	-.004	-.004	-.004	-.005	-.005	-.006	-.004	-.004	-.004	-.005
			12noon-3pm	-.005	-.004	-.005	-.005	-.003	-.005	-.005	-.005	-.004	-.004	-.004	-.005	-.005	-.005	-.004	-.004	-.004	-.005
			3pm-6pm	-.005	-.004	-.005	-.004	-.003	-.005	-.005	-.005	-.004	-.004	-.004	-.005	-.005	-.005	-.004	-.004	-.004	-.004
			6pm-9pm	-7.979E-05	-.001	-.004	-.005	-.003	-.005	-.002	-.005	-.004	-.004	-.004	-.004	-.002	-.003	-.004	-.004	-.004	-.003
			9pm-12midnight	.002	.004	-.004	-.006	.002	-.005	-.001	-.006	-.005	-.005	-.005	-.004	-.001	-.002	-.006	-.005	-.005	-.002
		Sunday	3am-9am and 9pm-12pm midnight	.013	.003	.021	.021	.000	.016	.020	.021	.023	.023	.023	.019	.021	.015	.022	.022	.022	.017
			9am-12noon	-.001	.000	.001	.002	.002	.002	-.001	.001	.001	.001	.001	.000	-.001	.000	.001	.001	.001	.000
			12noon-3pm	.001	-.001	.001	.001	-.001	.001	-.001	.001	.000	.000	.000	1.679E-05	-.001	.000	2.841E-05	.000	.000	.000
			3pm-6pm	-4.359E-05	6.792E-05	.000	.001	.001	.001	-.001	8.388E-05	.000	.000	.000	.000	-.001	-.001	.000	.000	.000	-.001
			6pm-9pm	-.004	.001	.005	.007	.005	.006	.002	.006	.005	.005	.005	.005	.002	.001	.005	.004	.004	.002
	Train Service Provision Density	Weekday	12am-6am		-.007	.000	-.002	-.014	-.005	.003	.000	.002	.002	.002	-.001	.003	.001	.004	.004	.004	.002
			6am-9am	-.007		.000	.001	.023	.006	-.004	-.001	-.004	-.004	-.004	.001	-.003	.001	-.004	-.005	-.005	.000
			9am-12noon	.000	.000		.002	.000	.002	.001	.002	.003	.003	.003	.002	.001	.000	.001	.001	.001	.001
			12noon-3pm	-.002	.001	.002		.003	.003	.001	.002	.002	.002	.002	.002	.001	.000	.001	.001	.001	.000
			3pm-6pm	-.014	.023	.000	.003		.011	-.007	-.001	-.006	-.006	-.006	.001	-.007	.001	-.006	-.007	-.007	-.001
			6pm-9pm	-.005	.006	.002	.003	.011		-.001	.002	.000	.000	.000	.002	-.001	.001	-.001	-.001	-.001	.000
			9pm-12midnight	.003	-.004	.001	.001	-.007	-.001		.001	.003	.003	.003	.001	.004	.002	.002	.002	.002	.001
		Saturday	12am-9am	.000	-.001	.002	.002	-.001	.002	.001		.002	.002	.002	.002	.001	.002	.002	.002	.002	.001
			9am-12noon	.002	-.004	.003	.002	-.006	.000	.003	.002		.004	.004	.002	.003	-5.824E-05	.002	.003	.003	.001
			12noon-3pm	.002	-.004	.003	.002	-.006	.000	.003	.002	.004		.004	.002	.003	-5.824E-05	.002	.003	.003	.001
			3pm-6pm	.002	-.004	.003	.002	-.006	.000	.003	.002	.004	.004		.002	.003	-7.114E-05	.002	.003	.003	.001
			6pm-9pm	-.001	.001	.002	.002	.001	.002	.001	.002	.002	.002	.002		.001	.001	.001	.001	.001	.001
			9pm-12midnight	.003	-.003	.001	.001	-.007	-.001	.004	.001	.003	.003	.003	.001		.002	.002	.002	.002	.001
		Sunday	3am-9am and 9pm-12pm midnight	.001	.001	.000	.000	.001	.001	.002	.002	-5.824E-05	-5.824E-05	-7.114E-05	.001	.002		.001	.000	.000	.002
			9am-12noon	.004	-.004	.001	.001	-.006	-.001	.002	.002	.002	.002	.002	.001	.002	.001		.004	.004	.002
			12noon-3pm	.004	-.005	.001	.001	-.007	-.001	.002	.002	.003	.003	.003	.001	.002	.000	.004		.004	.002
			3pm-6pm	.004	-.005	.001	.001	-.007	-.001	.002	.002	.003	.003	.003	.001	.002	.000	.004	.004		.002
			6pm-9pm	.002	.000	.001	.000	-.001	.000	.001	.001	.001	.001	.001	.001	.001	.002	.002	.002	.002	

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 18 (2.0%) nonredundant residuals with absolute values greater than 0.05.

Table 16: Reproduced Correlation Matrix (Public Transport Service Provision Factors)

Table 16: Reproduced Correlation Matrix (Public Transport Service Provision Factors) shows that most of the residual values are less than 0.05. These values are highlighted in green. Only 18 non-redundant residuals have residual values greater than 0.05, and they account for only 2% of all correlations. Therefore, it can be concluded that the correlations in the public transport service provision density factors model are significantly close to those in the observed data.

5.2.1.6 Public Transport Service Provision Density Factor Equations

Based on the factor loadings shown in Table 15: Rotated Component Matrix (Public Transport Service Provision Factors), public transport service provision factors can be described in terms of the following equations:

$$
\begin{aligned}
&\textit{Train service Provision Factor} \\
&\quad = .956T1_i + .956T2_i + .956T3_i + .955T4_i + .955T5_i + .955T6_i + .955T7_i \\
&\quad + .955T8_i + .954T9_i + .954T10_i + .953T11_i + .953T12_i + .953T13_i \\
&\quad + .952T14_i + .95T15_i + .949T16_i + .935T17_i + .952T18_i
\end{aligned}
$$

$$
\begin{aligned}
&\textit{Bus/ferry service Provision Factor} \\
&\quad = .946B1_i + .936B2_i + .932B3_i + .924B4_i + .917B5_i + .899B6_i + .882B7_i \\
&\quad + .881B8_i + .873B9_i + .869B10_i + .864B11_i + .859B12_i + .859B13_i + .856B14_i \\
&\quad + .852B15_i + .789B16_i + .772B17_i + .697B18_i
\end{aligned}
$$

Where:

T1		Train Service Provision_Density_PerKm2_Saturday 15-18hr
T2		Train Service Provision_Density_PerKm2_Saturday 9-12hr
T3		Train Service Provision_Density_PerKm2_Saturday 12-15hr
T4	=	Train Service Provision_Density_PerKm2_Weekday 9-12hr
T5	=	Train Service Provision_Density_PerKm2_Saturday 0-9hr
T6	=	Train Service Provision_Density_PerKm2_Sunday 12-15hr
T7	=	Train Service Provision_Density_PerKm2_Sunday 15-18hr
T8	=	Train Service Provision_Density_PerKm2_Sunday 9-12hr
T9	=	Train Service Provision_Density_PerKm2_Saturday 18-21hr
T10	=	Train Service Provision_Density_PerKm2_Weekday 12-15hr
T11	=	Train Service Provision_Density_PerKm2_Weekday 21-24hr
T12	=	Train Service Provision_Density_PerKm2_Saturday 21-24hr
T13	=	Train Service Provision_Density_PerKm2_Sunday 18-21hr
T14	=	Train Service Provision_Density_PerKm2_Sunday 3-9hr and 21-24hr
T15	=	Train Service Provision_Density_PerKm2_Weekday 0-6hr
T16	=	Train Service Provision_Density_PerKm2_Weekday 18-21hr

T17	=	Train Service Provision_Density_PerKm2_Weekday 6-9hr
T18	=	Train Service Provision_Density_PerKm2_Weekday 15-18hr
B1	=	Bus/Ferry Service Provision_Density_PerKm2_Sunday 18-21hr
B2	=	Bus/Ferry Service Provision_Density_PerKm2_Sunday 9-12hr
B3	=	Bus/Ferry Service Provision_Density_PerKm2_Weekday 18-21hr
B4	=	Bus/Ferry Service Provision_Density_PerKm2_Sunday 12-15hr
B5	=	Bus/Ferry Service Provision_Density_PerKm2_Sunday 15-18hr
B6	=	Bus/Ferry Service Provision_Density_PerKm2_Weekday 21-24hr
B7	=	Bus/Ferry Service Provision_Density_PerKm2_Saturday 15-18hr
B8	=	Bus/Ferry Service Provision_Density_PerKm2_Weekday 6-9hr
B9	=	Bus/Ferry Service Provision_Density_PerKm2_Saturday 12-15hr
B10	=	Bus/Ferry Service Provision_Density_PerKm2_Weekday 15-18hr
B11	=	Bus/Ferry Service Provision_Density_PerKm2_Saturday 0-9hr
B12	=	Bus/Ferry Service Provision_Density_PerKm2_Saturday 18-21hr
B13	=	Bus/Ferry Service Provision_Density_PerKm2_Weekday 12-15hr
B14	=	Bus/Ferry Service Provision_Density_PerKm2_Saturday 9-12hr
B15	=	Bus/Ferry Service Provision_Density_PerKm2_Weekday 9-12hr
B16	=	Bus/Ferry Service Provision_Density_PerKm2_Saturday 21-24hr
B17	=	Bus/Ferry Service Provision_Density_PerKm2_Sunday 3-9hr and 21-24hr
B18	=	Bus/Ferry Service Provision_Density_PerKm2_Weekday 0-6hr

5.2.1.7 Descriptive Analysis: Public Transport Service Provision Factors

		Train Service Provision Density Factor	Bus/Ferry Service Provision Density Factor	Average Public Transport Stops per Km2
Mean		.000	.000	11.244
Std. Deviation		1.000	1.000	8.106
Skewness		11.218	4.761	1.710
Std. Error of Skewness		.143	.143	.143
Kurtosis		152.944	37.235	8.486
Std. Error of Kurtosis		.284	.284	.284
Minimum		-.202	-.758	.138
Maximum		14.453	9.916	68.150
Percentiles	10	-.202	-.739	1.566
	25	-.202	-.602	4.990
	50	-.202	-.290	10.962
	65	-.202	.064	13.903
	75	-.202	.266	15.651
	80	-.202	.400	16.883
	90	.269	.903	20.386

Table 17: Descriptive Statistic on Public Transport Service Provision Factors

As shown in Table 17, the mean and standard deviation values for both train and bus/ferry service provision density factors have been standardized to 0 and 1, respectively. The mean value for average public transport stops per km2 is 11.24 and standard deviation is 8.11. This indicates that there is significant variation among average public transport stops per km^2 across the Perth metropolitan suburbs; this is confirmed by the variance of its minimum (0.14) and maximum (68.15) values. The respective skewness values are 11.218, 4.761 and 1.71 and it can be concluded that these variables are normally distributed despite being positively (left) skewed.

Perth has the highest train and bus/ferry service provision factor scores at 14.453 and 9.916, respectively. Suburbs without train stops have 0 or negative train factor scores. The percentile results indicate that more than 80% of the Perth metropolitan suburbs score below the average on train service provision factors. The highest train service factor score for 90% of these suburbs is 0.269, with a maximum train factor score of 6.08. Therefore, it can be concluded that good train service is only provided for less than 10% of the Perth metropolitan suburbs, reflecting the fact that 239 out of 292 suburbs (81.85%) do not have train stations.

The highest bus/ferry service factor score in 90% of the suburbs is 0.903, with a maximum of 9.916. Therefore, high bus/ferry service is provided in only 10% of the suburbs. The ten suburbs that have the lowest bus/ferry service provision factor scores are Cottesloe, Guildford, Woodbridge, Ashfield, West Leederville, Hazelmer, Neerabup, Kenwick, Jolimont and Wellard. Moreover, 65% of the suburbs have below average bus/ferry services provision factor scores.

The average public transport stops per km^2 in the Perth metropolitan suburbs is 11.206. While Northbridge has a maximum of 68 stops per km^2 and Bullsbrook has 0.14 stops per km^2, 60% of the suburbs have below average stops per km^2. From the value of the highest average stops per km2 (68.15) and the value at the 90th percentile (20.37), it can be concluded that public transport service stops are densely located in only 10% of the Perth metropolitan Suburbs. Furthermore, the variable's skewness is positive at 1.7, and its kurtosis value is relatively high at 8.43. The histogram in the following Figure 43 shows that the normal distribution curve of the average stops per km^2 in the Perth metropolitan suburbs is left-skewed and pointy. It is obvious that the majority of suburbs have 0-20 stops per km^2 and only a very few suburbs have more than 40 stops per km^2.

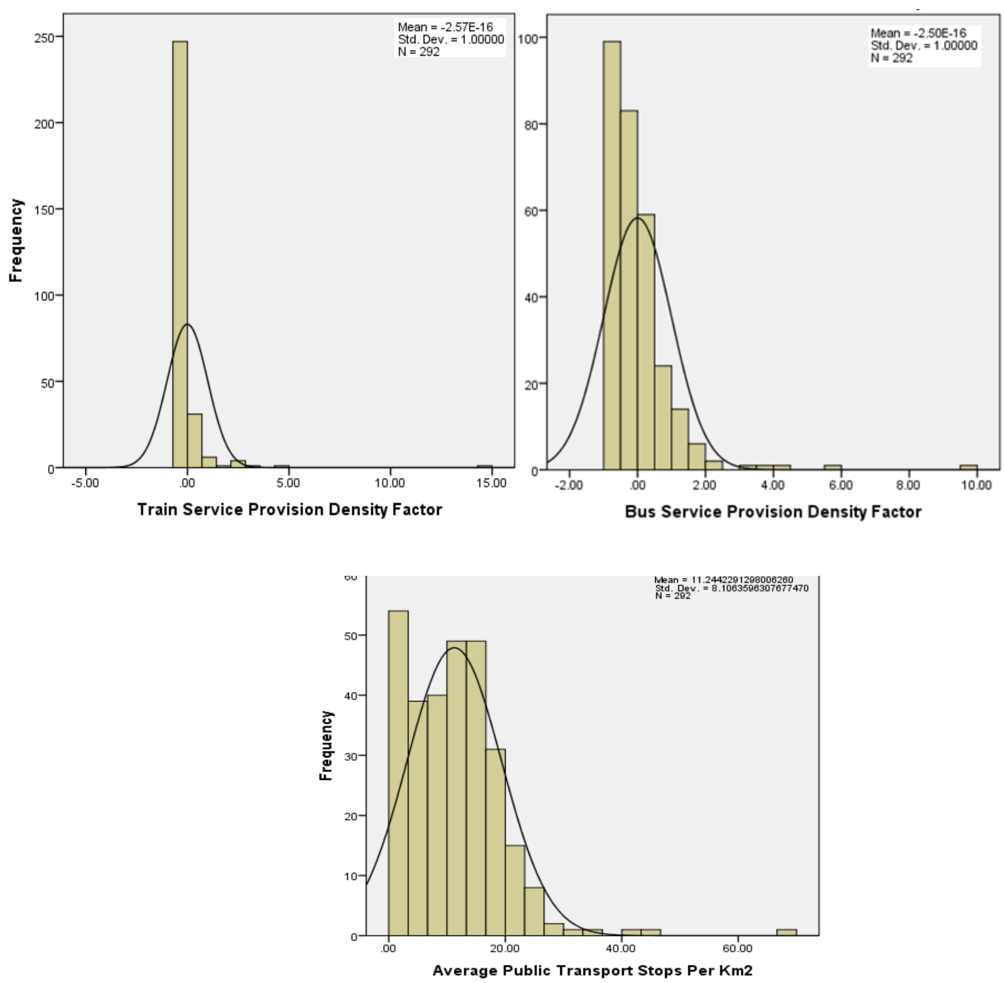


Figure 43: Normal Distribution Figures of Public Transport Service Provision Density Factor Scores and Average Public Transport Stops per Km^2

Figure 43 shows that the train service provision factor scores, bus/ferry service provision factor scores and average public transport stops per km2 are normally distributed but positively skewed. These public transport service provision factor scores are used in the regression analysis to develop a public transport usage model.

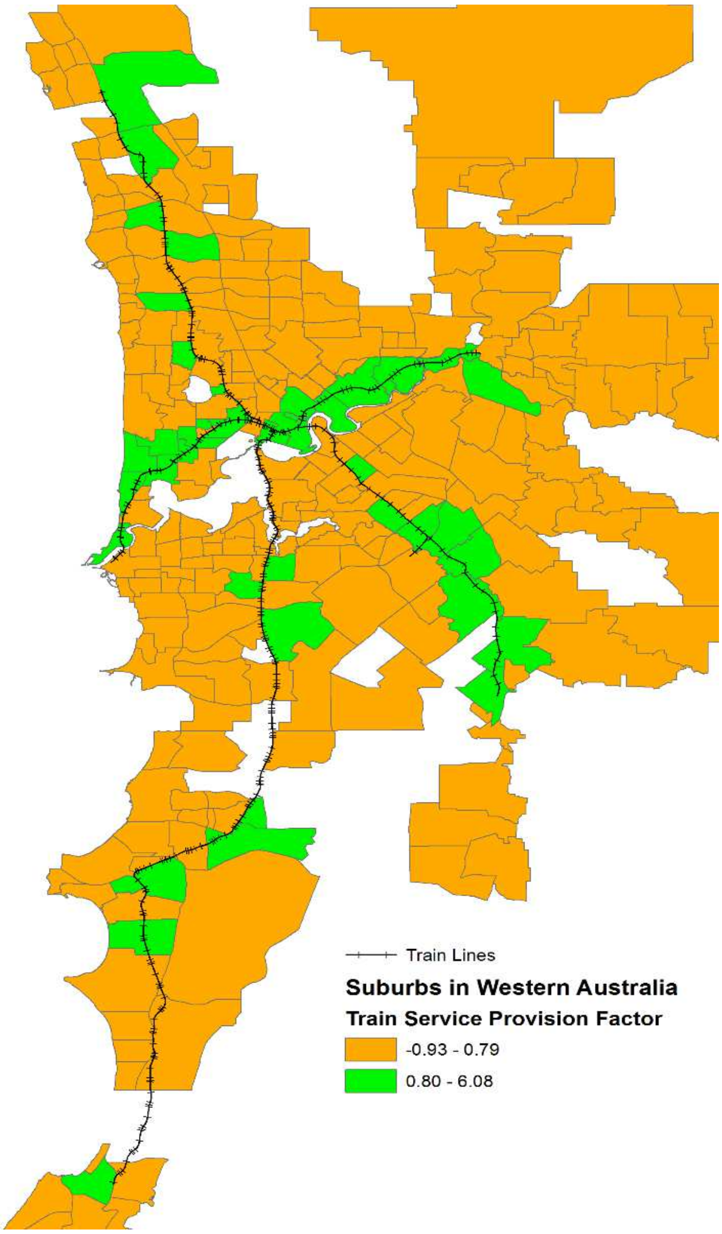


Figure 44: Train Service Provision Factor Score Map

Figure 44 illustrates the factor scores for train service provision in the Perth metropolitan suburbs. Service provision on the Fremantle, Midland and Armadale train lines is relatively better than on the Joondalup and Mundurah lines. The train service provision on the Joondalup train line is comparatively lower because there are fewer train stations along this line; this is reflected in lower factor scores. Similarly, there are only 9 train-stations (excluding Perth central train station) on the Mandurah line, resulting in lower train service provision factor scores.

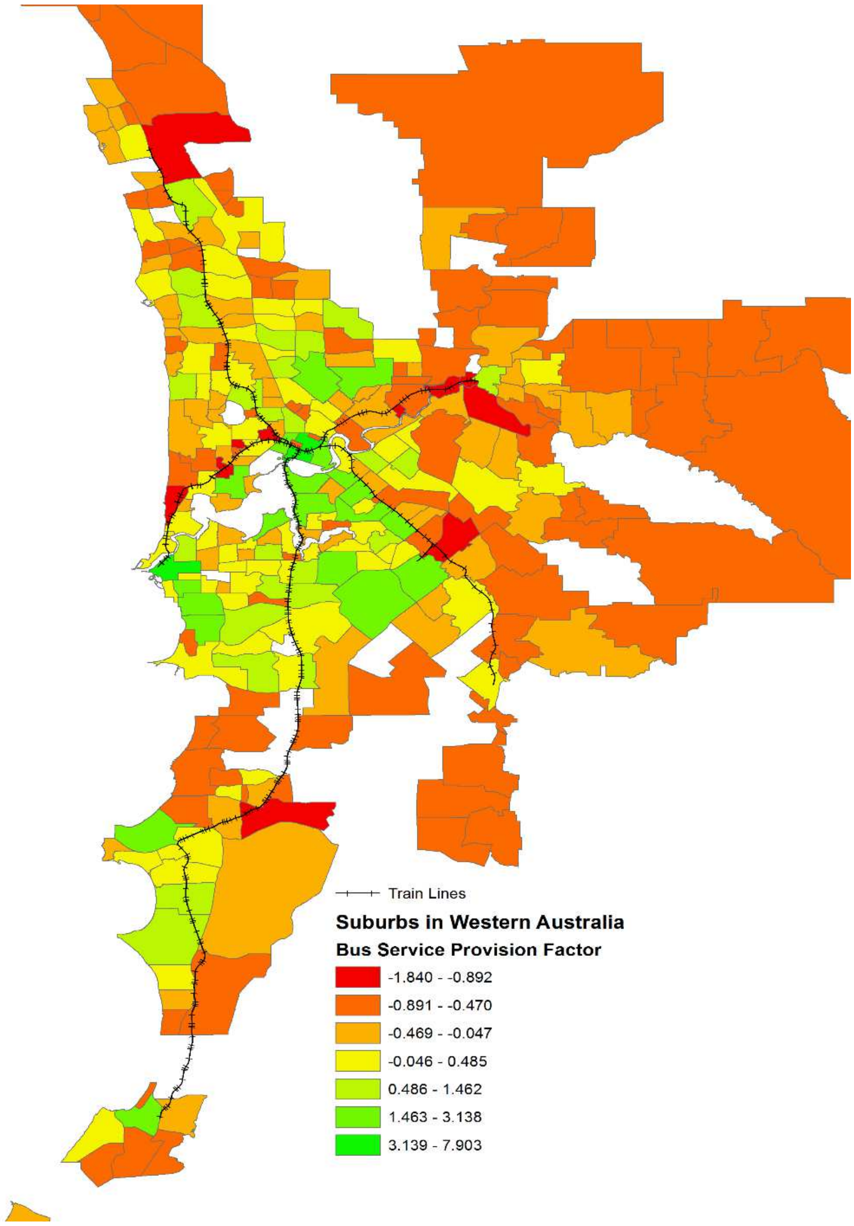


Figure 45: Bus/Ferry Service Provision Factor Scores Map

The map in Figure 45 shows the suburbs where good bus/ferry service is provided across the Perth metropolitan suburbs in 2009. The Perth Central Business District is one of the significant areas. Suburbs where universities exist such as Crawley and Bentley have high public transport usage and benefit from good bus/ferry service provision. Suburbs that have the advantage of good bus services are Joondalup, Kingsley, South Perth, Como, Applecross, St James, Wilson, Bullcreek, Wellington, Canningvale, Huntindale, Thornlie, Cannington, Murdoch, Leeming, Beeliar, Rockingham, Warnbro, and Mandurah. In addition, the public transport usage in Fremantle is the highest after Perth, and good bus service provision leads to high public transport usage in surrounding suburbs such as East/South Fremantle, Beaconsfield, White Gum Valley, Hamilton Hill, and Spearwood.

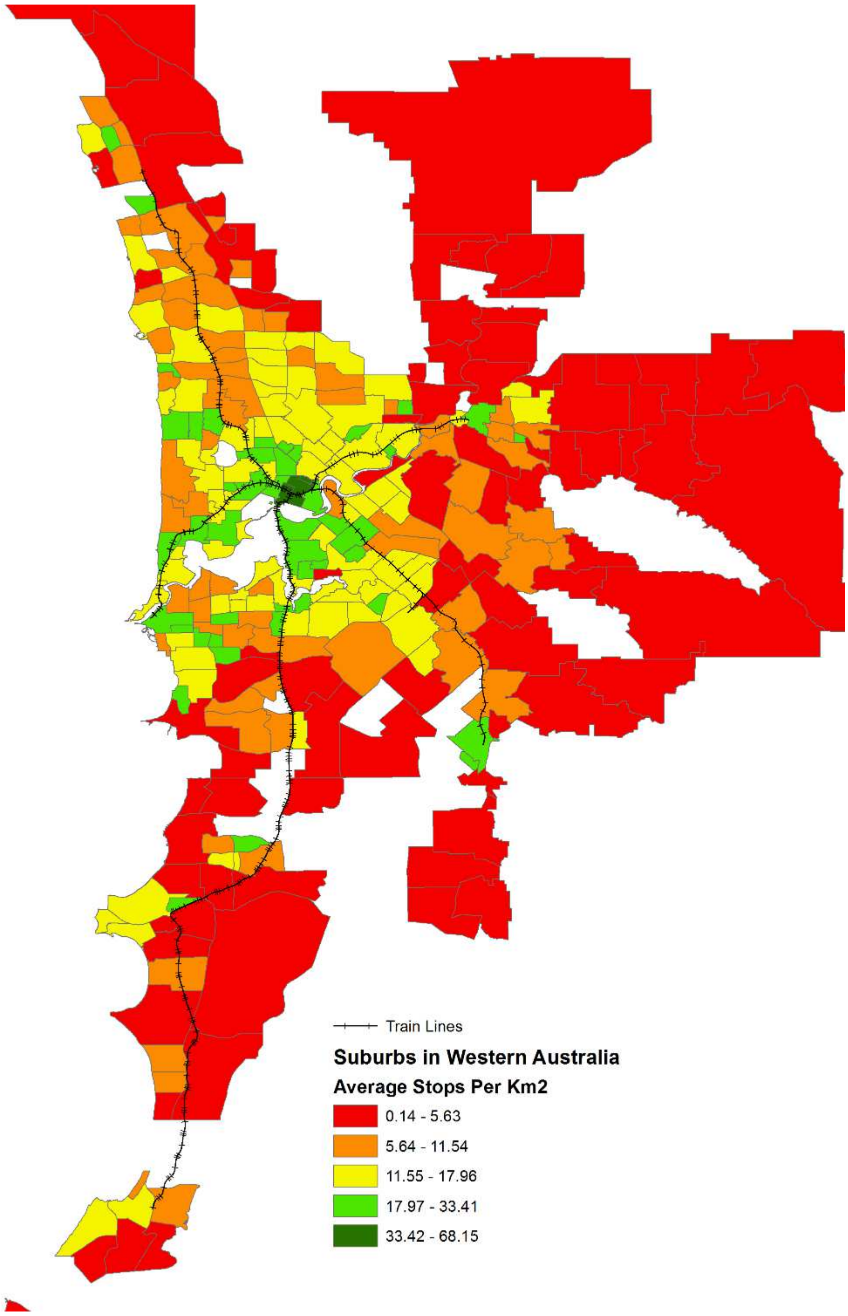


Figure 46: Average Stops per Km2 Map

Figure 46 shows the average public transport stops per km2 in the Perth metropolitan suburbs. It is noticeable that Perth Central Business District area has the highest number of average stops per km2, ranging between 33.42 and 68.15. Other suburbs that have high average stops per km2 are Ridgewood, Kinross, Scaborough, Doubleview, Innaloo, Joodanna, Yokine, Nedlands, Mount Claremount, North Perth, Willeton, Karrawara, Como, Manning, Armadale, White Gum Valley, Fremantle, Kardinya, Coolbellup, and Orelia. Comparing the maps of the different variables used in this study reveals that the average stops per km2 reflects the estimated resident population density per km2. Therefore, it can be confirmed that more public transport stops are provided in the areas with high estimated resident population densities.

5.2.2 Land Use Characteristics and Socio-Economic Factors

Factor analysis is also used to extract latent land use characteristics and socio-economic factors from the estimated resident population densities by age and gender, student (up to year 12) population density, number of residents from different weekly income groups, average rent, and average car ownership per household in each suburb.

5.2.2.1 Sampling Adequacy Test for Land Use Characteristics and Socio-Economic Factor Analysis

Field (2013) explains that the sampling adequacy can be measured by using the Kaiser-Meyer-Olkin method. This method represents the ratio of the squared correlation between variables to the squared partial correlation between variables. The resulting number can vary from 0 to 1. Field (2013) points out that a value close to 1 indicates the relative compactness among the patterns of correlations, which results in distinct and reliable factors. He advises that Kaiser recommended accepting values greater than 0.5 as barely acceptable.

Leech (2011) explains that Kaiser-Meyer-Olkin Measure of Sampling Adequacy should be greater than 0.7, which indicates sufficient items for each factor. She also recommends that the significance value of the Bartlett's Test of Sphericity should be less than 0.05, indicating that the correlation matrix is significantly different from an identity matrix (in which correlations between variables are all zero).

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.815
	Approx. Chi-Square	7254.127
Bartlett's Test of Sphericity	Df	105
	Sig.	.000

Table 18: KMO and Bartlett's Test (Socio-Economic Variables)

As shown in Table 18, KMO measure is 0.815 (greater than 0.7). Based on this result, it can be concluded that there are enough test cases predicted by each factor. Also, the Significance value of the Bartlett's Test of Sphericity is less than 0.05, meaning that the urban form and socio-economic variables are highly correlated enough to provide a reasonable basis for factor analysis.

5.2.2.2 Communality Test on Land Use Characteristics and Socio-Economic Factor Analysis

A Communality test will also be used to identify whether there is a problem of sample size. Leech (2011) points out that the initial communalities are derived from the squared multiple correlations between an item and all other items, and that they represent the relations between the variable and all other variables before rotation. She explains that there is a high chance that a small sample size can distort results if many or most communalities are lower than 0.30.

Communalities

	Initial	Extraction
Estimated Resident Population Density Age: 0-16 (Male)	1.000	.748
Estimated Resident Population Density Age: 0-16 (Female)	1.000	.761
Estimated Resident Population Density Age: 17-35 (Male)	1.000	.797
Estimated Resident Population Density Age: 17-35 (Female)	1.000	.841
Estimated Resident Population Density Age: 36-64 (Male)	1.000	.949
Estimated Resident Population Density Age: 36-64 (Female)	1.000	.949
Estimated Resident Population Density Age: 65 and over (Male)	1.000	.798
Estimated Resident Population Density Age: 65 and over (Female)	1.000	.719
Student Population Density	1.000	.274
No of Residents Whose Weekly Income Below 250	1.000	.914
No of Residents Whose Weekly Income Between 250 and 1000	1.000	.957
No of Residents Whose Weekly Income Between 1000 and 2000	1.000	.931
No of Residents Whose Weekly Income Between Above 2000	1.000	.594
Average Rent	1.000	.796
Average Car Ownership per Household	1.000	.656

Extraction Method: Principal Component Analysis.

Table 19: Communalities of Land Use Characteristics and Socio-Economic Variables

The extraction values in Table 19 represent the common variances and can be used to verify that eigenvalues greater than 1 are applicable, depending on the number of variables and test cases. The reliability of a test can be measured by its communality given by the sum of the loading squares on the extracted factors (Gray, 2012). Nine land use characteristics, six socio-economic variables, and 293 test cases are included in this analysis. According to Field (2005), when there are fewer than 30 variables, extraction values should be greater than 0.7. Similarly, when more than 250 test cases are analysed, their communalities should be greater than 0.6. All of the extraction values, with the exception of the one for the student population density variable, are greater than 0.6. Therefore, it can be concluded that eigenvalues (greater than 1) are applicable for the factor analysis of land use characteristics and socio-economic variables.

5.2.2.3 Factors Selection from Land Use Characteristics and Socio-Economic Factor Analysis

In this section, the factor analysis results for land use characteristics and socio-economic variables are discussed. Leech (2011) explains that the initial eigenvalues in the "Total Variance Explained" table refer to the variance accounted for in terms of the number of "items' worth" of variance each explains. Components 1 and 2 together explain almost as twice variance as all other components combined, as shown in Table 20.

Total Variance Explained

Component	Initial Eigenvalues			Loadings			Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.817	52.111	52.111	7.817	52.111	52.111	6.768	45.117	45.117
2	2.299	15.328	67.439	2.299	15.328	67.439	3.298	21.986	67.103
3	1.569	10.457	77.896	1.569	10.457	77.896	1.619	10.793	77.896
4	.921	6.138	84.034						
5	.760	5.066	89.100						
6	.568	3.789	92.889						
7	.484	3.226	96.116						
8	.328	2.189	98.304						
9	.105	.701	99.005						
10	.073	.487	99.492						
11	.024	.159	99.651						
12	.019	.125	99.776						
13	.014	.092	99.868						
14	.011	.074	99.941						
15	.009	.059	100.000						

Extraction Method: Principal Component Analysis.

Table 20: Total Variance Explained (Urban form and Socio-Economic Factors)

Table 20 shows how the variance is divided among the 15 possible factors. It is noticeable that only three factors have eigenvalues (greater than 1), which is the common criterion for the usefulness of a factor. The "% of variance from extraction sums of squared loadings" and "% variance of from rotation sums of squared loadings" columns show that the percentage of co-variation among items accounted for by each factor receives more equal contributions after rotation. The scree plot in Figure 42 verifies the number of factors that should be extracted based on their eigenvalues.

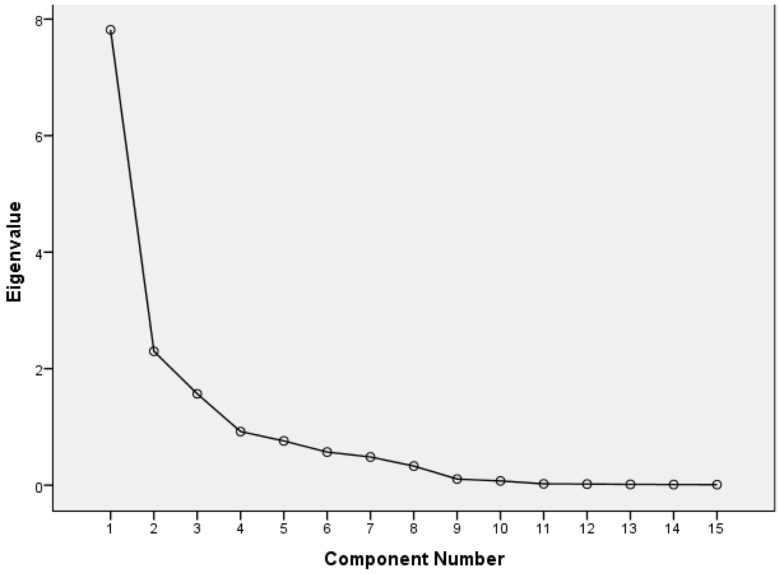


Figure 47: Scree Plot of Land Use Characteristics and Socio-Economic Factors

The scree plot of land use characteristics and socio-economic factors shown in Figure 47 confirms that there are three land use characteristics and socio-economic factors that explain 77.89% of total public transport usage density. The amount of variance accounted for (eigenvalue) by the successive components declines dramatically after the first three factors are extracted.

5.2.2.4 Land Use Characteristics and Socio-Economic Factor Loadings

Rotated Component Matrix[a]

	Component		
	1	2	3
Estimated Resident Population Density Age: 36-64 (Male)	.941		
Estimated Resident Population Density Age: 36-64 (Female)	.934		
Estimated Resident Population Density Age: 17-35 (Female)	.900		
Estimated Resident Population Density Age: 17-35 (Male)	.875		
Estimated Resident Population Density Age: 65 and over (Male)	.871		
Estimated Resident Population Density Age: 65 and over (Female)	.824		
Estimated Resident Population Density Age: 0-16 (Female)	.821		
Estimated Resident Population Density Age: 0-16 (Male)	.819		
Student Population Density	.503		
No of Residents Whose Weekly Income Between 250 and 1000		.954	
No of Residents Whose Weekly Income Below 250		.928	
No of Residents Whose Weekly Income Between 1000 and 2000		.923	
No of Residents Whose Weekly Income Between Above 2000		.557	.423
Average Rent			.879
Average Car Ownership Per Household			.712

Extraction Method: Principal Component Analysis.

a. Rotation converged in 5 iterations.

Table 21: Rotated Component Matrix of Land Use characteristics and Socio-Economic Factors

Table 21 shows how much each socio-economic variable contributes to each component of the socio-economic factors before applying the Varimax rotation method. It is useful to find out whether the Varimax rotation enhances the factoring of socio-economic variables. Before rotation, some of these variables are accounted for by more than one component, which makes the results harder to interpret.

5.2.2.5 Validity of Land Use Characteristics and Socio-Economic Factor Model

			Estimated Resident Population Density									No of Residents Whose Weekly Income					
			Age: 0-16 (Male)	Age: 0-16 (Female)	Age: 17-35 (Male)	Age: 17-35 (Female)	Age: 36-64 (Male)	Age: 36-64 (Female)	Age: 65 and over (Male)	Age: 65 and over (Female)	Student Population Density	Below 250	Between 250 and 1000	Between 1000 and 2000	Between Above 2000	Average Rent	Avg Car Ownership Per Household
Reproduced Correlation	Estimated Resident Population Density	Age: 0-16 (Male)	.748[a]	.754	.731	.755	.839	.842	.743	.700	.447	.376	.398	.458	.451	.187	-.249
		Age: 0-16 (Female)	.754	.761[a]	.733	.757	.844	.848	.745	.702	.450	.385	.409	.470	.462	.196	-.243
		Age: 17-35 (Male)	.731	.733	.797[a]	.819	.846	.836	.796	.756	.434	.274	.283	.295	.281	-.006	-.440
		Age: 17-35 (Female)	.755	.757	.819	.841[a]	.873	.863	.818	.777	.448	.287	.296	.312	.299	.006	-.442
		Age: 36-64 (Male)	.839	.844	.846	.873	.949[a]	.948	.855	.808	.499	.395	.415	.466	.454	.146	-.345
		Age: 36-64 (Female)	.842	.848	.836	.863	.948	.949[a]	.847	.799	.502	.405	.428	.488	.480	.180	-.312
		Age: 65 and over (Male)	.743	.745	.796	.818	.855	.847	.798[a]	.757	.439	.316	.327	.342	.316	.010	-.423
		Age: 65 and over (Female)	.700	.702	.756	.777	.808	.799	.757	.719[a]	.412	.298	.307	.316	.287	-.010	-.416
		Student Population Density	.447	.450	.434	.448	.499	.502	.439	.412	.274[a]	.159	.174	.225	.262	.167	-.114
	No of Residents Whose Weekly Income	Below 250	.376	.385	.274	.287	.395	.405	.316	.298	.159	.914[a]	.934	.881	.512	-.145	-.168
		Between 250 and 1000	.398	.409	.283	.296	.415	.428	.327	.307	.174	.934	.957[a]	.915	.551	-.104	-.141
		Between 1000 and 2000	.458	.470	.295	.312	.466	.488	.342	.316	.225	.881	.915	.931[a]	.651	.111	.004
		Between Above 2000	.451	.462	.281	.299	.454	.480	.316	.287	.262	.512	.551	.651	.594[a]	.394	.179
		Average Rent	.187	.196	-.006	.006	.146	.180	.010	-.010	.167	-.145	-.104	.111	.394	.796[a]	.570
		Average Car Ownership Per Household	-.249	-.243	-.440	-.442	-.345	-.312	-.423	-.416	-.114	-.168	-.141	.004	.179	.570	.656[a]
Residual[b]	Estimated Resident Population Density	Age: 0-16 (Male)		.226	-.077	-.059	-.003	.028	-.107	-.114	-.028	.010	.008	-.033	-.138	-.098	.081
		Age: 0-16 (Female)	.226		-.081	-.062	-.005	.024	-.097	-.106	-.026	.011	.008	-.034	-.131	-.090	.071
		Age: 17-35 (Male)	-.077	-.081		.166	.014	-.037	-.081	-.091	-.074	-.001	.020	.030	-.007	.044	.024
		Age: 17-35 (Female)	-.059	-.062	.166		.010	-.034	-.079	-.086	-.049	-.005	.020	.029	-.014	.027	.032
		Age: 36-64 (Male)	-.003	-.005	.014	.010		.034	-.023	-.047	-.051	-.014	-.004	.011	-.005	-.019	.017
		Age: 36-64 (Female)	.028	.024	-.037	-.034	.034		-.002	-.015	-.037	-.007	-.005	-.002	-.018	-.038	.031
		Age: 65 and over (Male)	-.107	-.097	-.081	-.079	-.023	-.002		.208	-.058	.025	-.004	-.022	.047	.041	.012
		Age: 65 and over (Female)	-.114	-.106	-.091	-.086	-.047	-.015	.208		-.084	.017	-.006	-.015	.065	.052	.005
		Student Population Density	-.028	-.026	-.074	-.049	-.051	-.037	-.058	-.084		.029	.014	-.015	-.044	-.068	-.032
	No of Residents Whose Weekly Income	Below 250	.010	.011	-.001	-.005	-.014	-.007	.025	.017	.029		.039	-.061	-.121	.048	.047
		Between 250 and 1000	.008	.008	.020	.020	-.004	-.005	-.004	-.006	.014	.039		-.026	-.105	.031	.045
		Between 1000 and 2000	-.033	-.034	.030	.029	.011	-.002	-.022	-.015	-.015	-.061	-.026		.059	-.010	-.036
		Between Above 2000	-.138	-.131	-.007	-.014	-.005	-.018	.047	.065	-.044	-.121	-.105	.059		-.008	-.219
		Average Rent	-.098	-.090	.044	.027	-.019	-.038	.041	.052	-.068	.048	.031	-.010	-.008		-.159
		Average Car Ownership Per Household	.081	.071	.024	.032	.017	.031	.012	.005	-.032	.047	.045	-.036	-.219	-.159	

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 34 (32.0%) nonredundant residuals with absolute values greater than 0.05.

Table 22: Reproduced Correlation Matrix of Land Use Characteristics and Socio-Economics Variables

Field (2013) suggests checking the percentage of ‘non-redundant residuals with absolute values greater than 0.05’ at the bottom of the *Reproduced Correlation* to determine whether this percentage is less than 50%. He explains that the correlations in the reproduced matrix differ from the ones in the R-matrix because they are calculated from the model rather than the observed data. The differences between the observed correlations and the ones based on the model can be used to assess the fit of the model. He recommends that these differences should be small and less than 0.05. Additionally, in a good model, the percentage of non-redundant residuals with absolute values greater than 0.05 should be low. As shown in the note of Table 22, there are only 34 non-redundant residuals with absolute values greater than 0.05, or 32%.

5.2.2.6 Land Use Characteristics and Socio-economic Factor Equations

As shown in Table 21, one urban form and two socio-economic factors can be extracted from the data. These factors are named as follows:

1. **Student and Mid-aged Dominant Resident Population Density Factor**: Estimated resident population for all age-by-gender groups and student population density contribute to this factor. The socio-economic variables contributing to this factor are:
    a. Estimated Resident Population Density Age: 0-16 (Male)
    b. Estimated Resident Population Density Age: 0-16 (Feale)
    c. Estimated Resident Population Density Age: 17-35 (Male)
    d. Estimated Resident Population Density Age: 17-35 (Female)
    e. Estimated Resident Population Density Age: 36-64 (Male)
    f. Estimated Resident Population Density Age: 36-64 (Female)
    g. Estimated Resident Population Density Age: 65 years and over (Male)
    h. Estimated Resident Population Density Age: 65 years and over (Female)
    i. Student Population Density
2. **Weekly Income below $2000 Earner Dominant Income Factor**: The main contributor to this factor is the earnings of different weekly income groups:
    a. Number of Residents (weekly income below $250)
    b. Number of Residents (weekly income $250-$999)
    c. Number of Residents (weekly income $1000-$1999)
    d. Number of Residents (weekly income equal of above $2000)
3. **Affluence Factor:** socio-economic variables captured by this factor are:
    a. Number of Residents (weekly income above 2000)
    b. Average Rent
    c. Average Car Ownership per Household

Based on Table 21, the equations for all land use characteristics and socio-economic factors can be formulated as:

$$\begin{aligned} &Students\ and\ Mid-Aged\ Dominant\ Resident\ Population\ Density\ Factor \\ &\quad = .941SE1_i + .934SE2_i + .9SE3_i + .875SE4_i + .871SE5_i + .824SE6_i \\ &\quad + .821SE7_i + .819SE8_i + .503SE9_i \end{aligned}$$

$$\begin{aligned} &Weekly\ Income\ Below\ \$2000\ Earner\ Dominant\ Income\ Factor \\ &\quad = .954I1_i + .928I2_i + .923I3_i + .557I4_i \end{aligned}$$

$$Affluence\ Factor = .423A1_i + .879A2_i + .712A3_i$$

where:

SE1	=	Estimated Resident Population Density Age: 36-64 (Male)
SE2	=	Estimated Resident Population Density Age: 36-64 (Female)
SE3	=	Estimated Resident Population Density Age: 17-35 (Female)
SE4	=	Estimated Resident Population Density Age: 17-35 (Male)
SE5	=	Estimated Resident Population Density Age: 65 and over (Male)
SE6	=	Estimated Resident Population Density Age: 65 and over (Female)
SE7	=	Estimated Resident Population Density Age: 0-16 (Female)
SE8	=	Estimated Resident Population Density Age: 0-16 (Male)
SE9	=	Student Population Density
I1	=	No of Residents Whose Weekly Income Between 250 and 1000
I2	=	No of Residents Whose Weekly Income Below 250
I3	=	No of Residents Whose Weekly Income Between 1000 and 2000
I4/A1	=	No of Residents Whose Weekly Income Between Above 2000
A2	=	Average Rent
A3	=	Average Car Ownership per Household

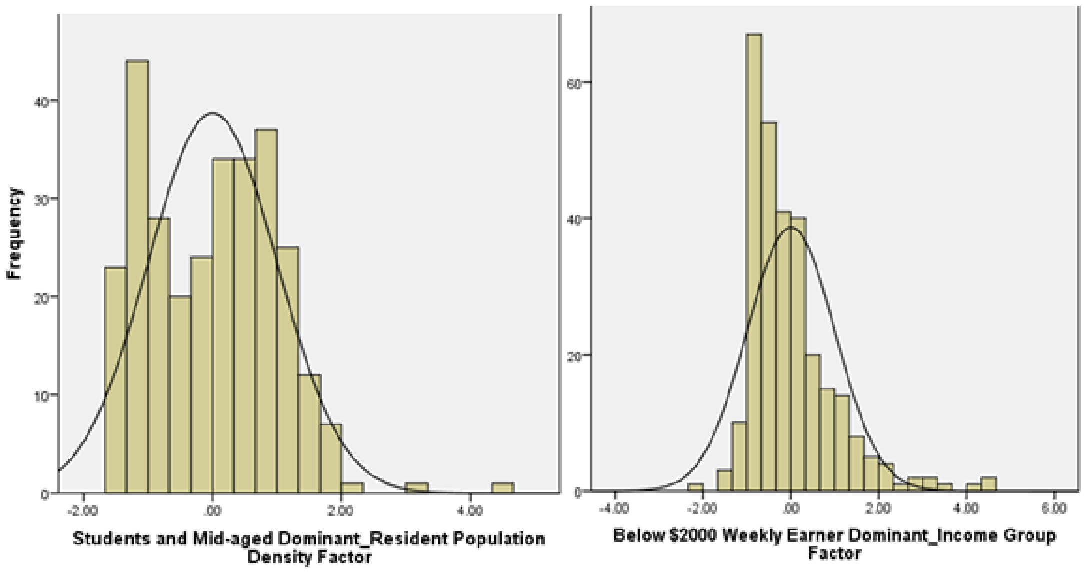
5.2.2.7 Descriptive Analysis: Land Use Characteristics and Socio-Economic Factors

	Students and Mid-aged Dominant Resident Population Density Factor	Below $2000 Weekly Earner Dominant Income Group Factor	Affluence Factor
Mean	.00	.00	.00
Std. Deviation	1.00	1.00	1.00
Skewness	.402	1.681	-.079
Kurtosis	.238	3.933	3.516
Minimum	-1.61993	-2.03011	-3.87535
Maximum	4.44343	4.49966	3.84650

Table 23: Descriptive Statistics on Land Use Characteristics and Socio-Economic Factors

As shown in Table 23, the mean values of all socio-economic factors are standardized to 0. The Histograms in Figure 48 show how they are distributed in terms of skewness and kurtosis. Figure 48 shows how the distributions of land use characteristics and socio-economic factors are derived from a normal distribution, as all of their skewness values are between -3 and 3. Further, the urban form and socio-economic factors do not have leptokurtic distribution, as their kurtosis values are not significantly high.

Even though the student and mid-aged dominant resident population density factor has a bimodal distribution, its skewness 0.402 (which is between -3 and 3 and also closer to 0) and low kurtosis 0.238, indicating that it is normally distributed. The below $2000 weekly earner dominant income factor has positive skewness, as most of the scores are clustered at the lower end of the scale. Moreover, this factor also has positive kurtosis (leptokurtic), visible from its pointy distribution. On the other hand, the affluence factor is normally and symmetrically distributed as compared to other factors, and its kurtosis value is 3.516, which is not too steep.



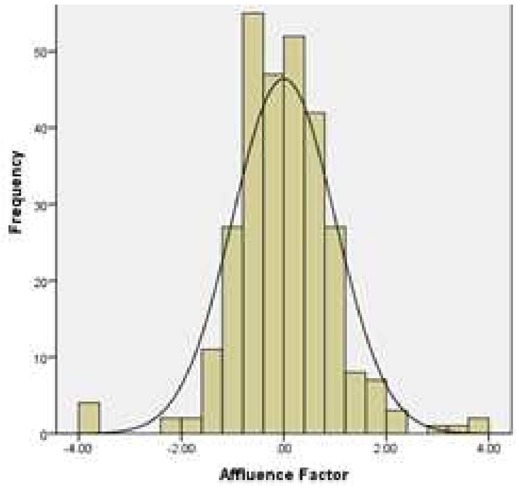


Figure 48: Histograms of Land Use Characteristics and Socio-Economic Factors

The analysis shows that all assumptions and conditions are met for all public transport service provision factors and land use characteristics and socio-economic factors to be used in multiple regression to construct the predictive model of public transport usage in the Perth metropolitan areas.

5.3 Multiple Regression

Nolan (2007, : p.156) define multiple regression as *a "statistical technique that includes two or more predictor variables in a prediction equation."* They mention that multiple-regression is more widely used than a simple linear equation because most dependent variables are best explained by more than one independent variable. In this research, service provision factors, land use characteristics, socio-economic, and urban form variables are used to predict the dependent variable (total public transport usage). Before running the regression, several tests were conducted to verify that all five of the required assumptions are satisfied, with data transformations performed as necessary. These five assumptions for multiple regressions are assumption of normality, assumption of homoscedasticity, assumption of linearity, assumption of non-multicollinearity and assumption of independence. In this section, the results from several tests are explained in details to confirm how all of these assumptions are satisfied.

5.3.1 Initial Checks for Normal Distribution

Descriptive Statistics

	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Public Transport Usage Per Km2	46282.321	177171.526	13.268	0.143	200.843	0.284
Train Service Provision Density Factor	0.000	1.000	11.218	0.143	152.944	0.284
Bus Service Provision Density Factor	0.000	1.000	4.761	0.143	37.235	0.284
Average Public Transport Stops Per Km2	11.244	8.106	1.710	0.143	8.486	0.284
Employment Density (Manufacturing/ Processing/ Fabrication Industry)	24.722	82.943	6.593	0.143	55.135	0.284
Employment Density (Storage/ Distribution/ Service Industry)	33.130	93.252	4.961	0.143	29.135	0.284
Employment Density (Shop/ Retail/ Other Retail/ Entertainment/ Recreation/ Culture Industry)	121.180	294.352	7.716	0.143	80.489	0.284
Employment Density (Office/ Business/ Residential/ Utilities/ Communications Industry)	227.982	928.354	8.437	0.143	83.774	0.284
Employment Density (Health/ Welfare/ Community Services Industry)	42.238	145.586	9.191	0.143	99.515	0.284
Students and Mid-aged Dominant_Resident Population Density Factor	0.000	1.000	0.402	0.143	0.238	0.285
University Student Population Density	101.8173	970.02039	12.611	.143	175.953	.284
Below $2000 Weekly Earner Dominant_Income Group Factor	.000	1.000	1.681	.143	3.933	.285
Affluence Factor	0.000	1.000	-0.079	0.143	3.516	0.285
Road Length per km2	9730.826	4539.136	-0.222	0.143	-0.524	0.284
Distance from City Center	18.908	13.873	1.583	0.143	2.966	0.284
Valid N (listwise)						

Table 24: Descriptive Statistics on Dependent and All Independent Variables

The first test is for the normality assumption. Table 24 shows that the distributions for public transport usage per km^2, employment densities (in all industries), and university student population are positively skewed. This indicates that there are some outliers in the data sets, making it necessary to transform the data. According to Figure 16 and Figure 17, the variances of public transport usage in Perth, Fremantle, Bentley, Murdoch and Joondalup differ significantly from the remaining suburbs. This is due to the nature of public transport usage, unequal distributions of employment densities, and the presence of universities in some suburbs. Therefore, these outliers need not be excluded from the multiple regression analysis. As mentioned, they will be transformed instead. In order to choose the right transformation method, curve estimates of the relationships between the dependent variable (public transport usage) and independent variables with high positive skewness are conducted to determine whether these are linear or curvilinear.

5.3.2 Curve Estimates between Dependent Variables and Predictors

Curve Estimate												
Dependent Variable: Public Transport Usage per Km2												
	Linear		Logarithmic		Quadratic		Cubic		Power		Exponential	
Independent Variable	R Square	Sig.	R Square	Sig.	R Square	Sig.	R Square	Sig.	R Square	Sig.	R Square	Sig.
Train Service Provision Density Factor+1.	0.317	0.000	0.078	0.000	0.700	0.000	0.836	0.000	0.117	0.000	0.114	0.000
Bus/Ferry Service Provision Density Factor	0.316	0.000	0.054	0.000	0.680	0.000	0.801	0.000	0.523	0.000	0.264	0.000
Average Public Transport Stops Per Km2.	.168	0.000	.054	0.000	.191	0.000	.304	0.000	.743	0.000	.530	0.000
Employment Density (Manufacturing/ Processing/ Fabrication Industry)+1.	.019	.019	.038	.001	.054	0.000	.063	0.000	.054	0.000	.008	.124
Employment Density (Storage/ Distribution/ Service Industry)+1.	.032	.002	.045	0.000	.082	0.000	.083	0.000	.141	0.000	.028	.004
Employment Density (Shop/ Retail/ Other Retail/ Entertainment/ Recreation/ Culture Industry)+1.	.283	0.000	.083	0.000	.302	0.000	.712	0.000	.517	0.000	.144	0.000
Employment Density (Office/ Business/ Residential/ Utilities/ Communications Industry)+1.	.609	0.000	.114	0.000	.790	0.000	.865	0.000	.420	0.000	.096	0.000
Employment Density (Health/ Welfare/ Community Services Industry)+1.	.561	0.000	.091	0.000	.721	0.000	.851	0.000	.441	0.000	.102	0.000
Students and Mid-aged Dominant_Resident Population Density Factor+2.	.047	0.000	.039	.001	.051	.001	.051	.002	.591	0.000	.501	0.000
University Student Population Density+1.	.023	0.010	.020	.015	.023	.034	.024	.069	.043	0.000	.026	0.000
Below $2000 Weekly Earner Dominant_Income Group Factor+3.	.000	0.812	.001	.694	.000	.967	.004	.747	.069	0.000	.067	0.000
Affluence Factor+4.	.006	.191	.001	.662	.007	.350	.012	.343	.001	.543	.035	.001
Road Length in km per km2.	.062	0.000	.034	.002	.096	0.000	.105	0.000	.654	0.000	.600	0.000
Distance from City Center+1.	.043	0.000	.162		.095	0.000	.151	0.000	.394	0.000	.276	0.000

Table 25: Summary of Curve Estimates between Dependent Variable and Predictors

Table 25 illustrates that the most significant relationships between public transport usage and its predictors are curvilinear (i.e. cubic, power and exponential curves). To conduct the curve estimates for these relationships, the minimum values of some variables are deliberately incremented to be greater than 0. Otherwise, their cubic, power and exponential relationships could not be identified.

The predictors that have most significant relationship with public transport usage per km^2 in cubic form are as follows:

1. Train Service Provision Density Factor
2. Bus/Ferry Service Provision Density Factor
3. Employment Density (Manufacturing/ Processing/ Fabrication Industry)
4. Employment Density (Shop/ Retail/ Other Retail/ Entertainment/ Recreation/ Culture Industry)
5. Employment Density (Office/ Business/ Residential/ Utilities/ Communications Industry) and
6. Employment Density (Health/ Welfare/ Community Service Industry).

The predictors that have a significant power-curve relationship with public transport usage per km^2 are:

1. Average Public Transport Stops per km^2
2. Employment Density (Storage/ Distribution/ Service Industry)
3. Student and Mid-aged Dominant Resident Population Density Factor
4. University Student Population Density
5. Below $2000 Weekly Earner Dominant Income Factor
6. Road Length in km per km2 and
7. Distance from City Center.

Only the affluence factor has an exponential relationship with the dependent variable.

These curvilinear relationships between public transport usage and its predictors confirm that it is necessary to transform the data so that the normality and linearity assumptions for multiple regression can be satisfied.

5.3.3 Data Transformation

Field (2013) suggests two ways to transform variables with positive skewness, positive kurtosis and lack of linearity: (1) log and (2) square root transformations.

	Mean	Std. Deviation	Skewness		Kurtosis	
			Statistic	Std. Error	Statistic	Std. Error
Public Transport Usage per Km2	46282.32	177171.53	13.27	0.14	200.84	0.28
Log (Public Transport Usage per Km2)	9.12	2.19	-0.95	0.14	1.19	0.28
√ (Public Transport Usage per Km2)	150.68	153.82	4.30	0.14	34.23	0.28
Train Service Provision Density Factor	94.38	464.22	11.06	0.14	149.20	0.28
Log (Train Service Provision Density Factor)	1.00	2.19	1.85	0.14	1.78	0.28
√(Train Service Provision Density Factor)	3.41	9.11	4.05	0.14	23.18	0.28
Bus/Ferry Service Provision Density Factor	2099.37	2903.89	5.49	0.14	47.40	0.28
Log (Bus/Ferry Service Provision Density Factor)	6.79	1.73	-1.38	0.14	1.96	0.28
√(Bus/Ferry Service Provision Density Factor)	38.89	24.27	1.23	0.14	4.33	0.28
Average Public Transport Stops per Km2	11.24	8.11	1.71	0.14	8.49	0.28
Log (Average Public Transport Stops per Km2)	2.03	1.11	-1.48	0.14	2.10	0.28
√ (Average Public Transport Stops per Km2)	3.11	1.27	-0.08	0.14	0.33	0.28
Employment Density (Manufacturing/ Processing/ Fabrication Industry)	24.72	82.94	6.59	0.14	55.13	0.28
Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry))	1.32	1.71	1.17	0.14	0.25	0.28
√(Employment Density (Manufacturing/ Processing/ Fabrication Industry))	8.87	12.24	4.12	0.14	22.91	0.28
Employment Density (Storage/ Distribution/ Service Industry)	33.13	93.25	4.96	0.14	29.13	0.28
Log (Employment Density (Storage/ Distribution/ Service Industry))	1.65	1.80	0.88	0.14	-0.37	0.28
√(Employment Density (Storage/ Distribution/ Service Industry))	7.87	7.71	2.28	0.14	9.32	0.28
Employment Density (Shop/ Retail/ Other Retail/ Entertainment/ Recreation/ Culture Industry)	121.18	294.35	7.72	0.14	80.49	0.28
Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry))	3.36	1.92	-0.22	0.14	-0.78	0.28
√(Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry))	3.22	4.78	2.39	0.14	6.62	0.28
Employment Density (Office/ Business/ Residential/ Utilities/ Communications Industry)	227.98	928.35	8.44	0.14	83.77	0.28
Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))	3.38	2.00	0.13	0.14	-0.22	0.28
√ (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))	2.51	4.30	2.86	0.14	10.39	0.28
Employment Density (Health/ Welfare/ Community Services Industry)	42.24	145.59	9.19	0.14	99.52	0.28
Log (Employment Density (Health/ Welfare/ Community Services Industry))	2.42	1.57	0.20	0.14	-0.21	0.28
√ (Employment Density (Health/ Welfare/ Community Services Industry))	4.35	4.84	3.64	0.14	20.98	0.28
University Student Population Density	101.82	970.02	12.61	0.14	175.95	0.28
Log (University Students Population Density)	0.18	1.17	6.76	0.14	45.14	0.28
√ (University Students Population Density)	1.32	10.02	8.95	0.14	88.39	0.28

Table 26 Data Transformation Results

Data transformation is conducted by applying these two methods and then selecting the one that best satisfies the normality assumption. Field (2013) also explains that the perfect normal distribution has skewness and kurtosis values of 0. The data are more normally distributed (though not perfectly so) when these values are (positively or negatively) closer to 0. Therefore, the method that can transform the data with skewness and kurtosis values closer to 0 is chosen for public transport usage and the other independent variables with high skewness and kurtosis values.

As Table 26 shows, the log transformation method normalizes the unequally distributed independent variables, bringing their skewness and kurtosis values closer to 0. Therefore, it can be concluded that this model satisfies the assumption of normality for predictors. The assumption of normality for residual values is discussed in the multiple regression model section.

5.3.4 Initial Checks for Multicollinearity

Field (2013) suggests that the correlation matrix is extremely useful for an initial investigation of the relationships between predictors and the outcome, as well as a preliminary verification for multicollinearity. This section begins with a discussion of these correlations.

		Log (Public Transport Usage Per Km2)	Log (Train Service Provision Density Factor)	Log (Bus Service Provision Density Factor)	Log (Average Public Transport Stops Per Km2)	Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))	Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry))	Log (Employment Density (Storage/ Distribution / Service Industry))	Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry))	Log (Employment Density (Health/ Welfare/ Community Services Industry))
Pearson Correlation	Log (Public Transport Usage Per Km2)	1.000	.365	.885	.860	.644	.715	.372	.229	.661
	Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))	.644	.292	.607	.570	1.000	.773	.742	.659	.650
	Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry))	.715	.278	.688	.646	.773	1.000	.610	.515	.692
	Log (Employment Density (Storage/ Distribution/ Service Industry))	.372	.289	.347	.287	.742	.610	1.000	.873	.474
	Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry))	.229	.239	.238	.172	.659	.515	.873	1.000	.362
	Log (Employment Density (Health/ Welfare/ Community Services Industry))	.661	.255	.599	.598	.650	.692	.474	.362	1.000
	Students and Mid-aged Dominant_Resident Population Density Factor	.708	.148	.705	.735	.414	.533	.077	-.025	.497
	Log (University Students Population Density)	.208	.145	.138	.079	.187	.151	.150	.131	.241
	Below $2000 Weekly Earner Dominant_Income Group Factor	.260	.035	.225	.196	.184	.312	.136	.103	.224
	Affluence Factor	-.187	-.118	-.197	-.180	-.280	-.182	-.279	-.297	-.103
	Log (Train Service Provision Density Factor)	.365	1.000	.230	.249	.292	.278	.289	.239	.255
	Log (Bus Service Provision Density Factor)	.885	.230	1.000	.808	.607	.688	.347	.238	.599
	Log (Average Public Transport Stops Per Km2)	.860	.249	.808	1.000	.570	.646	.287	.172	.598
	Distance from City Center	-.519	-.193	-.468	-.501	-.441	-.399	-.280	-.135	-.376
	Log (Road Length (in m) per Km2	.806	.171	.847	.858	.565	.659	.294	.186	.578

		Students and Mid-aged Dominant _Resident Population Density Factor	Log (University Students Population Density)	Below $2000 Weekly Earner Dominant _Income Group Factor	Affluence Factor	Distance from City Center	Log (Road Length (in m) per Km2
Pearson Correlation	Log (Public Transport Usage Per Km2)	.708	.208	.260	-.187	-.519	.806
	Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))	.414	.187	.184	-.280	-.441	.565
	Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry))	.533	.151	.312	-.182	-.399	.659
	Log (Employment Density (Storage/ Distribution/ Service Industry))	.077	.150	.136	-.279	-.280	.294
	Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry))	-.025	.131	.103	-.297	-.135	.186
	Log (Employment Density (Health/ Welfare/ Community Services Industry))	.497	.241	.224	-.103	-.376	.578
	Students and Mid-aged Dominant_Resident Population Density Factor	1.000	.041	.000	.000	-.498	.756
	Log (University Students Population Density)	.041	1.000	.036	-.086	-.085	.065
	Below $2000 Weekly Earner Dominant_Income Group Factor	.000	.036	1.000	.000	-.038	.231
	Affluence Factor	.000	-.086	.000	1.000	-.057	-.133
	Log (Train Service Provision Density Factor)	.148	.145	.035	-.118	-.193	.171
	Log (Bus Service Provision Density Factor)	.705	.138	.225	-.197	-.468	.847
	Log (Average Public Transport Stops Per Km2)	.735	.079	.196	-.180	-.501	.858
	Distance from City Center	-.498	-.085	-.038	-.057	1.000	-.441
	Log (Road Length (in m) per Km2	.756	.065	.231	-.133	-.441	1.000

Table 27: Correlation Matrix

Table 27 shows that the dependent variable (public transport usage) has a high individual correlation with some predictors such as log (bus/ferry service provision density factor), log (average public transport stops per km^2), log (road length in km per km^2), log (employment density-shops, other retail, entertainment, recreation and culture industry), student and mid-aged dominant resident population density factor, log (employment density-manufacturing, processing, fabrication industry) and log (employment density- office, business, utilities, communication and residential industry). The predictor that has a negative correlation with public transport usage density is distance from city center; this suggests that the usage density gets lower as the suburbs get farther away from the city. Table 27 shows that there are no correlations among the predictors with a value greater than 0.9. This confirms that the multiple regression model generated from these variables does not violate the assumption of multicollinearity.

5.3.4.1 Multiple Regression with the Standard Method

Multiple regression with the standard (Enter) method was applied to predict the log (public transport usage density) with fourteen predictors.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.930[a]	.866	.859	.81686	.866	127.209	14	276	.000	1.899

a. Predictors: (Constant), Distance from City Center, Log (Affluence Factor), Log (Below $2000 Weekly Earner Dominant_ Income Group Factor), Log (University Students Population Density), Log (Train Service Provision Density Factor), Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry)), Log (Road Length (in m) per Km2, Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry)), Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry)), Log (Bus Service Provision Density Factor), Log (Employment Density (Storage/ Distribution/ Service Industry)), Log (Students and Mid-aged Dominant_ Resident Population Density Factor), Log (Average Public Transport Stops per Km2)

b. Dependent Variable: Log (Public Transport Usage per Km2)

Table 28: Multiple Regression Model Summary (Standard Method)

Table 28 summarizes a predictive model in which fourteen predictors are included. The model is statistically significant (F=128.209, p<0.001) with an R square value indicating it accounts for 86.6% of the variation in public transport usage in 292 out of 309 suburbs in Perth metropolitan area. Further, the difference between adjusted R square and R square is 0.007 (0.866-0.859). Since this model is derived from the population rather than a sample, it would account for approximately 0.7% less variance in the outcome measure.

Field (2013) states that the residual terms should be independent and uncorrelated for any two observations. If this assumption of independence is violated, the confidence intervals and significance tests would be invalid. The Dublin-Watson test can be used to test this assumption by calculating the serial correlations between errors, especially whether adjacent residuals are correlated. The test result can vary from 0 to 4, and any test value less than 1 or greater than 3 indicates a violation of the assumption of independent errors. Field (2013) recommends using a value of 2 as a criterion. As shown in the above model summary (Table 28), the Dublin-Watson test value is 1.899, which implies that the observations in this model are independent from each other and the residuals are uncorrelated. It thus fulfils the assumptions of independence and suggests that the confidence intervals and significance tests are valid.

Therefore, it can be concluded that the multiple regression model derived in this study satisfies the assumption of independent errors.

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	1188.330	14	84.881	127.209	.000[b]
	Residual	184.162	276	.667		
	Total	1372.492	290			

a. Dependent Variable: Log (Public Transport Usage per Km2)

b. Predictors: (Constant), Distance from City Center, Log (Affluence Factor), Log (Below $2000 Weekly Earner Dominant Income Group Factor), Log (University Students Population Density), Log (Train Service Provision Density Factor), Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry)), Log (Road Length (in m) per Km2, Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry)), Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry)), Log (Bus Service Provision Density Factor), Log (Employment Density (Storage/ Distribution/ Service Industry)), Log (Students and Mid-aged Dominant Resident Population Density Factor), Log (Average Public Transport Stops per Km2)

Table 29: Analysis of Variance (ANOVA)

Gray (2012) suggests that Analysis of Variance (ANOVA) should be used to test for a linear relationship between the variables to determine whether the derived model can predict the outcome significantly better than the mean. The ratio of the mean square for the regression to the residual mean square is indicated by the F statistic in the ANOVA table. According to Field (2013), the sum of squares and degrees of freedom (df) for the model indicates whether the fitted regression line has more predictive power than the mean.

In Table 28, the sum of squared differences between the observed values and the value predicted by the mean is 1372.492. The total sum of squared differences between the observed values and the value predicted by the regression model is 1189.820. The residual sum of squares (184.162, df =276) represents the total difference between the regression model; its value here implies that the regression model is a better predictor of the observed values than the mean.

Field (2013) suggests that the improvement due to fitting the regression model is much greater than the inaccuracy within the model if the value of F is greater than 1. In Table 29: Analysis of Variance (ANOVA), the F ratio is 127.209 with a significance value of less than 0.001. This indicates that the derived model significantly improves the ability to predict the outcome variable, as compared to not fitting the model.

5.3.4.2 Public Transport Usage Density Predicting Model

Field (2013) explains that the standardised coefficient beta *ß* indicates the number of standard deviations that the outcome will change as a result of one standard deviation change in the predictor. Therefore, it indicates the importance of a predictor in the model. The *t* test values indicate whether the predictors make a significant contribution to the model. Field (2013) suggests that the smaller the value of sig (with larger value of t), the greater contribution of that predictor. The standard deviation of log (public transport usage density) in this model is 2.668. The *ß* values of log (train service provision density factor) 0.14 and log (bus service provision density factor) 0.54 are considerably higher than the other predictors. Their smaller significant values (0.000) and larger t values (>4) also indicate that they contribute more to predicting public transport usage than the other predictors (see in Table 30).

	Unstandardized Coefficients		Standardized Coefficients			Confidence Interval for B		Correlations			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
(Constant)	2.668	1.310		2.037	.043	.090	5.246					
Log (Train Service Provision Density Factor)	.140	.024	.141	5.916	.000	.093	.186	.365	.335	.130	.854	1.170
Log (Bus Service Provision Density Factor)	.540	.075	.423	7.233	.000	.393	.687	.885	.399	.159	.142	7.029
Log (Average Public Transport Stops Per Km2)	.298	.120	.151	2.475	.014	.061	.535	.860	.147	.055	.130	7.686
Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))	.104	.050	.095	2.082	.038	.006	.203	.644	.124	.046	.232	4.312
Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry))	.074	.049	.065	1.493	.137	-.024	.171	.715	.090	.033	.257	3.892
Log (Employment Density (Storage/ Distribution/ Service Industry))	.063	.064	.052	.984	.326	-.063	.188	.372	.059	.022	.174	5.732
Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry))	-.157	.062	-.123	-2.530	.012	-.279	-.035	.229	-.151	-.056	.205	4.871
Log (Employment Density (Health/ Welfare/ Community Services Industry))	.087	.047	.063	1.844	.066	-.006	.180	.661	.110	.041	.418	2.394
Log (Students and Mid-aged Dominant_Resident Population Density Factor)	.423	.201	.113	2.105	.036	.027	.818	.769	.126	.046	.168	5.965
Log (University Students Population Density)	.123	.043	.066	2.858	.005	.038	.208	.208	.170	.063	.901	1.109
Log (Below $2000 Weekly Earner Dominant_Income Group Factor)	.485	.182	.066	2.663	.008	.127	.844	.263	.158	.059	.786	1.272
Log (Affluence Factor)	-.047	.116	-.010	-.407	.684	-.275	.181	-.036	-.025	-.009	.838	1.193
Log (Road Length (in m) per Km2	.087	.165	.029	.529	.597	-.238	.413	.806	.032	.012	.164	6.083
Distance from City Center	-.008	.004	-.053	-1.933	.054	-.017	.000	-.519	-.116	-.043	.640	1.564

Table 30: Coefficients of Multiple Regression Model

Based on these *b* (unstandardised coefficient) values, the predicting model can be defined as

$$\begin{aligned} Log(Public\ Transport\ Usage) \\ &= 2.668 + .14T_i + .54B_i + .298PTS_i + .104E1_i + .074E2_i + .063E3_i - .157E4_i \\ &+ .087E5_i + .423SR_i + .123U_i + .485I_i - .047A_i + .087R_i - .008D_i \end{aligned}$$

where:

T	=	Log (Train Service Provision Density Factor)
B	=	Log (Bus Service Provision Density Factor)
PTS	=	Log (Average Public Transport Stops per Km2)
E1	=	Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))
E2	=	Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry))
E3	=	Log (Employment Density (Storage/ Distribution/ Service Industry))
E4	=	Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry))
E5	=	Log (Employment Density (Health/ Welfare/ Community Services Industry))
SR	=	Students and Mid-aged Dominant Resident Population Density Factor
U	=	Log (University Students Population Density)
I	=	Below $2000 Weekly Earner Dominant Income Group Factor
A	=	Affluence Factor
R	=	Log (Road Length (in m) per Km2
D	=	Distance from City Center

5.3.4.3 Interpretation on Regression Coefficient Estimates

Beta values for each predictor in the coefficient table can be interpreted as individual estimates of public transport usage by assuming that other independent variables remain constant. When other predictors are held constant,

1. A 1% increase in the train service provision density factor can lead to a 0.14 % increase in public transport usage density. The Train Service Provision Density Factor equation (as discussed in 5.2.1.6 Public Transport Service Provision Density Factor Equations) can be used to calculate how the changes in train service provisions densities during different time segments produce changes in standardized effects of this factor.
2. A 1% increase in bus/ferry service provision density factor can result in a 0.54% increase in public transport usage density. The Bus/Ferry Service Provision Density Factor equation (as discussed in 5.2.1.6 Public Transport Service Provision Density

Factor Equations) can also be used to calculate how these services would change during time segments to increase the standardized effects of this factor.

	Perth		2009		2015	
		Factor Loading	Service Provision Density	(Service Provision Density x Factor Loading)	Service Provision Density	(Service Provision Density x Factor Loading)
Bus/Ferry Service Provision Density	Weekday 0-6hr	.697	9.07	6.321127173	9.07	6.321127173
	Weekday 6-9hr	.881	742.44	653.7569165	800.00	704.4433405
	Weekday 9-12hr	.852	640.51	545.9180783	700.00	596.6230479
	Weekday 12-15hr	.859	599.48	514.8208739	599.48	514.8208739
	Weekday 15-18hr	.869	795.35	690.9028867	795.35	690.9028867
	Weekday 18-21hr	.932	274.04	255.5392312	274.04	255.5392312
	Weekday 21-24hr	.899	116.18	104.4839178	116.18	104.4839178
	Saturday 0-9hr	.864	224.59	193.9510263	224.59	193.9510263
	Saturday 9-12hr	.856	384.39	329.0060608	500.00	427.9564311
	Saturday 12-15hr	.873	380.72	332.4312255	380.72	332.4312255
	Saturday 15-18hr	.882	370.14	326.5724603	370.14	326.5724603
	Saturday 18-21hr	.859	252.45	216.9070668	252.45	216.9070668
	Saturday 21-24hr	.789	183.77	144.9951622	183.77	144.9951622
	Sunday 3-9hr_21-24hr	.772	50.96	39.35683092	50.96	39.35683092
	Sunday 9-12hr	.936	226.96	212.3512321	226.96	212.3512321
	Sunday 12-15hr	.924	234.09	216.3495341	234.09	216.3495341
	Sunday 15-18hr	.917	249.42	228.6990366	249.42	228.6990366
	Sunday 18-21hr	.946	145.33	137.4655667	145.33	137.4655667
	Bus/Ferry Service Provision Density Factor Score			5149.828234		5350.169998
	% changes in Bus/Ferry Service Provision Density Factor					4%
Train Service Provision Density	Weekday 0-6hr	.950	4.32	4.102420623	4.32	4.102420623
	Weekday 6-9hr	.935	47.08	43.99858409	50.00	46.73022458
	Weekday 9-12hr	.955	23.97	22.89412986	30.00	28.65286638
	Weekday 12-15hr	.954	26.13	24.93552814	26.13	24.93552814
	Weekday 15-18hr	.920	61.55	56.60570436	61.55	56.60570436
	Weekday 18-21hr	.949	25.91	24.59416847	25.91	24.59416847
	Weekday 21-24hr	.953	10.37	9.881687036	10.37	9.881687036
	Saturday 0-9hr	.955	23.11	22.068523	23.11	22.068523
	Saturday 9-12hr	.956	20.73	19.81437781	20.73	19.81437781
	Saturday 12-15hr	.956	20.73	19.81437781	20.73	19.81437781
	Saturday 15-18hr	.956	20.73	19.81494939	20.73	19.81494939
	Saturday 18-21hr	.954	16.20	15.45829649	16.20	15.45829649
	Saturday 21-24hr	.953	10.37	9.879975727	10.37	9.879975727
	Sunday 3-9hr_21-24hr	.952	20.52	19.52486724	20.52	19.52486724
	Sunday 9-12hr	.955	20.73	19.79426573	20.73	19.79426573
	Sunday 12-15hr	.955	20.73	19.7943437	20.73	19.7943437
	Sunday 15-18hr	.955	20.73	19.7943437	20.73	19.7943437
	Sunday 18-21hr	.953	16.20	15.43640744	16.20	15.43640744
	Train Service Provision Density Factor Score			388.2069506		396.6973276
	% changes in Train Service Provision Density Factor					2.19%
	Public Transport Usage Density					2792123.97
	Coefficient between Public Transport Usage Density and Train Service Provision Density Factor					0.14
	Coefficient between Public Transport Usage Density and Bus/Ferry Service Provision Density Factor					0.54
	Expected public transport usage density in 2015 based on changes in train service provision density factor					2800673.186
	Expected % increase in public transport usage density (2015) based on changes in train service provision density factor					0.31%
	Expected public transport usage density in 2015 based on changes in bus/ferry service provision density factor					2850779.26
	Expected % increase in public transport usage density (2015) based on changes in bus/ferry service provision density factor					2.10%

Table 31: Example 1- Predicting Changes in Public Transport Usage Based on Changes in Public Transport Service Provisions

The Table 31 is an example of how service provision factor equations are used to calculate factor scores, and how these factor scores and coefficient values from the predictive model can be used to predict public transport usage density based on changes in service provision density during a particular time segment. In this case the model is applied for Perth suburb.

In the example, the bus/ferry service provision densities increase from 742 (in 2009) to 800 (in 2015) for the weekday 6-9 am time segment; they increase from 641 (in 2009) to 700 (in 2015) for the weekday 9-12noon segment. Together, these two changes can result in a 2.1% increase in public transport usage density in Perth. Additionally, two train service provision density increases from 47 (in 2009) to 50 (in 2015), and 24 (in 2009) to 30 (in 2015), can lead to a 0.29% increase in public transport usage density in Perth.

3. A 1% increase in average public transport stops per km^2 can increase public transport usage density by 0.298%.
4. An increase in employment density (office/ business/ utilities/ communication/ residential industry) by 1% can lead to a 0.104% increase in public transport usage density.
5. An increase in employment density (shop/ other retail/ entertainment/ recreation/ culture industry) of 1% can lead to a 0.074% increase in public transport usage density.
6. An increase in employment density (storage/ distribution/ service industry) by 1% can lead to a 0.063% increase in public transport usage density.
7. An increase in employment density (manufacturing/ processing/ fabrication industry) by 1% can lead to a 0.157% decrease in public transport usage density.
8. An increase in employment density (health/ welfare/ community services industry) by 1% can lead to a 0.087% increase in public transport usage density.

9. A 1% increase in the students and mid-aged dominant resident population density factor can lead to a 22.2% increase in public transport usage. The rotated component matrix for estimated resident population density indicates that the population densities of males/females aged 36-64 and males/females aged 17-35 contribute more to this factor score than the other age/ gender groups. Higher population densities of these groups contribute to increases in their factor score, which can then lead to increases in public transport usage. These estimated resident population densities (age/gender) groups are have a significant impact on public transport usage density. The sensitivity of public transport usage to changes in this factor can be calculated using the equation discussed in section 5.2.2.6.

Perth		2009		2015	
	Factor Loading	Density	Density x Factor Loading	Density	Density x Factor Loading
Estimated Resident Population Density Age: 36-64 (Male)	.941	419.16	394.3964064	450.00	423.4144071
Estimated Resident Population Density Age: 36-64 (Female)	.934	259.36	242.1765766	259.36	242.1765766
Estimated Resident Population Density Age: 17-35 (Female)	.900	379.21	341.1475918	379.21	341.1475918
Estimated Resident Population Density Age: 17-35 (Male)	.875	508.56	444.7580485	508.56	444.7580485
Estimated Resident Population Density Age: 65 and over (Male)	.871	90.48	78.76903907	90.48	78.76903907
Estimated Resident Population Density Age: 65 and over (Female)	.824	58.95	48.58465363	58.95	48.58465363
Estimated Resident Population Density Age: 0-16 (Female)	.821	72.34	59.39784449	72.34	59.39784449
Estimated Resident Population Density Age: 0-16 (Male)	.819	65.65	53.74443157	65.65	53.74443157
Student Population Density	.503	500.00	251.4927053	500.00	251.4927053
Students and Mid-aged Dominant Resident Population Density Factor Score			1914.467297		1943.485298
% Increases Students and Mid-aged Dominant Resident Population Density Factor Score					1.52%
Public Transport Usage Density					2792123.97
Coefficient between Public Transport Usage Density and Student and Mid-aged Dominant Resident Population Density Factor					0.423
Expected public transport usage density in 2015 based on changes in Students and Mid-aged Dominant Resident Population Density Factor Score					2810025.681
Expected % increase in public transport usage density (2015) based on change in Students and Mid-aged Dominant Resident Population Density Factor Score					0.64%

Table 32: Example 2- Predicting Changes in Public Transport Usage Based on Changes in Estimated Resident Population Density by Age and Gender

Table 32 illustrates how a change, an increase in estimated resident population density by age and gender, leads to a change, an increase in public transport usage density in Perth suburb. In this example, there is only one change whereby the estimated resident population density by age 36-64 (Male) is increased from 419 (in 2009) to 450 (in 2015 as an example). This results in a 1.52% increase in the student and mid-aged dominant resident population density factor score that, in turn, predicts a public transport usage density increase of 0.64%.

10. A 1% increase in university student population density can lead to a 0.123% increase in public transport usage density.

11. A 1% increase in the below $2000 weekly earner dominant income factor can lead to a 0.485% increase in public transport usage. From section 5.2.2.6, it can be seen that the primary contributor to income factor scores is the number of residents whose weekly income is below $2000.

Perth		2009		2015	
	Factor Loading	No of Residents	No of Residents x Factor Loading	No of Residents	No of Residents x Factor Loading
No of Residents Whose Weekly Income Below 250	.954	840.00	801.36	924.00	881.496
No of Residents Whose Weekly Income Between 250 and 1000	.928	1928.00	1789.184	1928.00	1789.184
No of Residents Whose Weekly Income Between 1000 and 2000	.923	1013.00	934.999	1013.00	934.999
No of Residents Whose Weekly Income Between Above 2000	.557	415.00	231.155	415.00	231.155
Weekly Income Below $2000 Earner Dominant Income Factor			3756.698		3836.834
% Increases Weekly Income Below $2000 Earner Dominant Income Factor Score					2.13%
Public Transport Usage Density					2792123.97
Coefficient between Public Transport Usage Density and Weekly Income Below $2000 Earner Dominant Income Factor					0.485
Expected public transport usage density in 2015 based on changes in Weekly Income Below $2000 Earner Dominant Income					2821010.66
Expected % increase in public transport usage density (2015) based on change in Weekly Income Below $2000 Earner Dominant Income Factor					1.03%

Perth		2009		2015	
	Factor Loading	No of Residents	No of Residents x Factor Loading	No of Residents	No of Residents x Factor Loading
No of Residents Whose Weekly Income Below 250	.954	840.00	801.36	840.00	801.36
No of Residents Whose Weekly Income Between 250 and 1000	.928	1928.00	1789.184	1928.00	1789.184
No of Residents Whose Weekly Income Between 1000 and 2000	.923	1013.00	934.999	1013.00	934.999
No of Residents Whose Weekly Income Between Above 2000	.557	415.00	231.155	456.00	253.992
Weekly Income Below $2000 Earner Dominant Income Factor			3756.698		3779.535
% Increases Weekly Income Below $2000 Earner Dominant Income Factor Score					0.61%
Public Transport Usage Density					2792123.97
Coefficient between Public Transport Usage Density and Weekly Income Below $2000 Earner Dominant Income Factor					0.485
Expected public transport usage density in 2015 based on changes in Weekly Income Below $2000 Earner Dominant Income					2800356.04
Expected % increase in public transport usage density (2015) based on change in Weekly Income Below $2000 Earner Dominant Income Factor					0.29%

Table 33: Example 3- Predicting Changes in Public Transport Usage Based on Changes in Number of Residents with Weekly Income above $2000

Table 33 shows how changes in different income groups predict changes in public transport usage density. In the first example, the number of residents with weekly income below $250 increases by 10%, from 840 to 924. This results in a 2.13% increase in the weekly income below $2000 earner dominant income factor score. Public transport usage density could then be predicted to increase by 1.03%. In a second example, the number of residents with weekly income is above 2000 is increased by 10%, from 415 to 456. This results in only a 0.61% increase in the weekly income below $2000 earner dominant income factor score. Public transport usage density would then be expected to increase by 0.29%. These examples indicate that the same percentage change in different income groups can contribute to different percentage changes in the total factor score, leading to different predictions about changes in public transport usage density.

12. A 1% increase in the affluence factor can lead to -0.047% decrease in public transport usage density.

Perth		**2009**		**2015**	
	Factor Loading		x Factor Loading		x Factor Loading
No of Residents Whose Weekly Income Between Above 2000	.423	415.00	175.545	415.00	175.545
Average Rent	.879	471.52	414.46608	471.52	414.46608
Average Car Ownership Per Household	.712	1.01	0.71912	2.00	1.424
Affluence Factor Score			590.7302		591.43508
% Increases in Affluence Factor					0.12%
Public Transport Usage Density					2792123.97
Coefficient between Public Transport Usage Density and Student and Affluence Factor Score					-0.047
Expected public transport usage density in 2015 based on changes in Students and Affluence Factor Score					2791967.38
Expected % increase in public transport usage density (2015) based on change in Affluence Factor Score					-0.01%

Table 34: Example 4- Predicting Changes in Public Transport Usage Based on Changes in Average Car Ownership per Household

As shown in Table 34, the average car ownership per household in Perth (2009) is 1.01. If it increases to 2 cars per household in 2015, the affluence factor goes up by 0.12%. Here, public transport usage density can be predicted to fall by 0.01%.

13. An increase in Distance from City Center by one standard deviation (1 unit) can lead to a -0.8 % decrease in public transport usage. One standard deviation of distance from the city center is 18.9km. As a suburb gets farther from the city center by 18.9km increments, its predicted public transport usage density decreases by 0.8% per increment.

14. A 1% increase in Road Length (in km) per km2 can lead to a 0.08% increase in public transport usage.

5.3.4.4 Multiple Regression Model Validity

In this section, Case-wise diagnostics and residual statistics will be used to check the residuals for evidence of bias.

Case-wise Diagnostics

Case Number	Suburb Name	Std. Residual	Log (Public Transport Usage per Km2)	Predicted Value	Residual
191	MARTIN	3.018	7.17	4.7145	2.45522

a. Dependent Variable: Log (Public Transport Usage per Km2)

Table 35: Case-wise Diagnostics on Factor Regression Model

Table 35 lists all extreme cases that have standardised residuals less than -2 or greater than 2. Field (2013) recommends that 95% of the cases should have standardised residuals of about ±2. In the data analysed here, only 1 case out of 292 is extreme; therefore, only 0.3% of the total cases lie outside the acceptable limits. This confirms that the predictions of the derived factor regression model are fairly accurate.

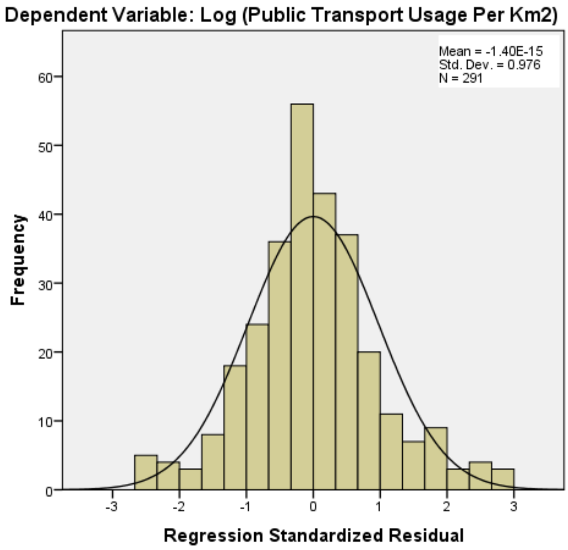
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2.9014	13.6551	9.1334	2.02554	291
Residual	-2.14758	2.41356	.00000	.79366	291
Std. Predicted Value	-3.077	2.232	.000	1.000	291
Std. Residual	-2.640	2.967	.000	.976	291

a. Dependent Variable: Log (Public Transport Usage per Km2)

Table 36: Residuals Statistics

Table 36 shows that the residuals' mean value is approximately in the middle between its minimum and maximum values, with a standard deviation of 0.79. These values indicate that the residual values are normally distributed. The standard deviations of standardised predicted values and standardised residuals are the same as 0. The results confirm that the public transport usage density model satisfies the assumption of normality. This can also be seen in Figure 49 below.



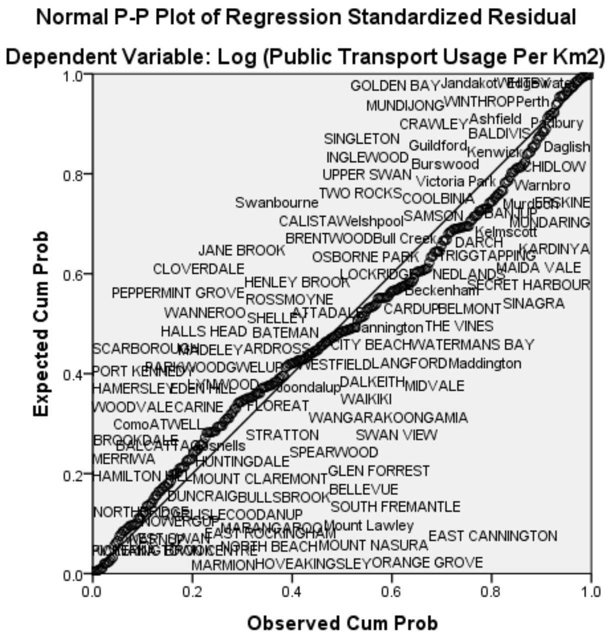


Figure 49: Histogram (Regression Standardised Residual)

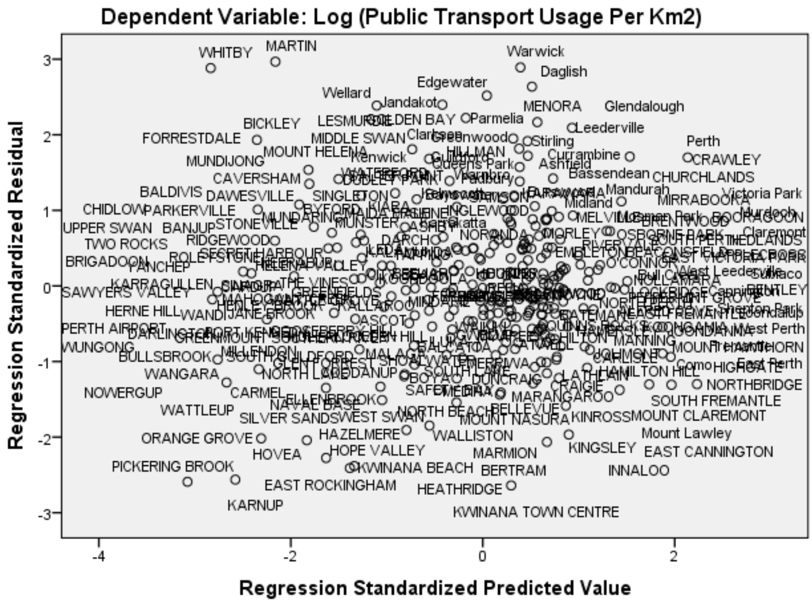


Figure 50: Scatter Plot (Regression Standardised Residual against Standardised Predicted Value)

Figure 50 shows that the spread of outcome scores is roughly equal at different values of the predictor variable. Therefore, this scatter plot confirms that the derived regression model satisfies the assumptions of linearity and homoscedasticity.

5.3.5 Cross-Validity of Derived Multiple Regression Model

5.3.5.1 Multiple Regression with Stepwise Method

The stepwise method is used to determine which predictors best explain public transport. Based on R (derived from multiple correlations between the predictor and outcome variables) and R^2 (a measure of how much of the variability in the outcome is accounted for by the predictors) it is possible to test how much the explanatory power of a model improves as more predictors are added (Field 2013). The aim of the stepwise method is to determine which combinations of variables best account for variations in the dependent variable—in this case, log (public transport usage density). Even though the stepwise results do not provide insight into the impact of all variables of interest, they do provide us with a foundation for understanding the influences of some key variables and the contributions of each predictor to variation in the dependent variable.

Model Summary[k]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.885[a]	.782	.782	1.01646	.782	1039.41	1	289	.000	
2	.900[b]	.810	.809	.95171	.027	41.658	1	288	.000	
3	.911[c]	.830	.828	.90257	.020	33.219	1	287	.000	
4	.918[d]	.842	.840	.87073	.012	22.373	1	286	.000	
5	.920[e]	.847	.844	.85852	.005	9.187	1	285	.003	
6	.923[f]	.852	.849	.84651	.005	9.150	1	284	.003	
7	.925[g]	.855	.851	.83891	.003	6.169	1	283	.014	
8	.927[h]	.859	.855	.82925	.004	7.630	1	282	.006	
9	.928[i]	.861	.857	.82259	.003	5.588	1	281	.019	
10	.929[j]	.864	.859	.81709	.002	4.791	1	280	.029	1.886

a. Predictors: (Constant), Log (Bus Service Provision Density Factor)

b. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor)

c. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor), Log (Employment Density (Health/ Welfare/ Community Services Industry))

d. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor), Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Students and Mid-aged Dominant Resident Population Density Factor)

e. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor), Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Students and Mid-aged Dominant Resident Population Density Factor), Log (Below $2000 Weekly Earner Dominant Income Group Factor)

f. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor), Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Students and Mid-aged Dominant Resident Population Density Factor), Log (Below $2000 Weekly Earner Dominant Income Group Factor), Distance from City Center

g. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor), Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Students and Mid-aged Dominant Resident Population Density Factor), Log (Below $2000 Weekly Earner Dominant Income Group Factor), Distance from City Center, Log (University Students Population Density)

h. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor), Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Students and Mid-aged Dominant Resident Population Density Factor), Log (Below $2000 Weekly Earner Dominant Income Group Factor), Distance from City Center, Log (University Students Population Density), Log (Average Public Transport Stops per Km2)

i. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor), Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Students and Mid-aged Dominant Resident Population Density Factor), Log (Below $2000 Weekly Earner Dominant Income Group Factor), Distance from City Center, Log (University Students Population Density), Log (Average Public Transport Stops per Km2), Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))

j. Predictors: (Constant), Log (Bus Service Provision Density Factor), Log (Train Service Provision Density Factor), Log (Employment Density (Health/ Welfare/ Community Services Industry)), Log (Students and Mid-aged Dominant Resident Population Density Factor), Log (Below $2000 Weekly Earner Dominant Income Group Factor), Distance from City Center, Log (University Students Population Density), Log (Average Public Transport Stops per Km2), Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry)), Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry))

k. Dependent Variable: Log (Public Transport Usage per Km2)

Table 37: Model Summary of Multiple Regressions with Stepwise Method

Table 37 shows how the model can offer better predictions by adding more predictors. Log (bus/ferry service provision density factor) is the most powerful predictor of variation in log (public transport usage density), which is 78.2% (R square= 0.782). This indicates that percentage changes in public transport usage in the Perth metropolitan suburbs is mainly driven by percentage changes in the bus/ferry service provision density factor.

Changes in employment densities in health, welfare, and community service industries have the highest impact on public transport usages, as compared to other industries. The student population and mid-aged dominant estimated resident population density factor is the third most influential predictor, followed by the income group factor. Furthermore, average public transport stops per km^2 and university student population density are also important factors.

It is significant that the differences between adjusted R square and R square values are very small: max 0.005, which indicates the good cross-validity of the derived models.

5.3.5.2 Robust Regression

As explained in the section 5.1.1, public transport usage in Perth, Fremantle, Murdoch, Bentley and Joondalup is substantially different from the rest of the suburbs in Western Australia. As a consequence, the public transport usage variable has high skewness and kurtosis values. These values are extreme outliers and could distort the regression results. Therefore, a log transformation is conducted to address these outliers and to normalise the distribution.

Following Barnett (1994), robust regression is also used to cross-validate the public transport usage density model that was derived from the standard (enter) regression method. The following graph plots the variable's leverage against its normalised squared residuals. Most of the observations are clustered together. There are only a few observations, such as Chidlow, Crawley, Karakatta and Martin with relatively high leverage; their squared residuals are comparatively high, but they are not significantly different from the majority.

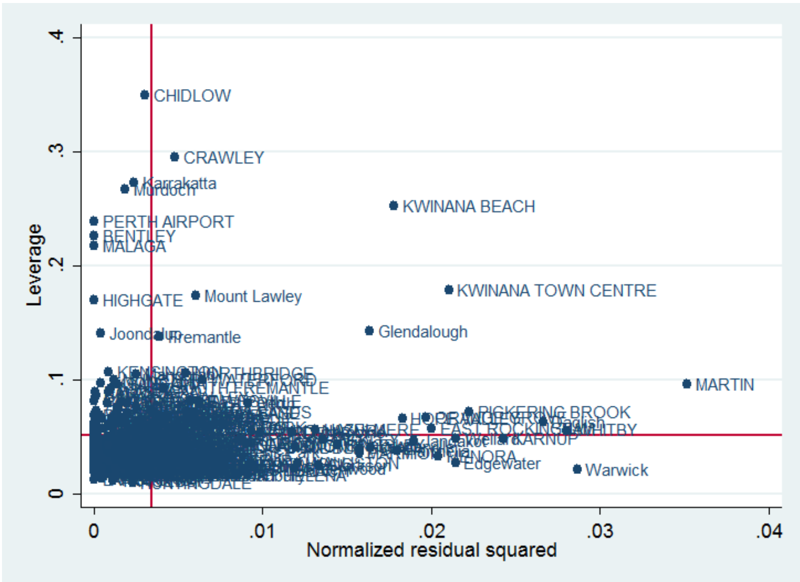


Figure 51: Scatter plot: Leverage vs Normalized Residual Squared (Robust Regression)

The software package Stata was used to generate a robust regression model. The results are as follows:

```
23 rreg log_Dens ln_Train ln_Bus log_AvgStop Log_DE1 Log_DE2 Log_DE3 log_DE4
   log_DE5 Log_ERPS log_Uni Log_Inco, Log_Affl, Log_Road, Distance, gen(weight)
   Huber iteration 1: maximum difference in weights = .64195266
   Huber iteration 2: maximum difference in weights = .08881194
   Huber iteration 3: maximum difference in weights = .02464455
   Biweight iteration 4:  maximum difference in weights =.27500406
   Biweight iteration 5:  maximum difference in weights =.03648498
   Biweight iteration 6:  maximum difference in weights =.03167046
   Biweight iteration 7:  maximum difference in weights =.017472
   Biweight iteration 8:  maximum difference in weights =.01509507
   Biweight iteration 9:  maximum difference in weights =.01161762
   Biweight iteration 10: maximum difference in weights =.01165596
   Biweight iteration 11: maximum difference in weights =.00713276

   Robust regression                           Number of obs =       292
                                                  F( 14,   276) =   142.31
                                                  Prob > F      =   0.0000
```

log_Dens	Coef	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_Train	.129820	.021879	5.93	0.000	.0867498	.1728917
ln_Bus	.539371	.0690923	7.81	0.000	.4033566	.6753862
log_AvgStop	.356169	.111363	3.20	0.002	.1369406	.5753983
Log_DE1	.073161	.0462571	1.58	0.115	-.0179004	.1642228
Log_DE2	.056505	.045763	1.23	0.218	-.033583	.1465948
Log_DE3	.079059	.0588596	1.34	0.180	-.0368108	.1949306
log_DE4	-	.0573281	-1.93	0.054	-.2237184	.0019933
log_DE5	.084174	.0436883	1.93	0.055	-.0018297	.1701796
Log_ERPS	.511210	.1858111	2.75	0.006	.1454234	.8769977
log_Uni	.131255	.0399545	3.29	0.001	.0526014	.20991
Log_Inco	.560990	.1686125	3.33	0.001	.2290607	.8929207
Log_Affl	-	.1071752	-1.26	0.209	-.3459324	.0760368
log_Road		.1529176	-0.45	0.652	-.3699665	.2320997
Distance	-	.0040095	-1.76	0.080	-.01494	.0008462
_cons	3.98789 9	1.21137	3.29	0.001	1.603199	6.372598

Where

Log_DE1 = Log(Employment Density(Office/Business/Utilities/ Communication/ Residential Industry)),

Log_DE2 = Log(Employment Density (Shop/Other Retail/Entertainment/Recreation/Culture Industry)

Log_DE3 = Log(Employment Density (Storage/ Distribution/ Service Industry)),

Log_DE4 = Log(Employment Density (Manufacturing/Processing/Fabrication Industry)),

Log_DE5 = Log(Employment Density (Health/ Welfare/ Community Services Industry)),

clist SSCName weight absr1 d1 in 1/10, noobs

SSCName	weight		d1
MARTIN	.01220883	3.276704	.0759785
Daglish	.05599827	2.798787	.0350293
Warwick	.06639201	2.841328	.0116933
WHITBY	.12409872	2.861236	.0323716
EAST ROCKINGHAM	.15830321	2.420708	.0237403
KWINANA BEACH	.17652596	2.561938	.1478009
Edgewater	.18437076	2.464587	.011223
KARNUP	.18624939	2.652829	.0238904
PICKERING BROOK	.19251097	2.571651	.0343344
Wellard	.19982141	2.493785	.0210935

Table 38: Robust Regression Model

As shown in Table 38, the d1 and d2 values of the F statistics are the same as the ones generated with the regression model using the standard (enter) method. The F change (142.31) is also not significantly different from the previous value (127.209). In both models, there is relatively low variation in each predictor-coefficient with the same sign (positive or negative). This is due to the weight assigned to each observation in the robust regression. These variations, however, are not significant. The results indicate that the public transport usage density model that was derived by the 'forced enter' method is stable and best fitted to the data. Therefore, this model will be used to analyse which factors have the most influence on the public transportation use. Policy implications of the findings are then discussed.

5.3.6 Multiple Regressions with Different Combinations of Observed Variables

Six different multiple regression models, using different combinations of observed variables, are thoroughly examined to gain a better understanding of which determinants are important in explaining public transport use, and of which contribute most to the development of a predictive model.

	Model A		Model B		Model C		Model D		Model E		Model F	
	Only Service Provision Factors		Land Use Characteristics		Land Use Characteristics + SocioEconomic Factors		Land Use Characteristics + Service Provision Factors		Land Use Characteristics + Service Provision+ Socio Economic Factors		Land Use Characteristics + Service Provision+ Socio Economic Factors + Urban Form	
R Square	0.828		0.742		0.755		0.86		0.864		0.866	
Adjusted R Square	0.826		0.736		0.747		0.856		0.858		0.859	
ANOVA (F)	461.11		116.53		96.03		172.71		147.11		127.209	
ANOVA (Sig)			.000[b]		.000[b]		.000[b]		.000[b]		.000[b]	
Durbin-Watson	2.008		1.586		1.614		1.99		1.941		1.9	
Coefficient			B	Sig.	B	Sig.	B	Sig.	B	Sig.	B	Sig.
(Constant)	2.730	.000	6.282	.000	5.975	.000	3.352	.000	3.108	.000	2.668	.043
Log(Train Service Provision Density Factor)	.157	.000					.136	.000	.142	.000	.140	.000
Log(Bus/Ferry Service Provision Density Factor)	.747	.000					.563	.000	.541	.000	.540	.000
Log (Average Public Transport Stops Per Km2)	.571	.000					.345	.004	.336	.004	.298	.014
Log (Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry))			.226	.001	.214	.001	.116	.019	.121	.015	.104	.038
Log (Employment Density (Shop/Other Retail/ Entertainment/ Recreation/ Culture Industry))			.230	.000	.172	.009	.103	.033	.070	.160	.074	.137
Log (Employment Density (Storage/ Distribution/ Service Industry))			.135	.114	.133	.111	.086	.178	.086	.173	.063	.326
Log (Employment Density (Manufacturing/ Processing/ Fabrication Industry))			-.187	.025	-.167	.041	-.189	.002	-.175	.004	-.157	.012
Log (Employment Density (Health/ Welfare/ Community Services Industry))			.129	.044	.118	.062	.094	.048	.083	.081	.087	.066
Log (Students and Mid-aged Dominant_Resident Population Density Factor)			1.824	.000	1.958	.000	.416	.012	.530	.002	.423	.036
Log (University Students Population Density)			.169	.004	.170	.003	.122	.005	.124	.004	.123	.005
Log (Below $2000 Weekly Earner Dominant_Income Group Factor)					.804	.001			.475	.008	.485	.008
Log (Affluence Factor)					-.277	.069			-.040	.732	-.047	.684
Log (Road Length (in m) per Km2)											.087	.597
Distance from City Center											-.008	.054

Table 39: Multiple Regression Models Comparison (with Different Combinations of Observed Variables)

The coefficient of determination (R Square) values in the above table show that Model F which is derived from all observed variables can explain public transport usage in the Perth metropolitan suburbs better than the others. The coefficient of determination values in Models A (only public transport service provisions are considered) and Model B (only land use characteristics are considered) can be interpreted as service provision factors explain more variation in public transport use than do land use characteristics alone. The R square value for Model C (both land use characteristics and socio-economic variables are considered) is slighter higher than for Model B, which indicates that combining socio-economic and land use characteristics can improve the model marginally. However, when service provision factors are considered along with land use characteristics, the predictive power of the model increases noticeably, from 0.74 to 0.86. Finally, the R square values in Model E and F indicate that including service provision factors along with land use characteristics, socio-economic and urban forms factors as explanatory variables accounts for the largest amount of variation in the dependent variable.

The results from the Durbin-Watson test also confirm that taking the service provision factors into account improves the model by better satisfying the assumption of auto-correction. When service provision factors are excluded, the Durbin-Watson test values are substantially less than 2 (closer to 1.5). However, when service provision factors are combined with the land use characteristics, socio-economic, and other urban from factors, the Durbin-Watson test value in Model F gets closer to 2 (at 1.9) indicating the improvement in fitness of model.

5.4 Conclusion

The findings from this study confirm that there are immense spatial variations in public transport usage in Perth CBD and other suburbs where universities are located such as Bentley, Murdoch, Joondalup, Crawley, Mount Lawley, Fremantle and Midland. The significantly high public transport usage is mainly driven by employed people and by students, who are the primary public transport patrons in the Perth metropolitan suburbs. Nevertheless, the temporal variation-by-month in public transport usage in Perth metropolitan suburbs is not very significant. There is a slight decrease in usage in December and January when university and schools have a break, and during public holidays. When public transport usage in 3-hour time periods is examined, it confirms that peak hours are during 6-9am and 3-6pm for weekdays. There are however no peak hours during the weekends when people also do not use public transport much until 12 noon. On weekends, there is relatively consistent usage between 12 noon and 12 midnight.

When factor analysis was conducted for service provision variables, two latent factors were derived: a bus/ferry service provision factor and a train service provision factor. Another factor analysis was conducted to find identify latent factors among land use characteristics and socio-economic variables. The derived factors are: students and mid-aged dominant resident population density; below $2000 weekly earner dominant income group; and an affluence factor. The scores of these latent factors were used, along with other land use characteristics such as employment densities in various industries, university student population densities, road length (in km) per km^2 and distance from city center, to generate a predictive model for public transport usage by way of standard multiple regression. The findings from the descriptive analysis of the explanatory and outcome variables show that the variations within these variables are immense across the Perth metropolitan suburbs. Therefore, the data were transformed before running the regression to make sure the developed model satisfies all of its assumptions. Additionally, other regression methods such as step-wise and robust regression are also applied to validate the robustness of the model. Based on the factor loadings of the service provision, land use characteristics, socio-economic factors and urban form factors along with the standardised coefficients (beta coefficients) calculated with the predicting model, a sensitivity analysis was conducted to calculate how changes in these factors can impact public transport usage across the Perth metropolitan suburbs.

Most importantly, this study finds that service provision factors are the primary determinants of public transport usage in the Perth metropolitan suburbs. To gain a better understanding and to develop a public transport usage model, land use characteristics and socio-economic factors need to be analysed together with service provision and urban form factors.

6 Conclusions

This chapter begins with an overview of the research objectives and findings. The summary of the findings is presented in the form of specific answers to the research questions listed in Chapter 2. Second, the chapter examines a range of policy recommendations and implications to facilitate and encourage an efficient and sustainable transportation system in Perth's Metropolitan Suburbs. Third, it addresses the limitations of this study in terms of data, methods, and results. Finally, the chapter discusses how new areas of research can be built and refined based on the limitations of this thesis.

The main objectives of this study can be can be summarised as follows:

3. First, to provide a comprehensive descriptive analysis of public transportation use patterns in the metropolitan suburbs of Perth, based on temporal and spatial factors and types of patrons.
4. Second, to analyse of the determinants of public transport use, and their synergistic influences on public transport usage with the aim of constructing a robust transportation-demand forecasting model to inform policy decisions.

6.1 Addressing the Research Questions

This section draws on conclusions from the findings reported in Chapter 5 in response to the research questions of the study (refer to Chapter 2 and below in this chapter). A multilevel regression model was developed for a thorough and rigorous analysis of public transport use. Before addressing the findings from this model, it is important to understand the results from the temporal and spatial descriptive analysis which allowed for its factors and variables to be identified. This allowed for a comprehensive picture to be constructed of public transport use variations across Perth's metropolitan suburbs according to types of patrons, origin suburbs of their journeys, and time of travel (weekdays, Saturdays, or Sundays) based on a full set of actual usage data for 2009.

As described in chapter 5, amongst all regression models, model F which is the public transport usage predicting model considering several determinants such as land use characteristics, social-economic and urban form factors and public transport service provision factors explains the largest amount of variance in public transport use, as attested by its high R-squared value. Therefore, model F was chosen as the multilevel regression function for this study. In the following section, the standardised coefficients from the model are used to describe the differential impact of its variables on public transport use, and to specify their respective effects. To construct such a model, the following specific tasks have been completed and the findings from each task are discussed as below:

1. To provide detailed analysis and spatial and temporal characterisation of transport usage patterns across the suburbs of Perth,

- The standard patron group contributes the highest percentage share (49%) of total public transport usage, followed by university students whose public transport usage accounts for 21% of total public transport usage. In addition, the usage from other students (up to year 12) accounts for 18% of total usage. However, the public transport usages from the health care beneficiaries and elderly people are relatively low.
- According to this study, there are distinctive spatial variations in public transport usage across metropolitan suburbs in Perth. First, total public transport usage in Perth suburb (and not to be confused with the total metropolitan area of the city referred to as Perth metropolitan suburbs or Perth) is the highest amongst all metropolitan suburbs due to its exceptionally high employment density amongst all the suburbs. Because of its significant land use characteristics and service provision factors, public transport use in Perth suburb is nine times higher than in Fremantle, which has the second highest use rate. Then it is followed by the suburbs with high university student populations such as Fremantle, Murdoch, Bentley, Joondalup and Crawley. It is also important to note that the usage per capita and service provision per capita in Karrakatta are high compared to other suburbs, indicating that high public transport usage could be driven not only by high density of activity population but also by service provision. The Perth metropolitan area is serviced by five train lines which terminate in Perth CBD, a wide network of buses and a short ferry line across the Swan River. The analysis showed that public transport use per capita becomes higher along the train lines for the suburbs which are located close to the city centre. In the same vein, per capita usage decreases as the suburbs get farther away from the central business district.
- The study also finds interesting temporal variations of public transport usage in Perth metropolitan suburbs. Although there are no seasonal fluctuations in the pattern of usages by standard patrons, senior, pensioner and health care patron, public transport usages of university students and students (up to year 12) fluctuate throughout the year depending on school/university study periods. It is noticeable that public transport usage on weekdays is significantly different from the usages on weekends in 2009. The only similarity between weekdays and weekends is that the public transport usage was the lowest during the 12am-6am peroid. The findings confirm that 6am-9am and 3pm-6pm are the peak hours for public transport use on weekdays. During weekends, the public transport usage during morning period 6am-9am is higher than 9am-12 noon. Then it increases significantly during 12noon-3pm and 3pm-6pm before falling during the evenings. The usage peaks again in the night period (9pm-till midnight). Patrons with

standard SmartRider cards and university students are the main users during weekends with both groups contribution being approximately the same.

- This analysis did not find a lot of seasonal variation in public transport use apart from a slight decrease in December and January. Although usage by students (up to year 12) and university students decreases during school holidays and semester breaks in the months of January, April, June, July, and December, overall there are no dramatic seasonal variations in public transport usage patterns for the Perth metropolitan suburbs.

2. To examine land use characteristics, such as resident population density by age and gender, and student (including university student) population densities, determinants of variations in public transport use across the Perth metropolitan suburbs. If so, which mixed land use characteristics factor has a greater potency in accounting for these variances?

   The findings are:

- Consistent with findings from previous studies, there is a significant relationship between density, especially resident population density (which is a land use characteristic) and public transport usage. The student and resident population densities factor is the third most influential determinant of public transport usage in the Perth metropolitan suburbs.
- The relationship between university student population density (another land use characteristic) and public transport use is also significant, even though the former's influence is not as strong as the "students and mid-aged dominant resident population density" factor.
- Many previous empirical findings demonstrate the existence of a significant relationship between density and public transport use (refer to Chapter 2), even though the size of this relationship varies depending on the locations, regions, and/or geographical scales investigated. The results from this research support the findings of prior studies. They also suggest that *not only the resident population, but also student (primary, secondary, tertiary) populations, should be considered in aggregated public transport demand modelling*, since patrons generate trips from the areas not only where they reside, but also where they study.

3. To analyse whether land use characteristics factors, such as the employment densities in various industries, significantly related to discrepancies in public transport use in metropolitan Perth suburbs.

   Below are the findings:

- Employment densities in particular industries do not have a significant influence on public transport use in Perth. Hence, the role of employment densities, along with types of industries, needs to be considered together with mixed land use development in order to formulate policies for integrated, sustainable transportation.

- The only industry category for which employment densities have a statistically significant influence on public transport usage is office, business, utilities, communication, and residential industries. While employment densities in other industries have positive relationships with public transport usage, the only industries with a statistically significant and negative effect on public transport use are manufacturing, processing and fabrication industries. On the other hand, employment densities in the storage, distribution, and service industries do not have significant relationships with public transport use no matter what other explanatory variables are considered in the regression models**.**
- Many previous studies find a significant relationship between employment densities and travel demand (see Chapter 2). These studies also promote the integration of mixed land use into the framework of sustainable transportation. The results from this study confirm such findings to the extent that there is a significant relationship between employment density and public transport use. This research shows that *industry-type should be considered when unpacking the relationship between mixed land use and public transportation*.

4. To investigate whether urban form factors, such as distance from the city centre and total road length per square kilometer, influence public transport use in Perth. If so, to what extent can they determine this public transport use,

- The empirical findings from this study show that the relationship between road length (in meters per square kilometer) and public transport usage is not statistically significant. On the other hand, distance from the city center is statistically significant. *As the centroid of a suburb gets farther away from the city center by 18.9km increments, its public transportation use density decreases by 0.8% per increment*.
- The result reflects lower public transport usage densities, along with poorer service provisions, in the outer suburbs of the Perth metropolitan areas.

5. To examine whether socio-economic conditions such as income, average rent as main household expenditure, and average car ownership per household, can be used as an explanatory factor in understanding the differences in public transport usages across the suburbs in Perth. If so, which socio-economic conditions have the most analytically salient and empirically cogent in explaining the variation,
   - The number of residents with weekly income below $2000 is a statistically significant factor in explaining the differences in usage density across Perth's suburbs. The extent to which this factor determines public transport use is noteworthy, as it is the second strongest of the observed variables in this regard.
   - On the other hand, the relationship between the affluence factor and public transport usage density is not statistically significant, and its negative effect remains unchanged in all model specifications. Recall that this factor was constructed by combining the number of residents whose weekly income is above $2000 with average rent and average car ownership per household.
   - These results confirm previous findings of statistical significance in positive income elasticity for public transport use, but this relationship is reversed when car ownership also enters into the equation. Thus, the results support the recommendations to *consider the relationship between high income and car ownership jointly when examining their total impact on public transportation use*.

6. To investigate whether it is possible to use the availability of public transport services to explain differential usages of public transport across the suburbs in Perth. If so, how would it induce differential patterns of public transport in Perth,
   - The relationships between all three service provision factors which train service provision density factor, bus/ferry service provision density factor and average (bus/ferry/train) stops per km^2 and public transport use density are particularly significant; the values of these relationships relatively remain the same regardless of which combinations of other variables are included in the model. While the three factors are among the top five most influential determinants of public transport usage density, bus/ferry service provision density is the most influential factor among all of them. This highlights the importance of providing network solutions that can take people as close to their destination place as possible.
   - Findings from all these three service provision factors also indicate that service network density (average stops per km^2) along with bus service provision quality (in terms of frequency) should be evaluated, improved and optimized, with a focus on modally-integrated public transportation systems to maximise usage.
   - Integration among all available public transport services is important to elevate its attractiveness for greater usage. Perl (2011) stress the importance of integrating local train and intercity train systems to increase accessibility and then to encourage more usage at higher levels of geographical scale (between cities).

The results from this research also support the importance of modal-integration among available public transport services (local bus and train systems) at more granular geographical scales (suburbs level).

- This result supports the findings from Cervero (2001b) that an increase in accessibility (time or walking distance) can lead to an increase in public transport demand. It actually *confirms the importance of accessibility's role in determining public transport demand*, as described by Polat (2012, p. 1217).

7. To analyse how sensitively changes in these observed variables impact on variations in public transport use,
   - If we rank all factors based on their relative elasticities of public transport usage, we observe high elasticities for the three service provision factors, population density factor, low-income group factor, and employment densities in office/business/ utilities/communication/residential industries. Bus/ferry service provision has the highest elasticity: *a one percent increase in this factor leads to 0.54 percent increase in public transport density*. The following table summarizes the sensitivity public transport use to the observed variables:

1% increase in …	**=X% increase in public transport usage density**
Weekly Bus/Ferry Service Provision Density Factor	0.54%
Below $2000 Weekly Earner Dominant Income Group Factor	0.48%
Students & Mid-aged Dominant Resident Population Density Factor	0.42%
Average Public Transport Stops Per Km2	0.29%
Weekly Train Service Provision Density Factor	0.13%
University Students Population Density	0.12%
Employment Density (Office/ Business/ Utilities/ Communication/ Residential Industry)	0.1%

8. To determine which factor has the highest explanatory power when land use characteristics, urban form, socio-economic, and public transport service factors are considered at once,
   - To answer this question, when these variables are considered together, the bus/ferry service provision density factor has the most explanatory power.
   - In addition, it is also important to acknowledge and consider the comparatively *strong influences the "below $2000 weekly earner dominant income group factor" and "students and mid-aged dominant resident population density factor"* because their influence on variations in public transport use are very close that of bus/ferry provision density.

9. To develop the most appropriate way to construct the model to predict the public transport usage based on all these factors.

- To conduct a more comprehensive analysis, it is necessary to include a wider range of explanatory variables regarding land use characteristics, urban form, socio-economic, and service provision factors. However, the presence of multi-collinearity among a wide spectrum of explanatory variables can violate the assumptions of regression analysis. Thus, factor analysis should be used to identify latent variables among these explanatory variables based on their correlations. This reduces the number of explanatory variables while maintaining the richness and granularity of the data.
- Therefore, the most appropriate way to construct the model is to *apply factor analysis to derive any latent variables among a large number of explanatory variables*, and then to conduct *multiple regression analysis in a sequential fashion* by including these latent variables. *It is very important to conduct necessary initial checks that the variables are normally distributed, and to perform transformations if necessary*.

Based on these findings, the main research question can be answered as below:

What are the primary determinants that explain the spatial and temporal variations in public transport usage in Perth for the year 2009?

- The bus/ferry service provision density factor is the most important factor in explaining the spatial and temporal variations in public transport use in Perth's metropolitan suburbs, along with socio-economic factors (below $2000 weekly earner dominant income group factor) and land use characteristic (students and mid-aged dominant resident population density factor).
- It is also important to note that we must account for the combined influences of land use characteristics, socio-economic factors, and service provision factors in shaping public transport use.

6.2 Policy Implications

These findings have a number of policy implications that encourage and promote the use of public transportation as a more environmental friendly and sustainable mode of travel, which can enhance the quality of life in Perth's metropolitan suburbs. As discussed at length in the literature review chapter, many studies have confirmed the role of land use characteristics, socio-economic factors, urban forms, and travel modes. However, depending on where and how the studies are conducted, the strength, direction, and significance of these variables' effects diverge substantially. The present study follows Handy (2005) recommendation that we consider a more comprehensive repertoire of variables to better understand their nuanced relationships with public transport demand. Accordingly, it examines a large number of variables concerning land use characteristics, urban form, socioeconomics, and service provisions. This more inclusive approach allows to gain a more nuanced and granular understanding of how their synergistic influences can account for the differential in public transport usages across Perth metropolitan suburbs. The policy implications based on these findings are discussed at three levels, namely strategic, tactical and operational. Based on the presented descriptive and multilevel regression analyses of the temporal and spatial variations in public transportation usage in Perth's metropolitan suburbs, the following factors need to be considered to improve service provision and encourage more use of public transport:

6.2.1 Strategic Level Recommendations

Handy (2005) stressed the importance of rigorous research methods as well as the need to improve research design. Again, this study was designed to be as rigorous as possible in terms of conceptualization and research techniques. Its main strength is that it improves on previous research designs by including a wider range of explanatory variables and sub categories[15], as well as its extraction and aggregation of data at more granular geographical scale (at the suburb level) than before. The findings from this study show that l*and use characteristics, urban form, socio-economic factors, and service provisions factors should not be considered in insolation since*, depending on which explanatory variables are considered in the model, the significance and magnitude of a given variable can be changed.

Regarding the influence of employment density on public transport use, Cervero (1988) states that mixed-use development can improve suburban mobility and can encourage the reduction of auto- dependency. Later, Cervero (1991) examines the relationship between land use characteristics and the percentage of work trips made by modes of public transportation,

[15] Instead of aggregated population density, population densities by age and gender are included in factor analysis. Both of students (year up to 12) and university students are also considered separately instead of aggregating them. The number of residents whose weekly income in four different groups are used as explanatory variables rather than aggregating them to get average income in each studied suburb because it is aimed to examine how each income group has various influence on public transport usages.

finding that the degree of mixed land use has a strongest relationship of the variables considered with public transport usage. In addition, Schimek (1996) quoted in Badoe (2000, , p. 238), asserts that high employment densities in CBD and inner suburbs are conductive to higher public transport usage. Furthermore, Frank (1994a) find that, of the variables they analysed, employment density has the strongest relationship (at 0.44) with public transport usage for shopping trips, followed by population density (at 0.16). The empirical findings from this study relating to employment density are noteworthy because *employment densities only have significant relationships with public transport use in certain industries*. Specifically, *employment densities in the office, business, utilities, communication and residential industry category are very strong determinants of public transport use*, reflecting their high employment densities in the CBD (22% of total employment in these industries are located in the CBD) and high service provision. Accordingly, this result confirms that the presence of high employment densities in the CBD, which is the target area for high provision of public transport service, is highly conductive to higher usage, thus highlighting the *importance of integrating mixed land use development and a sustainable transportation system*.

Prior studies speak to the role of socio-economic factors, especially income and average car ownership elasticities, in public transport usage. The majority of previous studies show a significant negative impact of car ownership on public transport. Yet, income-elasticities for public transportation vary depending on the locations of the studies, since in some areas, people perceive it as an inferior good, such that its income elasticity is negative. Meanwhile, the income elasticity to public transport use is relatively low, particularly where service quality is high and public transport is not perceived as an inferior good. Therefore, White (2004) recommend that income and car ownership be jointly controlled when examining their influence on public transport use. The findings from this study illustrate that the income factor (which is mainly dominated by the number of residents whose weekly income is below $2000) is the second most influential determinant of the public transport usage variations in Perth. On the other hand, the relationship between the public transport usage density and the affluence factor (constituted by high average car ownership, average rent, and the number of residents whose weekly income is above $2000) is negative but this relationship is not statistically significant. Therefore, this study recommends that *policymakers should also target the areas where there are a high number of residents whose weekly income is below $2000*. This resonates with the suggestions by Currie (2009) and Dodson (2004) that transport and urban planners need to address the social and economic disadvantages low-income households encounter in using public transport.

Therefore, at a strategic level, it is important that policy makers consider *integrating public transportation and land use planning to facilitate sustainable transportation outcomes in conjunction with social, economic and environmental benefits while reducing transportation and other disadvantages in the outer suburbs. Integrated public transportation (more frequent bus services integrated with train services) should also be provided to the areas with high* *student and resident population densities, as well as to the areas where there is a high density of people with weekly incomes below $2000.*

6.2.2 Tactical Level Recommendations

More importantly, it is essential to consider integrated public transport and land use planning to increase public transport usage, while providing more frequent bus service with higher service network density (enhanced accessibility), in conjunction with the provision of train services, to *areas where student and resident population density are high, and where more people whose weekly income below $2000 reside*.

The following factors should be taken into account at a tactical level:

a) Integrating modes of public transport in a system with feeder bus services for the suburbs that have very low public transportation usage per capita and no direct access to train services at farther distances from city center,
b) Evaluation of the public transport service stop-placement efficiency and service network density—especially bus stops, because the majority of public transport stops are bus stops in Perth metropolitan suburbs). This evaluation should be accompanied by improvements to bus service provision; policymakers should monitor and increase the use of public transport services, and,

6.2.3 Operational Level Recommendations

This study also indicates the importance of public policy regarding land use characteristics in promoting public transport use, because students (up to year 12) and resident population densities are among its most influential determinants. Indeed, this factor (Students and Mid Aged Dominant Resident Population Density Factor) is the most influential of the land use characteristics across all regression models. Another land use characteristics variable of note is university student population density. The analysis shows the consistency and robustness of this variable across various model specifications, throughout which it preserves its magnitude and significance. Thus, *it is essential to consider the role of university student population density in public policy* that plans for sustainable transportation.

In addition, the train service provision factor is one of the most influential determinants of variations in public transportation usage across the Perth metropolitan area. This indicates the role of public transport service provision in public policy for promoting sustainable transportation. Furthermore, it is also necessary to enhance bus service provision (frequency) and service network density (average stops per square kilometre) to ameliorate the accessibility and availability of services. This is explained by the high service provision elasticities for public transport usage density indicating that people respond significantly to changes in the accessibility and availability of services. Another recommendation is to *consider integrated modes of travel in public transport planning, providing more frequent feeder (electric) bus service to train stations and more densely distributed bus stops that increase accessibility.* When making policy prescriptions on the provision of bus service, it is also necessary to consider using environment-friendly electric buses instead of commonly used internal combustion engine (ICE) buses as a proactive approach to minimise the impact of oil depletion. According to Gilbert (2010, , pg.230), fuel use per person-kilometre in mega-joule by electric public transport (0.6) is considerably lower than that of ICE (diesel) local public transport (2.8) and fuel consumption can be reduced significantly by relying on electrified public transport modes. This is especially so if renewably-generated electricity is the source, as is the case in Calgary's LRT system called "Ride the Wind" which relies entirely on wind power for its electricity needs, Calgary Transit (n.d).

The following factors need to be considered to improve public service provision to attract or encourage more public transport usage at an operational level:

a) More frequent service provision based on the peak hours of patrons with standard smart cards, particularly on weekdays when school is in session,
b) More frequent service provision in the morning periods from 6-9am and 12noon until midnight during the weekends, which will encourage greater public transport use for weekend activities,

6.3 Future Research

Every study has limitations as a result of theoretical framework, such as choice of research method and paradigm, and practical constraints such as data availability, field setting, and different physical, socio-economic, and cultural backgrounds of the areas that are researched. Apart from these issues, the present study also faces a number of methodological limitations in terms of measurements and models. Therefore, these limitations need to be carefully identified so that we can refine future research agendas and expand our knowledge of public transport systems.

1. Analysis of Different Measures of Urban Forms

This study used total road length density as one of the urban forms in its model. Total road length density however as an aggregate does not capture the existence of different types of roads and the different relationships that these urban forms can have with public transport use. If total road length density can be sub-categorised based on the types of roads, along with their proximity to freeways and/or highways, it would allow for interesting comparisons of different relations between road availability and public transport use. It could also explain how different types of road availability make public transport use more or less attractive.

Moreover, other urban forms, such as parking availability, should be controlled to examine the extent to which limited availability encourages or discourages public transport use. In addition, it could provide better explanations of why employment density in some industries does not attract more public transport use—for example, employment density in retail, other retail, entertainment, and recreational industries were not significantly related to the public transport use, which could be due to the high amount of parking available in shopping centres. This is another area that future research should address in the interest of providing relevant policy recommendations regarding integrated land use development and transportation planning.

2. Application of Activity Based Public Transport Demand Modelling

One of the limitations of the research design used in this study is the nature of aggregated data. Recall that total numbers of public transport trips generated in each suburb were aggregated to calculate the public transport usage density. Therefore, the purposes of these trips could not be identified, such that an analysis of the determinants of public transport usage based on different trip purposes could not be conducted.

It is recommended that future researchers identify the stops with estimated journey activities, such as going to school or shopping or leisure activities or work, by determining whether these stops are next to landmarks such as schools, universities, shopping centres, or recreational areas.

3. Developing Patrons' Profiles Models

This study developed public transport usage prediction models based on 2009 public transport usage by existing patrons, including and socio-economic factors, associating land use characteristics, urban form factors and public transport service provision factors. To achieve public transport usage at the maximum possible level, it is essential to understand the travel needs of not only the existing patrons, but also the potential patrons. Therefore, future studies should be conducted to model and profile potential patrons. For example, journey to work datasets collected by the Australian Bureau of Statistics Census, Transperth SmartRider data (current public transport usage data) and public transport service provision data could be analysed together by applying data visualisation tools to model potential patrons' profiling to identify the suburbs or public transport service routes where public transport service provisions should be increased. Students including university and up to year 12 are also major users of public transport contributing 39% of the public transport usage in 2009. A collaborative research integrating data from the Western Australian Education Department and public transport service provision from Transperth could allow to model study trips profiling based on the location or suburbs where these students reside and the places where they study. Public transport service routes which need to be increased to be more attractive for new student patrons and also to encourage existing patrons to use public transport could be identified.

4. Need for the Development of a Comprehensive Panel Data for Causal Analysis

Another limitation of the present research is its use of cross sectional data for inference. For instance, Gim (2012) emphasises the need to consider temporal precedence, causality, and causal directions among explanatory variables and travel behaviour to better understand their interactions—e.g., density can encourage more public transportation demand, which can be incentive to provide more public transport services, which in turn can lead to more residential choices, etc. Because cross-sectional data primarily draws upon the relationship between a host of variables and public transport use within a given year, it is difficult to make these kinds of inferences. Thus, one of the immediate next steps would be to collect and compile a granular and exhaustive dataset on a host of variables in another calendar year so that we can conduct a panel study. This would allow to examine the role of specific policy interventions in shaping public transportation demand. Further, one of the recommendations from Willumsen (2011) is to introduce interventions such as subsidies for public transport use, increase in service provision and implementing bus priority lanes, to reduce the long term negative consequences of urban sprawl with low density development which can cause higher car congestions as well as difficulty and high operation cost to provide good quality public transport services.

Therefore, there is a need to collect and construct a panel database on the land use characteristics, socio economic variables, urban forms, and public transport service provisions and usage starting from the year 2009. Panel data is necessary to conduct a causal analysis and deepen the understanding of causal relationships among the explanatory variables and public transport usage density. This could be achieved by collecting all data sources described in Section 3.9 (Origin of Multiple Data Sources). Employment densities data from the Employment Survey conducted by the Western Australian Department of Planning may not be available, but place of work census data can be used as an alternative because the variation between these two surveys is not that significant—while Census 2006 reported 109,692 as total employment, the employment survey 2007-08 reported 106,999 as total employment with the difference between them not being significant. This could be an extremely meticulous task, but it would be very useful in making a rigorous causal inference about the effects of a specific variable on public transport use. It is also equally useful to examine whether there is any pattern of cumulative causality among the observed variables, explaining how change in one variable leads to successive changes in other variables. This would be very helpful in unpacking the effects of integrated land use on public transportation; for example, higher density areas may generate more public transport services, and the density in these areas may increase because people tend to choose places with better quality public transportation, which in turn increases public transport use.

5. **Application of Different Statistical Techniques**

In this research, only exploratory factor analysis was used in the development of a multiple regression model. One of the limitations when using such models for cross sectional data is that it is impossible to make causal claims about the explanatory variables. However, this method can identify the strengths and directions of important variables' effects on public transport demand, while controlling for a host of other variables.

With longitudinal data, structural equation modelling can be used to examine multiple causal directions among observed variables. Then, the circular cumulative causation theory developed by Myrdal (1956) can be deployed to determine how changes in one variable induce changes in other variables and, consequently, how these changes accumulate and perpetuate the initial change. Also, Liao (2014) suggest using structural equation modelling to examine endogenous relationships among socio-demographic~~s~~, attitudinal, and residential preferences. Structural equation modelling can be combined with confirmatory factor analysis, path analysis, and latent growth modelling to identify the latent variables and their interrelationships, as well as their causal effects.

6. Application of Data Mining Techniques to Examine Travel Pattern Regularity

Gim (2012) states that data mining techniques play a key role in transportation research by extracting actual travel behaviour patterns from multiple datasets. Ma (2013) performed a comparative analysis of the efficiency and accuracy of five data mining algorithms to classify patrons and measure their loyalty to public transport usage. The algorithms are: rough set-based algorithm, C4.5 decision tree, Navïe Bayes, K-NN, and Neural network. Their findings show that the rough set-based algorithm is the most accurate and efficient one for analysing travel pattern regularities. This algorithm was previously proposed by Pawlak (2007) for use in classifying vague and uncertain data, and to help expert systems to understand datasets and develop meaningful rules for classification. Moreover, if regularities are discovered for different types of patrons, they can be used with site selection criteria from various types of businesses to facilitate decision support systems. These systems use fuzzy analytic hierarchy processes and neural networks to identify suitable locations for businesses based on the places to which their targeted customers travel most.

7. Optimisation Modelling for a Public Transport Service Network

The findings from this research indicate that there is a positive relationship between public transport usage and public transport service provisions in regard to average stops per km^2 and frequencies with public transport use. Nevertheless, the provision of highly concentrated service into one single route with extremely high frequencies of every 30 seconds will not facilitate the maximum utilisation of service. Public transport service provision is a scarce resource and optimisation plays an important role to improve its efficiency and effectiveness. Therefore, there is need to conduct research to develop an optimised public transport service network. It is also necessary to investigate any critical points of diminishing returns in the relationship between public transport usage and service provisions. This will allow to identify how public transport usage can be increased with an efficient allocation of services across multiple routes while accounting for the differential needs of the resident, employment and student populations.

8. Application of Mixed Research Methods

Mixed research methods have been applied in some of the previous public transport research to gain a better understanding of the factors people consider in choosing between available transport modes or/and to identify which areas need to be improved to enable public transport modes to be more attractive. Some techniques used in the mixed research methods are: survey to collect the quantitative data and interview for qualitative data to achieve triangulation of the research findings. Due to time and volume constraints in completing this PhD research, it does not include interview or other qualitative data collection methods. It would be highly useful to apply mixed research methods in future research to validate or comprehend by triangulating the findings with qualitative analysis.

6.4 Concluding Remarks

In conclusion, this research is designed to expand knowledge of public transportation systems on both theoretical and practical levels.

The theoretical contribution of this study comes from its focus on links between three main analytical pillars: land use characteristics, socio-economic characteristics, and public transport service provision, so as to explicate their potential contributions to public transport use in the case of Perth's metropolitan suburbs. Further, compared to previous studies, the present one uses a more fine-grained geographical scale. The findings from this research also expand the understanding of the main determinants of public transport by highlighting the synergistic and interactive nature of these variables. Therefore, it calls attention to the limitations of parochial approaches that use too few variables.

As practical contribution, a predictive model of public transportation use has been developed based on the combined influences of land use characteristics, urban forms, socio-economic factors, and service provision. In terms of policy implications, it is recommended that modally integrated public transport systems be enhanced, in conjunction with land use development planning, by considering social inclusion, including student populations, and economic development of Perth's metropolitan suburbs, with the aim of creating a more sustainable future.

References

Achen, Christopher H. 1982. *Interpreting and Using Regression* Edited by John L. Sullivan: Richard G. Niemi, *Quantitative Applications in the Social Sciences*. Newbury Park, California,London: London : SAGE.

Adhikari, Pramod. 2006. "Socio-Economic Indexes for Areas: Introduction, Use and Future Directions." edited by Australian Bureau of Statistics, Canberra: Australian Bureau of Statistics.

Aditjandra, Paulus Teguh: Cao, Xinyu: Mulley, Corinne. 2012. "Understanding Neighbourhood Design Impact on Travel Behaviour: An Application of Structural Equations Model to a British Metropolitan Data." *Transportation Research Part A: Policy and Practice* 46 (1): 22-32. doi: http://dx.doi.org/10.1016/j.tra.2011.09.001.

Aiken, Jacob Cohen: Patricia Cohen: Stephen G. West: Leona S. 2002. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Edited by Jacob Cohen. Mahwah, N.J.: Taylor and Francis.

Andersen, Robert. 2008. *Introduction. Modern Methods for Robust Regression*. Thousand Oaks, CA: SAGE Publications, Inc.

Asensio, J. 2000. "The Success Story of Spanish Suburban Railways: Determinants of Demand and Policy Implications." *Transport Policy* 7 (4): 295-302. doi: 10.1016/S0967-070X(00)00030-5.

Australian Bureau of Statistics. 2 April 2014a. "3218.0 - Regional Population Growth, Australia, 2012-13 (Feature Article: Capital Cities: Past, Present and Future)." edited by Australia Bureau of Statistics, Canberra: Australia Bureau of Statistics.

———. 2 April 2014b. "3218.0 - Regional Population Growth, Australia, 2012-13 (Feature Article: Western Australia)." edited by Australia Bureau of Statistics, Canberra: Australia Bureau of Statistics.

———. 6 September 2011. "6530.0 - Household Expenditure Survey, Australia: Summary of Results, 2009-10." edited by Australia Bureau of Statistics, Canberra: Australia Bureau of Statistics.

———. 15 February 2006. "6530.0 - Household Expenditure Survey, Australia: Summary of Results, 2003-04 (Reissue)." edited by Australia Bureau of Statistics, Canberra.

———. 20 May 2011. Census Dictionary 2011. http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2901.0Chapter44402011.

———. 30 March 2010. Regional Population Growth, Australia, 2008-09 (Western Australia). Accessed 12 December 2013, http://www.abs.gov.au/ausstats/abs@.nsf/Products/3218.0~2008-09~Main+Features~Western+Australia?OpenDocument#PARALINK3.

———. 2007. "2905.0 - Statistical Geography: Volume 2 - Census Geographic Areas, Australia, 2006 ", Canberra.

Badoe, Daniel A.: Miller, Eric J. 2000. "Transportation-Land-Use Interaction: Empirical Findings in North America, and Their Implications for Modeling." *Transportation Research Part D: Transport and Environment* 5 (4): 235-263. http://www.sciencedirect.com/science/article/B6VH8-3YWX3K9-1/2/355635435caf24954213f2b9148dfc35

Bagchi, M.: White, P. R. 2005. "The Potential of Public Transport Smart Card Data." *Transp. Policy* 12 (5): 464-474. doi: 10.1016/j.tranpol.2005.06.008.

Bain, Stuart: Hensher, David A.: Li, Zheng. 2011. "R-Tresis: Developing a Transport Model System for Regional New South Wales." *Journal of Transport Geography* 19 (4): 615-622. doi: 10.1016/j.jtrangeo.2010.06.018.

Bar-Gera, Hillel: Boyce, David. 2003. "Origin-Based Algorithms for Combined Travel Forecasting Models." *Transportation Research Part B* 37 (5): 405-422. doi: 10.1016/S0191-2615(02)00020-6.

Barnett, Vic. 1994. *Outliers in Statistical Data / Vic Barnett and Toby Lewis*. Edited by Toby Lewis. Chichester: Chichester : Wiley & Sons.

Barton-Aschman Associates. 1981. *Traveler Response to Transportation System Changes*. Edited by: Co R. H. Pratt: Amp, : United, States: Washington, D.C.: U.S. Dept. of Transportation, Federal Highway Administration, Office of Highway Planning, Urban Planning Division.

Bass, Pablo: Donoso, Pedro: Munizaga, Marcela. 2011. "A Model to Assess Public Transport Demand Stability." *Transportation Research Part A: Policy and Practice* 45 (8): 755-764. doi: http://dx.doi.org/10.1016/j.tra.2011.06.003.

Bates, John. 2008. "History of Demand Modelling." In *Handbooks in Transport Volume 1: Handbook of Transport Modelling*, ed. D A Hensher: K J BUTTON. Oxford, United Kingdom: Elsevier.

Batey, Marjorie. 1971. "Concepturalizing the Research Process." *Nursing research* 20 (4).

Bellacicco, Antonio: Cossetto, Sandro: Wilson, Alan G. 1987. "Dynamic Models for Person Transportation and Their Relationship to Urban Structure and Change." *European Journal of Operational Research* 31 (2): 209-214. doi: 10.1016/0377-2217(87)90024-5.

Berenson, Mark L., ed. 2013. *Basic Business Statistics / Berenson ... [Et Al.]*. Frenchs Forest, N.S.W.: Frenchs Forest, N.S.W. : Pearson Australia Group.

Berenson, Mark L.: Levine, David M. 1999. *Basic Business Statistics : Concepts and Applications / Mark L. Berenson, David M. Levine*. Upper Saddle River, N.J.: Upper Saddle River, N.J. : Prentice-Hall.

Berry, William D. 1985. *Multiple Regression in Practice / William D. Berry, Stanley Feldman*. Edited by Stanley Feldman. Newbury Park, Califorina, London: Newbury Park, [Calif.], London : SAGE.

Bhat, Chandra R. 2005. "A Multiple Discrete–Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions." *Transportation Research Part B: Methodological* 39 (8): 679-707. doi: 10.1016/j.trb.2004.08.003.

Bhat, Chandra R.: Guo, Jessica Y. 2007. "A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels." *Transportation Research Part B: Methodological* 41 (5): 506-526. doi: 10.1016/j.trb.2005.12.005.

Boarnet, Marlon G.: Sarmiento, Sharon. 1998. "Can Land-Use Policy Really Affect Travel Behaviour? A Study of the Link between Non-Work Travel and Land-Use Characteristics." *Urban Studies* 35 (7): 1155-1169. doi: 10.1080/0042098984538.

Boarnet, Marlon: Crane, Randall. 2001. "The Influence of Land Use on Travel Behavior: Specification and Estimation Strategies." *Transportation Research Part A: Policy and Practice* 35 (9): 823-845. doi: 10.1016/s0965-8564(00)00019-7.

Bohnet, M.: Gertz, C. 2008. "Mode Specific Accessibility and Car Ownership" *International Conference on Mobility & Transport*, *Munich, Germany*: http://www.vsl.tu-harburg.de/Archiv/Bohnet_mobilTUM2008_Paper_Final.pdf.

Bowman, J. L.: Ben-Akiva, M. E. 2001. "Activity-Based Disaggregate Travel Demand Model System with Activity Schedules." *Transp. Res. Pt. A-Policy Pract.* 35 (1): 1-28.

Bresson, G.: Dargay, J.: Madre, J. L.: Pirotte, A. 2003. "The Main Determinants of the Demand for Public Transport: A Comparative Analysis of England and France Using Shrinkage Estimators." *Transp. Res. Pt. A-Policy Pract.* 37 (7): 605-627. doi: 10.1016/S0965-8564(03)00009-0.

Bresson, Georges: Dargay, Joyce: Madre, Jean-Loup: Pirotte, Alain. 2004. "Economic and Structural Determinants of the Demand for Public Transport: An Analysis on a Panel of French Urban Areas Using Shrinkage Estimators." *Transportation Research Part A: Policy and Practice* 38 (4): 269-285. doi: http://dx.doi.org/10.1016/j.tra.2003.11.002.

Bryman, Alan. 2004. *Social Research Methods / Alan Bryman*. Second Edition ed. New York: Oxford University Press, Inc.,.

Buchanan, Nick: Evans, Rick: Dodson, Jago. 2005. "Transport Disadvantage and Social Status: A Gold Coast Pilot Study." *Urban Research Program: Research Monograph 8* http://www.griffith.edu.au/__data/assets/pdf_file/0009/48645/urp-rm08-buchanan-et-al-2005.pdf.

Buehler, Ralph: Pucher, John. 2012. "Demand for Public Transport in Germany and the USA: An Analysis of Rider Characteristics." *A Transnational Transdisciplinary Journal* 32 (5): 541-567. doi: 10.1080/01441647.2012.707695.

Calgary Transit. n.d. "Ride the Wind!"- a Success Story. Accessed 23 April 2015, http://library.tac-atc.ca/proceedings/2002/calgary.pdf.

Cervero, R. 1991. "Land Use and Travel at Suburban Activity Centers." *Transportation Quarterly* 45: 479-491. http://www.uctc.net/papers/091.pdf.

———. 2001a. "Transport and Land Use: Key Issues in Metropolitan Planning and Smart Growth." *Australian Planner* 38 (1): 29-37.

———. 2001b. "Walk and Ride: Factors Influencing Pedestrian Acess to Transit." *Journal of Public Transportation* 3 (4): 1-23.

Cervero, R.: Hansen, M. 2002a. "Induced Travel Demand and Induced Road Investment - a Simultaneous Equation Analysis." *J. Transp. Econ. Policy* 36: 469-490.

Cervero, Robert. 1989. "Jobs-Housing Balancing and Regional Mobility." *Journal of the American Planning Association* 55 (2): 136-150. doi: 10.1080/01944368908976014.

———. 1996a. "Mixed Land-Uses and Commuting: Evidence from the American Housing Survey." *Transportation Research Part A: Policy and Practice* 30 (5): 361-377. doi: 10.1016/0965-8564(95)00033-x.

———. 2001c. "Integration of Urban Transport and Urban Planning." In *The Challenge of Urban Government: Policies and Practices*, ed. Mlila: Stren Freire, Richard. Washington, D.C: The World Bank Institute, Washington, D.C.

———. 2002b. "Built Environments and Mode Choice: Toward a Normative Framework." *Transportation Research Part D: Transport and Environment* 7 (4): 265-284. doi: 10.1016/s1361-9209(01)00024-4.

———. 2002c. "Induced Travel Demand: Research Design, Empirical Evidence, and Normative Policies." *Journal of Planning Literature* 17 (1): 3-20. doi: 10.1177/088122017001001.

———. 2003. "Road Expansion, Urban Growth, and Induced Travel: A Path Analysis." *Journal of the American Planning Association* 69 (2): 145-163.

Cervero, Robert 1988. "Land-Use Mixing and Suburban Mobility." *Transportation Quarterly* 45: 429-446. http://www.uctc.net/papers/003.pdf.

Cervero, Robert: Kockelman, Kara. 1997. "Travel Demand and the 3ds: Density, Diversity, and Design." *Transportation Research Part D: Transport and Environment* 2 (3): 199-219. doi: 10.1016/s1361-9209(97)00009-6.

Cervero, Robert: Radisch, Carolyn. 1996b. "Travel Choices in Pedestrian Versus Automobile Oriented Neighborhoods." *Transport Policy* 3 (3): 127-141. http://www.sciencedirect.com/science/article/B6VGG-3VWTBTF-8/2/8262fbde9fa943cc3ea9605f32635bd1.

Chang, Justins. 2006. "Models of the Relationship between Transport and Land-Use: A Review." *A Transnational Transdisciplinary Journal* 26 (3): 325-350. doi: 10.1080/01441640500468432.

Chapleau, Ka Chu: Robert. 2008. "Enriching Archived Smart Card Transaction Data for Transit Demand Modeling." *Transportation Research Record: Journal of the Transportation Research Board* 2063: 63-72. http://trb.metapress.com/content/072m86768385t412/?genre=article&id=doi%3a10.3141%2f2063-08.

Chiang, Wen-Chyuan: Russell, Robert A.: Urban, Timothy L. 2011. "Forecasting Ridership for a Metropolitan Transit Authority." *Transportation Research Part A: Policy and Practice* 45 (7): 696-705. doi: http://dx.doi.org/10.1016/j.tra.2011.04.016.

Clay, Michael J.: Johnston, Robert A. 2006. "Multivariate Uncertainty Analysis of an Integrated Land Use and Transportation Model: Meplan." *Transportation Research Part D: Transport and Environment* 11 (3): 191-203. http://www.sciencedirect.com/science/article/B6VH8-4JKRWJB-1/2/7008cdb62c6a646ffeb89364f98492c1

Comrey, A. L. 1973. *A First Course in Factor Analysis*. New York: New York: Academic Press.

Cozby, Paul. 2005. *Methods in Behavioral Research / Paul Cozby*. 9th Edition ed: McGraw-Hill Higher Education.

Cullinane, S. 2002. "The Relationship between Car Ownership and Public Transport Provision: A Case Study of Hong Kong." *Transport Policy* 9 (1): 29-39. doi: 10.1016/S0967-070X(01)00028-2.

Cullinane, Sharon: Cullinane, Kevin. 2003. "Car Dependence in a Public Transport Dominated City: Evidence from Hong Kong." *Transportation Research Part D* 8 (2): 129-138. doi: 10.1016/S1361-9209(02)00037-8.

Currie, G. 2010. "Quantifying Spatial Gaps in Public Transport Supply Based on Social Needs." *J. Transp. Geogr.* 18 (1): 31-41. doi: 10.1016/j.jtrangeo.2008.12.002.

Currie, Graham. 2009. "Australian Urban Transport and Social Disadvantage." *Australian Economic Review* 42 (2): 201-208. doi: 10.1111/j.1467-8462.2009.00549.x.

Curtis, Carey. 2008. "Evolution of the Transit-Oriented Development Model for Low-Density Cities: A Case Study of Perth's New Railway Corridor." *Planning Practice & Research* 23 (3): 285-302. doi: 10.1080/02697450802423559.

———. 2011a. "Integrating Land Use with Public Transport: The Use of a Discursive Accessibility Tool to Inform Metropolitan Spatial Planning in Perth." *A Transnational Transdisciplinary Journal* 31 (2): 179-197. doi: 10.1080/01441647.2010.525330.

Curtis, Carey: Holling, Carlindi. 2004a. "Just How (Travel) Smart Are Universities When It Comes to Implementing Sustainable Travel."

———. 2004b. "Travel Demand Management at Australian Universities: Recognition, Contemplation or Action?" In *ATRF04 27th Australasian Transport Research Forum, South Australia*. Transport Systems Centre, University of South Australia.

Curtis, Carey: Mellor, Roger. 2011b. "Anticipating a New Railway: The Relevance of Transit-Oriented Development to Businesses in a Car-Oriented City." *Urban Policy and Research* 29 (2): 141-165. doi: 10.1080/08111146.2010.540001.

Curtis, Carey: Scheurer, Jan. 2009. *Network City Activity Centres. Developing an Analysis, Conception and Communication Tool for Integrated Land Use and Transport Planning in the Perth Metropolitan Area, Research Monograph*. Perth (WA) Department of Planning and Infrastructure (DPI) and Curtin University of Technology. doi: 10.1016/j.progress.2010.05.001.

———. 2010. "Planning for Sustainable Accessibility: Developing Tools to Aid Discussion and Decision-Making." *Progress in Planning* 74 (2): 53-106. doi: 10.1016/j.progress.2010.05.001.

Curwin, Jon: Slater, Roger. 1997. *Quantitative Methods for Business Decisions*. 4th ed. London: International Thomson Business Press.

Dargay, Joyce M.: Hanly, Mark. 2002. "The Demand for Local Bus Services in England." *Journal of Transport Economics and Policy (JTEP)* 36 (1): 73-91.

Department of Climate Change: Australian Government. 2008. Climate Change Impacts. Department of Climate Change: Australian Government,. Accessed 8 June 2008, http://www.climatechange.gov.au/impacts/overview.html.

Department of Commerce. 2011. Fuelwatch Historical Prices. Accessed 2 Janurary 2015, http://www.fuelwatch.wa.gov.au/fuelwatch/pages/public/historicalPriceSearch.jspx.

Department of Environment-Australian Government. n.d-a. National Greenhouse Gas Inventory- Kyoto Protocol Accounting Framework. Accessed 7 February 2015, http://ageis.climatechange.gov.au/.

———. n.d-b. National Inventory by Economic Sector. Accessed 7 February 2015, http://ageis.climatechange.gov.au/ANZSIC.aspx.

Department of Local Government and Communities-Government of Western Australia. n.d. "Defining What We Mean by "Perth Metropolitan Area"." http://metroreview.dlg.wa.gov.au/OpenFile.ashx?Mode=446E37686749376A356D684D2B6E6D6D4D6E555273773D3D&ContentID=6C6B324C4E6D4F4E7873593D.

Department of Planning and Western Australian Planning Commission. August 2010. "Directions 2031 and Beyond." edited by Department of Planning and Western Australian Planning Commission.

Department of Transport. July 2011a. "Public Transport Plan for Perth 2031." edited by Department of Transport, Perth.

———. July 2011b. "Public Transport Plan for Perth 2031: Mapping out the Future for Perth's Public Transport Network, Summary." edited by Department of Transport, Perth.

Department of Transport: Government of Western Australia. August 2011. "Green House Gas Emissions from Household in Western Australia." edited by Department of Transport: Government of Western Australia, Perth, Western Australia.

Department of Treasury and Finance. n.d. "Western Australian, Economic Notes: Gross State Product, 2009-10." edited by Department of Treasury and Finance.

Dodson, Jago: Buchanan, Nick: Gleeson, Brendan: Sipe, Neil. 2006. "Investigating the Social Dimensions of Transport Disadvantage—I. Towards New Concepts and Methods." *Urban Policy and Research* 24 (4): 433-453. doi: 10.1080/08111140601035317.

Dodson, Jago: Gleeson, Brendan: Sipe, Neil. 2004. "Transport Disadvantage and Social Status: A Review of Literature and Methods " *Urban Policy Program: Research Monograph 5*. http://www.griffith.edu.au/__data/assets/pdf_file/0006/48642/urp-rm05-dodson-et-al-2004.pdf.

Dodson, Jago: Sipe, Neil. 2007. "Oil Vulnerability in the Australian City: Assessing Socioeconomic Risks from Higher Urban Fuel Prices." *Urban Studies* 44 (1): 37-62. doi: 10.1080/00420980601023810.

Doti, James: Adibi, Esmael. 1991. "A Model for Forecasting Public Transit." *Transportation Research Part A: General* 25 (6): 443-450. doi: http://dx.doi.org/10.1016/0191-2607(91)90021-H.

Douglas, G. 1991. *Planning on the Fringe: The Impact of Land Use Strategies on Congestion: Prepared for the Third National Conference Transportation on Solutions for Small and Medium Sized Areas*: Douglas & Douglas, Inc.

Enrique Fernández L, J.: de Cea Ch, Joaquín: Malbran, R. Henry. 2008. "Demand Responsive Urban Public Transport System Design: Methodology and Application." *Transportation Research Part A: Policy and Practice* 42 (7): 951-972. doi: http://dx.doi.org/10.1016/j.tra.2007.12.008.

European Commission. 1997. "Isotope (Improved Structure and Organisation for Urban Transport Operations of Passengers in Europe)." edited by Transport Research 4th Framework Programme Urban Transport, DGVII Luxembourg: Office of Official Publication of the European Communities.

Ewing, R.: Cervero, R.: Trb,. 2001. "Travel and the Built Environment - a Synthesis." In *Land Development and Public Involvement in Transportation: Planning and Administration*, 87-114.

Ewing, R.: Haliyur, P.: Page, G. 1994. "Getting around a Traditional City, a Suburban Planned Unit Development, and Everything in Between." *Transportation Research Record* 1466: 53-62.

Ewing, Reid: DeAnna, MaryBeth: Li, Shi-Chiang. 1996. "Land Use Impacts on Trip Generation Rates." *Transportation Research Record: Journal of the Transportation Research Board* 1518 (1): 1-6.

Fawcett, Jacqueline: Downs, Florence S. . 1992. *The Relationship of Theory and Research / Jacqueline Fawcett,* . 2nd Edition ed. Philadelphia: F.A. Davis Company.

Fertal, M J: WEINER, E: BALEK, A J: Sevin, A F. 1966. *Modal Split: Documentation of Nine Methods for Estimating Transit Usage*: Washington : U.S. Bureau of Public Roads, Office of Planning.

Field, Andy. 2005. *Discovering Statistics Using Ibm Spss Statistics*. Los Angeles: Los Angeles : Sage.

———. 2013. *Discovering Statistics Using Ibm Spss Statistics : And Sex and Drugs and Rock 'N' Roll / Andy Field*. Los Angeles: Los Angeles : Sage.

FitzRoy, F.: Smith, I. 1998. "Public Transport Demand in Freiburg: Why Did Patronage Double in Decade?" *Transport Policy* 5: 163-173.

FitzRoy, Felix R: Smith, Ian. 1997. *Passenger Rail Demand in 14 Western European Countries: A Comparative Time Series Study*.

Florian, Michael: Gaudry, Marc: Lardinois, Christian. 1988. "A Two-Dimensional Framework for the Understanding of Transportation Planning Models." *Transportation Research Part B* 22 (6): 411-419. doi: 10.1016/0191-2615(88)90022-7.

Frank, L. D.: Pivo, Gary. 1994a. "Impacts of Mixed Use and Density on Utilization of Tree Modes of Travel: Single Occupant Vehicle, Transit, and Walking." *Transportation Research Record* 1466: 44-52. http://www.reconnectingamerica.org/assets/Uploads/Frank-and-Pivo.pdf.

Frank, Lawrence Douglas. 1994b. "An Analysis of Relationships between Urban Form (Density, Mix, and Jobs: Housing Balance) and Travel Behavior (Mode Choice, Trip Generation, Trip Length, and Travel Time) : Transportation Research Part A: Policy and Practice." Ph.D, University of Washington. http://www.sciencedirect.com/science/article/B6VG7-3VWPM8W-2D/2/47a0b7fb75f86909dd7c682f7d8c1a9b

Gaudry, Marc. 1975. "An Aggregate Time-Series Analysis of Urban Transit Demand: The Montreal Case." *Transportation Research* 9 (4): 249-258. doi: http://dx.doi.org/10.1016/0041-1647(75)90066-0.

Gilbert, Richard: Perl, Anthony. 2010. *Transport Revolutions : Moving People and Freight without Oil*. Gabriola Island: Gabriola Island : New Society Publishers.

Gim, Tae-HyoungTommy. 2012. "A Meta-Analysis of the Relationship between Density and Travel Behavior." *Transportation* 39 (3): 491-519. doi: 10.1007/s11116-011-9373-6.

Gleeson, Brendan. 2006. *Australian Heartlands : Making Space for Hope in the Suburbs*. Crows Nest, N.S.W.: Allen & Unwin.

Goodwin, P. B.: Williams, H. C. W. L. 1985. "Public Transport Demand Models and Elasticity Measures: An Overview of Recent British Experience." *Transportation Research Part B: Methodological* 19 (3): 253-259. doi: http://dx.doi.org/10.1016/0191-2615(85)90008-6.

Gorsuch, Richard. 1983. *Factor Analysis*. 2nd edition ed: Psychology Press.

Government of Western Australia. n.d. Bigger Picture Wa: Major Projects, Max Light Rail. Accessed 6 Februrary 2015, http://getthebiggerpicture.wa.gov.au/max-light-rail/.

Gray, Colin D. 2012. *Ibm Spss Statistics 19 Made Simple / Colin D. Gray & Paul R. Kinnear*. Edited by Paul R. Kinnear. New York: New York : Psychology Press.

Greenwald, Michael J. 2006. "The Relationship between Land Use and Intrazonal Trip Making Behaviors: Evidence and Implications." *Transportation Research Part D: Transport and Environment* 11 (6): 432-446. doi: 10.1016/j.trd.2006.09.003.

Gwilliam, Ken. 2008. "A Review of Issues in Transit Economics." *Research in Transportation Economics* 23 (1): 4-22. doi: http://dx.doi.org/10.1016/j.retrec.2008.10.002.

Hair, Joseph F.: Black, William C.: Babin, Barry J. 2010. *Multivariate Data Analysis : A Global Perspective*. Upper Saddle River, N.J, London: London : Pearson Education, Upper Saddle River, N.J.

Handy, Susan. 1992. "Regional Versus Local Accesibility: Neo-Traditional Development and Its Implications for Non Work Travel." *Built Environment* 18 (253-267).

———. 1996. "Methodologies for Exploring the Link between Urban Form and Travel Behavior." *Transportation Research Part D: Transport and Environment* 1 (2): 151-165. doi: 10.1016/S1361-9209(96)00010-7.

———. 2005. "Critical Assessment of the Literature on the Relationships among Transportation, Land Use, and Physical Activity." *Transportation Research Board and the Institute of Medicine Committee on Physical Activity, Health, Transportation, and Land Use. Resource paper for TRB Special Report* 282. http://onlinepubs.trb.org/onlinepubs/archive/downloads/sr282papers/sr282Handy.pdf.

Hansen, Mark: Gillen, David: Dobbins, Allison: Huang, Yuanlin: Puvathingal, Mohnish. 1993. "The Air Quality Impacts of Urban Highway Capacity Expansion: Traffic Generation and Land Use Change." *University of California Transportation Center*.

Hanson, Susan: Hanson, Perry. 1981. "The Travel-Activity Patterns of Urban Residents: Dimensions and Relationships to Sociodemographic Characteristics." *Economic Geography* 57 (4): 332-347. http://www.jstor.org/stable/144213

Hendrickson, Chris. 1986. "A Note on Trends in Transit Commuting in the United States Relating to Employment in the Central Business District." *Transportation Research Part A: General* 20 (1): 33-37. doi: http://dx.doi.org/10.1016/0191-2607(86)90013-0.

Hensher, David A. 2001. *Transport : An Economics and Management Perspective / David A. Hensher, Ann M. Brewer*. Edited by Ann Brewer. Oxford, New York: Oxford University Press.

Holmgren, J. 2007. "Meta-Analysis of Public Transport Demand." *Transportation Research Part A: Policy and Practice* 41 (10): 1021-1035. doi: 10.1016/j.tra.2007.06.003.

———. 2013. "An Analysis of the Determinants of Local Public Transport Demand Focusing the Effects of Income Changes." *European Transport Research Review* 5 (2): 101-107. doi: 10.1007/s12544-013-0094-0.

Holtzclaw, J.: Clear, R.: Dittmar, H.: Goldstein, D.: Haas, P. 2002. "Location Efficiency: Neighborhood and Socioeconomic Characteristics Determine Auto Ownership and Use - Studies in Chicago, Los Angeles and San Francisco." *Transp. Plan. Technol.* 25 (1): 1-27. doi: 10.1080/03081060290032033.

Holtzclaw, John. 1994. "Using Residential Patterns and Transit to Decrease Auto Dependence and Costs." http://www.smartgrowth.org/library/articles.asp?art=190&res=1024.

Horn, M. E. T. 2002. "Multi-Modal and Demand-Responsive Passenger Transport Systems: A Modelling Framework with Embedded Control Systems." *Transportation Research Part A: Policy and Practice* 36 (2): 167-188. doi: http://dx.doi.org/10.1016/S0965-8564(00)00043-4.

Hunt, J. D.: Kriger, D. S.: Miller, E. J. 2005. "Current Operational Urban Land-Use-Transport Modelling Frameworks: A Review." *Transport Reviews: A Transnational Transdisciplinary Journal* 25 (3): 329 - 376. http://www.informaworld.com/10.1080/0144164052000336470.

Ichikawa, Hiroo. 1982. "A Comparative Study of Urban Form and Its Spatial Characteristics of Japanese and North American Cities ". Transportation Research Part A: General, University of Waterloo,, Waterloo, Ontario, Canada. http://www.sciencedirect.com/science/article/B6X3B-469GJNC-44/2/24c199c1a17a49c62bdac782b28b8a6a

John Taplin, Michael Taylor, Doina Olaru, and Sharon Biermann. 2014. "Transport Modelling Review: Independent Review." http://www.patrec.org/publication_docs/TMR_Final_report_June2014.pdf.

Kelsey, J.L.: Thompson W.D: Evans, A.S. 1986. "Cross-Sectional and Other Types of Studies." *Methods in observational epidemiology* 8: 187-194. http://edocs.library.curtin.edu.au/eres_display.cgi?url=dc60158347.pdf&copyright=1.

Kenneth J. Dueker: Zhong-ren, Peng. 2008. "Geographic Information Systems for Transport." In *Handbooks in Transport Volume 1: Handbook of Transport Modelling*, ed. D A Hensher: K J BUTTON. Oxford, United Kingdom: Elsevier.

Kennewell, Catherine: Shaw, Brian J. 2008. "Perth, Western Australia." *Cities* 25 (4): 243-255. doi: http://dx.doi.org/10.1016/j.cities.2008.01.002.

Kenworthy, J.: Laube, F. 1999a. "A Global Review of Energy Use in Urban Transport Systems and Its Implications for Urban Transport and Land-Use Policy." *Transportation Quarterly* 53 (4): 23-48. <Go to ISI>://CCC:000083545400005.

Kenworthy, Jeff. 1991. "The Land Use and Transit Connection in Toronto : Some Lessons for Australian Cities." *Australian planner* 29 (3): 149-154. http://link.library.curtin.edu.au/p?pid=CUR_ALMA51111988740001951

———. 2013. "Decoupling Urban Car Use and Metropolitan Gdp Growth." *World transport policy and practice* 19 (4): 8-21. http://www.eco-logica.co.uk/pdf/wtpp19.4.pdf

Kenworthy, Jeffery. 2008a. "International Review of the Significance of Rail in Developing More Sustainable Urban Transport Systems in Higher Income Cities." *World Transport Policy & Practice* 14 (2): 21-37. http://link.library.curtin.edu.au/p?pid=CUR_ALMA51110241150001951

Kenworthy, Jeffrey. 2007. "Urban Planning and Transport Paradigm Shifts for Cities of the Post-Petroleum Age." *Journal of Urban Technology* 14 (2): 47-70. doi: 10.1080/10630730701531708.

Kenworthy, Jeffrey R. 2008b. "The Sustainable Urban Passenger Transport Package Deal." http://sustainability.curtin.edu.au/local/docs/Kenworthy_indd.pdf.

Kenworthy, Jeffrey R.: Laube, Felix B. 1996. "Automobile Dependence in Cities: An International Comparison of Urban Transport and Land Use Patterns with Implications for Sustainability." *Environmental Impact Assessment Review* 16 (4–6): 279-308. doi: http://dx.doi.org/10.1016/S0195-9255(96)00023-6.

———. 1999b. "Patterns of Automobile Dependence in Cities: An International Overview of Key Physical and Economic Dimensions with Some Implications for Urban Policy." *Transportation Research Part A: Policy and Practice* 33 (7–8): 691-723. doi: http://dx.doi.org/10.1016/S0965-8564(99)00006-3.

Kenworthy, Jeffrey Raymond. 1986. "Driving Cycles, Urban Form and Transport Energy." http://murdoch.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwpV27DsIwDLSACbGAAPGU8gGgPkJDuyEBFQsbe9U8Krp0KO3A3xNDWI4b7IkcRYrufLYvAGQdIy-27Zi60uOcu1KtXKEclqxYwrisrBjuPanPQXXjbnNGsfySYon50yZbg7HVhKbm3DgSfly_f_pjMCHsQmf3UsvuQUNlfbB2eYopOhFXbDdbkDLncUaQGRKds5Oi8hIn6j50N4BNuD9tD0sTIVIotBWXCP2Yv48a1ctk-TBo9ugQWjqFVyMgHpP6HXGhCYbQkBj7SgpGZSIcyQOfOmMIfg4z-WPvFNpO4BtJYQatIi_VvL7UG1lek-I.

Kenworthy, Peter Newman: Jeffrey. 1989. *Cities and Automobile Dependence: A Sourcebook*. England: Ashgate Pub Co.

———. 1999c. *"Traffic Claming" from Sustainability and Cities: Overcoming Automobile Dependence*. Edited by Stephen Wheeler: Timothy Beatley, *The Sustainable Urban Development Reader*: Routledge.

Kim, Jae-on. 1978. *Factor Analysis : Statistical Methods and Practical Issues / Jae-on Kim, Charles W. Mueller*. Edited by Charles W. Mueller. Newbury Park, [Calif.],London: London : SAGE.

Kitamura, R.: Mokhtarian, P. L.: Laidet, L. 1997. "A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area." *Transportation* 24 (2): 125-158. doi: 10.1023/a:1017959825565.

Klinger, Thomas: Kenworthy, Jeffrey R.: Lanzendorf, Martin. 2013. "Dimensions of Urban Mobility Cultures – a Comparison of German Cities." *Journal of Transport Geography* 31: 18-29. doi: 10.1016/j.jtrangeo.2013.05.002.

Koushki, Parviz Amir. 1988. "The Effect of Socio-Economic Development on Household Travel Behavior in Saudi Arabia." *Socio-Economic Planning Sciences* 22 (3): 131-136. doi: 10.1016/0038-0121(88)90014-6.

Landgate. 2013. "Perth Central Business District." http://www.landgate.wa.gov.au/docvault.nsf/web/PS_FREEMAP/$FILE/Perth_CBD.pdf.

Leech, Nancy L. 2011. *Ibm Spss for Intermediate Statistics : Use and Interpretation / Nancy L. Leech, Karen C. Barrett, George A. Morgan*. Edited by George Arthur: Barrett Morgan, Karen Caplovitz, *Spss for Intermediate Statistics : Use and Interpretation*. London: London : Routledge Academic.

Liao, Felix Haifeng: Farber, Steven: Ewing, Reid. 2014. "Compact Development and Preference Heterogeneity in Residential Location Choice Behaviour: A Latent Class Analysis." *Urban Studies*. doi: 10.1177/0042098014527138.

Lim, Jin Young Park: Dong-Jun Kim: Yongtaek. 2008. "Use of Smart Card Data to Define Public Transit Use in Seoul, South Korea." *Transportation Research Record: Journal of the Transportation Research Board* 2063: 3-9. doi: http://dx.doi.org/10.3141/2063-01.

Limtanakool, Narisra: Dijst, Martin: Schwanen, Tim. 2006. "The Influence of Socioeconomic Characteristics, Land Use and Travel Time Considerations on Mode Choice for Medium- and Longer-Distance Trips." *Journal of Transport Geography* 14 (5): 327-341. http://www.sciencedirect.com/science/article/B6VG8-4GV9S8W-1/2/6110ece4ffbb1fba9e7ef6d3e62ba2cd

Lin, J.: Long, L. 2008. "What Neighborhood Are You In? Empirical Findings of Relationships between Household Travel and Neighborhood Characteristics." *Transportation* 35 (6): 739-758. <Go to ISI>://CCC:000259444400003.

Lin, Jen-Jia: Yang, An-Tsei. 2009. "Structural Analysis of How Urban Form Impacts Travel Demand: Evidence from Taipei." *Urban Studies* 46 (9): 1951-1967. doi: 10.1177/0042098009106017.

Louviere, Jordan J. 2000. *Stated Choice Methods : Analysis and Applications*. Edited by David A.: Swait Hensher, Joffre Dan. Cambridge (UK), New York (NY, USA): Cambridge University Press.

Love, Arnold J. 1991. *Evaluation Methods Sourcebook / Arnold J. Love, Editor*. Ottawa, Ontario: Canadian Evaluation Society.

Lowry, Ira S. 1964. "A Model of Metropolis." http://www.rand.org/pubs/research_memoranda/2006/RM4035.pdf.

Luk, James YK: Chung, Edward. 1997. *Induced Demand and Road Investment: An Initial Appraisal*.

Ma, Xiaolei: Wu, Yao-Jan: Wang, Yinhai: Chen, Feng: Liu, Jianfeng. 2013. "Mining Smart Card Data for Transit Riders' Travel Patterns." *Transportation Research Part C: Emerging Technologies* 36 (0): 1-12. doi: http://dx.doi.org/10.1016/j.trc.2013.07.010.

Mackett, R. 1990. "Exploratory Analysis of Long-Term Travel Demand and Policy Impacts Using Micro-Analytical Simulation." In *Developments in Dynamic and Activity-Based Approaches to Travel Demand (Oxford Studies in Transport)*, ed. Peter Jones, 384-405. Brookfiled: Gower Publishing Co.,.

Martinez, Francisco J. 2008. "Towards a Land-Use and Transport Interaction Framework." In *Handbooks in Transport Volume 1: Handbook of Transport Modelling*, ed. D A Hensher: K J BUTTON. Oxford, United Kingdom: Elsevier.

Mattsson, Kjell Jansson: Harald Lang: Dan. 2008. *Models for Public Transport Demand and Benefit Assessments*. Edited by David A.: Button Hensher, Kenneth John, *Handbook of Transport Modelling / Edited by David A. Hensher, Kenneth J. Button*. Amsterdam, London: Elsevier.

McFadden, D. 1997. *The Theory and Practice of Disaggregate Demand and Forecasting for Various Modes of Urban Transportation*, *Transport Economics : Selected Readings / Edited by Tae Hoon Oum ... [Et Al.]*. Amsterdam, The Netherlands: Amsterdam, The Netherlands : Harwood Academic Publishers in co-operation with the Korea Research Foundation for the 21st Century.

McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3 (4): 303-328. doi: 10.1016/0047-2727(74)90003-6.

McNally, Michael G. 2008. "The Four Step Model." In *Handbook of Transport Modeling*, ed.: Kenneth John Button David A. Hensher. Amsterdam ; London: Elsevier.

McNally, Michael: Kulkarni, Anup. 1997. "Assessment of Influence of Land Use-Transportation System on Travel Behavior." *Transportation Research Record: Journal of the Transportation Research Board* 1607 (-1): 105-115. doi: 10.3141/1607-15.

Mees, P. 2000a. *A Very Public Solution: Transport in the Dispersed City*: Melbourne University Press.

Mees, Paul. 2000b. *A Very Public Solution : Transport in the Dispersed City / Paul Mees*. Carlton South, Vic.: Carlton South, Vic. : Melbourne University Press.

Messenger, Todd: Ewing, Reid. 1996. "Transit-Oriented Development in the Sun Belt." *Transportation Research Record: Journal of the Transportation Research Board* 1552 (-1): 145-153. doi: 10.3141/1552-20.

Meyer, John: Straszheim, Mahlon 1997. *Transport Demand: The Basic Framework*, *Transport Economics : Selected Readings / Edited by Tae Hoon Oum ... [Et Al.]*. Amsterdam, The Netherlands: Amsterdam, The Netherlands : Harwood Academic Publishers in co-operation with the Korea Research Foundation for the 21st Century.

Mogridge, MJ. 1990. "Planning for Optimum Urban Efficiency: The Relationship between Congestion on the Roads and Public Transport." *Transportation Planning Systems* 1 (1).

Mokhtarian, Michael N. Bagley: Patricia L. 2002. "The Impact of Residential Neighborhood Type on Travel Behivour: A Structural Equations Modeling Approach." *The Annals of Regional Science* (36): 279-297. http://www.uctc.net/papers/607.pdf.

Morency, Catherine: Trépanier, Martin: Agard, Bruno. 2007. "Measuring Transit Use Variability with Smart-Card Data." *Transport Policy* 14 (3): 193-203. doi: http://dx.doi.org/10.1016/j.tranpol.2007.01.001.

Munizaga, Marcela A.: Palma, Carolina. 2012. "Estimation of a Disaggregate Multimodal Public Transport Origin–Destination Matrix from Passive Smartcard Data from Santiago, Chile." *Transportation Research Part C: Emerging Technologies* 24 (0): 9-18. doi: http://dx.doi.org/10.1016/j.trc.2012.01.007.

Murray, Alan T. 2001. "Strategic Analysis of Public Transport Coverage." *Socio-Economic Planning Sciences* 35 (3): 175-188. http://www.sciencedirect.com/science/article/B6V6Y-43MJR0C-2/2/6013ce10cfeaf45f9a01ad5be4d797ee

Myrdal, Gunnar. 1956. *An International Economy : Problems and Prospects*. [1st ed.]. ed. New York: New York : Harper.

Nazem, Mohsen: Trépanier, Martin: Morency, Catherine. 2013. "Integrated Intervening Opportunities Model for Public Transit Trip Generation-Distribution: A Supply-Dependent Approach." *Transportation Research Record: Journal of the Transportation Research Board* 2350 (-1): 47-57. doi: 10.3141/2350-06.

Neuman, William Lawrence. 2006. *Social Research Methods : Qualitative and Quantitative Approaches*. 6th ed. Boston, Mass.: Pearson/Allyn and Bacon.

Newman, Peter W. G.: Kenworthy, Jeffrey R. 1996. "The Land Use—Transport Connection: An Overview." *Land Use Policy* 13 (1): 1-22. doi: http://dx.doi.org/10.1016/0264-8377(95)00027-5.

Nijkamp, P.: Pepping, G. 1998. "Meta-Analysis for Explaining the Variance in Public Transport Demand Elasticities in Europe." *Journal of Transportation and Statistics* 1 (1): 1-14.

Nishiuchi, H.: King, J.: Todoroki, T. 2013. "Spatial-Temporal Daily Frequent Trip Pattern of Public Transport Passengers Using Smart Card Data." *International Journal of Intelligent Transportation Systems Research* 11 (1): 1-10. doi: 10.1007/s13177-012-0051-7.

Nolan, Susan A.: Heinzen, Thomas E. . 2007. *Statistics for the Behavioral Sciences* Edited by Thomas E. Heinzen. New York : [Basingstoke: New York : Worth Publishers.

Nurlaela, Siti: Curtis, Carey. 2012. "Modeling Household Residential Location Choice and Travel Behavior and Its Relationship with Public Transport Accessibility." *Procedia - Social and Behavioral Sciences* 54: 56-64. doi: 10.1016/j.sbspro.2012.09.725.

Olszewski, Piotr. 2007. "Singapore Motorisation Restraint and Its Implications on Travel Behaviour and Urban Sustainability." *Planning - Policy - Research - Practice* 34 (3): 319-335. doi: 10.1007/s11116-007-9115-y.

Oppenheim, Norbert. 1995. *Urban Travel Demand Modeling : From Individual Choices to General Equilibrium / Norbert Oppenheim*. New York: New York : Wiley.

Paez, A.: Trepanier, M.: Morency, C. 2011. "Geodemographic Analysis and the Identification of Potential Business Partnerships Enabled by Transit Smart Cards." *Transp. Res. Pt. A-Policy Pract.* 45 (7): 640-652. doi: 10.1016/j.tra.2011.04.002.

———. 2012. "Modeling Isoexposure to Transit Users for Market Potential Analysis." *Transp. Res. Pt. A-Policy Pract.* 46 (10): 1517-1527. doi: 10.1016/j.tra.2012.07.004.

Painter, Martin Terre Blanche: Kevin Durrheim: Desmond. 2008. *Research in Practice: Applied Methods for the Social Sciences*. Second Edition ed: UTC Press.

Paulley, N.: Balcombe, R.: Mackett, R.: Titheridge, H.: Preston, J. M.: Wardman, M. R.: Shires, J. D.: White, P. 2004. "Factors Affecting the Demand for Public Transport." http://www.abstracts.aetransport.org/paper/download/id/1944.

———. 2006. "The Demand for Public Transport: The Effects of Fares, Quality of Service, Income and Car Ownership." 295-306. ISSN 0967-070X.

Pawlak, Zdzisław: Skowron, Andrzej. 2007. "Rudiments of Rough Sets." *Information Sciences* 177 (1): 3-27. doi: 10.1016/j.ins.2006.06.003.

Pelletier, Marie-Pier: Trépanier, Martin: Morency, Catherine. 2011. "Smart Card Data Use in Public Transit: A Literature Review." *Transportation Research Part C: Emerging Technologies* 19 (4): 557-568. doi: 10.1016/j.trc.2010.12.003.

Pereira, F. C.: Rodrigues, F.: Polisciuc, E.: Ben-Akiva, M. 2015. "Why So Many People? Explaining Nonhabitual Transport Overcrowding with Internet Data." *Intelligent Transportation Systems, IEEE Transactions on* PP (99): 1-10. doi: 10.1109/TITS.2014.2368119.

Perkins, Carey Curtis: Tim. 2006. "Travel Behaviour: A Review of Recent Literature." http://urbanet.curtin.edu.au/local/pdf/ARC_TOD_Working_Paper_3.pdf.

Perl, Anthony: Calimente, John. 2011. "Environmental Reviews and Case Studies: Integrating High-Speed Rail into North America's Next Mobility Transition." *Environmental Practice* 13 (01): 47-57. doi: doi:10.1017/S1466046610000542.

Perl, Anthony: Pucher, John. 1995. "Transit in Trouble? The Policy Challenge Posed by Canada's Changing Urban Mobility." *Canadian Public Policy / Analyse de Politiques* 21 (3): 261-283. doi: 10.2307/3552088.

Perl, Richard Gilbert: Anthony. 2010. "The Post Carbon Reader : Managing the 21st Century's Sustainability Crises." In *Transportation in the Post-Carbon World*, ed. Richard: Lerch Heinberg, Daniel, 347 – 360. Healdsburg, Calif.: Healdsburg, Calif. : Watershed Media.

Pitombo, C. S.: Kawamoto, E.: Sousa, A. J. 2011. "An Exploratory Analysis of Relationships between Socioeconomic, Land Use, Activity Participation Variables and Travel Patterns." *Transport Policy* 18 (2): 347-357. doi: 10.1016/j.tranpol.2010.10.010.

Polat, C. 2012. "The Demand Determinants for Urban Public Transport Services: A Review of the Literature."

Pope, Jeff. 2005. *Transport Economics / Jeff Pope*. 1st edition.. ed. Guildford, WA: Guildford, WA, Vineyard Publishing.

Porta, Jan Scheurer: Carey Curtis: Sergio. June 2008. "Spatial Network Analysis of Multimodal Transport Systems: Developing a Strategic Planning Tool to Assess the Congruence of Movement and Urban Structure." http://www.sustainability.curtin.edu.au/local/docs/perth_snamuts_report.pdf.

Preston, John. 2001. "Integrating Transport with Socio-Economic Activity - a Research Agenda for the New Millennium." *Journal of Transport Geography* 9 (1): 13-24. http://www.sciencedirect.com/science/article/B6VG8-42991PW-2/2/7d279ddb16b71480c90c7f54339537ea

Prevedouros, Panagiotis Dimitrios. 1990. "Demographic, Social, Economic and Personality Factors Affecting Suburban Transport Behavior." Ph.D., Northwestern University, United States -- Illinois. http://proquest.umi.com.dbgw.lis.curtin.edu.au/pqdweb?did=745647261&Fmt=7&clientId=22212&RQT=309&VName=PQD

Pubic Transport Authority. n.d-a. Pulic Transport Authority: Annual Report 2008-2009- Transperth. Accessed 10 September 2014, http://www.pta.wa.gov.au/Portals/0/annualreports/2009/index.html.

———. n.d-b. Pulic Transport Authority: Annual Report 2009-2010- Transperth- Transperth Snapshop of 2009/2010. Accessed 10 September 2014, http://www.pta.wa.gov.au/Portals/0/annualreports/2010/transperth/transport-snapshot-of-200910/index.html.

Public Transport Authority. n.d. Metro Area Express. Accessed 6 February 2015, http://www.pta.wa.gov.au/max/AbouttheProject/tabid/700/language/en-AU/Default.aspx.

Pushkarev, B.: Zupan, J.M. 1976. *Urban Densities for Public Transportation*. New York, N.Y.: Tri-State Regional Planning Administration.

Raykov, Tenko. 2008. *An Introduction to Applied Multivariate Analysis / Tenko Raykov, George A. Marcoulides*. New York: New York : Routledge.

Reserve Bank of Australia. 2009. "Statement of Monetary Policy, 5 May." edited by Reserve Bank of Australia, Camberra.

Richardson, Ronny. 2011. *Business Applications of Multiple Regression*. New York: New York : Business Expert Press.

Riekko, Hans William. 2005. "Urban Form and Public Transit Use: A Study Examining the Relationship between Urban Form and Public Transit Use in Greater Toronto Area in 2001." M.PL., Queen's University at Kingston (Canada), Canada.

http://proquest.umi.com.dbgw.lis.curtin.edu.au/pqdweb?did=1130593051&Fmt=7&clientId=22212&RQT=309&VName=PQD

Rodier, Caroline J.: Johnston, Robert A.: Abraham, John E. 2002. "Heuristic Policy Analysis of Regional Land Use, Transit, and Travel Pricing Scenarios Using Two Urban Models." *Transportation Research Part D: Transport and Environment* 7 (4): 243-254. doi: http://dx.doi.org/10.1016/S1361-9209(01)00022-0.

Sappington, Jason Daniel. 2007. "The Influence of Urban Form, Socio-Demographics, and Transit-Oriented Development on Journey-to-Work Characteristis." Master Thesis, Graduate Faculty of the University of New Orleans, University of New Orleans. http://scholarworks.uno.edu/cgi/viewcontent.cgi?article=1643&context=td.

Schimek, Paul. 1996. "Household Motor Vehicle Ownership and Use: How Much Does Residential Density Matter?" *Transportation Research Record: Journal of the Transportation Research Board* 1552 (-1): 120-125. doi: 10.3141/1552-17.

Schwandt, T. A. 2007. *The Sage Dictionary of Qualitative Inquiry*: SAGE Publications, Inc.

Seaborn, Catherine: Attanucci, John: Wilson, Nigel H. 2009. "Using Smart Card Fare Payment Data to Analyse Multi-Modal Public Transport Journeys in London" *88th Annual Meeting of the Transportation Research Board*, *Washington*: http://hdl.handle.net/1721.1/69577.

Shannon, T.: Giles-Corti, B.: Pikora, T.: Bulsara, M.: Shilton, T.: Bull, F. 2006. "Active Commuting in a University Setting: Assessing Commuting Habits and Potential for Modal Change." *Transport Policy* 13 (3): 240-253. <Go to ISI>://CCC:000238010900006: http://www.elsevier.com

Shiftan, Y.: Outwater, M. L.: Zhou, Y. S. 2008a. "Transit Market Research Using Structural Equation Modeling and Attitudinal Market Segmentation." *Transport Policy* 15 (3): 186-195. doi: 10.1016/j.tranpol.2008.03.002.

Shiftan, Yoram. 2008b. "The Use of Activity-Based Modeling to Analyze the Effect of Land-Use Policies on Travel Behavior." *The Annals of Regional Science* 42 (1): 79-97. http://dx.doi.org/10.1007/s00168-007-0139-1.

Sipe, Jago Dodson: Neil. 2006. *Shocking the Suburbs: Urban Location, Housing Debt and Oil Vulnerability in the Australian City*. Griffith University Urban Research Program. http://www98.griffith.edu.au/dspace/bitstream/10072/12665/1/41353.pdf.

Sirkin, R. Mark. 2006. *Statistics for the Social Sciences*. 3rd ed. Thousand Oaks, Calif.: Sage Publications.

Sohn, J. 2005. "Are Commuting Patterns a Good Indicator of Urban Spatial Structure?" *Journal of Transport Geography* 13 (4): 306-317. doi: 10.1016/j.jtrangeo.2004.07.005.

Soltani, A.: Allan, A. 2006. "Analyzing the Impacts of Microscale Urban Attributes on Travel: Evidence from Suburban Adelaide, Australia." *Journal of Urban Planning and Development-Asce* 132 (3): 132-137. doi: 10.1061/(asce)0733-9488(2006)132:3(132).

Souche, S. 2010. "Measuring the Structural Determinants of Urban Travel Demand." *Transport Policy* 17 (3): 127-134. doi: 10.1016/j.tranpol.2009.12.003.

Stanley, Robert. 1998. "Continuing Examination of Successful Transit Ridership Initiatives." *TCRP Research Results Digest* (29).

Stead, Dominic. 2001a. "Relationships between Land Use, Socioeconomic Factors, and Travel Patterns in Britain." *Environment and Planning B: Planning and Design* 28 (4): 499-528. doi: 10.1068/b2677.

Stead, Dominic: Marshall, Stephen. 2001b. "The Relationships between Urban Form and Travel Patterns. An International Review and Evaluation." *European Journal of Transport and Infrastructure Research* 1 (2). http://www.ejtir.tudelft.nl/issues/2001_02/pdf/2001_02_01.pdf.

Susilo, Yusak: Maat, Kees. 2007. "The Influence of Built Environment to the Trends in Commuting Journeys in the Netherlands." *Transportation* 34 (5): 589-609. http://dx.doi.org/10.1007/s11116-007-9129-5.

Tabachnick, Barbara: Fidell, Linda. 2000. *Using Multivariate Statistics*. 4th edition ed: Boston, MA : Allyn and Bacon.

Taylor, Brian D.: Fink, Camille N. Y. 2003. "The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature."

Thompson, Gregory: Brown, Jeffrey: Bhattacharya, Torsha. 2012. "What Really Matters for Increasing Transit Ridership: Understanding the Determinants of Transit Ridership Demand in Broward County, Florida." *Urban Studies* 49 (15): 3327-3345. doi: 10.1177/0042098012443864.

Tolley, Rodney. 1996. "Green Campuses: Cutting the Environmental Cost of Commuting." *Journal of Transport Geography* 4 (3): 213-217. doi: 10.1016/0966-6923(96)00022-1.

Transperth. n.d-a. Transperth Fares. Accessed 16 October, http://www.transperth.wa.gov.au/Tickets-Fares/Fares.

———. n.d-b. Types of Smartrider Cards. Accessed 22 September 2014, http://www.transperth.wa.gov.au/SmartRider/Types-of-SmartRider.

———. n.d. Transperth Zone Map. Accessed 22 September, http://www.transperth.wa.gov.au/Tickets-Fares/Zone-Map.

Trépanier, Martin: Tranchant, Nicolas: Chapleau, Robert. 2007. "Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System." *Technology, Planning, and Operations* 11 (1): 1-14. doi: 10.1080/15472450601122256.

Utsunomiya, Mariko: Attanucci, John: Wilson, Nigel H. 2006. "Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning." *Transportation Research Record: Journal of the Transportation Research Board* 1971: 119-126. doi: http://dx.doi.org/10.3141/1971-16.

Vogt, W. Paul. 2005. *Dictionary of Statistics and Methodology : A Nontechnical Guide for the Social Sciences / W. Paul Vogt*. Thousand Oaks: Thousand Oaks : Sage Publications.

Vuchic, Vukan R. 2005. *Urbantransit : Operations, Planning, and Economics / Vukan R. Vuchic*. Hoboken, N.J.: Hoboken, N.J. : Wiley.

Waddell, Paul. 2002. "Urbansim: Modeling Urban Development for Land Use, Transportation, and Environmental Planning." *American Planning Association. Journal of the American Planning Association* 68 (3): 297. http://proquest.umi.com.dbgw.lis.curtin.edu.au/pqdweb?did=128776291&Fmt=7&clientId=22212&RQT=309&VName=PQD.

Wardman, M. 2000. "Using Existing Stated Preference Data to Analyse Bus Preferences (Working Paper 555)." http://eprints.whiterose.ac.uk/2070/1/ITS213_WP555_uploadable.pdf.

Weber, J. 2003. "Individual Accessibility and Distance from Major Employement Centers." *Journal of Geographical Analysis* 5: 51-70.

Webster, F. V.: Bly, P. H. 1982. "The Demand for Public Transport. Part Ii. Supply and Demand Factors of Public Transport." *A Transnational Transdisciplinary Journal* 2 (1): 23-46. doi: 10.1080/01441648208716480.

Western Australia Planning Commission. 2004. "Netowrk City: A Community Planning Strategy for Perth and Peel." Perth, Western Australia.

Western Australian Planning Commission. April 2005. "Transport Modelling: Review of Perth's Transport Models." edited by Department of Planning and Infrastructure, Perth: Department of Planning and Infrastructure.

White, Peter. 2009. *Public Transport*. Edited by Professor John Glasson. Fith Edition ed, *The Nature and Built Environment Series*. London and New York: Routledge- Taylor and Francis Group.

White, R Balcombe: R Mackett: N Paulley: J Preston: J Shires: H Titheridge: M Wardman: P. 2004. "The Demand for Public Transport: A Practical Guide." http://www.demandforpublictransport.co.uk/TRL593.pdf.

Willumsen, Juan de Dios Ortuzar: Luis G. 2011. *Modelling Transport* Edited by Luis G.: Wiley Willumsen, InterScience. Fourth Edition ed. Oxford: Oxford : Wiley-Blackwell.

Zeibots, Michelle: Petocz, Peter. 2005. "The Relationship between Increases in Motorway Capacity and Declines in Urban Rail Passenger Journeys: A Case Study of Sydney's M4 Motorway and Western Sydney Rail Lines'" *Transporting the future: Transport in a changing environment-28th Australasian Transport Research Forum*,

Zhang, Ming. 2004. "The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong." *Journal of the American Planning Association* 70 (3): 344-360. doi: 10.1080/01944360408976383.

———. 2005. "Exploring the Relationship between Urban Form and Nonwork Travel through Time Use Analysis." *Landscape and Urban Planning* 73 (2–3): 244-261. doi: 10.1016/j.landurbplan.2004.11.008.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Penelope Williams	1969	To determine the consequences of adopting low, standard fares.	Survey + analysis	London, UK	May-Aug 1965	400 regular car commuter interview	Demand for private transport		Average Journey - Travel distance	Cost Consideration		Therefore, either a reduction in fares, or an increase in motoring costs, should reduce traffic congestion, but a choice between these two systems would appear to rest as much on social and political considerations, as on economic considerations. The possible savings in congestion, accident and other costs are examined in conjunction with some suggestions on the sources of the necessary subsidies.
K. J.: Fowkes Button, A. S.: Pearman, A. D.	1980	To examine the influence exerted by local public transport service on urban car ownership levels in order to forecast future lvel of car ownership and to indicate whether policies of extensive public transport subsidy are likely to be effective as tools in transport planning.	Survey & Analysis (by median)	West Yorkshire, England, UK	Spring/Summer 1975	Extensive West Yorkshire Transportation Study - 7,812 households	Car ownership			Income	Public Transport Accessibility	The quality of local public transport does exert an influence both on average household car ownership levels and upon the decision of a house hold to become car owning. Households living in areas which have good access by public transport to work opportunities tend to both have a lower level of car ownership and to be less inclined to become car owning. It suggests the improvement of public transport provision by local transport agencies may be a useful tool in restricting car ownership growh.
Patrick S. McCarthy	1982	To re-examine the transferability of a pre BART model of work trip modal choice to a post the Bay Area RApid Transit (BART) environment in order to analyse the temporal characteristics of work trip behaviour by using before and after data sets associated with the BART impact travel study.	Mathematical method & survey	San Francisco Bay Area, USA	1) Nov. 1973 - Apr. 1974 2) 1975	1) Direct Survey - 1724 residents 2) Re-interview - 900 respondents	Travel time Travel cost weighted by the hourly wage Weighted parking cost Elasticity (own/cross travel mode)		Distance to work	Gender Marital status Race Driver license Income Age Household size		The general form and the coefficient estimates of a pre BART model are transferable in time. The model's predictive success and its implied elasticity measures are generally accuate based on the independent variables, relative to those implied by re-estimating the entire model on post BART data when updated to reflect BART's presence. Elasticity measures of the service related variables were found to increase over time as economic theory would predict.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
P. B.: Williams Goodwin, H. C. W. L.	1985	To report on and consider some recent discussions of British Studies of demand analysis which emerged over a one-day conference held in April 1984 at the University of Oxford.	Mathematical method + Analysis	South and West Yorkshire, UK	1980s		Fare Elasticity Service Elasticity		Urban Change	Car Ownership Income Unemployment	Fare Level of Service	It was widely recognized that the application of smtable methods and models in a range of planning contexts was limited by the availability of appropriate data sets, and the usual compromises of practice. Equally it was acknowledged that a wide range of empirical results was now available and that the latest evidence should be exploited in any comparative studies across bus operations. There was general support for improved publicity of the results of studies on demand analysis and disaggregate elasticities in particular. British bus industry faces a period of transition which will be characterized by a reduction in the subsidisation and cross subsidisation of services and a vulnerability of integrated networks.
Chris Hendrickson	1986	To examine long-term changes in the use of public transportation for commuting in the US between 1960 and 1980 based on census data for 25 metropolitan areas and a statistical analysis that are related to the number of employees in CBD in order to determine the relationship between transit commuting and CBD employment.	Census data	Pittsburgh, PA, U.S.A	1960, 1970 and 1980	Federal censuses	- Number of people commuting by public transport - Number of other travel	- CBD employment				Based on the data given, both the relative importance and absolute number of commuters by public transport were found to be declining although the service and amount of public transport had evolved. Also, the formula suggested transit use for work travel is strongly related to CBD employment totals, but not to overall metropolitan area size. Addtionally, fixed rail transit is not likely to be sufficient to overcome the adverse trend noted previously.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Robert Cervero	1988	This article examines the potential mobility benefits of developing mixed-use suburban workplaces, ones where offices, shops, banks, restaurants, and other activities are built side-by-side	Survey + analysis	USA	1987, 1983	1) Warner Centre office in LA 2) Orange County area	1)work trips by drive-alone mode 2) drive-alone work trip minus regional drive-alone 3) work trips by vanpool or carpool 4) work trip by walking or cycling	1) total floorspace in office use 2) total floorspace in retail square footage 3) num of company vans 4) employees per on-site company 5) employees per freeway interchange 6) total floorspace size of full-time work force 7) employees per freeway interchange 8) rideshare coordinator				Overall, the findings of this research confirm the hypotheses set forth regarding the affects of mixed-use environments on commuting. Single-use office settings seem to induce solo commuting, whereas work environments that are more varied generally encourage more ridesharing, walking, and cycling. Particularly important to ridesharing is the availability of consumerre tail services. Whilet he synehronizati0n Of job and housing growth around suburban centers could be expected to encourage more foot and bicycle travel, at the same time, ridesharing and vehicle occupancyle vels could be expected to fall off some. The benefits of jobs-housing balancing, therefore, relate more to the shortening of vehicular trips and the easing of local through-traffic conflicts than to inducing people to walk or cycle to work.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Robert Cervero	1989	To examine whether the widening separation of suburban workplaces and the residences of suburban workers are caused by fiscal and exclusioary zoning that result in an undersupply of housing; rents and housing costs that price many service workers out of the local residential market; and several demographic trends.	Mathematical method	Chicago/San Francisco, USA	1980	40 Major suburban employment centers regarding commuting (journey-to-work data)	- Travel mode - Freeway traffic condition - Zoning and Tax incentive	- Num of employed residents - Num of employees		- Occupations - Housecost - Vehicles -Travel time - Housing provision		Jobs-housing imbalances seem to be a root cause of many problems plaguing America's metropolises. Restrictive zoning and inaffordable housing created a gap between where people live and where they work. The overall composition of land uses and the match-up of job and housing grwth at the subregional level are more important in achieving balance than numerical parity of workers and households within jurisdictions. The balancing of job and housing growth could improve regional mobility as any mix of traffic management or roadway expansion programs. Planners must use land development as a lever to improve mobility while suburban landscape is being rapidly transformed to ensure that sufficient affordable housing is being provided near job centres.

Appendix 1: Literature Review Summary

Author	**Year**	**Research Question/Objective**	**Statistical Method**	**Studied Region/Country/Cities**	**Dataset time**	**Datasource**	**Transport Use Variables**	**Land Use Characteristics**	**Urban Form**	**Socio Economic**	**Service Provision**	**Findings**
R. Cervero	1991	To build upon past research by studying the relationship between land use and various indicators of travel demand for a number of office buildings at six different suburban activity centers across the United States.	Survey + analysis	6 suburban centers in US	1989	NCHRP report	1) Trip generation rates 2) Work-trips made by private automobile 3) Work-trips made by mass transit mode 4) Work trips made by walking 5) Travel demand elasticities	1) Available parking spaces per employee Degree of mixed land-use Type of tenant 2) Available parking spaces per employee Degree of mixed land-use 3) Number of parking spaces Degree of mixed land-use Number of stories of office building Auto occupancy level, number of persons per vehicle during AM peak. 4) Degree of mixed land-use work-trip made				The land-use environments of contemporary suburban workplaces appear to have a modest to lnoderate influence on commuting behavior. The absence of strong statistical relationships perhaps reflect the absence of truly dense, mixed-use work settings in America's suburbans more than anything. Given enough variation from which to measure how land-use mixing, density, and levels of parking attiect mode clmice and occupancy levels, better fitting models could have lmssibly been produced.

Appendix 1: Literature Review Summary

Author	**Year**	**Research Question/Objective**	**Statistical Method**	**Studied Region/Country/Cities**	**Dataset time**	**Datasource**	**Transport Use Variables**	**Land Use Characteristics**	**Urban Form**	**Socio Economic**	**Service Provision**	**Findings**
James: Adibi Doti, Esmael	1991	To develop a model that explains public transit ridership in Orange County, California over quarterly periods during the 1973:4-1989:1 period in order to forecast the number of passengers from 1989:2 to 1993:4	Mathematical method (Double-logarithmic expression and comparison with each year)	Orange County, Califonia, U.S.A	Apr.1973 - Jan.1989	Public transit ridership	- Transport usage - Public transit vehicle service miles - Average fares	- Population		- Income - Total wage and salary employment		The relatively simple model can be used for identifying and measuring the causal relationship between ridership and other factors. The simplified formular is to remove auto-corelation and to reduce the likelihood of multicollinearity. The model also can be used to generate forecasts/stimulate what-if scenarios depending on variables According to the formular, the potential number of users are postulated based on the external factors. Population/Income reflect the market size, and total wage and salary employement can be used to compare with the ridership based on the large and regular sampling. Depending on usage, service miles and fares could be adjusted.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Genevieve: Small Giuliano, Kenneth A.	1993	To examine the presumption that local imbalances between employment and residential sites strongly influence people's commuting patterns by finding the commuting pattern for the Los Angeles region in 1980 which would minimise average commuting time or distance, given the actual spatial distributions of job and housing locations.	Analysis - about relevant studies	Los Angeles, USA	1980	Journey-to-work information for 1146 zones	Excess commuting (Commuting Time)	Travel Distance Travel Time Employment		Income Living condition (own/rent)		Commuting distance and time are not very sensitive to variations in urban structure, and are far in excess of what can be explained by jobs-housing imbalances, even when occupational mismatches are accounted for. The main exception is that the extreme imbalance of the downtown Los Angeles employment centre does increase commuting times. The behavioural assumption of cost minimisation in the standard model is inadequate to explain commuting, and that large-scale changes in urban structure designed to promote jobs-housing balance would have only small effects on commuting. Attempts to alter the metropolitan-wide structure of urban land use via policy intervention are likely to have disappointing impacts on commuting patterns, even if successful in changing the degree of jobs-housing balance.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
L. D.: Pivo Frank, Gary	1994	To examine the impacts of land-use mix, population density, and employment density on the use of the single-occupant vehicle, transit, and walking for both work trips and shipping trips. Focusing on 1) whether there is a relationship between urban form and modal choice, 2) whether the relationship exists when controlling for non-urban form factors, 3) whether the relationship is linear or nonlinear, 4) whether a stronger relationship exist between modal choice and urban form when they are measured at both trip ends as opposed to either the origin or the destination	Mathematical method & analysis	Puget, Washington, USA	1989	- Puget Sound transportation panel -U.S. Census - Washington state department of economic security - Puget sound regional council - King County BALD file	Propotion of work trip by - single occupant vehicle - transit - walking	Population - Density Employment - Density Land-use Mix		Lifecycle stage Driver license		-Employment density and land-use mix were both related to percent single occupant vehicle (SOV) use, transit use, and walking whereas population density was not significantly related to percent SOV use. -Urban form (density of population, employment, and land-mix use) and mode choice are significantly related. -Urban form is significantly related to mode choice for SOV use, transit use, and walking when non-urban-form factors (census) are controlled by the significance of both urban-form and non-urban-form variables. -Average urban-form measures rather than measures taken at the origin or destination have the strongest ability to predict variations in mode choice. -Policies are suggested to encourage population densities to increase to levels below 13 persons per acre will have little effect on mode choice. -Measuring urban form at both trip ends provides a greater ability to predict travel choices than looking at trip ends separately.
Peter: Kenworthy Newman, Jeffrey: Vintila, Peter	1995	To examine three strands of fundamental opposition to physical planning and the cynicism about government involvemnet in making cities less automobile dependent.	Analysis & Debate (3 approaches)	Many Cities in Australia, USA, Canada European /Asian	1) 1991 2) 1992 3) 1992 4) 1992	1) USA GNP 2) Pucher and Clorer 3) Housing and location choice survey 4) Australian House of Representatives Committee strategy report	Car / Transit use			GNP Fuel price		By employing physical planning and economic instrument togeter, Urban reform programmes to reduce automobile dependence have been set up to show that alternatives can be created through physical planning instruments.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Susan Handy	1996	This paper reviews alternative approaches for exploring the link between urban form and travel behavior, outlines issues and complexities that this research must address	Analysis	Austin, TX, USA								In addition to reconsidering aooroaches to research on the links between urban form and travel behavior, it is important to reconsider the overall policy goal toward which much of this research is directed namely, the goal of reducing automobile travel. In given the extent of existing development and the relatively small increment that new growth represents. Certainly, it is important that any development that occurs be designed appropriately so as to minimize the need for automobile travel, but other strategies to manage transportation demand, such as pricing strategies, are also needed if communities hope to reduce automobile use in the short run.

Author	**Year**	**Research Question/Objective**	**Statistical Method**	**Studied Region/Country/Cities**	**Dataset time**	**Datasource**	**Transport Use Variables**	**Land Use Characteristics**	**Urban Form**	**Socio Economic**	**Service Provision**	**Findings**
Robert Cervero	1996	To explore the question that mixed land-uses encourage non-auto commuting from past research by investigating how the presence of retail activities in neighborhoods influences the commuting choices of residents	Survey + analysis	Boston Lawrence Lowell, MA-NH Dallas, TX Detroit, MI Los Angeles Long Beach, CA Fort Worth Arlington, TX Minneapolis- St. Paul, MN Philadelphia, PA-NJ Phoenix, AZ San Francisco Oakland, CA	1985	American Housing Survey	Probability of commuting by automobile Probability of commuting by public transit probability of commuting by walking or bicycling no.of automobile in the house hold distance from home to work	Land-use variables - overall neighbourhood density control variables - no of private automobile (persons) - annual income - four-lane highway, railroad, airport - public transportation - distance from home to work				In summary, this research found a fair amount of elasticity between land-use environments and commuting choices in 11 large U.S. MSAs. One public policy implication of these findings would be to encourage denser, mixed-use development, at least in those areas that are well-served by public transit, where there are reasonable options for walking and bicycling to work, and where non-auto commuting is an explicit policy selective (such as for air quality attainment purposes). Infill development and reurbanization of traditional centers represent one approach to creating viable mixed-use centers. Encouraging nεw mixed-use suburban enclaves and edge cities, interlinked by efficient mass transit services, might be another. The possible policy mechanisms for bringing about such changes in the built environment are numerous, rangmg 1rom cooramatea reg10na1 p1anntng 01 iransponadon ano iano-use“ to congestion pricing and parking cash-

Appendix 1: Literature Review Summary

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Robert: Kockelman Cervero, Kara	1997	The built environment is thought to influence travel demand along three principal dimensions —density, diversity, and design. This paper tests this proposition by examining how the '3Ds' affect trip rates and mode choice of residents in the San Francisco Bay Area.	Factor analysis, Regression models		1990	1990 travel diary data and land-use records obtained from the U.S. census, regional inventories, and field surveys	Travel Demand		Distance • Euclidean distance between centroids of trip's origin and destination traffic analysis zones • Euclidean distance: to downtown San Francisco; downtown Oakland; downtown San Jose.	Socio-demographics of trip-maker > Age > Gender: male status > Employment: full-time or part-time status; professional occupation > Race and ethnicity: racial-ethnic category; Caucasian status > Possession of driver's license 2. Household of trip-maker > Size: number of members; number of	Transportation supply and services • Transit service intensity: route miles of peak-hour revenue service divided by developed area of tract • Distance to: nearest freeway-on ramp; nearest BART station; nearest commuter rail station; nearest light rail station; and nearest ferry landing • Proportion of commercial-retail parcels	> Population density, land-use diversity, and pedestrian-oriented design generally contribute to decreases in trip rates and increases in non-private car travel. > Strong and positive associations between compact development and non-personal vehicle mode choices for personal business trips and non-work trips. > Having accessible retail activities within neighbourhood is closely associated with transport mode choice for work trips. > Elasticities between each built environment dimension and travel demand were statistically moderate. > Intensity factor (a combination of retail store density, activity centre density, retail intensity, walking accessibility, park intensity, and population density) had a fairly marginal impact on travel demand in the San Francisco Bay Area. > Intensity factor has positive association with all non-work trips, personal business trips and work trips
Marlon G.: Sarmiento Boarnet, Sharon	1998	To examine the link between land-use patterns at the neighbourhood level and non-work trip generation for a sample of 769 individuals with a regression analysis of non-work trip frequencies	Mathematical method & analysis	Los Angeles, CA, USA	1) 1993 2) 1990 and 1994	1) Southern California travel diary data 2) Census and the Southern California Association of Gov	- travel behaviour	- Population density - Retail employment - Service employement		- Gender - Race - Education - Income - Num of children - Num of cars - Num of workers in household - Work day		The results suggest that choices about how to measure the variables and how to specify the regressions can influence the conclusions from these studies in potentially important ways. Although the paper gives no evidence for the link, some of the results at least suggest the possibility that the presence or absence of strong evidence might be due in part to choices about geographical scale and regression modelling techniques. Hence, the possibility that land-use and urban design might influence non-work travel is important enough, but it is still not enough understood to inform policy.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Randall: Crepeau Crane, Richard	1998	This paper provides the first direct tests of these hypotheses within a consistent behavioral framework. An analysis of household travel diary and GIS data for San Diego, California, finds little role for land use in explaining travel behavior, and no evidence that the street network pattern affects either short or long non-work travel decisions.	Mathematical methods + Analysis	San Diego, USA	1986	US Census	Daily car trip frequency	Land use variables >Connected street pattern >Mixed street pattern >Street network density (heavy) >Residential share of census tract >Commercial share of tract >Vacant share of census tract	>Distance to downtown	Household resource and taste variables >Income less than $20,000 >Income is $20,000 to $40,000 >Mean household age >Number of children (kids) >Household size >Housing tenure >Housing type	Price	From a research perspective the big question at hand is how urban form aects travel behavior. The most straightforward answer is that we know quite little; most research has estimated ad hoc models, avoiding the kind of systematic behavioral analysis that would allow us to compare dif- ferent situations with dierent features in a generalizable manner. The empirical model is thus based on the idea that trip behavior can be explained as a function of trip costs and bene®ts, which in turn are the product of trip lengths, modes, purposes, and individual characteristics. While results may vary in other areas, the empirical argument for using land use as an element of regional air quality or other environmental plans remains to be demonstrated.
Gregory K. Ingram	1998	To provide an insight into some summary findings of a large comparative study of automobile dependence in cities, whose ongoing goal is to at least partly remedy this obvious lack of comparative urban transportation data.	Mathematical method & analysis	46 cities in the USA, Australia, Canada, Western Europe and Asia	1990/91	Various data items from the sample of cities in each distinct geographic region - World Bank / Kenworthy et al.(1999)	- Urban Density - Trip time - Trip distance - Percentage of workers using transit/bicycle/foot - Transit operating cost recovery - Road expenditure - Total cost of cars	- Population - Urbanised land area (city size)		- Car use per capita - Gross regional product per capita - Economic variables	- Transit service and use per capita	This study compared variables with each other to find what affects the most automobile dependence. The international comparison suggests increasing automobile dependence and declining transit and non-motorised mode use in cities are not inevitable, but they appear to be responsive to public policy which seeks to minimise such trends through effective land use planning, transportation infrastructure and service delivery policies directed more towards non-auto modes and through economic policies setting higher charges for auto ownership and use.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Satoshi: Garling Fujii, Tommy: Kitamura, Ryuichi	2001	To investigate how coopration can be facilitated in the real world dilemma of choosing to commute by automobile rather than by public transport.	Survey	Osaka, Japan	1998	Mail survey - 335 drivers	-Relative frequency of commuting by public transport -Expected and Actual commute times by public transport -Expected commute times for first and second commute by public transport -Frequency of continued choice of public transport related to difference between expected and actual commute times					A temporary structural change may be an important catalyst that triggers cooperation in a social dilemma. The relative frequency of automobile commuting before the closure was inversely related to the increase of public transport. The expected commute time by public transport was overestimated by automobile commuters, and to a higher degree when driving frequency was higher. Overestimates of commute times by public transport were corrected after the drivers' first public transport use during the closure. Thoese whose overestimates were corrected continued to use public transport to a larger extent than those not corrected. This correction may in turn increase public transport use.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Robert Cervero	2002	To overcome some of the deficiencies of past mode-choice analyses, particularly those that have focused on the effects of built environments, through an expanded specification of mode-choice utility by using a normative model that weighs the influences of not only three core dimensions of built environment - density, diversity, and design, but factors related to generalised cost and socio-economic attributes of travelers.	Mathematical method & analysis	Berkeley, CA, USA	1994	Work Trips - montgomery county	Comparative modal attributes: >Total travel time differential: transit , drive-alone >Total travel time differential: transit , group-ride >Total travel cost differential: transit , drive-alone >Total travel cost differential: transit , group-ride			Trip-maker attributes: >Vehicle ownership: number of auto- mobiles in household >Gender: 0 ¼ Male, 1 ¼ Female (specific to drive-alone) >Gender: 0 ¼ Male, 1 ¼ Female (specific to group-ride) >Driver's license: 0 ¼ No, 1 ¼ Yes (specific to drive-alone) >Driver's license: 0 ¼ No, 1 ¼ Yes (specific to group-ride)	Proportion of multi-family households in origin TAZ within one-half mile of metrorail station	The results argue for the explicit inclusion of land-use variables in the utility expressions of mode choice model in urbanised settings. They also reveal the importance of including economic attributes of competing modes, notably travel time and price variables, in the specification of models that test the influences of land-use factors on travel demand. A logical extension would be to account for the possible influences of self-selection on mode choice as a form of testing the influences of three blocks of variables - traditional travel time/cost and demographic variables, attitudinal and lifestyle preference variables, and built-environment factors.

Author	**Year**	**Research Question/Objective**	**Statistical Method**	**Studied Region/Country/Cities**	**Dataset time**	**Datasource**	**Transport Use Variables**	**Land Use Characteristics**	**Urban Form**	**Socio Economic**	**Service Provision**	**Findings**
Thomas W. Sanchez	2002	This article presents a wage inequality analysis for 158 large US metropolitan statistical areas (MSAs). The analysis is concerned with whether public transport has a detectable influence on 1990 levels of wage equality.	Mathematical method + Analysis	USA	1989-1991	Various	> Proportion of persons using transit for work commute	> Geographical size of MSA (square miles) > Number of jobs per 100 population aged 15–64 years Male to female ratio (aged 15–64 years) > Proportion of persons employed in agriculture > Proportion of persons employed in nance, insurance or real estate > Ratio of the proportion of MSA jobs in CC to proportion of MSA population		> Serious crimes reported per capita > Full-time-equivalent workers Proportion of female-headed households with children > Proportion of female-headed households with children > Home price index (percentage of mean from all MSAs) > Median household income > Median household income (predicted)	> Transit supply/density (directional miles per 100 square miles)"	This research focused on incorporating a public policy variable to predict levels of metropolitan wage inequality. The research hypothesis tested whether mobility increases from public transport in uenced the wage distribution of metropolitan areas. The results of three separate regression models suggest that social and demographic, economic and spatial characteristics are signi cant determinants of wage inequality. The public policy variable tested, public transport supply, was also a signi cant factor and had detectable effects on wage inequality across large US MSAs.
S. Cullinane	2002	To assess whether the provision of good, cheap public transport can discourage the purchase of, or desire to purchase, a car by using the example of Hong Kong where GDP is high but public transport sill dominates.	Survey - Interview	Hong Kong	2000	Face-to-face questionnaire survey - 389 Uni students	- Travel mode - Frequency - Reason for using transport - Concern about important issues	Attitude towards car and trip, Car ownership				If public transport is perceived to be both good and cheap, it can suppress the demand for cars whereas it may have little impact on mode choice if public transport is perceived to be poor since the measures are not sufficient in scale to have an impact on the choice decision of individuals. Although HK public transport is profitable and runs without subsidies; however, it is not the case for most Western countires. Generally as GDP goes up, the car demand increases; however, HK is not.

Appendix 1: Literature Review Summary

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Sharon: Cullinane Cullinane, Kevin	2003	1) To determine why people own cars in Hong Kong 2) To explore how dependent car owners are on their cars 3) To identify the policy implications	Survey - Interview	Hong Kong	2001, 2002, 2003	Direct Survey - 401 Car owners	- Anuual Mileage - Reasons of car ownership - Trip purpose	Attitude towards car and trip, Car ownership				To achieve greater sustainability, car ownership and use must be controlled since people become dependent on their car for all journey purposes once a car has been acquired to carry things and to save time. It suggests if controlling car use is an objective of transport policy, the best is not to improve congestion. Although the existence of good public transport can deter car ownership, but to deter car use, car ownership needs to be targeted. HK is still low car-dependent; hence, there is the opportunity to prevent car ownership from becoming the norm but the car ownership is already high.
Georges: Dargay Bresson, Joyce: Madre, Jean-Loup: Pirotte, Alain	2004	To introduce a panel data analysis of annual time series from 1975 to 1995 for 62 urban areas in France To investigate the implications for the estimated price, income and service elasticities of extending the observation period and including the structural variables.	Mathematical method & analysis	62 urban areas, France	1975 - 1995	Natitonal statistical organisation - CERTU/INSEE: 62 public transport areas - bus services / other public transport modes Price and income in real terms: 1995 prices	1) Elasticities 2) Public transport travel coefficient			1) Income Price (fare) Vehicle km 2) Num of trips Num of cars Home location Age groups		The estimates based on different time pedirods indicate that the elasticities are not stable over time. The Bayesian approach is an improvement compared to mere conventional fixed-coefficients specifications; hence, there is a considerable variation in the elasticities among areas. The variation of elasticities over time is examined by the comparison between log-log and semi-log specifications for the fare elasticity. The competition with automobile should be considered. = The downward trend in public transport patronage is mainly due to increasing car ownership, and that this effect will be less important over time since the growth of the car stock is decelerating. = The use of public transport is sensitive to the volume supplied and to its price making the financial equilibrium of this industry problematic. = The combination of demographic and economic approaches proves useful

Appendix 1: Literature Review Summary

Author	**Year**	**Research Question/Objective**	**Statistical Method**	**Studied Region/Country/Cities**	**Dataset time**	**Datasource**	**Transport Use Variables**	**Land Use Characteristics**	**Urban Form**	**Socio Economic**	**Service Provision**	**Findings**
Ming Zhang	2004	To analyse the influence of land use on travel mode choice	Choice Modelling	Boston, Hong Kong	1991, 1992	1991 Household Survery, Boston 1992 Travel Characteristic Survey, Hong Kong		-Land use and location -Distance to nearest transit station at origin (1000 ft.) -TAZ population density at origin (persons/acre) -TAZ job density at origin Gobs/acre) -TAZ population density at destination (persons/acre) -TAZ job density at destination (jobs/acre) -Connectivity: % non-cul-de-sac		-Age: <30 (walk or bike) -Full-time worker (drive alone) -Home owner (drive alone) -Female no children (transit) -Vehicles per worker (drive alone/ shared ride)		-Effects of land use attributes vary depending on the trip purpose (i.e. work related or non-work related) and whether public transport use is measured at travel origins or destinations. -Statistically significant relationship between train usage and population densities at both trip origins and destinations, -Demand for public transportation for non-work related trips in Boston is not significantly correlated with population density at the trip origin, or employment density at destinations. -It is critical to consider the composite effect of changes in land use characteristics on mode of transportation choices.

Appendix 1: Literature Review Summary

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Ming Zhang	2005	To expand an activity-based time-use analysis of the relationship between urban form and nonwork travel that is valuable to travel demand analysis and has important implications to the practice of environmental design, urban planning, and land development.	Survey & Mathematical method	Boston, USA	1991	Activity-Travel Survey in Boston	> Travel time > Activity-travel purpose > Activity-travel frequency	Employment Num of households	Accessibility to establishments	Gender Age Driver license Income Accessibility to establishments Num of Vehicles Time use Time share		Varying effects of modifying spatial accessibility on nonwork activity participation and travel among different activity categories. When accessibility to schools improved, children and adults were found to pay more visits to and spent more times in schools without generating additional total school travel. The increase in social travel was due to more frequent trip making resulting from higher accessibility, and the increase in travel time and frequency may lead to more undesirable transportation and environmental consequences. So the potential positive benefits of accessibility enhancement may be offset from a societal perspective. In order to reduce the offset, encouraging modal shift from driving to nondriving modes is suggested.

Appendix 1: Literature Review Summary

Author	**Year**	**Research Question/Objective**	**Statistical Method**	**Studied Region/Country/Cities**	**Dataset time**	**Datasource**	**Transport Use Variables**	**Land Use Characteristics**	**Urban Form**	**Socio Economic**	**Service Provision**	**Findings**
Jan: Burghouwt Ritsema van Eck, Guillaume: Dijst, Martin	2005	To compare a number of local spatial configurations of land use and transport facilities in a Dutch new town to address the question what impact these configurations have on the quality of life of different population categories.	Survey + analysis	Zoetermeer, Dutch	1997	Trip diary	1) feasibility of activity programs 2) travel time efficiency travel distance efficiency	1) Concentration dispersion density car-free heighbourhood 2) average car distance car travel time per journey				urban planners and policymakers should supplement their environmental goals with social goals. Spatial and transport policies aimed at changing modal splits in residential areas, should not only assess environmental gains but also the opportunities individuals are offered to participate in their prefeπed activities within their daily time budgets. We realize, that the different spatial configurations we have analysed cannot be realized overnight at any given location. Obviously, the best opportunities for implementing a desirable spatial configuiration occur in the design phase of a new residential area. However, existing built-up areas also offer opportunities to improve the mobility situation, for instance by a revision of parking policies, the provision of tailor-made trans- port services, the coordination of the location of schools, sport facihtles, etc. In all cases, it is clear that spatial policy can gain from a more

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Chandra R. Bhat	2005	To develop simple and powerful closed-form model for multiple discreteness by referring the econometric model - multiple discrete-continuous extreme model (MDCEV) and the multinominal logit (MNL) in order to analyse time-use allocation decisions among a variety of discretionary activities on weekends using data from the 2000 San Francisco Bay Area survey	Mathematical method & Survey (The collected data through survey are applied to Mathematical methods - MDCEV/MNL)	Austin, Texas, U.S.A	2000	the 2000 San Francisco Bay Area Travel Survey (BATS) - 15,000 households in the Bay Area for 2 day period	- Rate of satisfaction - Time invested	- Household location - Individual demographics - Employment characteristics		- Activity purposes - Total time available - Household demographics		The suggested models is used for multiple discreteness in demand that is derived from utility maximisation theory with diminishing marginal returns. Individuals in households with several other adults and in households with low incomes have a high propensity to participate in in-home recreation over the weekend whereas individuals in households with children with medium incomes, and with bicycles prefer out-of-home leisure activities... (omitted) The model can be used to assess the impacts of changing demographics and employment patterns on time-use patterns using the prediction process (in Sec 5.4.). Also, the predicted changes in time use patterns can then be used to examine the implied travel changes.
J. J.: Yang Lin, A. T.	2006	To analyse how the major concepts in the compact-city paradigm affects sustainable development - urban sustainability - by applying 'structural equation modeling' to 92 samples.	Mathematical method & analysis	92 medium/small-sized cities, Taiwan	2001	1) Statistical abstracts - local governments 2) Urban and regional development statistics 3) Population and Housing Census 4) Commerce and Service Census 5) urban land prices 6) Household finances investigation 7) DLA GIS data	- Compactness of city - Urban sustainability	- Density - Mix of land uses - Intensification: change of density - Environment: pollution		- Economy - Society: public service/crime/housing affordability		The influences of compact-city concepts on sustainability are both positive and negative. A high-density pattern and intensification negatively affect the environmental and social sustainability, but positively affect economic sustainability. Mixture of land use is beneficial to econoic sustainability, and does not significantly affect environmental or social sustainability. Complementary strategies that guide and promote the use of the compact-city paradigm toward the goals of sustainability are proposed to mitigate the negative influences of density on environmental and social sustainability and enhance the positive influence of the economy on environmental sustainability.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Stefanie: Grotz Fobker, Reinhold	2006	To investigate which living conditions best meet the needs of elderly people based on results from the FRAME(2002) research project examining the everyday mobility of persons aged 60 and older in different urban settings.	Survey	Bonn, Germany	1) 2002 2) 2004 3) 1998	1) FRAME 2) Bundesstadt Bonn Statistik stelle 3) Geospace	Shopping distance Short trips	Density		-Occupation -Welfare recipients -Living space -Subsidised housing -Criminal charges -Doctors -Cars -Age -Driving licence -Income		According to the variables, the behaviour is reflected by a reduction in activities and a shrinking of the activity space. Some advantages of life in the old urban development become apparent. A high share of short leisure trips is observed in this area facilitating a more intense use of slow modes since the city centre is accessible on foot. An active lifestyle is possible in all the study areas, but the preconditions for withdrawal into one's own neighbourhood vary between different residential areas. The importance of basic supply resources and leisure facilities within the residential environment and a preferably rail-bound public transport connection to the city centre need to be pointed out. Transport authorities and local authorities can support social activities through special offers and programmes for elderly people.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
H. Knoflacher	2006	To review the failure of conventional transport policies to address the many problems caused by private care use in cities in advanced nations through human behaviours. To suggest that restructuring parking provision can address the problems with the solution that strong incentives are provided for cars to be parked in garages that are only as accessible as public transport stops.	Non statistical (Dogma - Proof Based)	Vienna, Austria	1) 1955 - 1982 2) 1970s / 1990s	1) the development of the individual speed of Austrians 2) the travel survey data from various nations	- Number of pedestrians - Number of cars - Journey time - Body energy consumption depending on transport	- Travel time budgets				Changing in parking management could be the solution to reduce travel time and road congestion. The suggested transport policy aims to the prioritisation of public transport with the proposition that the population has cars parked in centralised garages that are only as accessible as public transport stops with strong incentives. Once provided, it would generate better environmental conditions, more flexibility and opportunities for nearby jobs, and recreational activities and social contacts. Regarding the incentives, the person who parks in the garage, would have to pay less, and the charge should be related to public transport fares and the financial and operational costs of the garage operators.
Birgitta: Uzzell Gatersleben, David	2007	To explore affective experiences that are limited to commuter stress in order to aim to persuade people to abandon their car for sustainable transport policy initiatives	Mathematical method & Survey (Questionnaire)	Guildford, England, UK	2000	389 University employees	-Travel time -Travel mode -Travel distance					Commuter journeys by car and public transport can be stressful because of traffic volume, the behaviour of other road users, and poor infrastructure provision. Boredom is another factor needed to be taken into account. Walking and cycling score positively on arousal as well as pleasure although taking the longest time so they seem an optimum form of travel from an affective perspective. Although other studies often ignore the use of bicycles and walking in focusing on utilitarian and cognitive evaluations, the car may not alway be the most optimum travel mode to use from an affective perspective.

Author	**Year**	**Research Question/Objective**	**Statistical Method**	**Studied Region/Country/Cities**	**Dataset time**	**Datasource**	**Transport Use Variables**	**Land Use Characteristics**	**Urban Form**	**Socio Economic**	**Service Provision**	**Findings**
Mizuki: Shen Kawabata, Qing	2007	To examine commuting inequality between cars and public transit and its connection with urban spatial structure in the San Francisco Bay Area to address two lines of questions.	Mathematical method & analysis	San Francisco, CA, USA	1990, 2000	Census Data	-Travel Mode -Travel Time	- Workers and Jobs - Population			Accessibility	While job accessibility for car users decreased, job accessibility for public transit users increased in most zones. Commuting time by drinving alone lengthened in most zones, whereas commuting time by public transit shortened in a number of zones. Job accessibility was inversely associatied with commuting time for driving along and for public transit. Also, the degree of the association was greater for public transit than for driving alone. Moreover, the inverse association strengthened for driving along but weakened for public transit.
Chandra R.: Guo Bhat, Jessica Y.	2007	To examine the presence of 'true' causality versus residential sorting-based 'spurious' associations in the land-use transportation connection. To develop a methodological formulation to control for residential sorting effects in the analysis of the effect of built environment attributes on travel behaviour-related choices.	Mathematical method	San Francisco, CA, USA	2000	1) San Francisco Bay Area Travel Survey (BATS) 2) Census 3) Metropolitan Transportation Commission in San Francisco	- BE characteristics	- Density measures - Land-use structure - Regional accessibility measures - Commute-related mearues		- Demographics and housing cost variables - Ethnic composition - Household demographic variables	- Local transportation network measures	- Built environment (BE) attributes affect residential choice decisions and car ownership decisions; hence, change in BE characteristics - policy has to be evaluated in the joint context of both decisions. - The commonly used population and/or employment density measures are actually proxy variables for such BE measures as street block density and transit accessibility. - Regarding car ownership decisions, both household demographics and BE characteristics are influential, but households have a more dominant effect. - There is variation in sensitivity to BE attributes due to both demographic and unobserved factors. Hence, ignoring such systematic and random variations in sensitivity to BE attributes will lead to inconsistent results which can lead to inappropriate policy decisions. - Household income is the dominant factor in residential sorting. Ignoring income effects in car ownership can

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Martin: Ley Danyluk, David	2007	to examine the relationship between gentrification and the transport mode selected for the journey to work.	Survey + analysis	Toronto, Montreal, Vancouver Canada	2001 and earlier	Census of Canada	Transport Mode			Gentrification index: possession of post-secondary education and employment in the quaternary sector of pro-fessionals, managers, administrators and technical workers		Findings showed that its residents are more likely to ride a bicycle to work than use the public transport, for a short or medium distance. residents of gentrified areas are more likely than other commuters to ride a bicycle to work, even when controls are introduced to remove the effects of distance to the downtown core. At the same time, they are less likely to be users of public transport, despite their political support for the notion of the public household. Relationships with travel mode are confused by internal divisions within the population of gentrifiers. Despite common agreement on being 'urban people' and disavowing a suburban way of life, there are ideological differences between cohorts associated with different stages in the gentrification process.
J.: de Cea Ch Enrique Fernández L, Joaquín: Malbran, R. Henry	2008	To present a methodology for solving the public transport network design problem and describe its application as the design study developed to propose a new structure for the transit system	Mathematical method	Santiago, Chile	2001	1) Household origin-destination survey 2) On board surveys 3) Origin-destination survey	- Value of time - Number of trips - Travel time			Income	- Number of service - KM of service - Time cost	The private operating costs and the social costs of the restructured system using higher standard buses, are lower than the costs of the current system. These cost reductions allow government authorities to introduce a number of modernising measures without subsidies and fare increases.

Author	**Year**	**Research Question/Objective**	**Statistical Method**	**Studied Region/Country/Cities**	**Dataset time**	**Datasource**	**Transport Use Variables**	**Land Use Characteristics**	**Urban Form**	**Socio Economic**	**Service Provision**	**Findings**
Sergio: Tirachini Jara-Díaz, Alejandro: Cortés, Cristián E.	2008	To examine the effect of having different levels of aggregation regarding demand information on the optimal values of frequency and vehicle size by comparison with models to explore improvements on the recommendations and conclusions obtained from classical microeconomic models To compare the analytical expressions obtained for the optimal designi variables under different demand aggregation levels.	Mathematical method & Analysis	Santiago, Chile	1998	Ministry of Transport and Telecommunicatios, Chile	1) Passenger boarding/alighting late 2) Operator Cost 3) Optimal frequency and capacity				1) Length of corridor Bus movement travel time Marginal pasenger bording time Trip rate between stations 2) Cost per vehicle-hour Cost per vehicle-KM 3) Running time between station (min] Distance between stations (km) Subjective value of waiting time Subjective	Established the optimal conditions for frequency on a public transport corridor with inelastic dernand, in cases where the demand data is only available at an aggregated level (at the level of an entire line, or ridership per direction of movement) as well as cases in which it is feasible to obtain more detailed information on the demand structure. Developed two analyses, one which is purely analytical and another based on the result of applying the different aggregation level models to two examples in which real public transport data for peak periods are available. From the analytical part we clearly identified those terms in the optimal frequency expression that generates the differences among the models. Suggest the underestimation of optimal frequency and overestimation of vehicle size when not accounting for users cost fully is even more important

Appendix 1: Literature Review Summary

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
J.: Long Lin, L.	2008	To present an alternative view of where people reside and how they travel across the US through innovative mining of public data sources. To investigate the interaction between neighbourhood environment and household travel in the US.	Mathematical method & analysis	USA	1) 2000 2) 2001	1) Census Transportation Planning Package 2) National Household Tranvel Survey	- Neighbourhood clusters	- Employment rate - Land use: density - Journey-to-work		- Household size - Household vehicle - Race/ethnicity - Income - Home tenure status - Age - Resident industrial type - Occupation - Household trip rate - Travel time		Household life cycle status and residential location are not independent. Transit availability at place of residence tends to increase the transit mode share regardless of household automobile ownership and income level, and job-housing trade-offs are evident when mobility is not of concern. Racial preference in residential location and contrasting travel characteristics among ethnic groups. Inadequate and unbalanced transit service, unequal mobility and accessibility among ethnic groups, job-housing trade-offs, and balances between family needs and residential location choice are unlikely to be addressed adequately in regional transportation planning.
Geraldine: Kaufmann Pflieger, Vincent: Pattaroni, Luca: Jemelin, Christophe	2009	To identify three features common to both past and present transport and urban planning policies - namely, contingency, reproduction and innovation.		6 European cities in France, Germany and Switzerland								

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Jen-Jia: Yang Lin, An-Tsei	2009	To examine the connections between urban form and travel demand at the aggregate level using traffic analysis zone data from Taipei, Taiwan	Mathematical method + Analysis	Taipei, Taiwan	2000		9 latent variables - density, diversity, design, trip generation, private mode split, transit service, private mode facility, social condition, economic condition -trip generation -> generation rate -private mode split -> modal split of car and motorcycle	density -> building/emploeyement diversity -> housing job / retail	design -> road density, grid network	social condition -> ratio of student/female economic condition -> income, car/motorcycle ownership	transit service -> sub stop density, distance to station, bus route, transit accessibility private-mode facility -> car/motorcycle accessibility, parking space density	density is positively related to trip generation and negatively associated with private mode split mixed land use reduces trip generation and indirectly increases private mode split and a pedestrian-friendly built environment significantly reduces private mode split. Empirical evidence indicates that land use density, diversity and design affect travel demand in urban areas both directly and indirectly.
Géraldine : Rozenblat Pflieger, Céline	2010	To demonstrate the mutual constitution of networks and cities. To show the positive contribution that network theory can make to the analysis of urban networks.										

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Emre: Wenglenski Korsu, Sandrine	2010	To explore, in the French context, the hypothesis that employment problems experienced by low-skilled jobseekers are partially caused by spatial urban factors.	Mathematical method & Analysis	France	1999	micro data from the 1999 French population census	Coefficient Estimated probability of long-term unemployment			Gender/Family Status Age Socio-professional status Educational attainment Nationality Job Accessibility Social environment of neighbourhood		The research we have developed with data from the Paris–Île-de-France metropolitan area, the French capital, yielded results supporting this hypothesis. In common with most pioneering French studies, our findings show that, all else being equal, a low-skilled worker faces higher risks of long-term unemployment if he/she suffers from poor job accessibility and if he/she experiences longterm exposure to high-poverty neighbourhoods. A large share of low-skilled workers in Paris have job accessibility far below the average and this level seems low enough negatively (although slightly) to affect their employment. Similarly, many low-skilled workers in Paris live for long periods of time in neighbourhoods with poverty and unemployment rates that are significantly above average and the concentration of poverty in these neighbourhoods seems strong enough to generate a social environment that damages the employment outcomes of resident workers.
Gordon Dabinett	2010	To analyse urban planning practices in South Yorkshire to reveal how EU strategic spatial ideas and values are reproduced. To examine how the notion of spatial justice was interpreted as the organising concepts within the European Spatial Development Perspective became situated within a territory severely affected by deindustrialisation in the 1980s, but subsequently a major beneficiary of EU Structural Fund programmes.	Analysis (Case study)	South Yorkshire, UK	1998	ONS		Emploiyement Population Area				The analysis reveals how policy-making at this scale used a construct of polycentric urban development that reasserted a model of economic growth based on the indigenous assets held in city centres at the expense of more redistributive measures targeted at the former coal-mining communities in the sub-region.

Appendix 1: Literature Review Summary

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Miquel-Àngel Garcia-López Ivan Muñiz	2010	To examine in the case of the Barcelona metropolitan region between 1986 and 2001, whether its employment is moving from polycentricity to scatteration and whether its employment location model is increasingly random and unstructured.	Mathematical method & analysis	Barcelona, Spain	1986-2001	Employment data	Employment decentralisation Employment deconcentration Employment polycentricity	Density Emploment Infrastructure land				In spite of the decentralisation and deconcentration processes, employment concentrated in centres still represents a significant percentage of total employment and new sub-centres have emerged in the periphery. Polycentricity and scatteration do not appear to be two alternative models but are in fact complementary. And the dispersion of employment does not follow an amorphous model, but is structured based on transport infrastructure. An increasing influence of employment sub-centres on employment location and density condition is shown; hence, polycentricity has been reinforced.
C. S.: Kawamoto Pitombo, E.: Sousa, A. J.	2011	To discover relations between the socioeconomic characteristics, activity participation, land use patterns and travel behaviour of the residents in the Sao Paulo Metroplitan Area (SPMA) by using Exploratory Multivariate Data Analysis (EMDA) techniques.	Analysis & Survey	Sao Paulo, Brazil	1997	An origin-destination home interview survey by METRO-SP. 98,780 individuals 389 TZs	- The use of tranportation 'credits' for Transit tariff - Activity participation variables			- Car ownership - Origin Cluster - Family Income - Occupation		Travel behaviour was represented by trip-chaining patterns, allowing a clear definition of inter-relations even using only exploratory techniques.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Wen-Chyuan: Russell Chiang, Robert A.: Urban, Timothy L.	2011	To analyse monthly ridership of the Metropolitan Tulsa Transit Authority in order to identify the relevant factors that influence transit use and create an accurate forecast (3 years) that will help justify the investment in public transit based on various planning scenarios.	Mathematical method	Tulsa, OK, USA	Oct.1998 - Aug.2008	1) Tulsa Transit 2) Midwest gas 3) Oklahoma food stamps 4) US Census Bureau	Ridership	Population		Income Gas price Tulsa Transit budget level Seasonality Fare Food stamps		Food stamps have a negative impact on the ridership. Budget have the significant positive impact on the ridership. Bus fare is a deterrent for riders. Seasonality has a larger impact on the ridership - less travel under hot climate. The price of gasoline has a small, but statistically significant, effect on ridership; hence, it is inelastic on ridership. A simple combination of these forecasting methodologies yields greater forecast accuracy than the individual models separately. A scenario analysis is conducted to assess the impact of transit policies on long-term ridership - the most-likely and worst-case scenarios.
Pablo: Donoso Bass, Pedro: Munizaga, Marcela	2011	To determine the probability of migrating from public to private transport at both aggregated and disaggregated levels to assist transport authorities in maintaining or increasing levels of public transport use.	Mathematical method & analysis	Santiago, Chile	1) 2010 2) 2007	1) the Santiago panel - an experimental sample of Catholic University employees surveyed regarding the public transport system change 2) Transantiago agency - customer satisfaction/service quality	- Fare evasion rate - Quality of changes - Num of user of transport			- Wealth - Travel time - Age	- Fare - Frequency compliance index	The suggested model predicted migration with 57% accuracy in the first preference recovery measure. It can help to focus policy and management measures and to increase the efficiency of public investment. - More sensitivity to quality of service than to fare changes. - Lower income users received the worst service but migrated less than higher-income users. - Higher evasion rates are observed for low-income users. - The suggested model can be an alternative to traditional customer retention models that can be applied to many instances.

Appendix 1: Literature Review Summary

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Manuel: Vance Frondel, Colin	2011	to investigate the determinants of public transit ridership with the aim of quantifying the effects of fuel prices, fares, person-level attributes, and characteristics of the transit system on transport counts over a five-day week	Survey + analysis + Mathmatical	German	2010	German Mobility Panel	Coefficient Marginal effects Robust standard errors			Real fuel price Fare Age income Minutes - to public transit Degree license employement gender	-Public transit density -Parking space	Our estimates reveal fuel prices to have a positive and substantial influence on transit ridership, though there is no evidence for a statistically significant impact of the fare. Research highlights fuel prices to have a positive and substantial influence on public transit ridership. urban form is also a strong correlate of ridership. There is no evidence for a statistically significant impact of the fare on ridership. females are more dependent on public transit than males. However, females and males show the same responsiveness to changes in fuel prices.
Stuart: Hensher Bain, David A.: Li, Zheng	2011	To introduce a initial demand modelling framework for the development of a regional transport and land use model system (R-Tresis) for NSW	Mathematical method	Sydney, NSW, Australia	2008	National visitor survey	Origin destination -traffic activity by mode -- trip distance	- population				The model system will be embedded in a decision support system that can be used to investgate a large number of economic, social and environmental policies. It is designed to offer a modelling capability for regional jurisdictions that is currently only available in Aus for the major metropolitan areas.

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
Calvin P.: Zandbergen Tribby, Paul A.	2012	To develop a high-resolution, GIS-based multimodal model that used travel time as the accessibility measurement in the form of travel time from origins to destinations by way of public transit. To demonstrate the applicability of the model to assess the change in travel times due to a resent addition of bus service in the study area.	Mathematical method & analysis	Albuquerque, NM, USA	2006	Bureau of Business and Economic research	-Travel time (saving) -Travel mode	Population		Income Age Car ownership		High-resolution spatio-temporal, GIS-based public transit network model to assess the changes brought about by new routes that can highlight the travel time saving for individual address points. There are no inequities in public transit provision would grossly over-simplify one of the most significant consideratios in travel: that of travel time, including walking, waiting, and transferring. The northern part of the study area did see a significant increase in travel time savings, increasing the likelihood of public transit being chosen over commuting by automobile - lead to a decrease in congestion and pollution. The unequal allocation of travel time saving to transit user concentrations based on socio-economic need is found.
Paulus Teguh: Cao Aditjandra, Xinyu: Mulley, Corinne	2012	To explore whether chcanges in neighbourhod characteristics bring about changes in travel choice. In order to reduce the gap regarding timeline and geographical areas, it expends the evidence using British data.	Mathematical method & Survey	Tyne and Wear, England, UK	1) 2004 2) 2001	1) Index of Multiple Deprevation (IMD) 2) British Census Data	Car ownership Travel attitudes Shopping accessibility Travel accessibility		Shopping accessibility	Income Household size Ethnicity/Race/Age	Travel accessibility	It is discovered that neighbourhood characteristics do influence travel behaviour after controlling for self-selection as well as car ownership. - All findings suggest that ladn-use policies at neighbourhood level can play an important role in reducing driving.

Appendix 1: Literature Review Summary

Author	Year	Research Question/Objective	Statistical Method	Studied Region/Country/Cities	Dataset time	Datasource	Transport Use Variables	Land Use Characteristics	Urban Form	Socio Economic	Service Provision	Findings
J. Holmgren	2013	To provide empirical estimates of how different factors, including price and car ownership affect the demand for local public transport. To investigate the effects of income changes on local public tranpsort demand.	Mathematical method & analysis	Sweden	1986 to 2001	Annual data of the urban traffic of the 26 Swedish counties	Demand for public transport Quality of service Transport capacity Vehicle-kilometres			Income Car ownership		Although the demand for public transport might decrease with increased income, there is no evidence of such effects even when the full effect of changes in income and changes in car ownership is taken into account as the total effect of income on public transport demand is virtually zero in elasticity terms. + Increasing transport capacity will increase the quality of service. + Increased understanding of the long-run effects of income changes on public tranport demand will improve the posibilities of making accurate forecasts and efficient planning.