Curtin University Sustainability Policy (CUSP) Institute

Agglomeration Economies in Australian Cities:
Productivity benefits of increasing urban density and accessibility

Roman Lubor Trubka

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

August 2011

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: ………………………………

Date: ………………………

Abstract

Agglomeration economies are a subject that has been gaining a significant amount of interest in the realms of policy and urban planning. The term refers to the externalities that arise out of the interactions of firms and employees, which are made possible by spatial proximity. Although empirical studies measuring the impacts of agglomeration economies on firm and employment productivity have been conducted for a number of nations around the world, no such study has yet has been conducted for Australia or Australian cities. The research embodied in this thesis seeks to measure the magnitude by which employment productivity in a range of industries in Australian cities is influenced by agglomeration and offers a method for these estimations that is suitable given the types of data collected and made available nationally. Furthermore, analyses are conducted on a wider range of industries than reported by existing works on the subject.

Analyses are carried out primarily on Sydney and Melbourne; however, one analysis incorporates all eight capital cities. The rationale behind conducting analyses on two cities is to allow comparisons to be made, thus providing a means for validating the city-specific results and contributing to an understanding of whether elasticity estimates can be generalized within the nation. Topics such as the relative importance of urbanization versus localization economies are addressed as well as the issue of endogeneity. Current state-of-the-art practices in incorporating the benefits of agglomeration economies in transport project appraisal in Australia are reviewed. Additionally, the outcomes of the empirical analyses are drawn on in a discussion of the relevance of agglomeration economies for sustainability and urban planning.

The findings show industry-specific employment productivities do benefit significantly from agglomeration and at magnitudes comparable to international studies. The devised econometric model proves effective at estimating agglomeration impacts and can be replicated for other Australian cities and regions – a suggested alternative to generalizing industry-specific elasticities as evidence exists that they are likely to differ for at least some industries. The evidence of agglomeration economies working in Australian cities becomes a powerful companion rationale for considering density and quality public transport services which are frequently at the centre of urban sustainability strategies.

Acknowledgements

It is difficult to know where to begin and at what point to end when thanking to those who had bearing and influence on my research and overall journey.

To begin, I would like to thank and acknowledge Professor Peter Nemetz whose course I took on sustainable business at the Sauder School of Business at UBC during the Canadian winter of 2005. The materials he covered and the messages he delivered were truly inspirational and breathed a sense of purpose into me, as I felt rather deflated by the standard business school curriculum that seemed so mindless of the environmental challenges that convention had created.

I am indebted to my supervisor, Distinguished Professor Peter Newman, who from our first meeting in March 2008 has put so much faith and trust in me. He has given me so many amazing opportunities to grow personally, academically and professionally. I have learned a great deal from him in the last several years and am moved by his passion and dedication to sustainability. Nothing I write can do justice to the extent of my appreciation for what he has done for me.

I must express a great deal of gratitude to the Distinguished Professor of Economics and the Curtin School of Business, Harry Bloch. He volunteered countless hours to guide me through my analytical work by making himself available in person or by responding promptly to my countless emails. He made economics and econometrics manageable and gave me wonderful support. Without his help, I simply do not see how I could have accomplished what I have.

On a similar note, I would like to thank the head of the Department for Econometrics and Quantitative Modelling at the Curtin School of Business, Professor Felix Chan. He also gave me a substantial amount of guidance on my analytical work, helping me understand econometrics in ways that textbooks would never teach.

I give special thanks Michael Chappell at Pracsys Economics, whose friendship and mentorship were incredibly valuable to me throughout this process. He was my link to the 'real world', a great proponent and supporter of my work, and in trying times took me in as family. To him I am very grateful.

I would also like to thank Florian Wendelin-Mayer, who gave me incredible instruction and guidance on the use and application of GIS.

For their assistance, support and friendship, I would like to extend thanks to the administrators and my peers at CUSP. Brenton Burger, Vanessa Rauland and Svetla Petrova deserve special mention.

Last but far from least, I have my mother, father, godmother and sister – all of whom were ever only a phone call away – to acknowledge. Thank you for believing.

Table of Contents

Chapter 1: Introduction

1.1 Introduction

This research empirically investigates the question of how accessibility affects productivity across a range of industries in Australia. The estimates of these gradients can, in turn, be used to improve the methods by which transport infrastructure projects in Australia are valued. In effect, however, this research does much more than this.

Broadly, it explores the question of why cities exist, acting as a powerful reminder of why we travel and do not chose arbitrarily and homogeneously to disperse across the regions where we live and work. A related question asks why firms choose to locate where they do. However, the inquiry does not end here; the process of formulating and addressing these questions reveals others inextricably associated with them: questions such as, "Does location matter given that our communication technologies have been so vastly improved in recent years?" “Can the automobile and the urban form that evolved because of it satisfy the transportation needs of our cities, given employment growth projections and the associated concerns about sustainability?” In addition, I ask, “How can we design and plan our cities to further economic growth to maintain prosperity, or the growth of prosperity, in Australian cities well into the future?”

This research is as relevant to the concerns of economic developers and land developers as it is to transport planners. It spans a number of disciplines, drawing on a variety of techniques, including those involved in econometric analysis, geographic information systems, finance and accounting and urban planning. This empirical work seeks to shed light upon the above questions but also sets the foundation for a greater narrative of interest to those responsible for the future development of cities. My hope is for researchers and practitioners to seek a direction for urban development that will make cities more sustainable, more resilient, more efficient, and ultimately more liveable places.

This thesis is not simply about Australian cities. I address universal principles that, if applied, could enable cities fundamentally to work better. These principles apply to developed cities everywhere. However, the data employed in the analyses of this work are based on Australian cities. Hence the focus is primarily on them. In terms of application, however, most developed cities need to address issues related to urban form – whether to build out or build up. This is the case even with the emerging cities of China and India whose strong growth and large populations are raising important questions about development patterns.

1.2 The Significance of this Research

Policy Significance

People, firms, organizations and institutions experience many benefits from their interactions with each other and their markets. This is why we have the CBDs, regional centres, activity centres, knowledge corridors, industrial parks, and similar concentrations of employment that constitute our urban topography. As these benefits have never been quantified for planning purposes, or historically been properly understood for planning purposes, this gap reveals research opportunity. This sort of investigation could yield valuable tools and insights for planners, consultants, and policy-makers to help create a more productive, competitive and sustainable local economy.

This has been an area of investigation in the UK for the past several years, embodied in the term "Wider Economic Benefits" (or WEBs) of transport infrastructure. Infrastructure Australia[1] has now recognized it as an important area for further research and has incorporated it in its economic reporting requirements for infrastructure funding (Infrastructure Australia 2009). In Australia, as a result, state and local governments (usually through consultants) have begun trying to estimate agglomeration economies as part of submissions for infrastructure funding (Newman, personal communication, 15 May 2011). Where this has been undertaken, however, it appears that assessments have not used local data or been carried out in a rigorous manner. Similarly, the CRC for Spatial Information[2]

---

[1] Infrastructure Australia is a statutory body established in 2008 to advise governments, investors and infrastructure owners on matters of identifying infrastructure needs, financing mechanisms, policy, pricing and regulation. Visit www.infrastructureaustralia.gov.au to learn more.

[2] The Cooperative Research Centre for Spatial Information (CRCSI) is an R&D centre comprising federal and state government agencies, universities, and private sector companies, organized to conduct user-driven research in the area of spatial information. Visit www.crcsi.com.au to learn more.

has recognized the value of quantifying agglomeration externalities to support development activity in Australia.

Academic Significance

While most of the empirical work on econometrically quantifying the effects of agglomeration economies has been pioneered in the UK and the US, only a few key figures have been responsible for the most prominent bodies of work. Only one has performed analyses on an industry-disaggregated basis. This research is an opportunity to undertake an investigation into the benefits of agglomeration in Australian, a country for which an extensive analysis of agglomeration economies has not yet been carried out. Further, it draws on cross-sectional census data which have not previously been applied to an analysis of this nature. This research also allows for: (1) comparisons with studies conducted using UK data; (2) insights about how agglomeration forces work differently in different countries; and (3) how a different approach to their measurement may affect the outcomes of their estimations. This research is unique in that it estimates agglomeration economies for a number of cities within one country, allowing for insights to be gained into whether agglomeration effects differ among industries intra-nationally. Finally, this research is an opportunity to extend the industry scope of this sort of econometric analysis by including industries not previously investigated, such as mining, retail, cultural services and medical services. Overall, it represents an opportunity to review the value of undertaking agglomeration economy work in general to determine if there is a potential to make it more useful and available in the planning flow of cities.

Sustainability Significance

In academic and practical terms, the notion of agglomeration economies feeds primarily into an understanding of productivity in cities. This is fundamentally an economic concept but increasingly there are overlaps with concepts related to sustainability in cities (Newman and Jennings 2008). Sustainability in cities is primarily about reducing the city's ecological footprint (which includes its resource consumption, land take and waste production), whilst simultaneously improving its liveability (which includes quality of life, elements of the natural environment, the health promotive capacity of the environment, living affordability, accessibility, and cultural capital). There is, however, a growing awareness that housing density is critical to achieving these objectives. Only by increasing density in centres will we see productivity increases related to sustainability and the emergence of what is sometimes referred to as the 'sustainability multiplier'. Whilst the study of the economic impacts of density and accessibility on productivity does not directly pertain to sustainability, it has the potential to serve as a companion rationale for sustainability strategies that seek to increase urban densities for the sustainability benefits that compact development facilitates.

1.3 Research Questions

This thesis addresses five main questions.

1. Can useful agglomeration elasticities be determined for Australian cities? This fundamental question not only asks whether there is a measurable productivity effect from economic density but also whether data sources exist to enable such analyses to occur whilst producing reliable results.
2. Are elasticity estimates amenable to being generalized across Australian cities? As there is a desire for agglomeration externalities to be incorporated into planning efforts across the nation, it is useful to know if generalizing results from one city across others makes sense.
3. Are elasticity estimates robust to changes in the geographic scale of analysis? As the complexity of analyses increases with more detailed datasets, it is useful to understand the impacts of employing different geographic scales.
4. How can these elasticities best be applied in an Australian context? This question aims to gain some understanding of the ability of current transport modelling systems in Australia to apply these elasticities to infrastructure projects.
5. What are the broader implications of agglomeration economies for urban planning and infrastructure? This question entails taking a step back from the essentially econometric endeavour and determining the significance of the subject for cities and planning.

1.4 Structure of the Thesis

The organization of this thesis is as follows: Chapter 2 begins by broadly discussing the importance of studying cities and then narrows the focus to the topic of transportation and how it influences urban form. The chapter then discusses the importance of agglomeration economies as the centripetal force that keeps development from dispersing homogeneously across economic regions, and reviews the empirical works that attempt to quantify its effects in cities and countries around the world. Chapter 3 discusses the sources of data drawn on for this research and explains the theoretical model applied to estimating the effects of agglomeration on labour productivity. Chapter 4 offers overviews, results, and conclusions for three separate analyses investigating the strength of agglomeration economies in a number of Australian capital cities, while for each using the statistical local area (SLA) as the geographic unit of reference. Chapter 5 is organized in the same manner as Chapter 4 but reports on a separate set of three analyses, all of which employ the much smaller geographic unit of the work destination zone (WDZ) as the spatial unit of reference. Among the six analyses, topics such as broad industry and industry-specific productivity elasticities with respect to employment concentration, the benefits of industry localization versus industrial diversity, and the impacts of endogeneity on the elasticity estimates are addressed. Chapter 6 draws the findings together by discussing the preferred results to be used as inputs into infrastructure valuation, how results can be applied to transport infrastructure valuation practices, the implications of the findings for planning in Australian cities and recommendations for further research and analysis. Finally, appendices elaborate on some econometric concepts, the empirical results and the computer code generated for this work. Appendix I consists of three published papers on the costs of sprawl that were written as part of the foundational work for this thesis.

CHAPTER 2: A Review of the Overarching Literature

2.1 Organization of this Chapter

Starting off the chapter, Section 2.2 titled "Why Study Cities?" gives a broader context for the significance of cities as a topic of study. It focuses its discussion on their growing importance as places to live and work, their economic significance, and the role they have been envisioned to play in both improving social welfare and being part of the solution to achieving greater levels of sustainability. Section 2.3 reviews a number of models that have been used to explain why cities exist, why they chose to locate where they do, and the determinants of the sizes that they achieve. The purpose for a theoretical account of the existence of cities is to show that there is a certain order, or logic, to how people organize spatially and temporally. An understanding of this is valuable for creating an awareness of how policy decisions may affect urban form outcomes. This issue is carried over into Section 2.4 that discusses how transport shapes cities. Transport has an integral role in determining urban form and as such can be linked to the problems and benefits that various urban forms are associated with. Section 2.5 delves a little deeper into transport and its infrastructure, discussing some conventions, issues, and innovations in its valuation. This is followed by a discussion of the sources of benefit that are embodied by the term 'agglomeration economies' in Section 2.6 and then a review of empirical works investigating the matter in Section 2.7. Section 2.8 briefly discusses some ways in which the micro-foundations of agglomeration have been investigated more closely, as conventional approaches to empirical studies of the phenomenon have typically addressed it in an aggregated fashion. Section 2.9 merges the domains of agglomeration economies and transport infrastructure investment by discussing the economic theory on the matter as established by Venables (2003), which instigated efforts to capture agglomeration economies in transport project appraisal. In Section 2.10 a variety of common measures and indices typically utilized in the study of urban economics and agglomeration economies are explored. Finally, Section 2.11 concludes the chapter.

2.2 Why Study Cities?

The world is becoming increasingly urbanized. In 2007, the population distribution between urbanized and rural areas was balanced for the first time in history, as the share of the world's population living in cities reached 50% (Burdett and Sudjic 2008). Currently, the world population is roughly 6.8 billion (Population Reference Bureau 2009). Thus, approximately 3.4 billion people now live in urbanized settlements. This trend towards urbanization is not expected to plateau at this distribution. Projections suggest that by 2050, 75% of the world's population will be living in cities (Burdett and Sudjic 2008). The world population is also expected to increase to 9 billion, given a medium-growth trajectory over the same period (Population Reference Bureau 2009). In this case, we can expect an urbanized population of approximately 6.5 billion worldwide in about 40 years' time. Accommodating this huge growth will require a massive amount of infrastructure investment in both current and newly emerging cities. It will also require planning for a significant number of new residential dwellings and the creation of employment opportunities. Theories of economic development, however, tell us that not all jobs are created equal. In ideal circumstances, these new jobs will comprise as much high-quality, export-driven employment as possible.

The responsibilities of planners, however, do not end here. Decisions affecting infrastructure investment and land-use must be made in conjunction with principles that seek to minimize anthropogenic global temperature rise and the impact on the natural environment. Australia's commitments require that greenhouse gases be reduced by at least 60% of 2000 levels by 2050 (Department of Climate Change and Energy Efficiency 2011), with similar targets being made internationally. In light of the projected population increases, the task ahead of greenhouse gas mitigation becomes significantly greater than it would be if we considered only present-day population.

The situation facing Australia is that it is a highly urbanized country: in 2008 approximately 75.2% of the population lived in the 17 major cities – that is, cities with populations of over 100,000. Nearly two-thirds (63.9%) lived in the capital cities (Infrastructure Australia 2010). Whilst they house the majority of the country's population, cities in Australia are predominantly characterized as low-density and sprawling. In 2010, Sydney and Melbourne ranked 63rd and 69th in the world in terms of population size, but

113th and 127th respectively in terms of population density (City Mayors Foundation 2010a; 2010b). With a present-day Australian population of approximately 22.5 million (Australian Bureau of Statistics 2011), projections are that by 2056 Australia's population will reach 35.5 million under a medium-growth scenario, with 72% of this growth occurring in the capital cities (Infrastructure Australia 2010). This concentrated growth in already congested urban locations creates concern over how these congestion levels will rise, as already the estimated avoidable cost of congestion in Australian capital cities in 2005 was $9.4 billion, expected to rise to $20.4 billion by 2020 (Green Building Council Australia 2011). This restriction on movement has material effects on productivity and economic health.

A number of state plans or strategies offer guidance for this capacity expansion in Australian cities: *Melbourne 2030*, the *South East Queensland Regional Plan*, or Perth's *Directions 2031*. These plans emphasize such principles as enhanced housing density, mixed-use development, land-use integration with high public transport servicing, and boundaries limiting urban growth. Whether these intentions materialize will be revealed in time, as currently there is some scepticism over this because of reluctance by planning agencies to require compliance with compact city policies (Buxton 2006; Goodman and Coote 2007). Often the driving force required for change is evidence that new methods or approaches will be economically advantageous, such that they either produce a greater profit or are more effective at avoiding unnecessary costs. Appendix I contains papers with findings applied across Australia that were prepared as part of this research. The research embodied in these works, however, mostly overlooks the productivity implications of centres as they arise from person interactions.

The bottom line is that it is difficult to change conventions and common practices, especially when perhaps wiser alternatives are supported by attractive financial figures and concerns with the level of risk cannot be eased or mitigated. Goodman and Coote (2007) argue that in corporately owned shopping centres, for instance, niche shops are typically not found because the mix of retail functions is likely to be tightly controlled by a low-risk formula to appeal to investors. Anecdotally, personal experience has taught that land developers like to adhere to a proven product because good returns can be made with low risk. Trying something new, which in many cases means building higher-density developments, can be seen as a risk if market viability has not been tested locally for a product that in some cases may incur higher costs than business-as-usual alternatives.

Similar issues exist in the public sector when governments choose projects that deliver short-term, noticeable gains that incite the least amount of controversy, to appease the constituency that elected them. Despite the challenges that planners and developers face in managing cities to thrive environmentally, socially and economically into the future, it is clear (historically and internationally) that cities and agglomerations have distinct productivity benefits over more dispersed, rural urban forms. This matter is discussed at length in this thesis.

Fujita and Thisse (2002) offer a thorough account of how productivity and productive activities are geographically concentrated within nations. In 1990, they explain, Japan accounted for 3.5% of East Asia's total area[3], 7.9% of the population, 72% of the GDP and 67% of the manufacturing-specific GDP. At a more detailed level, Japan itself is further dominated by the five prefectures that contain its three major metropolitan areas: Tokyo/Kanagawa, Aichi and Osaka/Hyogo. The spatial disparity of productivity is even greater in the metropolitan area of Paris, which accounts for 2.2% of the country's area, 18.9% of its population and 30% of its GDP. Similarly, Glaeser and Gottlieb (2009) explain that 68% of Americans occupy only 1.8% of the country's land area and relate this concentration of people to within-country variations in earnings. They find urbanites to earn significantly higher wages than their rural counterparts. Further, Venables (2003) finds that the inner city of London generates a level of per capita GDP that is nearly 80% above the national average (when occupational composition is not controlled for) and 34% to 41% when controls are imposed. The reporting of both of these figures – where occupational differences are left uncontrolled and accounted for – is of significant value. In the former case, higher value-adding activities are more likely to co-locate than the lower value activities. Thus, cities are conceivably a prerequisite for their existence. In the latter case, the productivity disparity shows that the same activity will generally perform better economically in a denser area than in a geographically less dense area. Considering these spatial variations in earnings, it is clear that geographic concentration plays a substantial role in the economic performance of cities and countries.

Fujita et al. (1999) offer another perspective on why we need to understand the economics of cities. Quoting *The Economist,* they argue that open trading systems (such as

---

[3] East Asia is viewed as comprising Japan, South Korea, Taiwan, Hong Kong, Singapore, Philippines, Thailand, Malaysia, Indonesia and China.

NAFTA and the EU) across countries act to level national production advantages but enhance the importance of cities. This view helps explain the incredible growth of cities in the past 100 years, with megacities, in particular, reaping much of this economic opportunity. In recent decades, it has also become characteristic of advanced countries to increasingly enter into the business of producing information rather than tangible goods. In facilitating face-to-face interactions and allowing knowledge spillovers to occur, cities are of vital importance and will remain so despite the common belief that ICT technologies can act as suitable substitutes (Glaeser 1998). Paul Romer's seminal work on endogenous economic growth emphasizes the importance of human capital as ideas fuel economic progress (Romer 1990), the opportunities for which are most manifested in cities where human interactions are greatest (Romer 1990; Glaeser 2000). The idea is that when situated around other talent, workers accumulate knowledge more easily and quickly than if isolated or around less experienced workers. The role of cities as tools for knowledge exchange is one point to emphasize and the other is the importance of knowledge itself. As Glaeser (2011) aptly puts it, "Infrastructure eventually becomes obsolete, but education perpetuates itself as one smart generation teaches the next" (p. 27).

There is also a great deal of research now that investigates the competitiveness of cities based on the premise of knowledge driving competitive advantage.[4] The University of Wales Centre for International Competitiveness, for instance, has released a number of reports comparing the knowledge competitiveness of cities. Their *World Knowledge Competitive Index* 2008 (Huggins, Izushi et al. 2008) compares 145 regions over 19 indicators with components drawn from human capital, knowledge capital, regional economy outputs, financial capital components and knowledge sustainability. As early as 1988, the Henley Centre for Forecasting in the UK estimated that 50% of occupations in Britain require brain skills rather than manual skills (Montgomery 2007). Arguments abound that cities offer a setting for civilized life and allow a degree of social aggregation that creates more possibilities than the sum of individuals could possibly achieve (Short 1991). Richard Florida, in *The Rise of the Creative Class* (2002), discusses how the ability to cultivate technology, talent and tolerance drives cities to grow economically and reinvent themselves

---

[4] See the works of Glaeser (1998; 2000), Glaeser, Kallal et al. (1992), Glaeser and Saiz (2004), Karlsson and Johansson (2004), and Simmie, Carpenter et al. (2006) for further discussion on the role of cities in knowledge creation.

in economically depressing times. Pearce and Barbier (2000) in *Blueprint for a Sustainable Economy*, argue that the relative importance of the three forms of capital – natural, physical and knowledge – shifts as economies advance. The most advanced economies make much of their economic gain in the cultivation and commercialization of knowledge, a form of capital that does not get "used up" as the other more conventional forms and has far fewer direct impacts on the environment and its sustainability than, say, heavy industrial or activities involving resource-extraction. Fujita and Tabuchi (1997), in studying Post-war Japan, see this transformation (from light to heavy industry and from heavy industry to high-tech services) as marking two distinct economic transformations that inevitably kept higher income generating knowledge-intensive services in the core of Tokyo, dispersing mass production activities to the periphery. This example of an evolution in a city's economic focus illustrates the central means by which cities experience economic growth – they do so through innovation which allows old activities to be performed in markedly different ways, all the while maintaining a focus on newly emerging industries where productivity levels will typically be higher (Burgess and Venables 2004).[5] Taking the knowledge economy perspective, the organization of cities not only affects the transportation of goods but the movement of people and ideas as well. Understanding the forces that drive spatial economic concentration – and those that limit it – can help guide decision-making practices to foster more economically prosperous and sustainable cities and regions.

Another important consideration for studying cities and the impacts of their transport infrastructure in particular, is that much of what planning authorities are concerned with is the 'liveability' for citizens within their jurisdictions. Infrastructure Australia summarizes the concept of liveability according to a number of indicators: health, amenity, housing, living affordability, and accessibility (Infrastructure Australia 2010). Urban form and transport affect liveability on all of these fronts. A dense, mixed-use urban

---

[5] The terms '1st Advantage' and '2nd Advantage' adapted from the use of the terms '1st Nature' and '2nd Nature' by economic geographers refer to the environmental conditions that assist the productive development of new activities, and then the reinforcing processes that take over to enforce growth, respectively (Krugman 1993; Burgess and Venables 2004). While a 1st Advantage may initially explain the reason for a city's existence or location, commonly the growth that subsequently occurs will be because of innovations that have little to do with a city's initial primary activity. A hierarchy emerges where growth is driven by innovation and innovation by knowledge, where the most educated cities are highly correlated with productivity growth and wealth (Glaeser 2011).

form – when designed intelligently – enables walking and cycling to be viable modes of transport that affect quality of life and health via increased levels of physical activity (Trubka, Newman et al. 2010c). Living affordability is affected by urban form because of the implications it has on private vehicle transportation, which in suburban development is highly vulnerable to price fluctuations of fuel (Dodson and Sipe 2008). Fuel constraints will continue to be an issue for cities as post-Peak Oil supply models show fuel supply currently declining by 1.5% to 6% per year (Dantas, Krumdiek et al. 2010), while fuel demand is increasing internationally. The International Energy Agency (IEA) reports that oil demand in India is expected to increase by 3.9% per year and the demand in China to increase by 3.5% per year until the year 2030, compared to a world growth rate of 1% (Sheppard 2009). This large growing international demand for oil on private vehicle commuters will not only affect household living affordability, but also citizens' decisions about where to work – an issue that affects productivity in cities where high-quality employment is highly centralized. The sprawling suburbs of the western world were initiated on the criterion that oil was cheap (Newman and Kenworthy 1991). This view involved little foresight into the planning challenges it would create for cities in the 21st century. For those who argue that the alternative to sprawl – that of building densely – is unaffordable, then they overlook that home prices, like other goods, are determined by supply and demand (Glaeser 2011). High prices in developments such as TODs merely indicate that these forms of dwelling are valued and not in great enough supply. Lastly, urban form has a direct impact on accessibility as certain density thresholds exist below which public transport is not viable (Newman and Kenworthy 1999). Additionally, it has been argued that high densities enhance accessibility by supporting a greater scale and diversity of amenity within a given area (Jacobs 1969).

The implications of decisions made today – as they impact on society, the economy and the environment – on city life leading into the future are vast. Cities are complex systems and as the world continues to urbanize, well-informed decisions will be vital for preparing them to accommodate the population increases they are expected to support.

2.3 Why Cities Exist and Determinants of Their Size and Location

An existing view of early cities stipulates that they were founded predominantly because of their ability to impose and maintain administrative control on surrounding countryside while reaping the additional benefit of manageability for defensive purposes (Meyer 2000). The advantages of specialization and offering surplus for trade and sustenance of a kingdom or empire's military forces are an economic efficiency identified even in the earliest of cities, though prior to industrialization the dominant reasons for city formation were most likely political and strategic rather than economically based (Meyer 2000). During the post-industrialization period true large cities began to emerge under economies of scale as machinery allowed production at unprecedented levels. Transportation of products and their durability during transport became relevant considerations in location choices. Factor endowments of extractable resources and geographic advantages fostered specialization whilst trade between and among cities would occur because of comparative advantage.

For many cities, initial site location is typically determined by the proximity to some environmental feature of strategic importance. This is what is sometimes referred to as a 1st Advantage by economists and implies in many cases the presence of a bay or river mouth where a port can be established or an endowment of forest or mineral deposit that can be harvested or extracted for economic gain. In such circumstances, settlement occurs because of location-based externalities. Jane Jacobs in *The Economy of Cities* (1969) discusses how cities have historically been the primary sources of innovation, initially founded on the basis of proximity to a valuable resource. Once established, settlements grow in scale as they assimilate production capabilities from sources from which they may previously have received imports. This process is called *import substitution* and can be construed as one of a number of ways in which cities can grow. Cities also grow because the cross-fertilization of ideas from the diversity of economic players in settlements and cities leads to the sporadic and often unforeseeable discovery of new products, production processes, and services. Jane Jacobs explains how inventions typically occur in cities and are subsequently transplanted into the hinterland where land is cheaper and more bountiful. According to Desrochers and Leppala (2010), it is the view of Jane Jacobs that specialization is a transitory phase that inevitably leads to a dead end; thus, the true role of cities and indeed their purpose is to foster diversity and breed creativity.

Many economists and economic geographers writing before and after Jacobs have tried to explain the existence and structure of cities through models to understand better how they form and organize themselves. Among the first was von Thunen (1826) – an economist and landowner in North Germany – who, for the purposes of his model, envisaged a rural isolated town set on a homogeneous area, void of any potential environmental advantages. He assumed that the town is supplied by a number of agricultural goods that can be cultivated at varying degrees of intensity and that vary in their cost of transport. Activities, according to his model, would organize themselves into concentric bands emanating from the town centre where land rents would be at their highest and dissipate gradually with distance. The result in a competitive market would be that the activities with the highest transport costs would locate closest to the centre and this location decision would inadvertently affect the transport costs of other goods producers that now have to locate further away.

Von Thunen's model, though basic and rather intuitive, offered substantial insight into economic organization for his time. A major shortcoming of his work, however, was that it failed to say anything about the determinants of the size or number of towns that compromise polycentric economic landscapes as occurs in the real world. Henderson (1974) begins to offer some solutions to these issues in his general-equilibrium model. He discusses the optimal city size being achieved by the balancing of the agglomeration forces that cause the formation of cities and congestion costs that prevent cities from getting too large. Any time congestion costs for commuters outweigh the wage (productivity) benefits of residing in a city, there exists an opportunity for a corporation to create a new "edge city", which is often a situation taken advantage of by land developers in many countries (Fujita, Krugman et al. 1999). As congestion costs affect everyone equally whilst economies of scale are industry-specific (or industry cluster-specific), Henderson's model sees the emergence of specialized cities where no industry resides that does not benefit from the scale of activity while simultaneously contributing to its congestion.

While Henderson's model was rather revolutionary for its time, its shortcoming was that it was essentially aspatial: it considered only monocentric urban forms and disregarded factors affecting distances between centres. From the field of geography emerged a significant contribution to this dilemma that predated Henderson's general equilibrium model (though was not considered in it) and came from the works of Christaller (1933) and Losch (1940). Their Central-place theory was a commentary on efficient urban systems formation, which differed from past conventions that typically studied cities as single entities. Central-place theory postulated that while agriculture was land-intensive and would be dispersive, the firms that serviced farmers and provided equipment and machinery to them would be subject to economies of scale. On a featureless plane, the resulting efficient organization of human activity would be in the form of 'central places', which would spatially organize into a latticed structure. Because of the presence of transportation costs, the market areas that emerged would be hexagonal. Their theory also considered a hierarchy of settlements that ranged from hamlets to regional capitals. Two forces in particular determined which hierarchical status a settlement would achieve: threshold and range. *Threshold* was the minimal market size required to bring about the provision of a particular good or service, while *range* was the maximum distance that consumers would be willing to travel for a particular good or service before the inconvenience outweighed the benefit. As settlements increased in size, they would also increase their number of functions and share of higher order services.

The shortfall of Central-place theory was that it served more as a description rather than an explanation of an efficient spatial economic structure. As such, it lacked a number of ingredients required for an economic model, in particular an account of the forces that result from the emergent behaviour of individuals seeking to maximize their welfare. Krugman (1991) published a paper that became the first of a body of works that is now known as New Economic Geography (NEG). NEG embraces elements of the works of von Thunen, Henderson, Christaller and Losch, while making other advancements to model the emergences of spatial economies. The seminal work of Krugman produced a core-periphery model that showed how increasing returns at the firm level in the presence of transportation costs and mobile productive factors could lead to the emergence of an industrial centre that provided goods to an agricultural hinterland. The emergence of a core-periphery model is a result of centripetal and centrifugal forces that cause people and firms to agglomerate and disperse. Centrifugal forces arise because agricultural land is an immobile factor of production and as such, so are the farmers that work the land. Hence, agriculture becomes a force causing development to disperse. Centripetal forces are more complex because of their circular nature. Increasing returns and mobile labour in manufacturing cause firms to agglomerate. Labour, as a mobile factor, is attracted to these centres because the increasing returns in agglomerations make firms more productive under monopolistic conditions. The increased productivity leads to higher wages that attract more employment, which causes the centres to grow, which in turn increases the varieties of differentiated goods. Because city-dwellers are both producers and consumers of their products, the growing centre creates a home-market effect. Because of transportation costs, it will be more economical for firms to locate where the markets are the largest and ship to smaller ones. Fujita and Mori (2005) summarize the centripetal forces as being a result of forward linkages, (whereby consumers enjoy a benefit of being located near producers), and backward linkages (whereby producers have the incentive of locating where markets are large). Via this process, influenced by the cost of transport, increasing returns and the mobility of productive factors, a dualistic development of an industrial core and agricultural periphery emerges.

Subsequent works by Krugman, Fujita and Mori have extended this methodology to create a general equilibrium model for polycentric urban formation. Fujita et al. (1999) introduce the concepts of *market potential* and *agglomeration shadow* to address issues of the size, spacing, and number of settlements in a spatial economy. Implying much of what was meant when Central-place theory utilized the terms 'threshold' and 'range', market potential refers to the number of potential consumers in a given area and *agglomeration shadow* can be construed as the space between industry-specific agglomerations where it is unprofitable for a firm to locate. Their model assumes development occurs on a featureless, uni-dimensional continuum and the evolution of the spatial economy unfolds as a storyboard process. As the economy's population gradually increases, the urban system self-organizes into a hierarchical system of settlements. The larger cities are deemed "higher order cities", as they include the presence of higher order activities that benefit most from larger agglomerations. As cities grow outward, peripheral populations offer a large enough market potential for lower order industries to relocate or establish new operations. Such activities might include low-value retail and basic amenities. These new settlements are considered tentative because they lack the critical mass necessary to achieve a "lock-in" effect. As a series of new lower order settlements emerge along a continuum, eventually one will become "upgraded" to a higher order settlement under the provided conditions that the market potential is sufficient to support the settlement's increase in size and it occupies a space beyond the agglomeration shadow of another higher order settlement. Urban growth is then predicated on market size and the proximity to other urban settlements, yet at the same time it evolves through a self-reinforcing process. In reality, the initial location for the foundation of a new city will be determined because of a strategic decision, typically because of the proximity to some favourable aspect of the location that has a catalytic role. However, via the self-reinforcing process of urban growth, eventually a certain scale will be reached where the initial advantages of the location will be overshadowed by the advantages of the agglomeration itself (Fujita and Mori 2005).

Such models, created to understand how the spatial economy evolves in terms of the number and size of settlements and the distances that separate them, are illuminating for developers and policy-makers responsible for growing metropolitan regions and economically activating centres within them. A common agenda item in sustainability and planning strategies, for instance, is to create a networked city where strategic activity centres are linked by efficient transport connections (both private and public) to remove pressure from central business districts harbouring the majority of high-quality employment opportunities.[6] Dispersing high-quality employment around a city relieves the burden on transport networks and the need for workers to commute great distances – both of which are significant contributors to greenhouse gas emissions. Further, dispersal improves the employment self-containment of sub-regions. Vogiatzis et al. (2009) demonstrate that despite strategies to decentralize employment in Adelaide, the historical trend of CBD employment growth continues. Understanding the determinants of centre size and growth in their relation to surrounding centres and the advantages of their location can help inform the procedures and strategies for achieving a more decentralized urban form.

2.4 How Transport Shapes Cities

Our cities would not have evolved into their current states if it were not for the automobile. Faster travel speeds enabled by the automobile allowed people to cover vast distances in comparison to active forms of travel, which dichotomised many of our city's

[6] See planning documents such as *Directions 2031* for Perth, *Melbourne 2030* and *Metropolitan Plan for Sydney 2036* for discussion of networked sub-centres to improve employment accessibility.

regions into areas of primarily (or even exclusively) housing and areas of amenity and employment. As travel is not a need in itself, but a need derived from the purposes we attribute to the destinations we venture out to reach, the incessant issues of congestion from city growth raise serious issues with what the automobile has facilitated to be the business-as-usual trajectory for development – namely sprawl.

Newman and Kenworthy in *Sustainability and Cities* (1999) describe how cities have evolved over time from 'walking cities' to 'automobile-dependent' cities. From the earliest city settlements in the Middle East until the 19th century, urban form was based on walking and characterized by densities of 100 to 200 persons per hectare. The land-use was mixed, streets followed the organic contours of the landscape and the geographic footprints of these cities were rarely larger than what could be traversed by foot in an hour (approximately 5 km across). Then from the 1860s, new transit technologies in the form of trains, trams, and street cars allowed development to push out while still maintaining travel-time diameters of roughly one hour. Sub-centres emerged along these transit routes that maintained walking-city characteristics and mixed-use qualities, while only slightly reducing their densities (down to 50 to 100 persons per hectare). During this era most American and Australian cities formed, many with certain areas that maintain characteristics of this time. In the 1930s, a U.S. consortium comprising General Motors, Firestone Tyres, Mack Trucks and Standard Oil, called National City Lines, bought out the transit systems in 45 American cities and subsequently closed them down. Subsequently, development was free to spread out with no encumbrance from density thresholds required to maintain the viability of transit networks. While cities around the world were affected by the introduction of the automobile, European cities did not experience these effects to the same degree as North American and Australian cities (Newman and Kenworthy 1981).

In addition to the closure of transit services by stakeholders in the diffusion of auto-dependency, other forces were also at work. While cities develop to improve economic efficiency by removing the friction of space (which, in effect, imposes a cost on economic activity), other matters require consideration. Cities fulfil a range of social, psychological, civic and administrative purposes, all of which operate under different optimum conditions. Writing half a century ago, Lampard (1955) contended that economists did not consider the economic role of cities in any comprehensive way. Rather, it was demographers, architects, planners, sociologists and political reformers who concerned themselves with the evolution of cities. He remarks that, "In an age of comprehensive, graduated transport services, people even questioned the necessity for living in cities at all" (p. 83). The people's genuine disgust for urban conditions – marred by crime, pollution, noise, congestion, etc. led to the romanticism of rural and suburban living.

Mumford (1961) discusses the motivations of the early creators of the suburbs. The allure came from individuals seeking to live a life on their own terms, where they could perform the roles of community life in the care of their own family and without having to moderate their behaviours for the appeasement of others. Mumford sees the suburb as a way to escape the defects of society and city life without forgoing the benefits of urban civilization. One could find reprieve, solitude, and an abundance of nature in a setting that allowed them to make a home in their own perfect image. The downfall of this alluring way of life, Mumford explains, was its widespread attractiveness in the 20th century that led to the flooding of uniform development and indistinct homes across the lands that left little nature in their wake, an argument also taken up in the writings of Stretton (1971) who gives a specifically Australian context. None of these places would have been accessible nor would this urban form have been able to evolve, if not for the automobile. No longer was there a restriction on where one could live while still having access to the types of amenities and opportunities deemed necessary for a modern way of life (i.e. benefits restricted primarily to urban settings). This movement towards urban sprawl in North America and Australia, however, did not occur without some help from federal and state planning authorities.

Significant catalysts for the proliferation of suburban living were the public housing programs and initiatives that followed the First and Second World Wars. The War Services Home Scheme in Australia was implemented immediately after the First World War and gave financial assistance to returning service personnel to buy or build new homes. The level of assistance peaked in 1921-1922 before slowing in the 1930s and 1940s. The programme was revived and intensified after the Second World War (Neutze 1977). North America had similar schemes around this time as well, led by the Federal Housing Administration (FHA). The dream of home ownership was made a viable and attractive option through a range of financial and institutional policies that fuelled the dream of owning a single detached unit in a quiet neighbourhood. At the same time, the inner city dwellings in Australian cities that were fundamentally labelled as 'slums' were slowly being replaced with development that suited the plans for the urban centres. Rapid suburbanization did not, however, go completely uncontested. Gleeson and Low (2000) give an example of Sydney's efforts in its County of Cumberland Planning Scheme of 1948 to establish a greenbelt to inhibit the range of sprawl. As the effect of the scheme was to rationalize rather than to prevent or slow suburbanization, it did not withstand the outcry of disapproval by private builders, developers, speculative builders and the Housing Commission who banded together in contestation. Such schemes threaten the 'Australian Dream' of single detached home ownership, the economic opportunities associated with housing supply, and the increasing pressure of housing provision for population growth. Even today, the convention of housing provision on the fringe represents a safe, low-risk and financially attractive form of development to many developers and investors.

That substantial portions of Australia's and North America's urban histories have been shaped by the automobile creates a significant challenge for planners. For one, it is difficult to change perceptions and values associated with a customary way of life. In *The Urban Wilderness: a history of the American City*, Warner (1972) contends that Americans "have no sense of where cities came from, how they grew, or even what direction the large forces of history are taking them" (p. 4). I hazard to say that any country without roots back to medieval times would not be any different. Without a sense of history, it is difficult to imagine how cities with a relatively short period of existence could revert back to a type of urban form that pre-dates the era in which they experienced a substantial amount of their growth – namely the era of cheap fuel and the automobile.

The effects of transportation (and the private vehicle in particular) on urban development and urban form can also be examined via econometric models that investigate the two-way relationship between road network congestion and road network provision. This is an empirical approach frequently embodied in the study of *induced travel demand*. Opponents of the view that cities can "build their way out of congestion" draw upon this concept to fuel their argument. It suggests that adding road network capacity will only spur new travel or divert trips from other routes and quickly return roads to their previously congested states, rendering efforts to relieve congestion futile.

One contributor to the induced travel demand effect is the impact of increased travel speeds on real estate development. Areas of good highway access are highly sought by real estate developers because of the profit opportunities they pose (Voith 1993; Boarnet and Chalermpong 2001). The effect is an endogenous process where accessibility increases land values that can be exploited for economic gain and the increased development in turn adds to the congestion levels of the road network. Cervero (2003) confirms this econometrically in his study investigating the occurrence of real estate development in corridors that have experienced road network improvements in California. His model explains two-thirds of the variance in corridor-specific development and shows a lagged effect in building construction of two years. This responsiveness of construction to past average road speeds suggests a causal link, giving evidence that improved driving speeds increase the attractiveness of properties to developers of well-serviced transport corridors. Much of this developmental activity will occur at the outer limits of cities, where the presence of cheap land can be exploited for significant profits after new transport projects improve access to the area. Land-uses in existing areas can also respond to increased travel speeds, represented by certain activities moving to more distant locations that can be accessed by commuters without their total journey times being affected (DeCorla-Souza 2000).

Another means by which transport investment will affect the form of the urban landscape is through the investment decisions made on the types of travel modes that infrastructure should support. Investment in roads, as opposed to public transit, reduces the generalized cost of travel for car users, thus attracting patronage away from transit services (Cervero 2002). Luk and Chung (1997) study the effects of the opening of a new major arterial freeway in Melbourne and find that over a mere 7-week period, passenger numbers on the Dandenong train line fell by about 14 per cent (a decline 10 per cent greater than what the region experienced as a whole over the same time period). Goodwin, Hass-Klau et al. (1997) study the opposite effect of how road capacity reduction can affect traffic levels. They find a 25 per cent reduction in traffic levels in areas across Europe, North America, Australia and Japan that reduced their network capacity for cars. As a result they make the assumption that these reductions occurred through fewer discretionary trips being made and commuters switching to alternative modes such as transit, cycling and walking.

The challenge for many areas in North America and Australia is that automobile dependency has set in so deeply that transit, cycling and walking have largely become unviable alternatives. Many suburbs simply do not have the densities to support transit servicing, or opportunities for residents to work or shop near their homes. This is not a situation that can easily be undone or altered in the short term. Transport infrastructure (and the form of development that spawns around it) is inherently difficult to alter once established. This works both ways, for sprawling suburbs or walkable centres. Urban forms shaped by the transport technologies of the past tend to be lasting. Densities in outer suburbs will continue to be low until the housing stock is replaced, which may take at least 50 or 60 years from initial construction. Similarly, we observe that old walkable centres of cities internationally maintain their historic characteristics over hundreds of years. Ultimately, planners should determine the kinds of urban forms they wish to achieve and maintain and then invest in suitable transportation choices to support them. Increasing road network capacity will relieve congestion; yet induced travel demand reveals that the effects will be short lived. Cervero (2003) sums up this situation nicely: "Congestion relief… does not necessarily make for a sustainable and liveable metropolis. Thus residents of places that are able to build themselves out of traffic congestion might not necessarily like what they get" (p. 159).

2.5 Managing Transport Costs – Conventions, issues and innovations

The common centrifugal force among the urban growth and organization models of von Thunen, Henderson and the originators of the NEG models is the existence of transport costs. Transport costs comprise a number of factors, some of which apply strictly to private transport and strictly to commercial transport separately and others that apply to both. Generalized cost of travel is calculated from a combination of 'out-of-pocket' costs and time costs. The former may include the cost of fare (in the case of public transit), the amortization of vehicle value, fuel costs, and any expected toll costs. Commuter travel time costs will vary according to whether journeys are expected to be discretionary or work-related and the traveller's estimated level of income. Generalizing the cost of travel on a per kilometre basis simplifies calculations dealing with the economic consequences of congestion and transport projects. The fundamental component of generalized costs of travel that varies with city size and location is the travel-time component.

In NEG, the shipment of goods typically takes on an iceberg effect, which assumes that products lose value over distances rather than considering transport costs separately from the value of the goods being shipped (Fujita and Mori 2005). The key implication of travel is that when transport networks are set up inefficiently, which is most prominently an issue in larger cities and centres, agglomeration diseconomies occur. That is, the benefits sourced from large concentrations of economic players dissipate, rendering proximity counter-productive because of frictions in the transportation network (Graham 2006).

The issue of mitigating the effects of congestion has a number of potential remedies that include adding lane-kilometres or building new routes entirely, applying demand management principles (which include actions such as congestion pricing, flexible parking and flexible work schedules), making changes to traffic control infrastructure to keep the flow of traffic running more efficiently, improving public transport services, or being proactive with land-use management and zoning. Generally, the most common approach is to add lane-kilometres or new routes, but congestion pricing and traffic control changes can often prove just as or even more effective. In fact, Duranton and Turner (2009) in their study of the effects of interstate highways, public transport and congestion pricing on reducing vehicle kilometres travelled (VKT) suggest that congestion pricing is the most effective solution. They find aggregate city demand for VKT with respect to lane-kilometres to be very elastic and the effects of adding public transport capacity to reduce road network congestion to be negligible.

In 2008, the U.S. Department of Transportation Federal Highway Administration released a report that studied the success of congestion-pricing schemes around the world.[7] Findings convincingly showed that congestion-pricing schemes effectively meet their principal objectives and tend to last for long periods, while the acceptability of such initiatives depends on the clarity and severity of the problems, and the most resonant need is not always congestion but can be another issue such as pollution. In a 1998 white paper by the U.K. Department of Environment, Transport and the Regions the "predict and provide" mantra was jettisoned because, as Noland and Lem (2002) comment, meeting the infrastructure needs of unconstrained growth is simply financially, environmentally and socially unsustainable. While the evidence suggests private transport infrastructure investment is not a panacea for traffic congestion, it cannot be avoided altogether because the productivity of centres relies on agglomeration economies which necessitate people to have access to these employment hubs. It is not the volume of people that is the issue, but

[7] See the report by Bhatt, Higgins et al. (2008) prepared for the U.S. Department of Transportation for further information.

the effect of their movements on reducing the accessibility for others. The greatest charge that critics of highway expansion deliver is that of induced travel demand eroding away the travel-time savings of any project, and many regional transport plans get held up in political and legal squabbles over this (Cervero 2003). Empirically this has been a hotly researched topic because, to the degree that travel forecasting models fail to account for induced demand effects, the travel-time savings and benefit-cost ratios (BCRs) of projects are likely to be upwardly biased (Cervero 2002).

The task of empirically proving the existence and estimating the magnitude of induced travel demand is riddled with challenges. To begin with, inconsistencies exist between studies from the outset because their definitions of the term can frequently vary (Lee, Klein et al. 1999); however, perhaps the greatest challenge is addressing the issue that the precise sources of increased traffic may be virtually impossible to ascertain without placing an electronic tag on each traveller affected by a new road improvement, which would be a monumental endeavour (Bonsall 1996). Confounding measurement even more is the existence of exogenous factors such as fuel prices, increased female participation in the workforce, increased car ownership, and an increase in the number of retirees that can all complicate empirical works. The most dramatic challenge, however, remains because redistributive versus generative trips are tough to disentangle.

Redistributive trips include instances where travellers may change their destination (such as where they work or shop) because a new route has been sped up, where travellers may simply change the route they take to get to their usual destination, or when the time of travel changes because it can be the case that in an individual's conventional routine they leave to work an hour early to avoid peak-hour congestion. DeCourla-Souza and Cohen (1999) offer a framework for identifying the sources of travel that should be included in the definition of induced trips and included in standard practice. These sources include the following:

1) Increase in person trip production related development

2) Increase in person trip attraction related development

3) Increase in number of daily motorized person trip productions and attractions per development unit

4) Increase in average motorized person trip distance

5) Increase in share of person travel by private motorized vehicles, and

6) Shift in vehicle travel to improved facilities within a corridor because of diverted traffic from other corridors.

Source 6 is construed as a diverted trip that would only be included in corridor-specific studies, whereas sources 1 through 5 apply to both corridor-specific and regional studies. More generally, what could be said about induced travel demand is that it should only include increased traffic volume that is directly a result of travel mode changes and new trips. DeCorla-Souza (2000) holds that land-use changes most significantly affect these and do so by a lengthening of trips.

The quantitative results that studies produce can vary according to a number of factors. One major factor affecting VKT elasticities is the timeframe of reference. Hansen, Gillen et al. (1993) for instance estimate an elasticity of .15 - .30 over a 4-year horizon, .30 - .40 over a 10-year horizon and .40 - .60 over a 16-year horizon from road network capacity expansion in the U.S. estimated at the corridor level. Fulton, Meszler et al. (2000) estimate a short-run VMT elasticity with respect to lane miles between 0.2 and 0.6 and disclose that they expect longer run elasticities to be even greater. Added capacity will initially increase speeds that will be partially eroded away in the short term by redistributed trips. Gradually, mode shift changes will occur and contribute to traffic levels while in the long term, destination changes, land-use changes and new development spawned by the capacity improvements will make additional contributions to VKT. At the metropolitan scale the estimates in Hansen, Gillen et al.'s study were larger. Generally, the elasticity estimates will be greater with larger areas of analysis because corridor-specific analyses do not pick up changes in traffic flows on feeder routes.

Estimates may also vary by the size of the metropolitan population. Noland and Cowart (2000) find that induced traffic effects are largest in medium-sized metropolitan areas while finding no difference in induced traffic effects between highly and minimally congested cities. This latter point is of particular interest because some sceptics of induced demand may argue that merely adding capacity is not enough to draw new traffic, but rather constrained supply in highly demanded routes must exist. A quantitative difference also exists when comparing urban to rural areas as Goodwin (1996) finds a short-term increase in unexplained traffic volumes of 5.7% in urban areas and 13.3% in rural areas. Furthermore, the type of facility will also have some influence over the magnitude of induced travel estimated as new facilities will draw new traffic from all hours of the day while existing facilities will likely only draw new traffic during peak periods (Ruiter, Loudon et al. 1980).

While induced-demand effects will vary on a case-by-case basis, it can be said with confidence that they exist and that they accumulate over time. The specific circumstances and conditions under which they occur, however, remain open to debate and further research. Because of this ambiguity and to an even greater extent because the effects of induced demand are only marginally included in traffic forecasting models, there is good reason to doubt the validity of giving travel-time savings so much weight in transport infrastructure project appraisal. In Britain, for example, travel-time savings have comprised approximately 80% of total transport infrastructure benefits in cost-benefit analyses of major projects (Metz 2008). Conventional appraisal does not fully capture the ephemeral nature of travel-time savings, nor does it account for numerous peripheral benefits derived from transport network improvements.

As partly captured by the account of sources of induced travel, a common criticism of weighting travel-time savings so heavily is that it assumes travel time is actually saved, but this is not entirely true. This may be the case in the short term for a particular corridor or route, but there is a substantial amount of evidence that shows that travel-time savings are actually conserved for other travel.

With the focus on travel-time savings it is assumed that travellers wish to reduce the number of daily trips that they make, travel shorter distances, and reduce their overall travel time (Jara-Diaz 2000). There is, however, little evidence to support this proposition because in the long-run, travellers will use the travel-time savings experienced from transport infrastructure improvements to make trips to further or additional destinations (Metz 2008). In this sense, time is conserved rather than saved because of transport infrastructure projects. This observation is further supported by studies from Schafer (2000) and Zahavi and Talvitie (1980) which have shown that over time, with changes in transport infrastructure, average daily travel time remains roughly constant at 1 to 1.1 hours per day. In the U.K., data records have shown that the average annual hours spent in transit is 385 and this has changed relatively little over the past 30 years (U.K. Department for Transport 2006), while research by the Standing Advisory Committee for Trunk Road Assessment (SACTRA) (1994) has found that journey-to-work travel times have remained roughly stable for the past six centuries. If there has been any change in mean travel time it has been in the form of a gradual increase, not decrease, as evident in travel-time data collected in the U.S.A. which has shown an average annual increase of 2 minutes of daily travel per person between the years of 1983 and 2001 (Metz 2008). What this leads one to believe is that people generally budget an amount of time that they are willing to dedicate to daily travel and the true benefit of a transport improvement is the destination opportunities available to them. In this sense, it is not travel-time savings that people value, but improved accessibility, which enables one to reach more places, or more distant places, without a significant impact on daily hours of travel. This is the view presented by Metz (2008) when he states, "the entire economic benefit [of travel-time savings] arises from activities at the new destinations and none from time savings" (p. 326).

To illustrate Metz's debated view, one could take the example of an individual who drives from an outer suburb for 30 minutes to get to work. For the purpose of this illustration, assume that this individual is over-qualified for their position, yet it is the best position they could find in a catchment within 30 minutes of where they live. Now let us assume that a road network expansion occurs that allows the individual to get to their destination in 20 minutes. It could then be very well possible that within an additional 10 minutes of travel the person could find another job to which they are better suited, where they would be more productive, and where they would earn a better income. Conventional appraisal would say that this employee saved 20 minutes of travel time per day, where in reality there was no savings of time but a productivity benefit.

This is one of the wider economic benefits that were identified in a report by the U.K. Department for Transport (DfT) in 2006 titled, *Transport, Wider Economic Benefits and Impacts on GDP* (UK Department for Transport 2006). The purpose of the study was to examine ways of improving the appraisal methodology of transport infrastructure projects by identifying missing benefits and setting out methods for their estimation. Out of the study emerged four so-called Wider Economic Benefits (WEBs). They included *agglomeration economies* (WEB 1), *increased competition as a result of better transport* (WEB 2), *increased output in imperfectly competitive markets* (WEB 3), and *economic welfare benefits arising from improved labour supply* (WEB 4).

Agglomeration economies (WEB 1) refer to the benefits that arise out of firms being located in close proximity to one another. They are the centripetal forces that cause economic activity to cluster in the urban growth and formation models of von Thunen, Henderson, and the NEG theorists. In general, the greater the economic density, the more productive incumbent employees and firms become because of their proximity to input providers and consumers of their products or services, the larger labour markets that improve labour matching and ensure a constant availability of workers, and the knowledge spillovers that occur because of knowledge networks embedded in the area. The manner in which transport infrastructure can affect the degree to which an area is agglomerated is via the 'effective density' of the area. Effective density is not only a measure of the economic density of a given area, but also its accessibility as it considers its proximity to surrounding economic mass. The implication of this measure is that with the distance between two economic players being held constant, a transport network improvement could increase their proximity by reducing the time it takes for them to access each other.

The consideration for the benefits of increased competition (WEB 2) to be included in appraisal is derived from the economic theory that non-competitive firms will tend to sell too little and at too high a price. They also lack the impetus to improve their products and services. As shorter transport time and lower transport costs effectively extend the geographic "reach" of firms they are able to compete in each other's markets. The downside of such a situation is that non-competitive firms will be at risk of new entrants forcing them out of business, which is a possibility. Since the DfT assumed the benefits and detriments would balance out, this WEB was not suggested for inclusion in any final appraisal reform recommendations.

The benefit of increased output in imperfectly competitive markets (WEB 3) is a benefit derived from the response to a reduction in business operating costs. If a delivery person can make more deliveries in a day, or an auditor can visit more clients, then a firm has the option of increasing its output or holding output constant while reducing its level of labour. Either way, the unit cost of output is reduced and, characteristic of imperfectly competitive markets, firms can pass on some cost savings to customers while experiencing increased profits.

The final WEB of economic welfare benefits arising from improved labour supply comprises a number of labour-market situations. Longer travel times (or higher commuting costs) can lead to individuals forsaking higher quality, higher paid jobs for less-preferable alternatives situated closer by. In some cases, as with stay-at-home mothers who are deciding whether or not work a few days a week, having suitable work opportunities easily accessible may sway them in favour of re-entering the workforce. A third effect may be that employees may end up working longer hours if they can get home in a shorter period of time. Labour market supply effects are inclusive of all three of these situations.

These WEBs are innovations that constitute an effort to improve upon conventional appraisal practices of transport infrastructure projects. The one of particular interest to this research is that of agglomeration economies. The recent recognition of their applicability to transport infrastructure appraisal can have interesting implications for the types of projects that get approved and built. Until now, this thesis has discussed agglomeration economies in the context of urban formation and as forces driving co-location. The next sections will discuss them in more detail and then review studies that have sought to quantify them.

2.6 Agglomeration Economies – Why we stick together

Agglomeration economies are not a recent area of study but in fact have been researched by economic geographers over the past several decades. They can broadly be defined as the externalities that result from the spatial concentration of economic activity. The foundations of agglomeration economies can be traced back to the work of Alfred Marshall (1890) when he made some critical observations on industrial organization. He documented that vertically and horizontally specialized firms, when clustered, will benefit in each other's presence via three specific channels. First, geographic concentration enhances the efficient provision of intermediate inputs by lowering transaction costs. It also leads to a greater variety of inputs and outputs because it supports the growth of related enterprises. Second, thicker labour markets improve the matching of specialized labour with firms in need of specialized expertise as well as ensure a more constant availability of employment to reduce hiring costs. Third, spatial proximity enhances the frequency and ease of formal and informal interaction that helps facilitate the efficient exchange of knowledge. More generally, the benefits of proximity can be simplified down to a number of cost savings that arise because of the absence of space between economic actors. These benefits are considered externalities because when a firm makes its decision on where to locate, it does so with its own benefit in mind and not that of other firms.

Marshall's specification of agglomeration economies has historically been the one most readily accepted. They apply to industry-specific agglomerations and thus are only present when firms share a line of business or interact in the same supply chain. These externalities are considered external to the firm but internal to the industry in which they operate. They are commonly referred to as *localization economies* or *Marshall's scale economies* (or even *MAR economies* after Marshall, Arrow, and Romer).

A second form of agglomeration economies, which emerged from the works of Jane Jacobs and has been receiving much more attention in recent years, is based on the premise that productivity is enhanced by the diversity of economic activity. For Jane Jacobs, productivity, economic resiliency and economic sustainability are sourced from cities hosting a great variety of firms and activities. At the time of her first writings, her views were not widely accepted because they challenged those of mainstream economics, which generally disregarded the importance of changing times, preferences and creativity (Desrochers and Leppala 2010). Commonly known as *urbanization economies* or *Jacobs's scale economies*, the benefits of diverse economic agglomerations arise from the existence of local public goods, the overall scale of markets, the proximity of input-output sharing, and inter-industry interactions that can result in the cross-fertilization of ideas (Graham 2007a). O'Sullivan (1996) and Scott (1998) explain them as efficiencies gained from firms being able to share common business and public services and city-wide labour market economies that arise from the scale of skills, training, and enhanced labour-market information. External to the firm and the industry, the benefits of urbanization economies are considered to be internal to the city.

Although economic debate often sets these two forms of agglomeration economies as competing paradigms, they both play an important role in the spatial economy. Duranton and Puga (2001) formalize a dynamic general-equilibrium model that treats diverse cities as nurseries that cultivate innovation and the development of new products. Once these products reach a mature state, however, their manufacture can be relocated to specialized centres that offer greater gains from industry localization. In such centres, firms can benefit from lower land rents and lower congestion costs while also experiencing cheaper inputs as greater market size lowers the prices of inputs from monopolistically competitive suppliers. Diversity, on the other hand, plays an important role in the learning stages of firms as they seek to source a wide variety of inputs and learn from a wide variety of processes. Their model is validated by evidence on firm relocations in employment centres in France. Duranton and Puga (2005) then show that at more mature stages, firms may decide to functionally specialize. When the costs of remote management are low enough, certain activities within a firm may be split up between diverse and specialized centres. The existence of localization and urbanization economies in this sense may not be competing but simply more applicable to different functions within firms.

Whether a cluster exists because of urbanization or localization economies is not always evident. While the micro-foundations of agglomeration economies have been theoretically formalized, little progress has been made on identifying the discrete sources of benefit and their relative importance (Andersson, Burgess et al. 2007; Graham 2007a; Glaeser and Gottlieb 2009). In some cases it can be that highly diversified centres may have localized linkages between firms, all the while nest many industries that do not relate through direct linkages (Chinitz 1961; Fujita and Tabuchi 1997). In other cases the mere identification of a cluster is not enough to indicate the precise type of agglomeration economy, whether urbanization or localization, at work. Searle and Pritchard (2005) conduct a study on the ITT cluster of North Sydney and find the motivation for firms to locate in the area is more attributed to the proximity to high-order business services in the central business district than to any intra-industry linkage. They do, however, find sporadic relationships that would suggest some evidence of localization economies and conclude that the existence of the cluster is a hybrid product of localization and urbanization economies. Studies such as this tend to be quite common as many cities and countries take an interest to replicating the success of Silicon Valley in their own jurisdictions. Saxenian (1996) documents the history of Silicon Valley and identifies five pertinent themes: *regional space, social construction, interplay of competition and collaboration, and innovation and risk*. What she describes are the ingredients for localization economies. Research on other technology clusters around the world, however, shows that often the actual presence of localization economies can be overplayed (Coe 1998; Lyons 2000). The ambiguity suggests that the relative importance of urbanization versus localization economies does not lend itself to being generalized very easily, but rather would need to be addressed on a case-by-case basis.

While methods are being advanced to identify and measure the relative importance of the individual sources of benefit arising from agglomeration, evidence to date does give support to the existence of all those identified (Rosenthal and Strange 2004). Of particular interest has been the topic of the mechanisms through which knowledge spillovers occur. They have become such a popular topic because modern cities are no longer predominantly industrial centres but have their cores dedicated to high-intensity knowledge-driven activities. While sources show that this trend of cities to increasingly specialize in knowledge-based activities is global, this is most strongly the case for OECD countries (Karlsson and Johansson 2004).

The importance of knowledge and skill levels in cities is supported by a number of research studies. Findings by Glaeser, Scheinkman et al. (1995), for instance, suggest that skill levels can predict economic growth. Glaeser and Mare (2001) find that new workers to arrive in cities do not receive a wage premium immediately but experience faster growth rates over time, suggesting that human capital accumulates faster in urban areas where people can learn more easily from one another. Glaeser, Kallal et al. (1992) find that urban employment growth is driven not by specialization, but by knowledge spillovers across differentiated industries and activities. Romer (1990) discusses how output in the United States is 10 times greater than what it was 100 years prior to the study and how most of this growth is due to technological change driven by the level of human capital. What all this evidence suggests is that cities are where knowledge creation most effectively takes place and can be disseminated, speeding up growth an innovation.

Threatening the importance placed on face-to-face interaction is the view that in the age of ICT technologies, physical density need not be a prerequisite to facilitate personal interaction as networks can act as capable substitutes (Johansson and Quigley 2004). On the other hand, this view is strongly contested by evidence suggesting that cities will be becoming ever more important for facilitating exchanges in knowledge, delivering an argument that nothing can substitute for face-to-face contact (Glaeser, Kallal et al. 1992; Glaeser 1998; Glaeser 2000). Karlsson and Johansson (2004) make an important distinction between information and knowledge in support of this. They describe information as consisting of uncomplicated messages that are, in essence, standardized data. The transfer of information can be characterized as being low in friction, which means it can be easily transferred over the Internet or other such media. Knowledge, on the other hand, is described as being intrinsically indivisible and as such is typically difficult to transfer without face-to-face interaction to enable all parties to calibrate and recalibrate their explanations.

As Teece (1998) describes it, knowledge is inherently "sticky" and it takes time and experience to accumulate it. It cannot be transferred as easily as information.

Cities, with their ability to bring many people of different and like backgrounds together with relative ease, help facilitate the transfer and progression of knowledge. Glaeser (2011) also emphasizes the importance of physical presence, especially when working with foreign cultures as physical responses in communication can relay a great deal of information that electronic media cannot. Physical presence also aids in building trust, which is vital for maintaining good working relationships. It is more realistic to view ICT technologies as being complements of, rather than substitutes for, personal contact, as business relations are unlikely to be serviced as effectively by physical or electronic interactions alone. It is also via the thick labour market effect that knowledge can be diffused as greater employment flexibility and lower hiring costs enable knowledge exchange to occur when employees switch between employers (Malmberg and Maskell 2001; Scott and Storper 2003; Eriksson, Lindgren et al. 2008).

This discussion of agglomeration economies has shown that there are a variety of benefits that arise from the co-location of firms, some that work toward increasing productivity in narrowly defined industries and some that operate across industries. Depending on the industry to which a firm belongs, the function of a department within a firm, or the stage in a firm's product's lifecycle, the relative importance of the numerous externalities may vary. The fundamental point to note, however, is that despite this variability there is a strengthening role for cities in modern economies. The increasing urbanization of countries as they gain wealth and develop, as well as the disproportionate contributions that major cities make to economic output, give strong evidence of this. While this section has discussed the theoretical reasons for why people will be more productive and create more wealth when located near others, the next section discusses the body of literature that seeks to empirically measure the magnitude of this relationship.

2.7 Empirical Evidence of Agglomeration Economies

As the theory behind agglomeration economies postulates, firms experience a degree of benefit from the spatial concentration of economic activity. This leads to increasing returns with industrial (localization) and urban (urbanization) scale. Estimating industry production, cost or wage functions while incorporating some variable that represents the density, size, or accessibility of observations in a dataset has achieved the econometric estimation of agglomeration forces. A general specification of this can be given by

$$Y_{it} = G_{it}(\cdot)F(X_{it})$$

where Y_{it} is a level of output that could be represented by sales revenue, turnover or gross value-added (GVA) in a production function for unit i at time t. In the case of a wage function, Y merely denotes the wage level. The term $F(X_{it})$ is a technology function that comprises a number of factor inputs that typically include total expenses, total number of employees, payroll information to calculate average wage, and a sum of all assets including fixed assets, current assets, debts owed to the firm and current liabilities to represent 'total assets' (Graham 2007c). In the case of a wage function, factor inputs are replaced with wage-determining factors such as age (a proxy for experience), level of educational attainment, occupation type, and sex. The term $G_{it}(\cdot)$ is known as a 'shift term', or some representation of agglomeration economies. While the precise inputs into the econometric models may vary because of the availability of data, care must be given to avoid bias created by omitting crucial variables whilst being aware that high correlations between agglomeration and control variables may confound parameter estimates (Melo, Graham et al. 2009). Cumulatively, the various components described above constitute a theoretical model that is the basic framework by which agglomeration economies have been estimated.

When surveying the body of empirical works on agglomeration economies a number of aspects stand out, differentiating between the studies. Some of these include the following:

- Geographic scale: Geographic units of analysis range from the comparisons of states and countries down to micro-units that can include employment zones, postal codes and wards.
- Geographic location of examination: Studies have been carried out in cities and countries around the world including the United States, Japan, Brazil, the United Kingdom and EU regions. This bears some significance as there is no a priori

reason for the effects of agglomeration to hold across differing countries and political entities (Graham 2005).

- Geographic unit of examination: The type of geographic unit can have some effect on results as well since government agency data are commonly provided according to units of political or administrative delineation. More meaningful units of analysis are those such as *employment*, *travel,* or *work destination zones* in Australia, *employment areas* in France or *local labour markets* in Italy. The challenge is that not all required types of collected data for wage or production function estimations are available at all geographic units of classification.
- Industrial scope: Early analyses of agglomeration economies have focused on the manufacturing sector while for the most part ignoring the service sector (Graham 2005). There is reason to believe that the benefits of agglomeration accrue to service-based industries as well and perhaps to an even greater extent than in manufacturing, as we observe these types of industries in the most densely developed areas.
- Specification of the agglomeration variable: Earlier studies merely use the employment size of the geographic units of analysis. While this measure does give an estimate of the returns to industrial or urban scale, it says little about urban form. Second generation studies tend to use employment density while more recent studies adopt accessibility indexes that account for the economic mass of geographic units and their proximities to other masses.
- Type of agglomeration economy measured: Empirically it is possible to test for the effects of urbanization and localization on productivity separately, or concurrently. As the effects of the two economies do coexist, it is beneficial to control for both since the two are likely to be highly correlated and the omission of one could misrepresent the magnitude of the economies of scale generated by the one included. Some studies differ in terms of whether both agglomeration economies are controlled for or only one or the other.
- Type of data used: Analyses can be carried out using data that take on either a cross-sectional or panel data structure. Cross-sectional data refer to observations made on units at one point in time. Panel data have cross-sectional properties but includes a time dimension as well, thus constitute data on which unit observations have been

repeated over time. This allows for the application of a number of techniques and controls that cannot be applied to models that use data with a mere cross-sectional structure (such as fixed-effects and random-effects estimators). Melo, Graham et al. (2009) show that studies using a panel data structure typically report elasticities that are approximately 2 percentage points lower – a likely result of them being able to control for time-invariant unobserved effects.

- Treatment of endogeneity issues: As the initial site location of cities and centres are typically exogenous, the processes that take over as they grow are endogenous. This means that not only does geographic concentration enhance productivity but also productive areas will attract more people and firms to locate there. This is a common issue in circumstances where two related equations are determined simultaneously, resulting in one or more of the independent variables in an equation being correlated with the error term. This reverse-causality is in some cases hypothesized to bias elasticity estimates upwards (Artis, Miguelez et al. 2009); however, it has also been shown that the bias can work in both directions and be industry-variant (Melo, Graham et al. 2009), and that the effect can be small and in some cases negligible (Graham 2007b; Melo and Graham 2009). The most common method for addressing the endogeneity issue is to use Two-Stage Least Squares (2SLS) and instrumental variables. See Appendix A for a mathematical explanation of the implications of simultaneous equation bias.

While the results of past empirical works vary in the magnitudes of agglomeration forces that they report – most likely because of any number of the aspects that affect research design mentioned above – none of the works surveyed report negative average effects of returns to urban scale or provide counter-evidence of the existence of agglomeration economies.

Nakamura (1985) studies the effects of localization and urbanization in Japanese cities simultaneously using cross-sectional data on 2-digit manufacturing industries. The discrimination of localization versus urbanization effects is achieved by estimating industry-level production functions. He finds urbanization economies are more important among light manufacturing industries and localization effects to dominate in heavy manufacturing industries. Associated unweighted average elasticities for urbanization and localization economies are 0.034 and 0.045 respectively, suggesting a greater importance of localization effects.

Also focusing on manufacturing industries, Henderson (1986) estimates agglomeration economies for U.S. and Brazilian urban areas. Urbanization and localization economies are estimated simultaneously using total and industry-specific employment size respectively. Conclusions are made that manufacturing industries receive more benefit from localization economies and that the benefits are greatest in medium-sized cities that tend to be more specialized. The benefits from employment size are found to diminish in larger cities. He estimates an elasticity of 0.19 with respect to industry size.

Ciccone and Hall (1996) improve on the city size measure by estimating agglomeration economies in manufacturing industries with respect to urban density, which is considered an improvement on the employment size metric as it emphasizes proximity, not just scale. Their analysis utilizes state-level productivity data in the U.S. and suggests that a doubling of employment density increases productivity by 6%. They address the issue of endogeneity using instrumental variables techniques, instrumenting with historical development patterns under the assumption that lagged population levels will be uncorrelated with current day productivity levels – a common approach in dealing with the simultaneity bias issue. Ciccone (2000) then carries out a similar analysis on EU regions, getting only a slightly smaller elasticity estimate of 4.5%.

Rice and Venables (2004) make the next novel advancement in the measurement of economic mass when estimating the effects of agglomeration economies in NUTS3 subregions in the U.K. A key issue that they address is that externalities are not constrained within administrative boundaries. With the use of employment size or density is the inherent assumption that neighbours cannot benefit one another across arbitrarily defined delineations between spatial units. This, of course, is not the case as externalities arise over continuous space (Duranton and Overman 2005). Rice and Venables address this by estimating the population of working age within a series of predetermined private-vehicle travel times that form into travel-time bands. They regress these 'employment potential' estimates on productivity and employment composition indexes to distil the influence that agglomeration economies have on productivity levels as indicated by average worker gross value-added (GVA). They estimate the returns to agglomeration with an elasticity of 3.5%

and find that agglomeration effects cease to be important beyond a journey time of 80 minutes.

Combes, Duranton et al. (2008) take an approach to identifying agglomeration economies that allows worker heterogeneity and region-specific non-worker endowments to compete for the explanatory contribution to labour productivity. The sorting of skilled workers across the spatial economy could be attributed to the location choices of firms that have different labour mixes. As such, productivity represented by wages could merely be indicative of the presence of firms requiring high levels of specialized employment. The non-human endowments argument attributes productivity premiums to the presence of physical capital – natural or human-made – such as rivers, ports, climates suitable to economic activity, or a specific technology or infrastructure. Their findings suggest that up to half of the wage disparities identified among French employment zones can be attributed to the sorting of skilled labour with a lesser but still substantial impact from agglomeration economies. Endowments had a minor effect, which supports the argument that once cities are established, growth takes on an endogenous process while the initial reasons for settling an area become less important. They estimate a return to urban density between 0.03 and 0.065, with their preferred estimate being 0.03.

Melo and Graham (2009) also investigate the presence of agglomeration economies using wage functions and address the concern of the sorting of skilled labour to denser areas. Their study employs data on U.K. workers and uses the fixed-effect estimator to measure the impacts of both employment density and market potential on wage levels across an aggregate of industry types as well as a number of broad industry subcategories. The application of the fixed-effects estimator removes the influences of time-invariant unobservable worker heterogeneity on the elasticity estimates produced. Their results show that controlling for the sorting of skilled labour can roughly halve the elasticity estimates produced, while controlling for reverse causality lowers the elasticity estimates when agglomeration is represented by employment density but raises them when represented by market potential (effective density). Their aggregate industry elasticity estimated with respect to market potential is 10.1%.

A summary of the findings from a number of empirical studies including those discussed up to this point, appears below in Table 2.1.

Table 2.1: Survey of Results from Empirical Works on Agglomeration Economies

Author(s)	Elasticity Estimate	Country	Independent Variable
Nakamura (1985)	.03	Japan	City size (population)
Moomaw (1985)	.07	USA	City size (population)
Henderson (1986)	.19	USA	Industry size (employment)
Ciccone and Hall (1996)	.06	USA	Employment density
Ciccone (2000)	.05	EU Regions	Employment density
Henderson (2003)	.03	USA	Industry size (plants)
Rice and Venables (2004)	.04	UK	Proximity / travel time
Combes et al. (2008)	.03	France	Employment density
Graham (2007a)	.12	UK	Effective density
Artis et al. (2009)	.02-.06	UK	Proximity / travel time
Melo et al. (2009)	.10	UK	Market potential

The direction of empirical works on agglomeration economies made a noticeable shift when Venables (2003) showed that the benefits of agglomeration could generate substantial gains when applied to transport infrastructure investments. These gains arise because transport projects effectively raise urban densities and strengthen the agglomeration economies experienced by firms, thereby improving productivity levels and creating a level of benefit not captured in conventional transport project appraisal. This theoretical proposition incited interest within U.K. transport planning authorities and led to a number of significant advancements on the subject matter, primarily attributed to the developments of Graham (2005; 2006; 2007a; 2007b; 2007c).

In his initial work, Graham (2005) uses a production function framework in a translog system to estimate the impacts of effective density – a composite measure of density and distance – on firm level productivity. He carries out his analysis on a number of 2 and 3-digit SIC industries using U.K. ward-level data and across a range of industry types that fall into both the manufacturing and service sectors, which was a first among empirical works on this matter. Studies had traditionally focused on manufacturing industries or aggregate employment regardless of industry type. The rationale behind including service sector industries in analysis is that, as demonstrated by the high concentrations of services in dense urban environments, agglomeration economies apply equally (if not more) to services as they do to manufacturing firms. As transport investments affect all industries and in regions that differ in their employment mixes, carrying out industrially disaggregated analyses allows the impacts of transport network improvements on productivity to be estimated for regions with a greater degree of accuracy. The study's service-industry results appear below in Table 2.2.

Table 2.2: Service Sector Elasticity Estimates Using UK Ward-level Data

Industry	Elasticity Estimate
Finance and insurance	.294**
Real estate activities	.084**
Computer and related activities	.072**
Business and management consultancy activities	.176**
Architecture and engineering services	.244**
Advertising	.353**
Labour recruitment and provision of personnel	.125*
Motion picture and video activities, radio and television	.447**

Source: Graham (2005)

In another similar report, Graham (2007a) produces a number of elasticity estimates across a variety of industry types but this time at a higher level of industrial aggregation than in the study discussed above. These more general results are shown below in Table 2.3.

Table 2.3: Broad Industry Sector Elasticity Estimates Using UK Ward-level Data

Industry	Elasticity Estimate
Manufacturing	0.077
Construction	0.072
Distribution, hotels, & catering	0.153
Trans, storage & communications	0.223
Real estate	0.192
IT	0.082
Banking, finance, & insurance	0.237
Business services	0.224
Whole economy	0.119

Source: Graham (2007a)

Results generally show agglomeration benefits to accrue to a greater extent in service-type industries, which is consistent with observations of denser urbanized centres being dominated by knowledge-intensive activities.

Another feature aspect of agglomeration economies is that while disparities exist across different industry types, they also exist across different scales of development. Graham (2006) investigates this situation by creating two alternate measures of effective density – one that weights the proximity of economic mass by the generalized cost of travel and the other by Euclidean (linear) distance. While the former is sensitive to congestion and travel constraints, the latter is not as it is void of any temporal parameters. The result is that when comparing the two measures across various scales of development, effective density weighted by the generalized cost of travel (ED_G) will grow less than proportionally with effective density weighted by linear distance (ED_L) as one moves to more urbanized areas. If ED_G is the superior measure of agglomeration then a model using ED_L will be biased because it will contain an additional error term, $\beta_{ED}\left(ED_G - ED_L\right)$, with which it will be correlated. Since ED_L increases at a greater rate with increased urbanization than ED_G, ED_L will be negatively correlated with $\beta_{ED}\left(ED_G - ED_L\right)$ and the elasticities estimated using ED_L will be downwardly biased. In this Graham shows that at higher effective densities the differences between the two measures can be as much as a little over 30%, and when increasing returns to agglomeration are present, this suggests an opportunity for the investment in transport infrastructure to reduce travel-time constraints and increase urban productivity.

On another matter, recent works have more intently approached the issue of distinguishing between urbanization and localization effects to identify the one that is likely to dominate in improving productivity levels across industry types. Duranton and Puga (2005) discuss the challenge associated with this because the micro-foundations of agglomeration economies are mutual to both narrowly defined and inter-industry scopes. It is because of this that identifying the individual effects of urbanization and localization economies may be difficult.[8] Furthermore, because some industries will be highly localized in areas that tend to also be highly urbanized, the correlations between measures of localization and urbanization will likely result in high levels of multicollinearity in econometric estimations. Measuring localization and urbanization economies in the same unit form could exacerbate this issue.

There is evidence, however, that addressing the separate agglomeration effects by applying a different unit of measure to each can reduce the confounding impacts of multicollinearity. Graham (2007c) employs a method of representing urbanization economies by an effective density (or market potential) index and using a distance-band approach to measure localization effects. The premise behind the latter is that as localization effects are understood to attenuate rather quickly with distance (Rosenthal and Strange, 2003; 2008) and take place over small scales (Duranton and Overman (2005), estimating industry-specific employment levels within a series of distance bands of a plant or geographic unit can be a suitable means to reducing the impacts of multicollinearity when the measurement of urbanization takes another form. The results identify localization

---

[8] This can be effectively illustrated by the findings of Searle and Pritchard (2005) in their study on the localized ITT cluster in North Sydney. Here is a situation where looking at employment data would suggest that the localized cluster exists because of localization economies; however, according to their findings the firms in the cluster are rather poorly integrated and are more likely to exist because of the proximity to the CBD. As such, urbanization economies may be more responsible for their choice of location. Because of the high degree of centrality and as a centre of major economic mass, looking at employment data for North Sydney would also score the area high in terms of effective density. Not only does this affect multicollinearity in econometric estimations, the precise types of interactions existing in the area would not reveal themselves through the analysis of financial and employment data.

effects to accrue to 13 of the 27 industry-groups analysed and urbanization effects to accrue to 14 of them. Localization elasticities are reported at weighted averages of 0.03 and 0.01, with urbanization elasticities at 0.07 and 0.19 for manufacturing and services sectors respectively. Consistent with pre-existing evidence, the study suggests that localization effects are restricted to being within 10 km of a localized cluster – although confounding influences of multicollinearity still arise.

The spatial decay of agglomeration benefits is of interest because as transport investments can affect the accessibility and temporal-proximity of places, little is known about how far-reaching these effects are and if they differ across industries. Rice and Venables (2004) give us an understanding of how manufacturing wages decline with distance from economic mass while Rosenthal and Strange (2003) and Duranton and Overman (2005) – though not directly addressing the impacts on productivity – indicate that localization effects attenuate sharply with distance. Employing distance bands to measure the rate of decay of agglomeration economies, Rosenthal and Strange (2008) estimate the relationship of agglomeration to human capital and wages, arriving at an elasticity of 0.045 with respect to employment size within 5 km. Industry-specific distances, however, are not taken into account. Graham, Gibbons et al. (2009) investigate this in some detail by considering a more flexible form of the market potential, or effective density, index. They consider the following equation:

$$A_i = Ln \sum_{\theta} L_{\theta i} d_{\theta}^{-\alpha}$$

where a given measure of agglomeration or urbanization A_i at location i equals the log-sum of labour in distance-band θ and location i, weighted by the distance between area i and the centre of band θ. The variable α, which in standard market potential indices is set at -1, is a distance-decay factor that is estimated using non-linear least squares.

The study's results estimate decay factors of 1.0 for manufacturing, 1.8 for consumer and business services and 1.6 for construction. After estimating average elasticities of 0.02, 0.03, and 0.08 for the broad industry categories of manufacturing and consumer services, construction, and business services respectively, they conclude that while the benefits of agglomeration are generally greater for services than for manufacturing, the impacts on the former decay more rapidly.

While the literature on empirical studies may produce agglomeration economies elasticities that vary to some extent, the evidence of agglomeration economies leaves little room for doubt as to whether or not they exist. This being said, perhaps the most common competing paradigm to explain the positive productivity gradient associated with the concentration of economic activity is that of 'firm selection'. Melitz and Ottaviano (2008) study the effects of the trade integration of markets on the "toughness" of competition between firms, which is endogenously related to firm productivity and the number of competing firms. They show that trade and market size affect the degree of competition facing firms, which in turns forces less productive players out of the market. This suggests a rationale of Darwinian selection, or survival-of-the-fittest, to explain the productivity differences between firms in larger and smaller cities or centres. Coombes, Duranton et al. (2009) examine the possibility of this even further by nesting a seminal firm selection model into a standard model of agglomeration and find spatial productivity differences to exist primarily because of agglomeration economies and very weak evidence of firm selection. In fact, they find no systematic evidence of larger cities facing greater effects of selection. They conclude by postulating that poorly integrated markets might show a greater effect of firm selection as markets adjust to new levels of competition. Eventually the effects of firm selection are likely to converge between foreign and home countries, at which point the end result may be no spatial differences in firm selection. They also add that firm selection may only be evident at small spatial scales for consumer services activities where close proximity and high substitutability will have greater consequences in terms of the level and form of competition.

Overall, economic geography has had a great deal of success at quantifying agglomeration economies and measuring the relative importance of localization and urbanization economies, as well as the rate at which the benefits diminish from a source. This being said, the micro-foundations of agglomeration economies – namely proximity to producer and consumer markets, the existence of thick labour markets, and knowledge spillovers – have not been measured directly (Graham 2007b). This is to say, very little is known about the balance of importance between these externalities for different industry sectors and the manner in which they arise. We can say that manufacturing industries rely more heavily on the benefits of reduced transportation costs for sourcing inputs and delivering goods, but by how much in comparison to the benefits of having access to skilled labour? Even more difficult to determine is the movement of knowledge. Transport investments and urban design can bring firms and people closer together, but the specific mechanisms through which the benefits are finally manifested are extraordinarily difficult to measure at an aggregate level. In particular, the urbanization economies that are described in the works of Jane Jacobs are difficult to track because they arise out of a diversity of interactions and engagements that by their very nature are often unpredictable. It is because of these measurement issues that agglomeration economies have commonly been described as existing in a "black box".

2.8 Agglomeration Economies – Beyond the "Black Box"

While the majority of studies on agglomeration economies have focused on the theoretical or empirical investigation of the subject at the aggregate level, some have attempted to better understand the micro-processes generating the benefits. This means both better understanding the relative importance of the three broad benefits of agglomeration – those of labour market pooling, backward and forward linkages, and knowledge spillovers – and also understanding how these benefits are generated.

Still at the aggregate level but beginning to disentangle the relative influence of the benefits of agglomeration, Ellison, Glaeser et al. (2007) conduct a study on coagglomeration patterns of industries in the UK. They carry this out by regressing an industry coagglomeration index on key industry attributes that represent each of Marshall's three agglomeration benefits. For input-output linkages they have data on the share of industry inputs coming from other industries, for labour market pooling they use data on 277 different occupation types, and as a proxy for knowledge spillovers they use data on patent citations, all of which are used to see how needs shared among industries affect coagglomeration. As a result of their study they conclude that agglomeration patterns are far from random and find all three benefits to be significant, yet input-output linkages to be most important followed by labour market pooling and then knowledge spillovers.

Desrochers and Leppala (2010) attempt to glean some understanding of cross-fertilizing activities from the experiences of Canadian entrepreneurs and inventors by conducting a series of interviews. As argued by Jane Jacobs, inventors and innovators either apply a novel solution to a conventional problem, or a standard solution to a novel task. Resonating with this, the authors identify three broad processes for innovation that arise as prominent themes in their interviews.

These processes include: 1) moving know-how or materials between different lines of work; 2) making observations of know-how or materials in one context and applying them in another setting; and 3) formal or informal collaborations with individuals of different skill specializations. While all are found to be important, the one found to be most beneficial across entrepreneurs and inventors that had experienced both diversified and less-diversified environments was the first process. The overall message, however, is that being able to locate in one place whilst having access to a wide variety of specialized workers[9] and suppliers who themselves have links to others is invaluable to the process of innovation. Furthermore, having the opportunity to gain a wealth of experiences by switching between lines of work is also of great importance.

A series of other studies have also attempted to investigate the importance to innovators of having access to knowledge or social capital while adopting a combination of empirical and qualitative techniques, with a focus on high-tech industries. Among the main findings is that firms do not automatically benefit from agglomerations but have to engage in network relations (Burger, Oort et al. 2009). This feeds into the definition of social capital, which can be defined as the resources embedded in social networks that can be accessed for actions. Boshuizen, Geurts et al. (2009) investigate the effects of network activity on the employment growth in high-tech industries. In doing so they count direct and indirect inter-industry network ties to represent the Jacobs hypothesis of urbanization benefits and intra-industry ties to represent the Marshall hypothesis of localization benefits. Additionally they control for the prevalence of business associations, which include those based on facilitating pure knowledge exchange and those that aid in export activity and the commercialization of knowledge. Their results suggest a negligible benefit from contacts outside of a firm's own industrial scope, a minor benefit from same-industry contacts, and

[9] Works such as those by Stigler (1951), Baumgardner (1988) and Holmes (1999) show theoretically and empirically how market size and industry localization are forces that increase the specialization of labour. The process operates such that larger markets enable labour and firms to specialize and in specialization there are certain efficiencies and benefits to productivity that arise. In this sense, geographic concentration encourages vertical disintegration.

the greatest benefit to occur because of association memberships. Associations, however, seemed to comprise mainly members contained within regional boundaries, thus geographic proximity holds in this example as a facilitator of knowledge exchange.

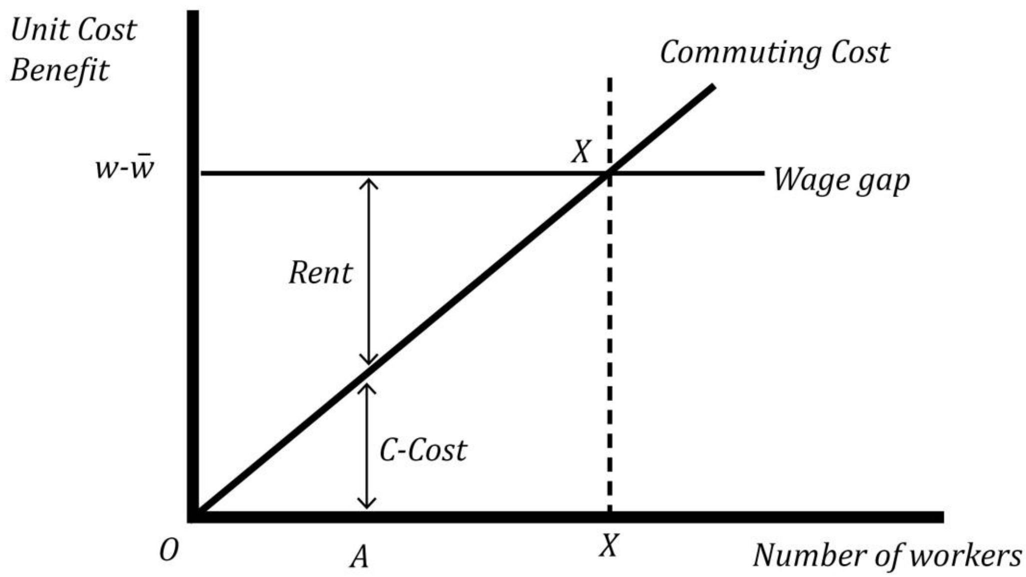
A similar study is carried out by Brandt, Hahn et al. (2008) in which they analyse the network size, density, centrality, cohesion and connectivity within innovation-oriented industries and research-focused firms in northern Germany. By way of network analysis they argue that one can better understand the flows of knowledge in these industries and classify interactions as being long-term strategic, short-term strategic, or human resources related. They argue for network analysis as being a powerful means to revealing opportunities for improving interactions for growing the knowledge economy in cities and regions.

While the relative influential significance of the benefits within the broader context of agglomeration externalities are not the focus of this thesis, these studies are examples of some ways in which the microfoundations of agglomeration are being further studied. The remainder of the thesis will focus on the effects of agglomeration in the aggregate sense and how the benefits, in any balance of proportions, synergize to create a real benefit that can be quantified for use in the appraisal of transport infrastructure projects and planning schemes.

2.9 Agglomeration Economies and Transport Infrastructure

Venables (2003) can be credited with setting out the theoretical foundations for merging the elasticity of productivity estimates conducted with respect to city size with the transport investment decisions that affect commuter travel speeds. In doing so he argues that real income changes can occur as a result of transport improvements. Figure 2.1a gives an illustration of a city in equilibrium, where point X determines the size of the city. The diagram depicts a situation where the wage-gap between city and non-city workers is eroded away by the presence of transport costs. It also depicts the trade-off between rents and commuting costs, as living closer to the city centre will increase living costs.

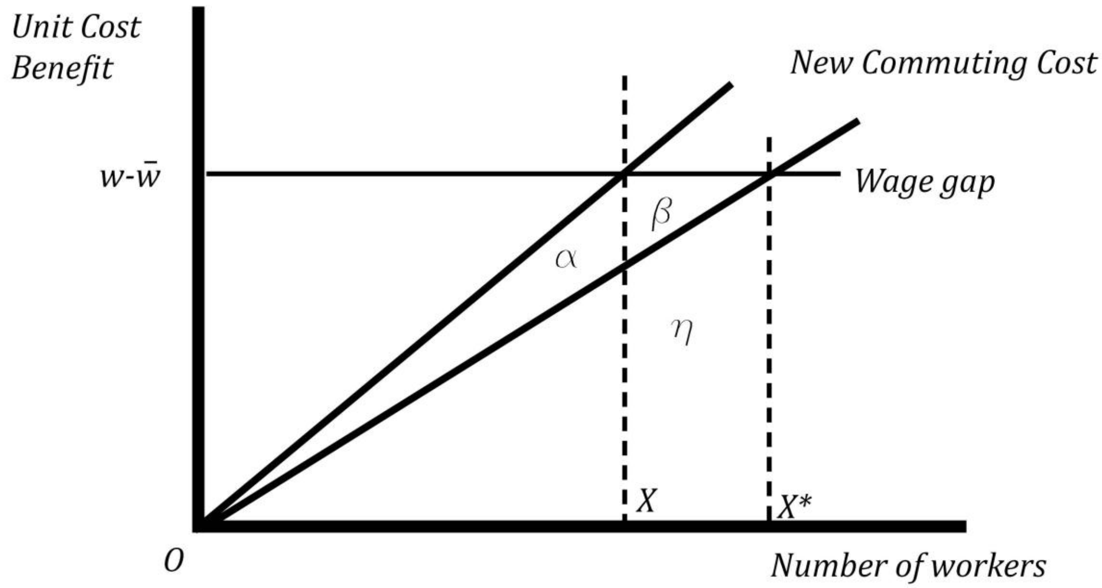
Figure 2.1a: City Size in Equilibrium



Source: Redrawn from Venables (2003)

Next, Figure 2.1b shows the effects of a transport infrastructure improvement on commuting costs, causing a downward shift in the commuting cost curve. Because travelling to the city from outer areas has become cheaper, more people will commute to reap the benefits of higher pay. The effect is that city size increases to X* and the output of the city increases by the amount $\beta + \eta$. The total change in commuting resources is represented by $\eta - \alpha$ which, when combined with the increased output $\beta + \eta$, generates a real income benefit of magnitude $\alpha + \beta$.

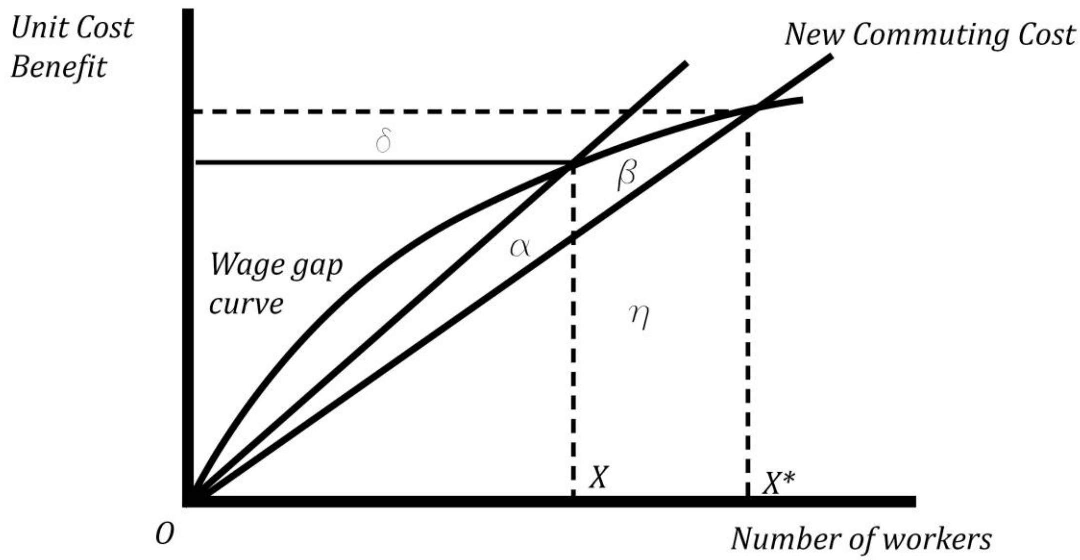
Figure 2.1b: Effect of Transport Improvement on City Size



Source: Redrawn from Venables (2003)

Figure 2.1c introduces the effect of increasing returns because of the existence of agglomeration economies, thus productivity gradient is no longer constant and is represented by a concave line. Equilibrium is established at the point where the wage-gap equates to commuting costs and the externalities of agglomeration increase productivity by the amount δ. Thus, actual real income gain can be represented by $\delta + \alpha + \beta$. The impact on productivity, δ, is what econometric works on the subject of agglomeration economies attempt to estimate. Ultimately, it is not just the size of cities that affect productivity, but also the efficiency of movement within cities that will affect the size of the returns to urban scale (Melo, Graham et al. 2009).

Figure 2.1c: Productivity Effect Resulting from Transport Improvement



Source: Redrawn from Venables (2003)

2.10 Measures, Indices and Indicators of Accessibility and Industrial Composition

The topic of accessibility has a number of applications, such as in travel demand forecasting, land-use planning, and assisting in location decisions for firms and public services. Federal and state government entities are also interested in the matter for planning budget and resource allocations to various local government areas. There are many ways that geographic locations can be compared, and the implications of these are of interest to planners. In the case of econometric works investigating agglomeration economies, the nature of the accessibility parameter specified has significant implications for how applications of the outputs can be used. This section will discuss a number of metrics used for representing accessibility and discuss some of the benefits and concerns for each.

Coombes and Raybould (2001) note that while metrics such as population size and population density are easily calculated, easily used, and universally accepted, their mere simplicity has something to say about how reliable they are at representing the significance, context, and degree of urbanization of geographic areas. Hansen (1959) aptly defines accessibility as "the potential of opportunities for interaction" (p. 73). More specifically, accessibility is represented by the spatial distribution of activities about a point with adjustment for the viability of people overcoming spatial separation. Many of the early works on agglomeration economies compared geographic units because of size. Accessibility, represented in this sense, operates under the assumption that larger markets or areas of employment enhance interactions. Size, however, may be a necessary condition, but it is not ideal for satisfying the definition given above. A prominent issue with size is that says little about the dominant urban form, the context of the geographic units in the broader spatial economy, or the distance between people.

The common unit of measurement, density, ameliorates one of these issues in that it begins to say something about urban form, but it stills fails to account for the context of a unit within a greater aggregation. Another issue with this metric is that it involves some ambiguity with its denominator that must be resolved – that of the specification of geographic area by which population or employment are to be weighted. Administrative units and political boundaries will often include undeveloped land, which may include places occupied by parks, forests, or rivers. Additionally, some measures of density will include the geographic space occupied by transport infrastructure, such as freeways and other roads, while others will omit it. Population and employment densities calculated by adjusting land area for these considerations are often referred to as *urbanized land areas*. Urbanized land area calculations for rural areas, however, do not apply as well as they do for urban areas, because productive land in the former sense does not need to be 'built-up' but could exist primarily in the form of pastures and fields. Even after working through these details, the density measure does not account for the situation of a geographic unit in a larger region.

Accessibility metrics begin to get more advanced as distance is included in their construction. Rice and Venables (2004) use a composite measure of population of working age and distance, where the former is calculated for each of a number of concentric ring bands emanating from a centre. They use population of working age as opposed to population or employment to represent the employment potential for their analysis areas. Distance in their index is represented by private vehicle travel times. In an econometric estimation of the effects of agglomeration on employment productivity, this specification of accessibility is incorporated into a wage function model as illustrated below.

$$\ln y_i = \beta_0 + \sum_b \alpha_b p_{bi} + \sum_j \beta_j x_{ji} + e_i$$

Here, after controlling for worker differences, income (y) in location i is determined by the accessibility of the area, represented as the sum of the working-age population (p) in band b. The populations within each band are estimated according to the following formula

$$P_b = \sum_z \left(\frac{A_b}{A_z} \cdot E_z \right)$$

where the population (P) of band b is determined by calculating the share of each zone's area that experiences overlap with band b, which is $\left(\frac{A_b}{A_z} \right)$, and then multiplying this share by each respective zone's population before summing this result over all the zones for each band.

Variants of this approach to representing the accessibility of an area can be amended by using linear distance instead of travel times to gauge proximity and employment instead of working-age employment, as done by Graham (2007b) and Rosenthal and Strange (2008). While this measure is a vast improvement over using mere size or density, it does involve some shortcomings. For one, the distance bands – depending on their size – do not show much continuity spatially. The effect of accessibility on productivity will be averaged in each band and as such will be sensitive to the band widths selected. Choosing bands that are too narrow will lower the reliability of the estimated populations that reside within, because the calculations of each band's employment is based on the assumption that economic activity is evenly distributed within each zone. Drawing data out from smaller spatially disaggregated units will help this issue but will not completely remedy it. In addition, accessibility will not be equal in every direction because movement cannot occur in all directions with the same ease. On the other hand, choosing bands that are too large will dilute any evidence of agglomeration economies. One particular benefit of using bands, however, is that they do give an account of how quickly agglomeration effects dissipate from a source.

An alternative popular accessibility measure in econometric analyses of agglomeration economies is the 'market potential index', which is synonymous with a 'gravity model'. Also known as an 'effective density index', its formulation is such that it measures the potential of location i in terms of the weighted sum of the intensity of activity in all surrounding areas, where the weights are a declining function of distance (Fujita, Krugman et al. 1999). The term 'market potential' has generally been used in analyses seeking to explain the location decisions of firms – as initially done by Harris (1954) for manufacturing industries in the US – whereas 'effective density' has been used to address the benefits of employment concentration as in the rhetoric of agglomeration economies. The index can be calculated by weighting proximity by linear distance according to the following formula

$$ED_i = \frac{E_i}{\sqrt{\left(A_i / \pi\right)}} + \sum_{j=1}^{n} \frac{E_j}{d_{ij}}$$

where the effective density of area i (ED_i) is equal to area i's own employment divided by its own radius (assuming that the analysis zones are roughly circular), plus the sum of all surrounding areas (j) weighted by their Euclidean distance. The exponent on distance as it stands is -1, however, there is a potential for the benefits of having access to certain industries to decay more or less rapidly than what the current linear specification presupposes. Hansen (1959) explains that a decrease in the exponent means that distance becomes less of a restrictive factor. In other words, a low exponent on distance will mean the impact of employment on accessibility will diminish less rapidly, thus accessibility (ED) will be shifted upwards and reflective of activities or industries for which people are willing to travel longer distances. Conversely, a high exponent will reduce the contributions of surrounding areas to accessibility (ED) and shift the overall measure downward, being reflective of industries or activities for which the benefits diminish more rapidly.

A more flexible specification of the effective density index allows the exponent on distance to vary and can be estimated by non-linear least squares. Graham, Gibbons et al.

(2009) estimate broad industry decay factors and find values for manufacturing to be lower than those for business and consumer services (approximately 1.0 for the former and 1.8 for the latter). This suggests that the agglomeration benefits attenuate more rapidly for business services than for manufacturing, which is reasonable because of knowledge requiring close proximity to be shared effectively and manufacturing relying more heavily on transport networks that are often fairly efficient in cities. Hansen (1959) reports that empirical tests on the magnitude of the decay factor on distance for a variety of activities have ranged between 0.5 and 3.0.

Song (1996) carries out an analysis where he tests the effectiveness of 9 different accessibility measures, including the linear and variable specifications of market potential, in explaining population distribution in Reno-Sparks, U.S. His results suggest that distance to the CBD is the most ineffective at explaining population distributions, most likely because it assumes all employment is located in the city centre and that it is the only destination of any value. Conversely, he finds that the specification of market potential, or effective density, that weights surrounding employment by d_{ij}^{-1} is not statistically bested by any other measure. Also included in the study were cumulative accessibility measures, which do not discount the contributions of employment within a predetermined boundary; average distance and weighted average distance indexes; an exponential distance decay function; and a Gaussian function (which declines gradually at first and then increases with distance).

Graham (2006) shows how the effective density index can be improved by substituting linear distance for the generalized cost of travel. Since the latter includes travel time, it would more accurately measure the level of agglomeration experienced by firms because it would take into account the relative efficiency of transport infrastructure in gaining access to surrounding areas. In other words, congestion and travel speeds would affect accessibility – which is an actuality in many urbanized areas. For the purposes of measuring elasticities of productivity with respect to effective density, Graham (2009) argues in favour of using linear distance because conventional appraisal already includes the benefits of travel time savings. He expresses concern that combining travel-time savings with productivity benefits estimated from a travel-time-weighted effective density could result in double counting.

Another measure useful for measuring industry concentrations, but generally not used in representing accessibility in econometric estimations of agglomeration economies, is the ‘location quotient’ used in economic base theory – the application of which helps one understand the function of a local economy. The theory postulates that the economic activities of a local economy can be classified as being ‘basic’ or ‘non-basic’. The basic sector is made up of firms and employees entirely reliant on external factors, while the non-basic sector depends on local economic conditions. Such is the foundation of ‘base multiplier analysis’, that a region’s economic base is the local economy’s raison d’être and all non-basic sector employment rely on the performance of the former. For instance, take X to represent direct income to a region generated from export activities and a to represent the share of this income spent locally. A second round of local earnings would then occur at the amount of aX, a third-round at $a(aX)$, or simply a^2X, and so on. The cumulative effect can be estimated by the following equation

$$Y = \frac{1}{1-a}X$$

where Y is regional income, X is the first-round income generated from export activities, and a is the share of export income spent locally. As the economic growth generated from export activities leads to regional population growth, new products and services can be provided locally because the market size will allow for their efficient supply. This leads to an increase in a and subsequently greater growth and earnings for the local economy.

The ‘location quotient’ is a measure or tool that assists in approximating the amount of basic employment within a geographic unit. It compares the share of employment in a particular industry for a specific geographic area, such as a local government area, to that of a larger economic region, such as the state or country in which it is located. It is calculated as follows:

$$LQ = \frac{E_{ri}}{E_{rT}} \Big/ \frac{E_{Ni}}{E_{NT}}$$

where the employment in region r and industry i (E_{ri}) divided by region r’s total employment T (E_{rT}) represents the region r’s share of employment in industry i, and the national employment in industry i (E_{Ni}) divided by the total employment in the country

(E_{NT}) gives the national share of employment in industry i. Dividing the former share by the latter produces the location quotient. When the result is a value in excess of 1, it is said that this represents the percentage of employment in industry i that can be considered basic, and thus dependent on non-local factors. A location quotient of 1 means that the employment activity is at a level at which it produces enough of that good or service to satisfy local demand. If the location quotient takes on a value of less than 1, this suggests that the local area is an importer of goods or services provided by industry i. Apart from its conventional application, the location quotient can be an effective means of identifying local industry clusters because it reveals relative industry concentration. It can also be used to identify areas that may have a particular productivity-enhancing endowment, whether human-made or naturally occurring, that causes a specific industry to concentrate there.

Another potentially useful measure or index applied in econometric works of agglomeration economies is one that gives insight into the degree of industry specialization or diversity of a given area a. Generally, empirical agglomeration works would just assess the overall scale of employment, undifferentiated by industry mix, on productivity or include measures for both urbanization and localization effects. A diversity index would not reveal whether a specific industry is localized in a given area, but rather comment on an area's overall degree of specialization. The representation of this only relies on one measure because specialization and diversity both reside in the same spectrum, merely at opposite ends. Combes, Duranton et al. (2008) and Melo and Graham (2009) make use of an inverse Hirshman-Herfindahl index to represent a given area's industrial diversity. Such an index takes the form of

$$DI_r = \left(\sum_i \frac{E_{ri}}{E_r} \right)^{-1}$$

which takes the inverse-sum of each industry's employment share of total employment for a particular region, r. A region fully specialized in one industry will return a value of 1, since its share of employment in the one industry will comprise 100% of the employment in the area. An increase in diversity results in an increase in the index. One can correct this index for differences at the national level by computing a relative-diversity index. This involves summing the absolute values of the differences between region-specific and national employment shares over all sectors. This is expressed formally as follows:

$$RDI_r = \left(\sum_i \left| \frac{E_{ri}}{E_r} - \frac{E_{Ni}}{E_N} \right| \right)^{-1}$$

This index increases the more that region r's employment mix matches that of the national economy.

These measures and indices represent a number of tools that have effectively been applied to study the spatial economy and, in particular, the existence of agglomeration economies in cities.

2.11 Conclusions

This chapter has served to give the background for the research undertaken in this thesis. Each topic covered, from the significance of cities and why they exist, to the role of transportation in shaping them, to the review of works undertaken to estimate the productivity benefits experienced by living in them, can be expounded and explored more deeply. The aim has been to give sufficient justification for the measurement of agglomeration externalities in Australian cities and industries and to argue for their relevance and consideration in planning policies and strategies. Hopefully it has been made clear that planners have a substantial task ahead of them in preparing for growing populations in cities and that the impacts of sprawling versus concentrating on densification and infill development can be significant and material. It has been argued that investment in transport infrastructure can be a powerful tool in directing how cities structurally evolve to achieve density or dispersal. The impacts of transport infrastructure and urban form have also been discussed, primarily through their links to productivity, which is a powerful connection to make when fears loom that working towards sustainability will likely have economic consequences.

Agglomeration economies constitute one of several real benefits that arise when planning for sustainable cities. They are the fundamental reason why cities exist and suggest that by building up instead of out, integrating land-use with public transport, focusing on mixed-use development, and designing centres with a human scale in mind, we can be more productive and more sustainable. The chapters that follow cover the empirical research conducted for this thesis where the productivity benefit of employment density and accessibility is estimated for Australian cities. The empirical investigation consists of six analyses that will hopefully provide the inputs for direct use in planning processes in Australia, if not at least provide a platform for future work to be conducted.

CHAPTER 3: Components of the Empirical Work

3.1 Introduction to the Empirical Analyses

Here in Chapter 3 we begin to get into the formal analytical work investigating the existence and impacts of agglomeration economies in Australia. A series of analyses are reported on that gradually increase in complexity whilst also using data at finer geographic scopes, reflecting the learning process that was involved. First, the data requirements and their sources will be discussed in overview. The details of the handling of this data will be provided in the subsequent sections that discuss the particulars of each analysis. After the discussion of data, a theoretical model will be given as justification for the general empirical method applied for estimating the productivity impacts of agglomeration economies in Australia. Following this will be accounts of six analyses that were undertaken for the purpose of this study. These will be discussed in Chapters 5 and 6.

3.2 The Data

For the purposes of the analyses carried out in this thesis, two types of data were utilized. For one, the construction of accessibility and industry concentration indices required data in the form of spatial layers, containing digital information on geographic boundaries and road networks. Secondly, suitable data for assembling industry-specific wage functions was required to control for firm or worker differences. The type of function to use, whether a production or wage function framework, was eventually established on the basis of the availability of data on Australian cities. This will be discussed first and then followed by an account of the sources of spatial data. Box 3.1 below outlines key and flexible/optional aspects of the data requirements for production and wage function estimations.

Box 3.1: Key and Optional Wage Function Dataset Characteristics

Production Function Approach Key Dataset Characteristics

- Spatially referenced small-area data (smaller the better)
- Industrially disaggregated data (2 or 3-Digit ANZSIC is preferable)
- Data on individual firms (preferable) or data that has been aggregated and averaged over spatial units
- A measure of unit output such as sales revenues or industry gross value added (GVA)
- Controls for observable heterogeneity such as capital stock data (fixed assets, current assets, current liabilities), employee data (no. of employees, wages)

Wage Function Approach Key Dataset Characteristics

- Spatially referenced small-area data (smaller the better)
- Industrially disaggregated data (2 or 3-Digit ANZSIC is preferable)
- Data on individuals (preferable) or aggregated and averaged over spatial units
- A measure of worker output such as wage, worker gross value added or income
- Controls for observable heterogeneity such as education level, experience (proxied by age), occupation, gender
- Data reported by place-of-work (not place-of-usual-residence)

Beneficial yet Optional Dataset Characteristics

- Panel data format (data on same observations repeated over 2 or more periods)
- Instrumental variable(s) data on long-lagged historical employment/population levels (most common instrument used) or some other suitable variable
- Data are ideally provided at the scale of employment zones or a similar classification. This tends to be a more 'meaningful' unit scale than ones based on administrative or political boundaries.

A prerequisite for spatial econometric analyses such as those applied in this thesis is that the data used must be geographically referenced, which tends to offer a number of challenges as one goes to smaller and smaller geographic units. For one, the wealth of data becomes less as one opts for more spatially disaggregated data. In Australia, for instance, industry financial data has traditionally only been collected at the national level and sometimes state levels to make comparisons between economic performances over past years or between states, rather than making these comparisons between smaller geographic units such as Local Government Areas (LGAs), Statistical Local Areas (SLAs), postal codes, suburbs, or employment zones. Another issue that arises is that of confidentiality. While nations such as France, the U.S. and the U.K. collect and grant access to sensitive data (albeit confidentialized in some way), in Australia gaining access to spatially-referenced firm-level and employment data are not a question of authority or purpose that enables this data to be made available – it is simply not collected with detailed geography or location in mind. Here I give some examples of this.

The Australian Bureau of Statistics (ABS) conducts a detailed statistical survey on small and medium-sized businesses called the Longitudinal Business Survey (LBS) that gathers microdata in a panel structure. This survey has been conducted over the years of 1993/94 through to 1997/98, and then again over the years of 2003/04 through to 2007/2008, each time covering a 4-year period. Because of reasons of confidentiality, however, the respondents historically have never been queried on the location of their establishments. This survey data provides a wealth of information on a sample of Australian firms including the industry in which they operate, the age of their business, their expenditure on research and development, their employee details, and their financial information such as profits or losses, the value of assets or liabilities and capital expenditure. This data would be incredibly useful for the purposes of this research; however, because of the absence of spatial referencing it cannot be used.

An alternative to the LBS data would be the Australian Industry (AI) data collected by the Australian Bureau of Statistics (ABS); however, it does not provide any geographic disaggregation below the state and territory level which means it is not detailed enough to make any judgments on the relationship between urban form (or transport infrastructure) and productivity. Moreover, the sample sizes would be far too small given that there are only eight states and territories in the country. Strictly for manufacturing industries, the ABS also collects and aggregates financial data on firms down to the SLA level. While smaller geographic units are always preferable, SLAs are small enough to make some intra-city comparisons on firm or labour productivity. Upon requesting these data, however, it was warned that if the number of firms in a SLA were too few, the data would have to be confidentialized, which means data for those spatial units will be blocked as being 'not for publication'. Upon further inquiry into having this data provided for the GMA of Perth, it was mentioned that the data had not been previously requested and thus would require a consultant to extract it at a cost far higher than could be afforded for this work. Not only was the price of the data substantial but it would not have been ideal, lacking in geographic detail and being restricted to only manufacturing industries when services are of interest as well.

Because of these data issues, estimating agglomeration elasticities for Australian cities using a production function framework was dismissed. At this stage the alternative of using a wage function approach was considered and spatially disaggregated employment data was sought. This endeavour proved to be equally challenging, but not without eventual reward. The ideal circumstance for using a wage function approach would have been to use data on wages from surveyed individuals' primary sources of earnings. The alternative is to use income, which would suffice but not be ideal. Income, as measured by the ABS in census surveys, encapsulates total earnings that may come from a number of contributing sources (such as investments, welfare transfers or second jobs). While in most circumstances people generally only have one source of income, that some are likely to have alternative sources suggests that any econometric estimations using income may have a slight increase in error variance. Controls for employment heterogeneity would not likely be able to predict contributions of alternative income sources. In light of this, spatially and industrially disaggregated data on employee earnings was first sought.

One avenue explored in the search for employment data was the Labour Force Survey (LFS), which is conducted monthly and quarterly by the ABS. The survey contains a host of information including, but not limited to, the respondent's sex, age, level of education, industry of employment, and occupation. The panel data structure of the survey would make it ideal for use in wage function estimations if it were not for the lack of questioning of respondents about their place of work or their earnings – two vital components to the econometric analysis. An alternative to this would be to use aggregated employment data collated by SLAs. The Australian Tax Office (ATO) provides the ABS with industry wage data but it does so at the place-of-usual-residence, not the place-of-work where it is needed.

A number of other potential sources of data were investigated, such as the Australian Chamber of Commerce and Industry (ACCI), state-specific chambers of commerce, various state and territory business entities, and the Australian Business Register, but none had any data available to provide the types of inputs necessary for this type of spatial economic analysis. In fact, any data they possess or rely on is generally sourced from the ABS. To my knowledge, the only remaining alternative was to use census survey data, which is collected every 5 years by the ABS.[10]

The eventual datasets purchased for use in the analyses laid out in this thesis drew on data from the 2006 census. The benefit of using census data is that it draws on an extremely large sample and it can be provided by place-of-usual residence or place-of-work, in which case the latter was needed and utilized for this study. Census data can also be organized and provided at a wide range of spatial scales, ranging from the state and national levels down to the employment zone level (for place-of-work data). Furthermore, as respondents are queried on their main industry of employment and occupation, census data also allows for the provision of highly industrially and occupationally disaggregated data.

Industry data used is classified under the Australia-New Zealand Standard Industrial Classification 2006 (ANZSIC 06) format, which can at its broadest level aggregate industries at the 1-digit level. This organizes industry data into one of 16 divisions. Moving to the 2-digit level there are roughly 90 industry subdivisions, at the 3-digit level there are approximately 300 groups, and finally at the 4-digit level there are just over 700 classes. There is a trade-off, however, between using less and more detailed industry groups.

By using more detailed industry groups, one could potentially be estimating up to 719 econometric functions, as the census data are aggregated and averaged over spatial units. Thus, the counts of respondents making up the averages become very small and can often be zero. These small cell counts are of some concern because the ABS perturbs the data to protect the anonymity of respondents and the effects of this become more significant with more detailed data. Perturbation refers to the randomization of reported and supplied ABS data such that the reported values are close to the true figures and the components of each table still add up to the true totals. The effect of using highly industrially disaggregated data with small area units is that many cell counts may record positive mean values when in fact no employment exists in the area at all. Conversely, cells may report a value of zero when in fact industry-specific employment does exist in the geographic unit. This effect is exacerbated when combining highly industrially disaggregated

[10] An excerpt from the Infrastructure Australia report titled *State of Australian Cities 2010* supports this limitation of poor data. They write, "Unfortunately, datasets measuring productivity and 'multi-factor' productivity are not available at an Australian city level, where cities are treated as a discrete economic entity in order to measure this" (p. 56).

data with highly spatially disaggregated data. Consultants at the ABS strongly advised that if smaller spatial units were desired then there should be a reduction in the detail of the industry groupings.

The benefit, however, of greater industrial disaggregation is that industry-specific effects of agglomeration do not get as 'washed out' in the econometric analyses. For instance, the 1-digit division of 'Manufacturing' is disaggregated into 15 subdivisions at the 2-digit level and 69 groupings at the 3-digit level. At the 3-digit level, for example, we might expect there to be significant differences in the agglomeration benefits accruing to *Pharmaceutical and Medicinal Product Manufacturing 184* and *Log Sawmilling and Timber Dressing 141*. Further industrial disaggregation allows the agglomeration externalities to be estimated more accurately.

In light of these trade-offs between spatial unit scale and industrial scope, the three analyses in Chapter 4 use 3-digit ANZSIC data aggregated at the geographic scale of the Statistical Local Area (SLA), and then the three analyses in Chapter 5 use 2-digit ANZSIC data aggregated at the smaller geographic scale of the work destination zone (or 'travel zone' as referred to in Sydney). Analyses using SLAs allow for greater industrial differentiation and as such we can better separate the benefits of agglomeration to industries that are high and low in knowledge content. For instance, *central banking* and *depository services* are differentiated at the 3-digit ANZSIC level but not at the 2-digit level. Similarly, 2-digit ANZSIC amalgamates all professional services into one category, yet at the 3-digit level separates out scientific research, architectural and engineering, advertising, market research and statistical, management and consulting, and veterinary services. These detailed industry scopes can be extremely interesting and useful to planners and policy-makers. The use of datasets with the smaller geographic units provides larger sample sizes, which in turn are likely to generate more efficient estimates of agglomeration economies yet do so at the expense of industrial scope. The benefit of using a larger sample size also means that there are more degrees of freedom to work with. As such, more flexible functional forms of the estimating equations can be applied without significantly influencing the robustness of the statistical results. Carrying out analyses at both spatial and industrial scopes should give us a better understanding of the magnitudes of the trade-offs therein while enabling some manipulations on functional form.

In addition to considerations of industrial and geographic scope, data was sourced from the 2006 census that would allow for the control of occupation, level of education, and level of experience. For the Chapter 4 analyses, occupational effects were controlled for by using data on the Australian and New Zealand Standard Classification of Occupations 2006 (ANZSCO 06), which allocates census respondents into one of the eight following categories:

1. Managers
2. Professionals
3. Technicians and Trades Workers
4. Community and Personal Service Workers
5. Clerical and Administrative Workers
6. Sales Workers
7. Machinery Operators and Drivers
8. Labourers

Education was controlled for by sourcing data on the numbers of industry-specific employees in each geographic unit that have their level of educational attainment fall into one of the two following categories:

1. Tertiary Education
2. No Tertiary Education

Finally, a control for experience was established by sourcing data on industry-specific mean age for each geographic unit. When sourcing data on these same controls for Chapter 5's analyses, which used data at the smaller geographic scale of the employment zone, the categories of these controls needed to change to some extent. This had to occur because of perturbation issues with the frequent reporting of low values in geographic units. Work Destination Zones, which can often be as small as a few hundred metres in diameter, will in many cases have industry-specific employment counts of fewer than a dozen workers. To further disaggregate these low numbers into several occupational categories, for instance, could make the reported values resulting after perturbation highly unreliable. To address this issue, the occupational categories for the latter three analyses were collapsed into the following three categories:

1. Managers and Professionals
2. Technicians, Trades Workers and Labourers
3. Community, Personal Service, and Sales Workers

The categories of education and mean age remained unchanged between the two sets of analyses. It should also be noted that the dependent variable in both datasets is mean income, as this is the only figure of earnings generated by the census. As previously mentioned, it is not as ideal a measure as wage, since it consists of the aggregate of all sources of earnings an individual may have; however, for the most part this should not be much of an issue as wages will for most individuals be the sole or majority contributor to total income.[11] From henceforth, the details of the datasets and particular analyses within Chapter 4 will be referred to as 'Part 1' analyses, while those pertaining to Chapter 5 will be referred to as 'Part 2' analyses.

Additional to the particulars of the control variables sourced for the two sets of analyses is the regional scope of interest. The first analysis in Part 1, which happens to investigate the presence and magnitude of agglomeration economies at their broadest level, is conducted on all eight capital cities in Australia: Sydney, Melbourne, Perth, Brisbane, Canberra, Adelaide, Darwin, and Hobart. The next two analyses are restricted to Sydney and Melbourne, largely because the first analysis does not incorporate controls for worker heterogeneity, which then had to be purchased for the second and third analyses. The cost incurred because of this restricted the number of cities that could be analysed. The analyses in Part 2, being those conducted at the geographic scale of the Work Destination Zone, also only look at Sydney and Melbourne. The rationale for keeping two capital cities in the analyses is because the results of one can then be compared with the results of the other. This may provide some insights as to whether industry-specific agglomeration effects can be generalized across Australia cities, although I will also acknowledge that differences

[11] A breakdown of income sources for Australian residents could not be found to give support of this claim, though anecdotal evidence of this claim was given by a reliable source at the Curtin Business School. The ABS produces a report on household income and income, which provides some additional insight into the matter. The 2009 report states that households with middle and high income levels get most of their income from wages but that lower income households get most of their income from pensions (which would not impact this analysis) and allowances. What this suggests is that if lower income earners receive higher levels of subsidy and areas of poor accessibility and low density are the least productive, then the estimated industry-specific productivity gradients are likely to be underestimated and give conservative results.

could merely arise because of estimation error. The decision to focus the analyses on Sydney and Melbourne was based on several aspects. Firstly, the two cities have the greatest number of residents of all the cities in Australia and thus the effect of perturbation on the data will hopefully be less. Secondly, since the analyses within Part 1 and Part 2 should share subject cities to enable comparison between analyses conducted at different spatial scales, Sydney and Melbourne are the best suited for this because they have among the most SLAs and WDZs within their boundaries which affects sample sizes. Thirdly, the two cities are the oldest, most established and most urbanized in Australia and thus the older centres within them that maintain development characteristics from before the widespread use of the automobile will hopefully produce more interesting results. Finally, cost and time taken to source spatial data was a limiting factor that prevented the analysis from extending to more than the two cities.

To expand on the sample size issue, it is useful to note that the number of geographic units within datasets varies quite dramatically between cities and the differences between the two geographic scopes are much greater still. At the SLA level, the number of geographic units and hence sample observations in the capital cities are: Sydney (64), Melbourne (79), Perth (37), Brisbane (215), Canberra (108), Adelaide (55), Darwin (41), and Hobart (8). Whether the entire sample size is available for the estimation of industry-specific wage functions depends on whether industry-specific employment exists in all of these areas. In comparison, the number of Travel Zones and Work Destination Zones in Sydney and Melbourne are 3,098 and 2,083 respectively. While most of the industry sample sizes will not come close to these maximum possible figures, the sample sizes with Work Destination Zones will still be significantly larger than when using Statistical Local Areas. Box 3.2 below gives a summary of all the details of the two datasets.

Box 3.2: Dataset Details for Part 1 and Part 2 Analyses

Part 1 Analyses

- Data sourced from the 2006 Census and provided by place-of-work
- Use the geographic scale of the Statistical Local Area (SLA)
- Cover all Australian capital cities for the first analysis and then Sydney and Melbourne for the rest

- Control for occupation using 1-digit ANZSCO employment data (8 categories) (Not applicable to the first analysis) [Data for Sydney and Melbourne only]
- Control for education using the categories of 'tertiary' and 'non-tertiary' educational attainment (Not applicable to the first analysis) [Data for Sydney and Melbourne only]
- Control for experience using 'mean age' (Not applicable to the first analysis) [Data for Sydney and Melbourne only]
- Data for all controls of worker heterogeneity are provided on an industry-specific basis at the 3-digit ANZSIC level (60 industries)
- Number of SLAs in the capital cities: Sydney (64), Melbourne (79), Perth (37), Brisbane (215), Canberra (108), Adelaide (55), Darwin (41), and Hobart (8)
- Income is used as the measure of output

Part 2 Analyses

- Data sourced from the 2006 Census and provided by place-of-work
- Use the geographic scale of the Travel Zone (TZ) in Sydney and Work Destination Zone (WDZ) in Melbourne
- Cover the capital cities of Sydney and Melbourne
- Control for occupation using a collapsed form of 1-digit ANZSCO employment data (3 categories)
- Control for education using the categories of 'tertiary' and 'non-tertiary' educational attainment
- Control for experience using 'mean age'
- Data for all controls of worker heterogeneity are provided on an industry-specific basis at the 2-digit ANZSIC level (30 industries)
- Number of WDZs in Melbourne are 2,083 and the number of TZs in Sydney are 3,098

The second type of data required was determined by the need for providing a variable or index to represent the degree to which employment is spatially and contextually concentrated in the units of analysis. Employment numbers were already present in the datasets provided by the ABS. The spatial data with which these employment figures had to be combined to produce indices of agglomeration were provided by other sources.

The Transport Data Centre (TDC) in Sydney was very helpful in providing much of what was necessary for generating agglomeration indices for the Sydney analyses. They were able to supply peak a.m., peak p.m. and daytime inter-peak travel times between pairs of statistical local areas and travel zones to match the period of analysis – the year 2006. They were also able to provide a spatial layer containing all the travel zones (TZs) in New South Wales in the TAB and ESRI Shapefile formats. The former was for use in MapInfo and the latter for use in a spatially enabled database called Postgres that contained the PostGIS add-on. These spatial files were provided in the MGA94 zone 56 coordinate projection system. Unlike for Melbourne, road centreline spatial files were not obtained for Sydney because its effectiveness as an instrumental variable was not as high as anticipated when trialled with Melbourne data. More information on the outcomes of the instrumental variables estimations and the effectiveness of the IV's trialled can be found in Part 2.

For the Melbourne analyses, VicRoads kindly provided ESRI Shapefiles containing the digitized boundaries of Victoria's Work Destination Zones and main roads infrastructure. To get a spatial layer of the entire road network, comprising state (main) and local roads, I had to approach the Department of Sustainability and Environment (DSE) in Victoria. All these spatial files were provided in the MGA94 zone 55 coordinate projection system. Unfortunately, travel-time data could not be obtained as easily from Victorian sources. As this data was not considered essential for the purposes of estimating industry productivity elasticities with respect to employment concentration (for the emphasis lay on Euclidean distance to represent proximity), after several emails and phone calls without any response the efforts to obtain this data ceased.

All this spatial data was provided free of charge, yet under the condition that proof of enrolment in an academic institution as a PhD candidate could be provided. The remaining spatial layers, however, had to be purchased from the ABS at a fee. These layers included the digitized boundaries for states, statistical divisions, local government areas, and statistical local areas for all of Australia. These were provided in ESRI Shapefile format in the GDA94 coordinate projection system.

The reason that the ABS uses one projection for all of Australia and that New South Wales and Victoria use their own is that state government planning authorities produce their own spatial layers on work destination and travel zones that they then provide to the ABS to 'fill in' with data. SLAs are a construct devised by the ABS at the national level for statistical purposes. Generally speaking, a projection system is an algorithm that facilitates a three-dimensional topographical area to be represented on a 2-dimensional plane. This enables distance calculations to be made via straight lines that still take into account changes in elevation. State-devised projection systems are likely to better reflect the local topography than any other. The details of these spatial data and their sources are summarized below in Box 3.3.

Box 3.3: Spatial Data Types and Sources

New South Wales Data

- SLA origin-destination a.m. peak, p.m. peak, and daytime inter-peak travel times (Source: TDC)
- TZ origin-destination a.m. peak, p.m. peak, and daytime inter-peak travel times (Source: TDC)
- TZ digital boundaries, MGA94 zone 56 format (Source: TDC)

Victoria Data

- WDZ digital boundaries, MGA94 zone 55 format (Source: VicRoads)
- Main roads centreline digital data, MGA94 zone 55 format (Source: VicRoads)
- Main + local roads centreline digital data, MGA94 zone 55 format (Source: DSE)

Australia-wide Data

- State digital boundaries, GDA94 format (Source: ABS)
- Statistical division (SD) digital boundaries, GDA94 format (Source: ABS)
- Local Government Area (LGA) digital boundaries, GDA94 (Source: ABS)
- Statistical Local Area (SLA) digital boundaries, GDA94 (Source: ABS)

3.3 The Theoretical Model

The justification for using worker wages or income within a certain region and industry to reflect the level of productivity in a region is grounded in the basic economic principle that wages reflect the marginal productivity of labour. As Puga (2010) states, "If firms and workers are mobile and wages and land rents differ across space, higher wages and land rents in large and dense urban environments must reflect some productivity advantage" (p. 204). The theoretical justification for using a wage function to estimate agglomeration-related productivity benefits, following Combes, Duranton et al. (2008), is described below.

Let us consider the Cobb-Douglas specification of a firm's production function where output y is determined by labour l and capital k in industry-sector i and region r as indicated below.

□

$$y_{ir} = A_{ir}\left(s_{ir}l_{ir}\right)^{\alpha} k_{ir}^{1-\alpha}$$

In this specification, A represents total factor productivity and s_{ir} gives the relative efficiency or effectiveness of labour. The profit-maximizing firm will produce at a level where marginal revenue will equal marginal cost, or rather, maximize the difference between total revenue and total cost as represented in the equation that appears below where p is the price of output, w is the wage rate and c is the cost of inputs. At competitive equilibrium, the price of the input factors should equal the value of their marginal products.

$$\max \pi_{ir} = \max p_{ir}\left(A_{ir}\left(s_{ir}l_{ir}\right)^{\alpha} k_{ir}^{1-\alpha}\right) - \left(w_{ir}l_{ir} - c_{ir}k_{ir}\right)$$

Taking the first-order derivatives with respect to wages and capital to get their prices when in competitive equilibrium gives us the following expressions.

$$w_{ir} = \alpha p_{ir} A_{ir} s_{ir}^{\alpha}\left(\frac{k_{ir}}{l_{ir}}\right)^{1-\alpha} \qquad \text{and}$$

$$c_{ir} = (1-\alpha) p_{ir} A_{ir} s_{ir}^{\alpha}\left(\frac{k_{ir}}{l_{ir}}\right)^{-\alpha}$$

After rearranging the marginal input price expression c_{ir} to isolate for the capital to labour ratio and substituting it into the marginal cost of labour expression, we get the following:

$$w_{ir} = \alpha(1-a)^{(1-\alpha)/\alpha} s_{ir}\left(\frac{p_{ir}A_{ir}}{c_{ir}^{1-\alpha}}\right)^{1/\alpha}$$

This final expression tells us much of what we need to know about the influences of agglomeration economies on the wage rate. Wages are positively influenced by the quality of labour (s_{ir}), which will be controlled for in the econometric model; the price of the output (p_{ir}); and the technological efficiency of the local economy (A_{ir}). Additionally, wages are negatively affected by the cost of inputs (c_{ir}). The term A signifies many of the advantages of agglomeration such as the occurrence of knowledge spillovers, labour market pooling, input-output sharing, and other such externalities that increase the efficiency of activity. Similarly, c_{ir} can be moderated by the proximity and density of economic activity in more highly agglomerated areas, for one can expect greater competition and the presence of more substitutes to lower input prices whilst proximity would lower transaction costs. As mentioned by Melo and Graham (2009), these parts of the expression cannot be estimated separately but only as a whole, as it would be extremely difficult to unravel the separate benefits of agglomeration from one another. This relates to the discussion of agglomeration economies in section 2.7 where they are described as operating in somewhat of a "black box", where the micro-foundations have been well defined in theoretical discourse but not extensively disentangled in empirical works.

CHAPTER 4: Part 1 Analyses

4.0 Introduction to Part 1 Analyses

In entirety, the analytical work undertaken for this thesis is comprises six separate analyses that have been partitioned into two groups: Parts 1 and 2. The analyses of Part 1 will be covered here in Chapter 4 while those of Part 2 will be covered in Chapter 5. That which differentiates them the most is their treatment of geographic scale. In Part 1, the estimation of agglomeration economies occurs using the statistical local area (SLA) as the geographic unit of observation. A couple of reasons exist for why this level of geographic detail was selected first before moving to a much more spatially disaggregated dataset.

Likely to be the foremost reason was cost. Containing far fewer observations, the datasets using SLAs allowed analyses to be carried out for more cities than if using work destination zones (WDZs), while keeping costs down. Moreover, given that funds were limited and the shortcomings of using census data that existed, it was safer to test out the methodology on relatively cheaper datasets. A secondary reason for selecting SLAs was the complexity of making calculations involving more complex specifications of agglomeration variables. When using SLAs, the required calculations could be done with relative ease in standard, easily operable and readily available software such as Microsoft Excel. Advancing to the use of WDZs meant handling datasets and a number of calculations that Microsoft Excel could no longer handle. For this reason, the decision to move to datasets aggregated to small-area units seemed both daunting and exciting. It required learning how to use a couple different types of software and a language of script writing to import, organize, augment, extract, and export spatial and employment data for eventual use in an econometrics software package. This all took some time; however, the process led to a capacity enhancement that opened up new opportunities and possibilities.

The benefits of using Work Destination Zones are several. For one, it is a much more meaningful unit of measurement than the Statistical Local Area, which is more of an administrative unit than one based on actual economic activity. Secondly, their small size helps remedy many of the issues with identifying 'built-up' areas and the types of urban forms, which is a problem with larger geographic units. Additionally by being more numerous, employment zones offer much larger sample sizes that improve the precision of parameter estimates by reducing the variance of the estimators.[12] This in turn increases the values of test-statistics[13] and lowers the critical values in hypothesis testing[14], increasing the statistical power of such tests. Lastly, a larger sample allows for the application of more flexible functional forms and the addition of more control variables. In statistics, hypothesis testing uses the concept of 'degrees of freedom' to control the shape of the t-distribution, which is calculated according to the formula $N-K$, where N is the total sample size and K is the number of variables in the model. Ideally one wants to keep the degrees of freedom high for narrower confidence intervals, while more explanatory variables and adding quadratic or interaction terms will lower it. More observations in this sense will always be preferable because adding terms will not come at much cost to statistical power.

Without a great deal of certainty that the methodology of estimating the effects of agglomeration economies using census data would prove effective, beginning with SLA data was the safer option as it demanded less of an investment in time and money. The results, however, proved favourable and as such the movement towards using employment zone data was made. This was partially to validate the findings that used SLAs, but also to see if a major difference would arise from using a more refined spatial unit and to make use of the greater flexibility made available by the larger dataset. The sections that follow will expound on the details and the results of these analyses where Analysis 1-1 investigates an aggregate-industry productivity impact of agglomeration, Analysis 1-2 investigates the agglomeration impact on a 3-digit ANZSIC basis and Analysis 1-3 does the same while including a control for localized industry concentration. It should be noted that when results are reported, only those pertaining to the agglomeration variables and the overall model are given. If one

---

12 The variance of a parameter (or coefficient) estimate is calculated according to $\text{var}(\beta_1) = \sigma^2 / \sum (x_i - \bar{x})^2$, where the value of the denominator will increase with the number of observations. This in turn lowers the variance and increases the precision of the estimate.

13 When conducting a 'test of significance' on a parameter estimate, essentially you are asking if the estimate is different from zero. A t-statistic to test for this is determined according to $t = \hat{\beta}_1 / \sqrt{\text{var}(\hat{\beta}_2)}$. As such it will increase with a smaller variance. A larger absolute t-statistic value makes it easier, or more likely, for the estimate to pass a test of significance.

14 A 'critical value' for a predetermined level of confidence (typically 95%) in a test of significance is the value that the test statistic must exceed in order to conclude with a degree of certainty that the estimate is significantly different from zero. A larger sample increases the 'degrees of freedom' in a test, which in turn lowers the magnitude of the 'critical value'.

would wish for the rest of these results to be provided they should contact me. The emphasis of this thesis is on the policy significance and implications of agglomeration economies and due to the hundreds of regressions made over the course of the research, it was believed that extending the discussion to all the control variables used would detract from the work.

4.1.0 Analysis 1-1 Overview: Aggregated industry elasticity estimations using SLA data

This first analysis was very much an introductory attempt at estimating agglomeration economies in Australian capital cities. It did so by estimating the influences of employment density and employment size on a productivity index as defined by Rice and Venables (2004). The index takes on the following form:

$$q_i = \sum_k w_i^k \bar{\lambda}^k \qquad [4A]$$

where q_i is the value of the index for location i, w_i^k is the average income (or wage) of an employee in location i and industry k, and $\bar{\lambda}^k$ is the average share of employment in industry k across all capital cities. The index is constructed using 3-digit ANZSIC data and assumes that each statistical local area in all of the Australian capital cities shares the same employment composition and as such, any variation in the index will be a result of industry-specific variations in earnings. This in effect controls for sectoral mix and reflects any advantages of agglomeration experienced by a given area via the link to employee earnings.

Using ordinary least squares (OLS)[15], simple log-log regressions were then carried out by regressing the productivity indices of each capital city's SLA on their respective employment densities and employment sizes. This is expressed econometrically as follows:

$$\ln(q_i) = \beta_0 + \beta_1 \ln(A_i) + e_i \qquad [4B]$$

15 OLS is the most basic form of regression analysis, which is used in estimating the unknown parameters of a relationship where a dependent variable (Y) can be predicted by one or more independent variables (X). The method fits a line to a series of plots (observations) that minimizes the sum of the squared deviations between the observed and predicted values.

where the natural log of given area's productivity index (q_i) is predicted by the natural log of the employment concentration of the respective area (A_i), where the latter can be represented by employment size (as typically done in early studies) or employment density (as is more generally accepted). The resulting coefficient β_1 is interpreted as an elasticity, which in this context would be the percentage change in productivity index given a 100% increase in A_i.[16]

In this analysis, employment size was determined strictly by summing the total number of persons employed in a given SLA. Employment density was estimated by dividing a SLA's total employment by its respective total area, thus no adjustments were made in this measure to account for non-urbanized spaces.

4.1.1 Analysis 1-1 Results

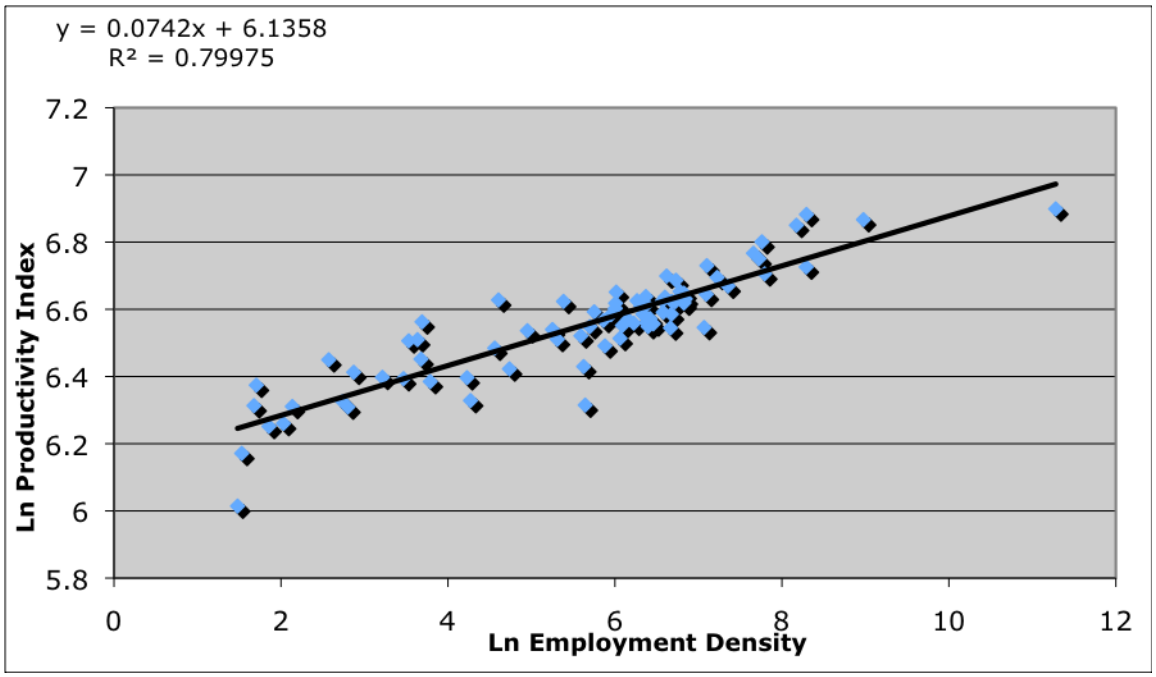
The results of the analysis varied significantly for the major cities, producing elasticities and R-squared values that did not reveal uniform relations between productivity and employment size and density across the nation. This to some degree can be expected because of a number of urban form characteristics that greatly differentiate the cities and because of the multitude of other contributors to urban productivity that cannot be captured in the econometric specification used. The inconsistency of the results among cities suggests that we may not be able to generalize the benefits of agglomeration across all the capital cities. On the other hand, large discrepancies in the geographic sizes of the SLAs across the capital cities may have a significant bearing on the magnitudes and strengths of the results as well, which could be the dominant factor behind the differences in cross-city results.

The results show that employment density is by far the best predictor of productivity in the city of Melbourne where it explains 80% of the variance in the productivity index (see Figure 4.1.1). The corresponding elasticity of productivity was 7.4%, implying that doubling the density of employment in a SLA in Melbourne would result in

[16] See Appendix B for a mathematical explanation of how estimating a log-log equation (one that is linear in logs) returns a constant elasticity estimate after a partial derivative is taken.

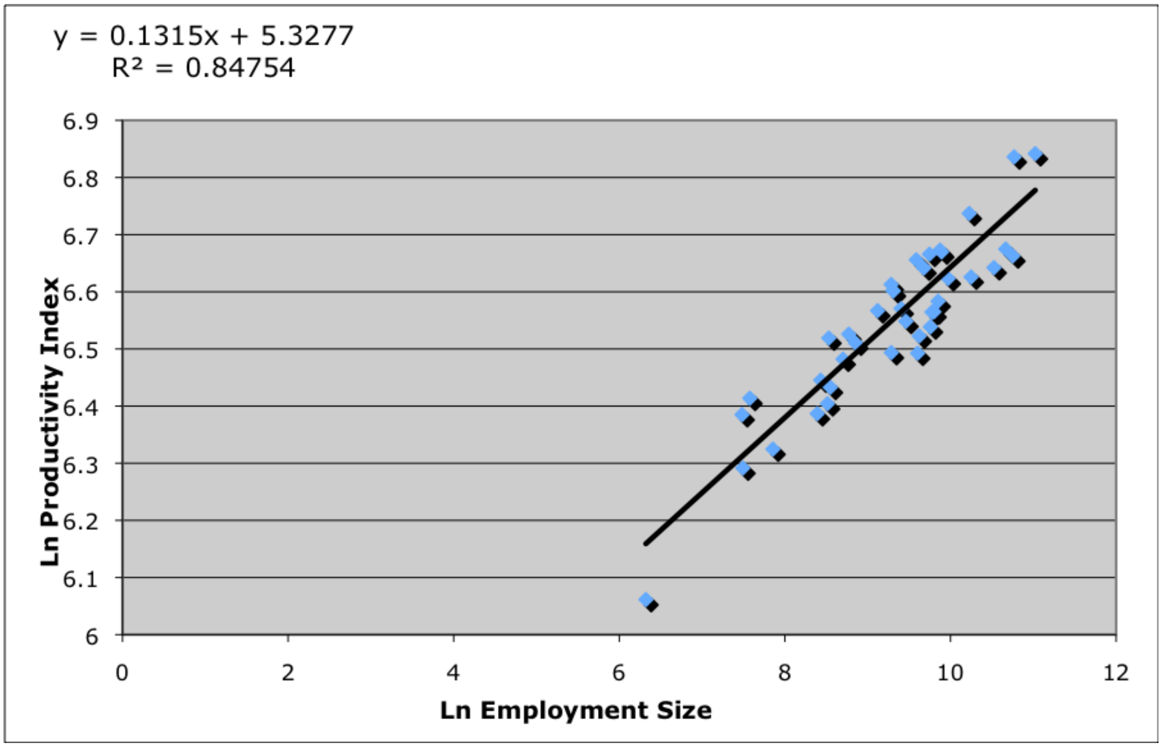
an average wage (labour productivity) increase of 7.4% across the existing occupations. When measured with respect to employment size rather than density, the R-squared value was not strongly affected as it dropped by merely one percentage point to 79%, yet the elasticity estimate jumped to 15%. The results for the remaining capital cities differed considerably, giving rise to a number of questions.

Figure 4.1.1: Regression Results for Melbourne – Elasticity of productivity estimate (measured with respect to employment density)



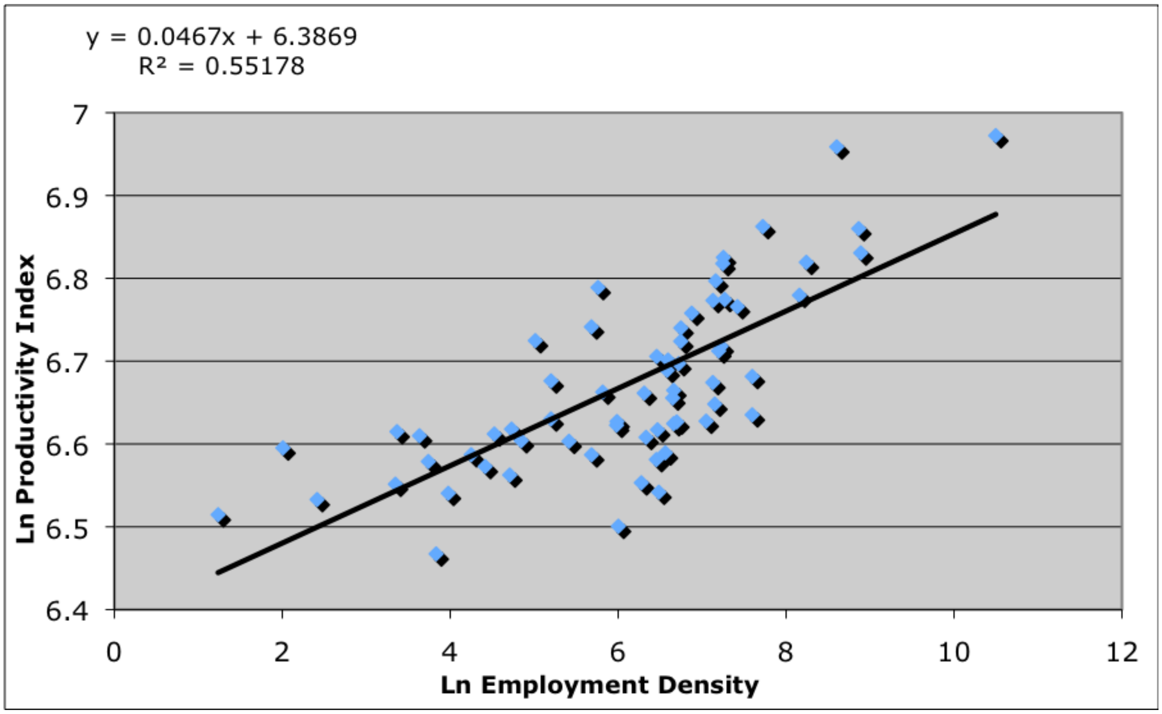
For Perth, the resulting elasticity was 3.5% when measured with respect to employment density, however, with an R-squared value of 32% much of the variance is left unexplained. A staggering difference results when the analysis is carried out with respect to employment size (Figure 4.1.2). The elasticity increases to 13% and the R-squared value to 85%. This would seem to imply that labour productivity might be better predicted as a function of employment size rather than density. The dramatic difference in estimated elasticities between the two independent variables can be traced back to their low correlation for this particular city (36%). This can be compared to a correlation in Melbourne of 85%, which likely explains why the difference in results between the two employment measures was not as great. It becomes quite clear that the relationship between employment size and geographic area in Perth is not as close as it is in Melbourne. This may be because of underlying differences in urban form as, unlike Perth, Melbourne experienced much of its growth prior to the widespread adoption of the automobile.

Figure 4.1.2: Regression Results for Perth – Elasticity of productivity estimate (measured with respect to employment size)



The analysis for the rest of the capital cities also generated interesting results. In Sydney, an elasticity of productivity of 4.7% was estimated with respect to employment density with 55% of the variance explained (Figure 4.1.3). Being Australia's first major population centre to emerge, one would expect the results to be closer to those for Melbourne, even though Melbourne grew more rapidly to become one of the world's largest cities in the late 19th century. The effects of agglomeration, however, may be slightly washed out in Sydney because of the relatively small number of SLAs constituting the major statistical region. The sample in Sydney consisted of 15 fewer SLAs than in Melbourne, even though Sydney's workforce and geographic area exceeds that of Melbourne's by roughly 200,000 people and 4,000 km^2 respectively. The effect may be that Sydney's sample areas are too large to measure the agglomeration externalities within with as much accuracy as in Melbourne. Despite these potential explanations for Sydney producing comparably less convincing results than Melbourne, they are still consistent with existing studies that return aggregate industry elasticity estimates in the range of 0.03 to 0.10 and R-squared values in the range of 0.3 to 0.8.

Figure 4.1.3: Regression Results for Sydney – Elasticity of productivity estimate (measured with respect to employment density)



Another city to show particularly unusual results worth discussing to some extent is Canberra. In both cases, when elasticity was calculated with respect to employment density and size the outcomes were quite similar (as was the situation with Melbourne). Analysis returned exceptionally high elasticities, however, of 0.37-0.40 with 71% of the variance explained. Considering that most existing studies report elasticities of productivity ranging between 3% and 19% for international cities and countries, this comes as a great surprise. To some extent, this gross difference can be explained by addressing the unique circumstances surrounding the capital city. Canberra is virtually void of all manufacturing type industries and heavily established by those in professional, scientific, technical and public administration services. These are among the industries where agglomeration externalities have their greatest impacts. Furthermore, Canberra is very much a planned city with only a few major centres where most employment is concentrated, which may justify the high productivity elasticities to some degree. The more probable reason for the extremely high elasticity estimates is likely to be because of the size of the geographic units and the mix of industries actually existing within the SLAs, both bearing some influence over the productivity index values.

Table 4.1.1 below displays an overview of the results of the analysis while Table 4.1.2 is a tabulation of some differentiating measures of the capital cities that may assist in giving possible explanations for the contrasting city results.

Table 4.1.1: Capital City Regression Results

	Employment Density		Employment Size		Correlation between E_D & E_S
	Elasticity	R^2	Elasticity	R^2	
Perth	0.0348 (.0085)	0.3199	0.1315 (.0092)	.8475	0.3676
Sydney	0.0467 (.0053)	0.5518	0.1026 (.0152)	.4224	0.4243
Melbourne	0.0742 (.0042)	0.7998	0.1537 (.0091)	0.7891	0.8514
Brisbane	.1300 (.0106)	.4118	0.2524 (.0098)	0.7591	0.6102
Adelaide	0.1093 (.0108)	0.6625	.1807 (.0139)	0.7652	0.8073
Canberra	0.3671 (.0228)	0.7100	0.3958 (.0246)	0.7109	0.8575
Darwin	0.0395 (0.0369)	0.1602	0.3987 (0.0296)	0.8226	0.3345
Hobart	0.2045 (0.0602)	0.2325	0.1784 (0.0178)	0.9432	0.3796

Note: P-values for the employment density and employment size parameters are indicated in brackets.

Table 4.1.2: Descriptive Figures of the Capital City Statistical Divisions

	Total Area (km^2)	Total Employment (2006)	Total ED (jobs/sq-km)	Sample Size (# of SLAs)	Avg SLA size (km^2)
Perth	5,422	613,841	113	39	139
Sydney	12,428	1,714,395	138	64	194
Melbourne	8,097	1,526,364	189	79	102
Brisbane	5,905	784,327	133	215	27
Adelaide	1,826	465,893	255	55	33
Canberra	814	176,929	217	108	8
Darwin	3,135	47,863	15	41	76
Hobart	1,357	84,949	63	8	170

Reviewing some of the figures from the two above tables gives us some understanding of why such dramatically different results may come from analysing the capital cities separately using SLAs and the productivity index. It seems as though the cities that have higher correlations between employment size and employment density display more consistent results in their R-squared values with respect to the two variables. On the other hand, the average sizes of the SLAs vary immensely between the cities, where generally cities with larger SLAs have shown that much of the variance in productivity is left unexplained when measured with respect to density. Thus weak associations between productivity and employment density may be because dense centres are being washed out and appear to be of lower density because of the large statistical boundaries characteristic of some of the cities. At this point, it becomes unclear if this is an urban form phenomenon or a drawback of inconsistent SLA sizes, or both.

There are also a number of possible reasons why we observe elasticity estimates ranging between .03 to .36 when measured with respect to employment density and between .10 and .40 when measured with respect to employment size. It could be that productivity does improve so steeply with increases in the employment variables, but it is more likely that we can trace the cause of the differences back to the varying sizes of the SLAs and the productivity index. The smaller the geographic size of a SLA, the more unlikely it is that its employment composition will contain all the employment types characteristic of the larger regional or national economy. Furthermore, the more distant a SLA is from the central business district (CBD), the less likely it is that it will harbour a major concentration of knowledge-intensive activities that generally pay higher wages. As the productivity index holds employment composition constant, there is no way to predict the wages for job types that do not exist in a given area. Thus, the contribution of non-existent industries to the index will be zero times the average share.[17] The less diverse a geographic unit is in employment composition, the less the index is able to control for productivity differences and the more biased it becomes. This may be why we observe such large elasticities in locations such as Canberra and Brisbane, which have the smallest average SLA sizes, but it does not do well to explain the results for Hobart. Another possible explanation for cross-city differences in elasticity estimates may be that the benefits of agglomeration could be non-constant, reflective of congestion in larger city-centres reducing the benefits of agglomeration. This could be why Hobart reports very different results from the larger capitals such as Sydney and Melbourne. In light of all these considerations, there is reason to believe that the estimated elasticities for Perth, Sydney and Melbourne may be most accurate.

4.1.2 Analysis 1-1 Conclusions

This initial analysis followed a simple methodology to make some first-cut estimates of agglomeration economies in Australian capital cities. The approach was able to control for differences in the industrial mix of capital city SLAs while using some rather crude measures of agglomeration. A few of the results, particularly those from Melbourne, Sydney and Perth, were consistent with findings from a number of international studies. The rest of the results did not align with expected outcomes and this deviation was most likely a result of the productivity index giving biased results in geographic units that are too small and lack industrial diversity. This issue could partially be resolved by calculating the index while using more aggregated sectoral data, such as 1 or 2-digit ANZSIC data rather than the 3-digit data used here. This, on the other hand, would have other undesirable effects in the

[17] To explain the situation further, the productivity index requires a given location's mean wage for an industry to be multiplied against the national share of employment in that industry. If that industry does not exist in the given location then its mean wage will be zero and productivity will appear much lower. In this circumstance, which becomes more exacerbated with smaller geographic units, the productivity index fails to control for employment composition effects.

way of not distinguishing between high and low-value activities that may fall under the same broad sectoral heading – such as basic and specialized types of manufacturing.

A number of other limitations exist in this analysis, which includes the application of simplistic measures of agglomeration economies and a lack of controls for worker differences that exist apart from them being employed in different industries. In terms of the former issue, using employment size as an independent variable – and even employment density when applied to large geographic units – gives no consideration to the specifics of urban form, making it very difficult to interpret results across cities that differ in this respect. Similarly, calculating employment density by merely dividing employment by total geographic area can 'wash out' the effects of urban form by including areas that are not considered to be 'built-up'. This will tend to impede the interpretability of the results as well. In terms of the latter issue, incorporating other worker characteristics in the model, such as occupation type and levels of educational attainment, can prevent these factors from influencing the elasticity estimates of agglomeration. The next analysis will address these issues while investing the influences of agglomeration across individual industry types.

4.2.0 Analysis 1-2 Overview: Industry-specific elasticity estimations using SLA data

Moving forward, all analyses will be estimating agglomeration economies on an industry-specific basis and solely for the capital cities of Sydney and Melbourne. Following the theoretical model described in section 3.3, industry-specific wage functions are econometrically estimated while controlling for labour characteristics likely to influence wages. These wage functions are estimated for 60 3-digit ANZSIC industries while including a control variable to represent the degree to which an area is agglomerated, namely a measure of 'effective density'. The dependant variable, which now is industry-specific mean income, is Cobb-Douglas in the wage-determining factors that include effective density (*U*), occupation type (*Occ*), education level (*Edu*) and experience (*Exp*). This formulation is shown in the following equation where income (*I*) in industry k and location m is determined by the above-specified factors.

$$I_{k,m} = U_m^{\beta} \cdot \prod_i Occ_{i,k,m}^{\delta_i} \cdot \prod_j Edu_{j,k,m}^{\gamma_j} \cdot Exp_{k,m}^{\psi} \quad [4C]$$

The controls addressing observable worker heterogeneity were sourced from census data that gave aggregate figures at the SLA level. As such, data on education and occupation were provided in the form of total numbers of employed persons in a given location and industry that fell into one of two education categories and one of eight occupation categories. Algebraically, the share of employment within industry k, location m, and occupation group o can be expressed as shown below.

$$Occ_{k,m,o}^{Share} = \frac{E_{k,m,o}}{E_{m,o}} \quad [4D]$$

where E denotes the number of people employed.

Similarly, the share of employment with an educational attainment of level a in industry k and location m can be expressed as follows:

$$Edu_{k,m,a}^{Share} = \frac{E_{k,m,a}}{E_{k,m}} \quad [4E]$$

The data on industry-specific mean age of employment for a given geographic area did not need any augmentation as it was provided directly by the ABS consultant. The approach to calculating the effective densities of the SLAs within Sydney and Melbourne, on the other hand, requires some more detailed explanation. This will be provided in the following section.

Calculating Effective Density for Sydney and Melbourne SLAs

To reiterate what was explained above in section 2.9, effective density is a measure of economic concentration that not only takes into account a given area's employment density, but also factors in the area's location in the context of the regional spatial economy. It does so by calculating own-area employment density and adding the sum of employment in all surrounding units that has been weighted by Euclidean distance, a generalized cost of travel, or travel time. In the case of this analysis, two effective density indices were specified for Sydney: one weighting employment by travel time and the other by Euclidean distance. The index was calculated for Melbourne solely by weighting surrounding employment by linear distance, thus providing only one version of the index.

To take into account that the settlement of economic activity does not necessarily occur directly in the centre of each SLA, an employment-weighted centroid for each SLA was first estimated. This meant calculating employment densities at the WDZ or TZ level and determining a threshold below which a geographic unit would be classified as having 'insufficient economic intensity' to have its area considered in an SLA's density and centroid calculation. After the assessment of the employment zone employment patterns, this threshold was determined to be at a level of 0.1 employees per hectare. The geographic areas of the SLAs were then adjusted by omitting zones that fell below this threshold and recalculating their geographic centre-points. This was done in a GIS program called MapInfo, in which employment data was merged with spatial layers of SLA and WDZ digital boundaries.

In order to calculate linearly weighted effective density indices for Sydney and Melbourne SLAs, the geographic coordinates of the adjusted SLA centroids were recorded so that the Euclidean distance between each pair of SLAs could be calculated. These calculations were carried out in Microsoft Excel but used the Haversine formula for estimating the linear distance between two geographic coordinates. The formula is derived from the 'spherical law of cosines' and as such takes into account the curvature of the earth. The formula appears as written below.

$$d_{ij} = a\cos\left(\sin(lat_i)\cdot\sin(lat_j) + \cos(lat_i)\cdot\cos(lat_j)\cdot\cos(long_j - long_i)\right)R \qquad [4F]$$

where *R* represent the radius of the earth, which was set at 6371 km. To add, all latitudes and longitudes first had to be converted to radians before being placed into the formula.

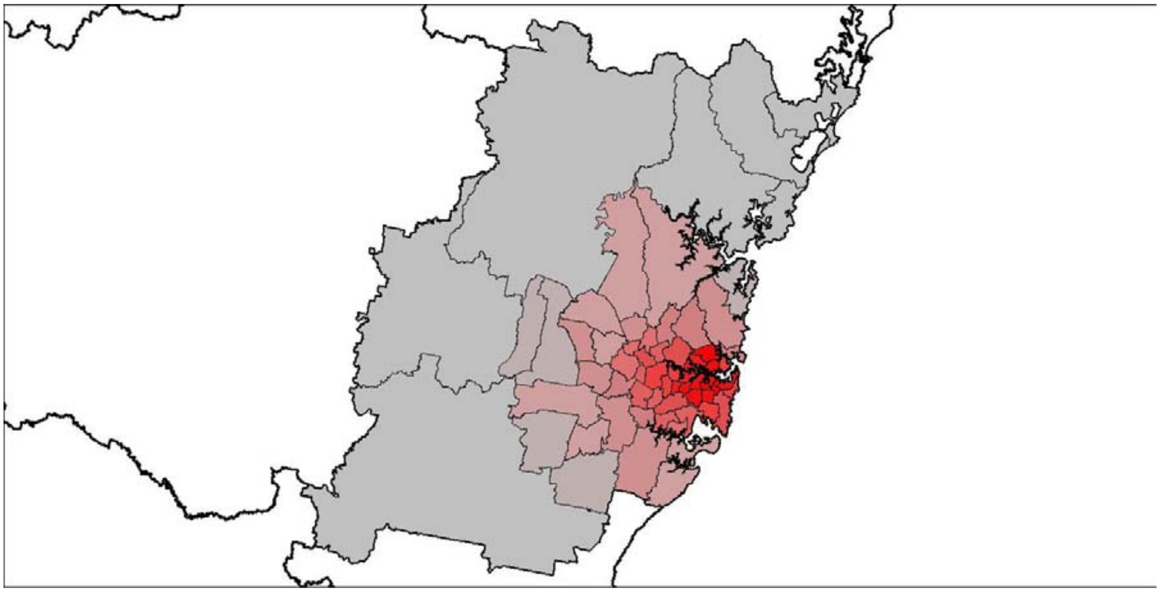
Once the linear distances between pairs of SLAs were determined, their values were set up in a cross-table of origin-destination zones in Excel and matched with total SLA employment data, from which effective densities were calculated according to the formula shown below.

$$U_i = \frac{E_i}{\sqrt{(A_i/\pi)}} + \sum_{j=1}^{j} \frac{E_j}{d_{ij}} \qquad [4G]$$

Effective density is represented here as U_i to emphasize that urbanization economies are being captured because of the index's calculation using total rather than industry-specific employment.

The alternative form of the effective density index for Sydney, where proximate spatial units were weighted by travel time, was calculated in a similar fashion. After adjusting the SLA areas of Sydney by omitting employment zones that were determined to be of insufficient economic intensity, the centremost travel zone of each SLA was recorded and supplied in a list to the Transport Data Centre. There, a transport consultant estimated and provided the a.m. peak, daytime inter-peak, and p.m. peak travel times between the pairs of designated zones. These travel times were then used as weights to replace d_{ij} in the above specification of effective density. The travel-time radius of own-area *i* was estimated by calculating the time it would take to travel the radius of each SLA if moving at an average speed of 40 km/hr. Figure 4.2.1 below shows a thematic map of Sydney representing effective density levels, where redder regions are characterized by greater levels of effective density.

Figure 4.2.1: A Thematic Map of Effective Density in Sydney SLAs



Once the employment shares of education and occupation were calculated and the effective density indices were prepared, the econometric estimation of the wage functions across the 60 selected industries could be carried out.[18] By taking logs of both sides of equation 4C and adding a disturbance term we get the econometric specification depicted below.

$$\ln I_{k,m} = \alpha_k + \beta \ln U_m + \sum_{i=1}^{i} \delta_i \ln Occ_{i,k,m} + \sum_{j=1}^{j} \gamma_j \ln Edu_{j,k,m} + \psi \ln Exp_{k,m} + e_{k,m} \qquad [4H]$$

This formula represents a functional form where all coefficients are linear in logs and as such, can be interpreted as elasticities. The industry-specific regressions were first carried out using ordinary least squares (OLS), which were then followed by conducting Breusch-Pagan tests for heteroskedasticity. In instances where the test rejected the null

[18] As all occupation and education shares take on values ranging from zero to one, taking their natural logs will in most cases produce negative values up to a maximum value of zero. To address this issue for each share, before taking its natural log, each was increased by a value of 1, creating a monotonic shift in the dataset. The result is that categories with a 0% share of employment will still have a value of zero after taking the log [Ln(1) = 0] while categories with a 100% share of employment will take a post-log value of 0.693 [Ln(2) = 0.693…].

hypothesis of the errors having constant variance (using an α of 10%), the generalized least squares (GLS) method was used instead. The term α is the probability of committing a Type-I Error, or in other words, the probably of rejecting the null hypothesis when in fact it should have been maintained.

The issue of heteroskedasticity arises when the variance of the errors (the deviations between predicted and observed values) is not constant across a sample. This issue is extremely common in cross-sectional analyses and while its effect is not of biasing the parameter estimates, its implication is that the OLS estimator no longer provides the 'best' parameter estimates and the standard errors will be incorrect (and consequently so will any estimated confidence intervals or hypothesis tests). Heteroskedasticity can be addressed by opting to compute 'robust standard errors', which is an option given by statistical software packages. Doing so will improve on the latter issue – the inefficiency of the parameter estimates because of larger standard errors – but will not address the former issue. The alternative is to use GLS instead of OLS, which allocates weights to the independent variable values that are inversely proportional to the predicted variances of the disturbances derived from an auxiliary regression. The motivation for using GLS was to improve both the accuracy and efficiency of the parameters being estimated.

4.2.1 Analysis 1-2 Results

In this section, the outcomes of the individual regressions estimating industry-specific wage functions for 60 selected industries are reported. Unfortunately, results could not be generated for (70) Oil and Gas and (109) Other Mining Services in Sydney because of insufficient sample sizes. Similarly, the results for industry 109 in Melbourne were derived from using OLS with robust standard errors instead of GLS because of too few observations. Results are first given on the Sydney and Melbourne regressions, followed by an account of the outcome of pooling the data on the two cities. Some comparisons will be made along the way to industry-specific results from production function estimates produced by Graham (2005; 2006; 2007a) using UK data. Finally, concluding remarks will be given along with a discussion of some limitations of this analysis.

Sydney and Melbourne Results

The findings reported here are of the agglomeration elasticity estimates (ε_U) derived from estimating industry wage functions. More specifically, these results reveal the effect that employment concentration and proximity have on the productivity of labour. Tables 4.2.1 and 4.2.2 below contain information on the industry ANZSIC code, the industry name, the elasticity of productivity estimated with respect to effective density, the standard error of the effective density parameter estimate, and the F-Value of each industry regression along with their adjusted R-squared values and sample sizes for the city of Sydney, conducted with linear effective density and travel time effective density, respectively. Spaces are left between groupings of like industries to assist in comparing estimates within like and across different industry sectors.

First addressing the wage function results estimated with the linear specification of effective density, 27 of the 58 industries returned significant parameter estimates of the effective density variable. The largest estimated significant elasticities are for the industries of *Data Processing, Web Hosting, & Electronic Information Storage Services* (0.294); *Financial Asset Investing* (0.237); *Other Professional, Scientific and Technical Services* (0.227); *Specialized Industrial Material Wholesaling* (0.226); and *Television Broadcasting* (0.201). At the opposite end of the spectrum, the lowest elasticities are reported for the industries of *Water Transport Support Services* (-0.114); *Supermarket and Grocery Stores* (-0.063); *Warehousing and Storage Services* (-0.049); *Depository Financial* (-0.031); and *Cafes, Restaurants and Takeaway Food* (-0.004). Of all the industries with negative reported elasticities, the only one to be statistically significant (and only marginally so) was *Supermarket and Grocery Stores*.

In general, the larger positive estimates support the view that knowledge-driven industries based on employment with high-quality knowledge content benefit most from agglomeration effects. These include scientific, technical, professional, financial, and media-related services. It was observed that manufacturing industries, as traditionally argued, do experience a significant benefit from co-location; however, the magnitude of this benefit is somewhat restricted to a lower elasticity range of roughly 5% to 7%. Lower-order industries comprising the retail sector, to my knowledge, have not had agglomeration effects estimated before in published literature to-date. The results generated here for this sector give point-elasticity estimates near or around zero, which integrates well with models of urban growth and urban formation such as New Economic Geography and Central Limit Theory. They postulate that lower-order industries, such as retail, are the first to leave a well-agglomerated centre to service a growing residential population on the edge of city limits. This would support why retail receives such low estimates, as clustering near other employment will have little impact on productivity and if it does, the benefit may be eroded away by high rents. The industry of *Public Order and Safety* is in a similar situation where its location is predominantly determined by population at large and not necessarily by the locations of firms. Hospitals and medical services on the other hand are population-driven services that seem to experience fairly strong agglomeration effects.

That 31 of the 58 industry wage-function estimations returned insignificant effective density parameter estimates is not an issue that raises the question of whether the econometric model or the data are inadequate. On the contrary, the use of aggregated employment data on statistical local areas seems to prove rather effective. In the great majority of cases, the adjusted R-squared values are over 0.50 and are often as high as 0.80 to 0.95. This may to some extent reflect that the data are provided as averages. Thus, much of the disturbance in the data may be 'smoothed out'; these values are rather high and suggest that a great deal of the variance in wages can be accounted for by the controls imposed. More important than the adjusted R-squared values, however, are the p-values from the F-tests. The F-test is a test of significance for the entire regression. It involves a comparison of the sum of squared errors from an original (unrestricted) regression with the sum of squared errors from a regression model in which the null hypothesis is assumed to be true. In this case, the null hypothesis is a joint one and assumes that all parameters, excluding the constant, are equal to zero. In other words, it tests whether the combined variables in the specified model do better to explain the variance in income than having none of them at all. In all instances, except for the regression for the industry, *Metal Ore Mining*, the p-values for the F-tests suggest that the models being estimated are highly significant as indicated by very small p-values.

The insignificant effective density parameters for a number of industries in most cases can be attributed to their point estimates being rather close to zero. This means that even if their standard errors are small, the confidence intervals around their parameter estimates are likely to include zero and thus result in the conclusion that they are insignificant. A deviation from this where a parameter is given a large yet insignificant estimate would be an industry such as *Water Transport Support Services*, where the coefficient estimate is -0.114 and the standard error is 0.085. Given that the adjusted R-squared value for this industry is very high at approximately 0.99, this suggests that the rest of the variables in the model aptly account for variations in wage in this industry with a rather wide range of possible influence from agglomeration. In an industry such as this, there are likely to be externalities arising from human or natural endowments that benefit the presence of firms rather their employment concentration per se. This would mean that a river mouth or port would be the fundamental determinant of location, rather than the presence of economic mass. On the other hand, there were no controls for endowments in the models and with such a high R-squared value one may assume a high level of multicollinearity to be present with another variable in the equation.

Among the many industries having positive elasticity estimates with respect to effective density, it is interesting to note the occurrence of a few negative elasticities, such as in the industries of *Water Transport Support Services*, *Supermarkets and Grocery Stores*, and *Depository Financial*. The former industry had such a large error term associated with it that it is very possible for the true estimate to be positive yet small. Alternatively the negative elasticity could be explained because of the dependence on physical endowments rather than some effect of agglomeration diseconomy. For industries such as the latter two, the causes of a negative elasticity estimate are likely to differ. Still there is the potential for the true values to be positive yet small but have negative estimates because of estimation error, but there may also be other possible explanations. They could be an indication of the presence of an endogeneity issue where, as larger urban areas host a great number of employment opportunities and initially pose a productivity uplift that raises wages, they in turn act as a great attracter of employment. Then the relatively low requirement of labour to be highly skilled and educated in these industries could bid down wages if competition for these jobs is high. Alternatively, the negative estimates could simply be because the model specification estimates a constant elasticity, whereas in the case of these industries the relationship between effective density and productivity may actually be convex. As such, productivity in these industries may be increasing at lower levels of effective density and decreasing at higher levels. The average of this relationship may be negative or close to zero with a degree of error making up for the rest of the negative magnitude.

When estimating industry elasticities in Sydney with respect to travel-time weighted effective density, 45 of the 58 results were of greater absolute magnitude than when measured with respect to linearly weighted effective density. These results are consistent with the findings of Graham (2006) who estimates returns to agglomeration across nine broad industry sectors in the UK. He finds higher elasticities in eight of the sectors when estimated with a generalized cost of travel effective density as opposed to linear effective density. While some of these differences may not be statistically significant, they are an indication that congestion in more highly urbanized areas of the city may be effectively reducing economic density and as a result restricting the productivity-enhancing benefits of agglomeration. It is difficult to make firm judgements on the actual magnitudes of differences between linear and travel-time weighted effective density elasticity results because the SLA estimates produced contain relatively large errors; however, the general pattern suggests a productivity-reducing effect of congestion levels in the city.

Table 4.2.1: Sydney Results Using SLA Data (ED weighted by linear distance)

Ind ID	Industry Name	ε_U	S.E.	P-Value (F)	Adjusted R^2	Sample Size (N)
70	Oil and Gas Extraction	N/A				12
80	Metal Ore Mining	0.354	0.396	0.181806	0.288	21
101	Exploration	0.085	0.175	0.348	0.136	23
109	Other mining support services	N/A				12
135	Clothing and Footwear Product Manufacturing	**0.167	0.066	3.54E-06	0.489	62
184	Pharmaceutical and Medicinal Product Manufacturing	0.073	0.053	1.00E-07	0.653	51
241	Professional and Scientific Equipment Manufacturing	*0.049	0.029	2.31E-12	0.750	59
242	Computer and Electronic Equipment Manufacturing	0.054	0.051	8.63E-14	0.799	57
246	Specialized Machinery and Equipment Manufacturing	0.056	0.050	3.83E-13	0.832	49
C00	Manufacturing, nfd	**0.062	0.026	1.26E-12	0.723	64
301	Residential Building Construction	0.051	0.050	5.01E-09	0.608	64
310	Heavy and Civil Engineering Construction	*0.050	0.029	9.99E-34	0.960	64
320	Construction Services, nfd	0.089	0.078	0.516853	-0.013	57

341	Specialized Industrial Material Wholesaling	***0.226	0.018	3.18E-26	0.950	56
391	Motor Vehicle Retailing	0.008	0.039	9.94E-26	0.933	60
411	Supermarket and Grocery Stores	*-0.063	0.032	1.72E-20	0.867	64
425	Clothing, Footwear and Personal Accessory Retailing	0.035	0.022	1.22E-32	0.956	64
427	Pharmaceutical and Other Store-Based Retailing	0.008	0.024	5.88E-18	0.832	64
440	Accommodation	***0.066	0.022	4.79E-23	0.900	63
451	Cafes, Restaurants & Takeaway Food	-0.004	0.031	3.36E-25	0.914	64
452	Pubs, Taverns & Bars	0.011	0.033	2.62E-08	0.579	64
461	Road Freight Transport	0.021	0.031	2.87E-07	0.532	64
462	Road Passenger Transport	0.024	0.062	2.18E-02	0.197	63
521	Water Transport Support Services	-0.114	0.085	4.24E-21	0.990	35
529	Other Transport Support Services	***0.132	0.040	8.07E-13	0.755	60
530	Warehousing and Storage Services	-0.049	0.064	6.21E-06	0.482	61
541	Newspaper, Periodical, Book, & Directory Publishing	**0.111	0.053	1.08E-10	0.673	63
551	Motion Picture & Video Activities	0.028	0.041	4.15E-36	0.976	60
562	Television Broadcasting	*0.201	0.106	1.57E-03	0.432	44
580	Telecommunications Services	*0.055	0.031	1.45E-20	0.868	64
591	ISPs & Web Search Portals	*0.107	0.062	1.01E-16	0.941	42
592	Data Processing, Web Hosting, & Electronic Information Storage Services	***0.294	0.047	5.85E-13	0.884	43
620	Finance, nfd	***0.140	0.035	1.21E-09	0.631	64
622	Depository Financial	-0.031	0.046	2.38E-10	0.655	64
624	Financial Asset Investing	***0.237	0.058	3.04E-06	0.543	55
631	Life Insurance	0.027	0.052	5.73E-19	0.978	36
632	Health & General Insurance	0.043	0.057	9.70E-05	0.395	63
641	Auxiliary Finance & Investment Services	0.059	0.037	6.47E-12	0.704	64
642	Auxiliary Insurance Services	0.064	0.040	2.50E-15	0.833	55
670	Property Operators & Real Estate Services, nfd	0.113	0.118	0.614454	-0.037	57
671	Property Operators	0.033	0.020	1.44E-10	0.669	63
672	Real Estate Services	0.026	0.022	2.16E-30	0.946	64
690	Prof, Sci & Tech Services, nfd	***0.176	0.058	3.26E-12	0.712	64

691	Scientific Research Services	0.011	0.048	5.69E-27	0.951	57
692	Arch, Eng & Tech Services	*0.044	0.023	5.79E-19	0.847	64
693	Legal & Accounting Services	0.030	0.040	1.45E-06	0.497	64
694	Advertising Services	***0.157	0.048	1.06E-14	0.773	64
695	Market Research & Stat Services	0.030	0.038	5.97E-32	0.956	63
696	Management & Consulting Services	**0.067	0.025	1.67E-19	0.855	64
699	Other Prof, Sci & Tech Services	***0.227	0.066	0.011833	0.221	64
700	Computer System Design	***0.093	0.021	6.44E-41	0.979	64
751	Central Government Administration	***0.086	0.027	1.16E-12	0.751	60
752	State Government Administration	0.032	0.026	3.84E-08	0.572	64
753	Local Government Administration	**0.040	0.018	3.90E-35	0.967	63
754	Justice	0.061	0.051	1.47E-15	0.940	39
771	Public Order and Safety Services	0.021	0.013	4.03E-34	0.962	64
810	Tertiary Education	***0.051	0.012	1.14E-21	0.881	64
840	Hospitals	***0.061	0.021	5.74E-06	0.484	61
851	Medical Services	***0.055	0.015	4.90E-11	0.678	64
853	Allied Health Services	***0.074	0.018	5.29E-08	0.566	64

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

Table 4.2.2: Sydney Results Using SLA Data (ED weighted by travel time)

Ind ID	Industry	ε_U	S.E.	P-Value (F)	Adjusted R^2	Sample Size (N)
70	Oil and Gas Extraction	N/A				12
80	Metal Ore Mining	0.212	0.483	0.220645	0.246	21
101	Exploration	0.091	0.236	0.360479	0.126	23
109	Other mining support services	N/A		□		12
135	Clothing and Footwear Product Manufacturing	**0.227	0.089	3.51E-06	0.489	62
184	Pharmaceutical and Medicinal Product Manufacturing	0.069	0.067	1.45E-07	0.646	51
241	Professional and Scientific Equipment Manufacturing	*0.093	0.047	9.86E-12	0.733	59
242	Computer and Electronic Equipment Manufacturing	0.095	0.069	3.30E-12	0.760	57
246	Specialized Machinery and Equipment Manufacturing	0.107	0.077	4.14E-15	0.869	49
C00	Manufacturing, nfd	**0.079	0.036	1.97E-12	0.718	64

301	Residential Building Construction	0.085	0.066	3.82E-09	0.612	64
310	Heavy and Civil Engineering Construction	0.026	0.033	1.21E-20	0.869	64
320	Construction Services, nfd	0.111	0.105	0.534926	-0.017	57
341	Specialized Industrial Material Wholesaling	***0.310	0.030	2.28E-32	0.974	56
391	Motor Vehicle Retailing	-0.017	0.061	7.08E-26	0.934	60
411	Supermarket and Grocery Stores	***-0.113	0.032	8.54E-29	0.938	64
425	Clothing, Footwear and Personal Accessory Retailing	*0.053	0.031	1.60E-31	0.952	64
427	Pharmaceutical and Other Store-Based Retailing	0.023	0.031	4.72E-18	0.834	64
440	Accommodation	***0.096	0.033	5.84E-22	0.889	63
451	Cafes, Restaurants & Takeaway Food	0.013	0.038	3.19E-25	0.914	64
452	Pubs, Taverns & Bars	-0.046	0.032	6.60E-16	0.797	64
461	Road Freight Transport	0.026	0.044	3.02E-07	0.531	64
462	Road Passenger Transport	0.002	0.091	0.02288	0.195	63
521	Water Transport Support Services	-0.081	0.062	7.21E-26	0.996	35
529	Other Transport Support Services	0.129	0.071	2.24E-07	0.565	60
530	Warehousing and Storage Services	-0.041	0.087	7.28E-06	0.478	61
541	Newspaper, Periodical, Book, & Directory Publishing	***0.170	0.044	6.90E-21	0.877	63
551	Motion Picture & Video Activities	0.083	0.052	6.49E-31	0.960	60
562	Television Broadcasting	*0.245	0.134	1.72E-03	0.427	44
580	Telecommunications Services	**0.109	0.045	2.84E-21	0.877	64
591	ISPs & Web Search Portals	*0.162	0.087	2.18E-16	0.938	42
592	Data Processing, Web Hosting, & Electronic Information Storage Services	***0.276	0.093	2.65E-14	0.906	43
620	Finance, nfd	***0.213	0.071	1.20E-05	0.447	64
622	Depository Financial	-0.013	0.058	2.88E-10	0.653	64
624	Financial Asset Investing	***0.320	0.059	5.53E-09	0.672	55
631	Life Insurance	0.004	0.072	2.82E-16	0.962	36
632	Health & General Insurance	0.068	0.076	8.90E-05	0.398	63
641	Auxiliary Finance & Investment Services	*0.088	0.049	4.82E-12	0.707	64
642	Auxiliary Insurance Services	0.044	0.031	9.53E-34	0.961	55

670	Property Operators & Real Estate Services, nfd	0.148	0.157	0.617246	-0.037	57
671	Property Operators	0.025	0.029	1.33E-15	0.797	63
672	Real Estate Services	0.044	0.031	9.53E-34	0.961	64
690	Prof, Sci & Tech Services, nfd	***0.200	0.063	8.81E-12	0.700	64
691	Scientific Research Services	*0.077	0.041	1.53E-26	0.948	57
692	Arch, Eng & Tech Services	***0.098	0.033	3.09E-24	0.906	64
693	Legal & Accounting Services	*0.085	0.085	5.42E-07	0.519	64
694	Advertising Services	***0.211	0.064	7.33E-16	0.796	64
695	Market Research & Stat Services	0.087	0.065	3.65E-27	0.932	63
696	Management & Consulting Services	**0.073	0.031	2.93E-19	0.851	64
699	Other Prof, Sci & Tech Services	***0.312	0.090	0.011755	0.222	64
700	Computer System Design	***0.107	0.021	1.43E-34	0.963	64
751	Central Government Administration	***0.113	0.031	2.17E-13	0.769	60
752	State Government Administration	0.048	0.036	3.36E-08	0.574	64
753	Local Government Administration	***0.075	0.027	4.37E-18	0.840	63
754	Justice	0.075	0.072	1.63E-14	0.928	39
771	Public Order and Safety Services	*0.031	0.017	4.92E-33	0.958	64
810	Tertiary Education	***0.074	0.019	1.48E-25	0.917	64
840	Hospitals	***0.095	0.028	1.78E-06	0.512	61
851	Medical Services	***0.081	0.020	1.60E-11	0.692	64
853	Allied Health Services	***0.096	0.024	4.51E-08	0.569	64

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

Now shifting the focus to the Melbourne productivity elasticities estimated with respect to linear effective density, results returned statistically significant coefficients for 37 of the 60 industries for which wage functions were estimated (compared to only 27 in Sydney). This is likely because Melbourne has more SLAs, thus the industry sample sizes can be expected to be slightly larger, even after accounting for zones that contain nil employment in the industry being investigated. Table 4.2.3 summarizes all these results, where the largest estimates were for *ISPs and Web Search Portals* (0.343); *Financial Asset Investing* (0.293); *Construction Services, nfd* (0.257), *Computer System Design* (0.220); and *Life Insurance* (0.219). *Metal Ore and Mining* also returns a larger elasticity (0.288) that is significant in Melbourne's case, though the standard error is quite large suggesting that the true value could deviate from this point estimate by a fair bit. *Data Processing, Web Hosting, & Electronic Information Storage Services* also produces a large elasticity of 0.531; however, the model as a whole is insignificant with an F-value in excess of 0.10.[19] Disregarding the mining industries, lowest coefficients were estimated for *Supermarkets and Grocery Stores* (-0.104), *Computer and Electronic Equipment Manufacturing* (-0.043); *Justice* (-0.028); *Depository Financial* (-0.016); and *Cafes, Restaurants and Takeaway* (-0.014).

In general, estimates for Melbourne are slightly larger than in Sydney, but the relative magnitudes of the estimates between industries appear fairly consistent. A couple of exceptions to this are for the industries of *Television Broadcasting*, *Specialized Industrial Material Wholesaling,* and *Computer and Electronic Equipment Manufacturing*. The former's estimate dropped from 0.201 in Sydney to 0.004 in Melbourne while in the case of the second industry mentioned, the estimate dropped from 0.226 to 0.102. The latter industry's estimate dropped from 0.054 to -0.043. Without more information, it is difficult to determine whether these large discrepancies are the result of issues in the datasets or if they can be explained away. It is possible, for instance, to hypothesize that the nature of the activities within these industry classifications differs considerably between the two cities, thus agglomeration economies may accrue differently between them. Alternatively, there may be an unobserved variable that is correlated with effective density in one of the cities that is influencing the estimates in one of the cases. To see if the wage function specification generates elasticity estimates that can be generalized to the two capital cities, we turn to the results of the pooled industry regressions.

Table 4.2.3: Melbourne Results Using SLA Data (ED weighted by linear distance)

Ind ID	Industry	ε_U	S.E.	P-Value (F)	Adjusted R^2	Sample Size (N)
70	Oil and Gas Extraction	-0.014	0.135	0.101799	0.510	20
80	Metal Ore Mining	*0.275	0.142	1.14E-02	0.623	23
101	Exploration	-0.071	0.210	0.316561	0.098	27
109	Other mining support services	-0.086	0.416	6.68E-06	0.276	18

[19] The reason for this outcome is unclear as it is the only industry in Melbourne other than *Oil and Gas Extraction* to have an insignificant wage function estimate. The data was checked and re-run with the same outcome resulting.

135	Clothing and Footwear Product Manufacturing	0.176	0.025	6.57E-49	0.980	73
184	Pharmaceutical and Medicinal Product Manufacturing	*0.117	0.069	1.41E-06	0.686	42
241	Professional and Scientific Equipment Manufacturing	0.013	0.078	2.30E-04	0.364	64
242	Computer and Electronic Equipment Manufacturing	-0.043	0.083	4.33E-02	0.169	61
246	Specialized Machinery and Equipment Manufacturing	0.093	0.067	3.31E-04	0.368	60
C00	Manufacturing, nfd	**0.100	0.041	1.03E-09	0.545	79
301	Residential Building Construction	**0.063	0.024	2.68E-29	0.890	79
310	Heavy and Civil Engineering Construction	***0.100	0.025	1.96E-52	0.978	79
320	Construction Services, nfd	***0.257	0.064	2.37E-10	0.822	42
341	Specialized Industrial Material Wholesaling	0.102	0.082	3.20E-05	0.445	60
391	Motor Vehicle Retailing	***0.103	0.033	3.22E-16	0.763	71
411	Supermarket and Grocery Stores	***-0.104	0.024	1.98E-31	0.906	79
425	Clothing, Footwear and Personal Accessory Retailing	0.015	0.030	7.30E-30	0.895	79
427	Pharmaceutical and Other Store-Based Retailing	*0.046	0.046	6.04E-27	0.870	79
440	Accommodation	**0.079	0.033	1.74E-10	0.571	79
451	Cafes, Restaurants & Takeaway Food	-0.014	0.040	2.89E-14	0.677	79
452	Pubs, Taverns & Bars	**0.041	0.019	1.45E-30	0.903	78
461	Road Freight Transport	***0.085	0.020	1.68E-16	0.725	79
462	Road Passenger Transport	0.029	0.047	1.86E-16	0.735	77
521	Water Transport Support Services	***0.101	0.023	7.90E-39	0.999	30
529	Other Transport Support Services	**0.145	0.060	3.42E-20	0.849	67
530	Warehousing and Storage Services	0.101	0.066	6.14E-04	0.333	64
541	Newspaper, Periodical, Book, & Directory Publishing	***0.151	0.030	2.01E-30	0.919	73
551	Motion Picture & Video Activities	0.118	0.088	4.01E-11	0.667	66
562	Television Broadcasting	0.004	0.066	1.86E-25	0.991	39
580	Telecommunications Services	0.069	0.051	1.27E-11	0.628	75
591	ISPs & Web Search Portals	***0.343	0.099	3.19E-29	0.986	45
592	Data Processing, Web Hosting, & Electronic Information Storage Services	***0.531	0.184	0.252085	0.076	45

620	Finance, nfd	**0.135	0.067	1.13E-19	0.811	73
622	Depository Financial	-0.016	0.048	1.18E-12	0.635	79
624	Financial Asset Investing	***0.322	0.064	5.79E-08	0.557	65
631	Life Insurance	***0.219	0.064	1.95E-06	0.776	33
632	Health & General Insurance	*0.050	0.027	2.01E-20	0.832	71
641	Auxiliary Finance & Investment Services	***0.169	0.052	3.89E-06	0.401	78
642	Auxiliary Insurance Services	0.054	0.071	2.27E-03	0.306	60
670	Property Operators & Real Estate Services, nfd	0.105	0.068	2.17E-24	0.930	57
671	Property Operators	***0.176	0.035	7.00E-17	0.759	74
672	Real Estate Services	0.037	0.038	5.87E-09	0.517	79
690	Prof, Sci & Tech Services, nfd	***0.161	0.030	6.16E-12	0.673	67
691	Scientific Research Services	0.028	0.048	7.82E-06	0.451	65
692	Arch, Eng & Tech Services	**0.105	0.040	7.94E-13	0.640	79
693	Legal & Accounting Services	***0.109	0.030	5.10E-23	0.829	79
694	Advertising Services	**0.173	0.078	2.83E-12	0.652	74
695	Market Research & Stat Services	***0.185	0.040	1.14E-37	0.939	79
696	Management & Consulting Services	***0.130	0.039	7.70E-12	0.613	79
699	Other Prof, Sci & Tech Services	***0.214	0.041	1.91E-21	0.841	72
700	Computer System Design	***0.220	0.032	2.39E-31	0.905	79
751	Central Government Administration	0.020	0.015	5.06E-25	0.897	68
752	State Government Administration	***0.155	0.051	8.33E-04	0.277	75
753	Local Government Administration	***0.054	0.018	2.88E-28	0.889	77
754	Justice	-0.028	0.410	0.337546	0.095259	27
771	Public Order and Safety Services	0.037	0.025	1.16E-06	0.426	78
810	Tertiary Education	***0.067	0.021	4.62E-07	0.445	78
840	Hospitals	***0.108	0.018	2.20E-53	0.984	75
851	Medical Services	***0.085	0.020	1.70E-05	0.364	79
853	Allied Health Services	***0.101	0.015	4.26E-19	0.773	79

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

Sydney and Melbourne Pooled Dataset Results

To test whether the magnitudes of industry-specific effects from agglomeration can be generalized to Sydney and Melbourne, insofar as is revealed by the SLA data, a Chow test was conducted on the pooled city dataset. The Chow test tests for whether a structural break exists between the two datasets or, in other words, whether it makes sense to estimate consistent parameters for Sydney and Melbourne. Testing for this is achieved by first estimating a restricted model with the pooled data, which is a regression that runs the standard model specification that does not distinguish between either of the two cities. It is called a restricted model because it does not allow the parameters to differ between the two cities. Following this is an estimation of an unrestricted model that includes all the variables of the standard model specification whilst including a city dummy variable and a number of interaction terms between the city dummy and all other controls. The Chow test then compares the explanatory power of these two models after adjusting for degrees of freedom, and essentially gives a probability that the unrestricted model fares better at predicting income than the unrestricted model.

In the pooled dataset, a city dummy variable for Sydney (D_S) is defined such that the restricted and unrestricted models appear as below.

Restricted Model Specification:

$$I_{k,m} = \alpha_{k,m} + \beta U_m + \sum_{i=1}^{i} \delta_i Oc_{i,k,m} + \sum_{j=1}^{j} \gamma_j Ed_{j.k.m} + \psi Ex_{k,m}$$

Unrestricted Model Specification:

$$I_{k,m} = \alpha_{k,m} + \beta U_m + \sum_{i=1}^{i} \delta_i Oc_{i,k,m} + \sum_{j=1}^{j} \gamma_j Ed_{j.k.m} + \psi Ex_{k,m} + D_S + D_S \beta U_m + D_S \sum_{i=1}^{i} \delta_i Oc_{i,k,m} + D_S \sum_{i=1}^{i} \gamma_j Ed_{j.k.m} + D_S \psi Ex_{k,m}$$

After running OLS on the pooled industry samples and conducting Chow tests on the industry-specific results, if no structural difference between the datasets was detected then GLS was run on the data (in the presence of heteroskedasticity) and the pooled regression results recorded. If, on the other hand, the null hypothesis of "no structural difference" between the two datasets was rejected at a p-value level of 0.10 then further examination for the source of the structural difference was carried out. The Chow test is effective at detecting a difference between the parameter estimates between two samples, but does not offer any insight as to which of the control variables this difference can be attributed. Only one of the interaction dummies or the city dummy needs to be significantly different from zero, or a number of controls to be jointly significant, to have the Chow test reject the null hypothesis. Thus if this was the case, individual p-values on the interaction and dummy variables were further examined. If an interaction term between the Sydney dummy and the effective density variable was insignificant then it was omitted from the model and subsequently OLS or GLS were run (depending on if heteroskedasticity was present) and the pooled results were recorded. If an interaction term was significant, then it was concluded that the elasticity of productivity with respect to effective density in the two cities differed and no industry-specific result was recorded.

The results of this process revealed a structural difference between the two city datasets for 30 of the 60 industries. Of the 30 industries for which a difference was detected, only seven were concluded to experience a different agglomeration effect. These industries included *Road Freight Transport*, *ISPs and Web Search Portals*, *Computer System Design*, *Central Government Administration*, *State Government Administration (*but not *Local Government Administration*), *Hospitals*, and *Medical Services* (but not *Allied Health Services*). A total of 35 industries were estimated with significant effects from agglomeration.

In industries for which a pooled elasticity could be recorded, the general effect was a reduction in the magnitude of standard error, as one would expect from having a larger sample size. This resulted in the industries of *Profession and Scientific Equipment Manufacturing*, *Specialized Machinery and Equipment Manufacturing, Motion Picture and Video Activities, Television Broadcasting,* and *Property Operators and Real Estate Services nfd* having statistically significant effects from agglomeration, which were previously insignificant in the two separate samples. The complete set of results is reported below in Table 4.2.4.

As for the industries for which pooled elasticity estimates could not be recorded because of detected differences in the parameters between the two cities, the causes or reasons cannot be determined conclusively. Possible reasons can either be estimation error or an actual existing difference in same-industry activity between the two cities. In the former case, when using a confidence level of 0.10 one would assume that in 10% of the cases the detection of a structural difference or the concluded statistical significance of a coefficient would be erroneous. Thus, one could attribute the incompatibility of the elasticity estimates between the two cities to be due to the chosen level of error to be tolerated in the tests. It could also be that some particular omitted variable is influencing the results in one or both of the cities, leading to a level of bias in the estimates. Alternatively, the scope of the activities within the industries may differ in each city such that firms benefit more from agglomeration in one city than in the other. A possible example of this may be the case where a greater composition of *computer system design* firms in Sydney, for instance, may service bigger businesses and more strategically oriented clients, whereas firms in Melbourne may concentrate more on non-business or smaller business services for which value-added is lower. If this was the case, the proximity to an internationally connected dense centre such as the CBD and the greater content of strategic knowledge involved in the activities of the firms in Sydney would result in greater benefits from agglomeration. The organizational structures between firms in the two cities may also differ and have some bearing on the exposure to agglomeration economies that firms may experience. Henderson (2003) for instance finds that in studying high tech and manufacturing firms, single-plant firms benefit from agglomeration economies to a much greater extent than corporate firms much less dependent on external environments.

Which of these potential reasons explains the significantly different elasticities estimated for the seven industries is uncertain. In fact, other industries may also be experiencing different returns from agglomeration between the two cities and they simply may not be "different enough" for a difference to be detected, given that the standard errors on the parameter estimates are in some cases relatively large. The importance ascribed to arriving upon a single industry elasticity estimate is dependent on the relative importance of having an estimate that can be generalized across multiple Australian cities. If city-specific elasticities can be estimated then they will likely better reflect internal conditions, especially if larger sample sizes can be acquired by using smaller spatial units. This matter will be addressed in Part 2's analyses; however, having pooled the SLA data does give us an idea of how comparable the industry-specific effects of agglomeration might be at this level of spatial analysis.

Table 4.2.4: Results for Sydney and Melbourne Using Pooled SLA Data (ED weighted by linear distance)

Ind ID	Industry	ε_U	S.E.	P-Value (F)	Adjusted R^2	Sample Size (N)
70	Oil and Gas Extraction	0.010	0.107	0.009823	0.472	32
80	Metal Ore Mining	*0.148	0.078	2.15E-29	0.988	44
101	Exploration	-0.093	0.123	0.637	-0.048	50
109	Other mining support services	**0.081	0.034	2.20E-12	0.963	30
135	Clothing and Footwear Product Manufacturing	***0.217	0.029	1.34E-46	0.890	135
184	Pharmaceutical and Medicinal Product Manufacturing	***0.090	0.027	5.76E-28	0.888	93
241	Professional and Scientific Equipment Manufacturing	*0.062	0.033	1.49E-21	0.712	123
242	Computer and Electronic Equipment Manufacturing	-0.013	0.039	1.96E-26	0.723	118
246	Specialized Machinery and Equipment Manufacturing	***0.149	0.037	1.11E-44	0.933	109
C00	Manufacturing, nfd	***0.068	0.018	2.70E-44	0.861	143
301	Residential Building Construction	0.039	0.028	3.03E-28	0.671	143
310	Heavy and Civil Engineering Construction	***0.097	0.017	3.80E-121	0.993	143
320	Construction Services, nfd	***0.142	0.049	5.57E-07	0.360	99
341	Specialized Industrial Material Wholesaling	0.131	0.027	1.19E-29	0.767	116
391	Motor Vehicle Retailing	*0.045	0.023	3.60E-45	0.892	131
411	Supermarket and Grocery Stores	***-0.095	0.016	1.20E-73	0.957	143
425	Clothing, Footwear and Personal Accessory Retailing	0.019	0.021	4.11E-56	0.879	143
427	Pharmaceutical and Other Store-Based Retailing	**0.033	0.016	1.08E-63	0.936	143
440	Accommodation	***0.072	0.020	3.42E-54	0.908	142
451	Cafes, Restaurants & Takeaway Food	0.008	0.033	1.43E-78	0.965	143
452	Pubs, Taverns & Bars	0.008	0.018	4.38E-25	0.633	142
461	Road Freight Transport	Sig. Diff. Coeff.				143
462	Road Passenger Transport	0.029	0.027	5.69E-23	0.678	140
521	Water Transport Support Services	0.121	0.072	2.70E-30	0.943	65
529	Other Transport Support	***0.139	0.045	7.68E-18	0.561	127

	Services					
530	Warehousing and Storage Services	0.029	0.046	8.51E-08	0.395	125
541	Newspaper, Periodical, Book, & Directory Publishing	***0.128	0.023	2.62E-47	0.891	136
551	Motion Picture & Video Activities	***0.149	0.050	1.02E-37	0.811	126
562	Television Broadcasting	***0.129	0.045	1.50E-20	0.777	83
580	Telecommunications Services	**0.055	0.027	5.80E-28	0.742	139
591	ISPs & Web Search Portals	Sig. Diff. Coeff.				87
592	Data Processing, Web Hosting, & Electronic Information Storage Services	**0.179	0.077	1.11E-10	0.533	88
620	Finance, nfd	***0.089	0.031	3.37E-76	0.949	137
622	Depository Financial	-0.005	0.027	3.07E-70	0.927	143
624	Financial Asset Investing	***0.268	0.035	4.29E-17	0.569	120
631	Life Insurance	0.073	0.064	3.39E-08	0.531	69
632	Health & General Insurance	0.062	0.026	7.41E-26	0.733	134
641	Auxiliary Finance & Investment Services	***0.078	0.028	4.64E-23	0.604	142
642	Auxiliary Insurance Services	*0.055	0.032	0.00E+00	1.000	115
670	Property Operators & Real Estate Services, nfd	***0.174	0.045	8.39E-09	0.380	114
671	Property Operators	***0.129	0.023	1.57E-37	0.780	137
672	Real Estate Services	0.037	0.026	3.63E-18	0.594	143
690	Prof, Sci & Tech Services, nfd	***0.172	0.032	1.79E-22	0.624	131
691	Scientific Research Services	0.047	0.034	3.63E-43	0.902	122
692	Arch, Eng & Tech Services	***0.105	0.038	1.02E-31	0.774	143
693	Legal & Accounting Services	**0.045	0.021	1.84E-72	0.933	143
694	Advertising Services	**0.104	0.044	8.45E-22	0.596	138
695	Market Research & Stat Services	**0.121	0.047	8.65E-45	0.821	142
696	Management & Consulting Services	***0.094	0.022	3.82E-31	0.704	143
699	Other Prof, Sci & Tech Services	***0.219	0.034	1.31E-19	0.566	136
700	Computer System Design	Sig. Diff. Coeff.				143
751	Central Government Administration	Sig. Diff. Coeff.				128
752	State Government Administration	Sig. Diff. Coeff.				139
753	Local Government Administration	***0.042	0.013	5.12E-74	0.960	140
754	Justice	0.017	0.128	1.00E-122	0.165	66

771	Public Order and Safety Services	0.009	0.013	7.80E-12	0.393	142
810	Tertiary Education	***0.054	0.011	2.29E-39	0.833	142
840	Hospitals	Sig. Diff. Coeff.				136
851	Medical Services	Sig. Diff. Coeff.				143
853	Allied Health Services	***0.097	0.012	1.06E-28	0.739	143

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

4.2.2 Analysis 1-2 Conclusions

This analysis was a first-cut at estimating industry-specific elasticities of productivity with respect to effective density in the Australian cities of Sydney and Melbourne. Using statistical local areas as the geographic unit of observation and a wage-function framework, the estimated industry elasticities are of comparable magnitude to prior work conducted by Graham (2005) who uses a production-function framework on industries in the U.K., and are aligned with urban growth and formation theories such as NEG and Central Place Theory. This analysis also included industries previously not assessed in an econometric framework, providing elasticity estimates for the mining, retail, and health sectors as well as several more refined industry classifications in the other more conventionally examined broader industry headings of finance; professional, scientific and technical; and media-related services. The findings generally suggest broad-sector elasticities[20] to be around 0.08 for manufacturing, 0.11 for construction, near zero for retail, 0.05 for transport, 0.17 for media, 0.12 for finance, 0.09 for real estate services, 0.12 for professional services, 0.07 for government, and 0.08 for medical services.

The estimation of agglomeration impacts with respect to travel-time weighted effective density was conducted only for Sydney and verified the findings of Graham (2007), who finds the use of generalized cost of travel in the effective density index to increase the estimated returns from agglomeration. The pooling of Sydney and Melbourne data reveals that structural differences exist in half of the industries analysed; however, in most cases the differences are not attributable to divergent impacts from agglomeration. One can only speculate what the precise causes of the structural differences between the

[20] Calculated as broad-industry averages from combined Sydney and Melbourne results.

estimates for the two cities are, though sampling error sits as a viable explanation as well as differences in within-industry specialization. The relatively few instances where the effective density variable drives the structural difference imply that agglomeration externalities on an industry-specific basis will in most cases benefit firms equally in both cities. This, however, awaits validation in Part 2, where employment zone data will enable more efficient parameter estimation, which in turn will give greater contrast to city estimates if they are likely to exist. Part 2 analyses will improve on this section's procedures by utilizing larger samples, being more flexible in its model specifications and addressing the endogeneity issue. Before moving onto this, Analysis 1-3 will maintain the use of the SLA dataset to test the effects of adding a control for local-industry concentration.

4.3.0 Analysis 1-3 Overview: Estimating localization and urbanization effects using SLA data

In this analysis, a progression from Analysis 1-2 is made to control for the effects of localized industry concentration in addition to the urbanization effects captured by the effective density measure. The presence of localization economies, or in other words the externalities that arise out of the co-location of like activities, is captured here by a measure not trialled in the existing literature – namely the employment concentration factor (ECF) or location quotient (LQ).

The ECF, as one might recall from section 2.9, is a measure of relative industry concentration that uses the greater regional economy or national economy as a unit of reference. It is estimated by the following formula,

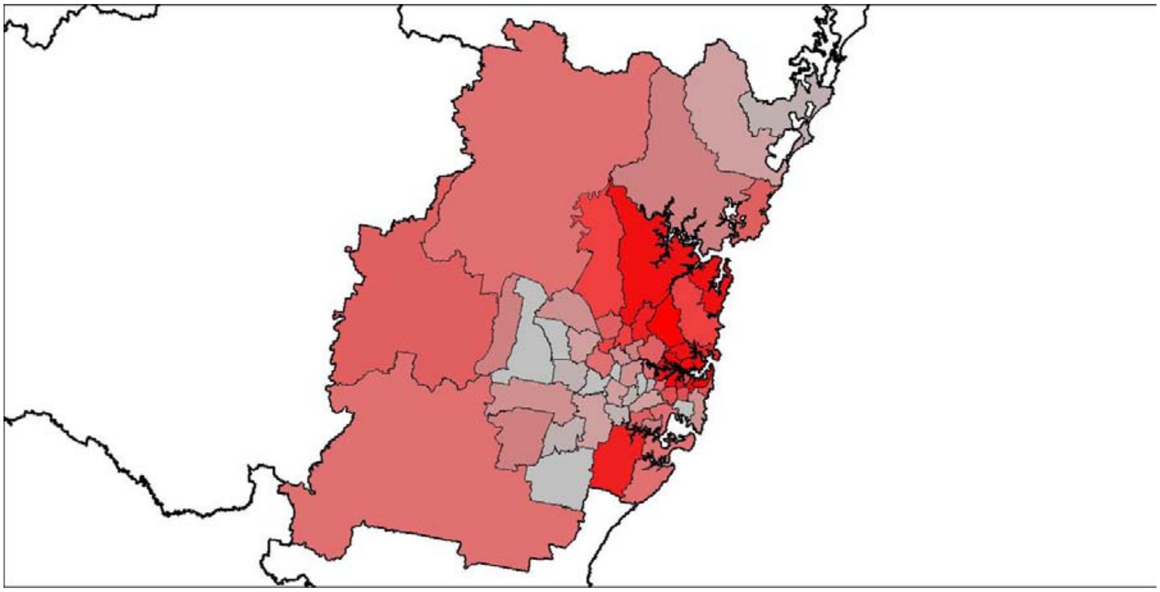
$$ECF = \frac{E_{ri}}{E_r} \Big/ \frac{E_{Ri}}{E_R} \qquad [4I]$$

where the employment level of industry i in subregion r as a share of the subregion's total employment is weighted by the employment share in the same industry of the greater regional or national economy. The estimating equation then becomes as indicated below.

$$\ln I_{k,m} = \alpha_k + \beta \ln U_m + \sum^{i} \delta_i \ln Occ_{i,k,m} + \sum^{j} \gamma_j \ln Edu_{j,k,m} + \psi Exp_{k,m} + \phi ECF_{k,m} + e_{k,m} \qquad [4J]$$

The implication of using the EFC is slightly different from that of the using employment size. Its inclusion in a wage function is more accurately expressed as measuring the effect of the relative magnitude of employment in a particular industry, rather than mere employment level. As such, industry concentration will still go up with increased employment in a given industry and location but it will be the attractiveness of that particular location compared to others in the region that says something about that location's productivity advantages. For instance, knowing that a geographic unit has 100 people employed in a given industry might mean little unless one knew how this compares to other locations. The comparison can be made by observing one value (the ECF) rather than a whole range of values for which some descriptive statistics would have to be provided to give some objectivity to an interpretation. Figure 4.3.1 gives a thematic map of Statistical Local Area ECF values across the Sydney Statistical Division for industry *692 - Architectural, Engineering and Technical Services*. As one can see by comparing the map to Figure 4.2.1 that displays SLA effective density in Sydney, geographic concentrations of industry 692 differ somewhat from effective density patterns. Identifying this gives justification for simultaneously incorporating measures for the both types of agglomeration economies as there may be spatial considerations other than overall employment concentration that give rise to labour productivity increases.

Figure 4.3.1: A Thematic Map of ECF values for Industry 692 in Sydney



The downside of using an ECF in the context of econometrically estimating industry wage functions is that the interpretation of the elasticity is a little more challenging and cannot be readily compared to other studies previously controlling for localized industry concentration. It should also be noted that while the ECF says much about the attractiveness of a given area at a given point in time, its application to a time-series analysis becomes unstable as it relies on regional or nation-wide employment shares in a given industry remaining constant. As such, ECF values for a given area can differ across time periods simply because of changes in other-industry employment numbers while holding own-industry employment constant. In the context of using this measure to estimate the impact of localized industry concentration on industry productivity, this would mean that location-specific productivity changes could be shown to occur without a change in a location's employment levels and merely as a result of a shift in the employment mix of the greater region. This, however, is not of concern here where the data are purely cross-sectional, thus involving data sourced from only one period in time. While this analysis trials this measure as a control for industry localization, Part 2 of the analysis will address localization economies in a more conventional fashion by using employment bands.

4.3.1 Analysis 1-3 Results

The effect of adding an ECF to the Sydney industry regressions improved the adjusted R-squared values, at least marginally, in 36 of the 58 industries. In 15 of these cases, R-squared values improved by more than 0.10, suggesting that localization in these industries adds substantial explanatory power to wage disparities across the city. These industries are indicated in Table 4.3.1 where their adjusted R-Squared values are reported in bold. The most prominent sectors to experience these effects are retail, finance, professional services and government administration. Additionally, while the number of industries to have significant effective density coefficients estimated remained the same at 27, the model specification with the ECF produced wage-function estimations in which 32 industries experienced significant effects from industry localization.

Perhaps the most interesting outcome of including the ECF in the regressions was that many industries that previously reported no effects (or very insignificant effects) from urbanization economies now show rather strong influences from localization. One industry to exemplify this is *Public Order and Safety Services*, which continues to show a weak effect from urbanization but reports a strongly significant positive productivity elasticity with respect to the ECF parameter of 0.085. Other industry results to respond in a similar fashion are *Computer and Electronic Equipment Manufacturing*, *Motor Vehicle Retailing*, *Warehousing and Storage Services*, *Depository Financial*, *Health and General Insurance*, *Real Estate Services*, *Legal and Accounting Services*, and *State Government Administration*. In the case of these industries, the results suggest that localization effects do much better to explain the spatial variation of labour productivity than urbanization economies. For an industry such as *Motor Vehicle Retailing*, this rationalizes well as car dealers often cluster in places along transport corridors or in commercial developments such as auto-malls that are not necessarily in the densest locations. Even bank branches can be observed to cluster together in a wide range of location types across greater city metropolitan regions, not just in the densest areas. A manufacturer of technical products would also likely operate in a cluster away from denser areas if its operations were fairly standardized. Without a more detailed understanding of the operations of these industries, however, it is difficult to validate the motivations for location decisions against the results of this analysis beyond giving speculative explanations.

Also of particular interest in the results are instances where industries now seem to show a sharing of the benefit to productivity from a combination of urbanization and localization effects. The simultaneity embodied in this analysis by estimating both forms of agglomeration economies together could in some cases lead to an econometric issue because, in circumstances where two variables are highly correlated, regression analysis can potentially produce inaccurate results by essentially not knowing to which variable to attribute the explanatory power. The result could be that both variables might come out insignificant, or that one dominates the other by taking on a larger coefficient at the other's expense. The effect of this, however, would be fairly obvious in a counterintuitive result, such as if a significant agglomeration coefficient was estimated in Analysis 1-2 but not in Analysis 1-3 for either variable, or if one parameter receives a questionable estimate because of its magnitude. Without being able to say with absolute certainty, this does not seem to be the case for any of the results reported in Table 4.3.1. As one can see in the table, 15 industries are estimated with significant influences from both urbanization and localization controls.

By controlling for both types of agglomeration effects, the largest significant coefficients on effective density are reported for *Data Processing, Web Hosting and Electronic Information Storage* (0.288), *Other Professional, Scientific and Technical Services* (0.224), *Specialized Industrial Material Wholesaling* (0.209), *Clothing and Footwear Product Manufacturing* (0.188), and *Financial Asset Investing* (0.150). These same industries appeared as reporting the top five estimates for the model specified in Analysis 1-2, save for *Clothing and Footwear Manufacturing* that replaced *Television Broadcasting* because the latter's coefficient became insignificant. Of these industries, *Financial Asset Investing* was the only one to also display a significant coefficient on the ECF (0.471), which was accompanied by the largest reduction of the effective density coefficient (down by 0.087). Also showing highly significant simultaneous estimates of urbanization and localization parameters are the industries of *Finance, nfd* (0.122 and 0.189), *Management and Consulting Services* (0.070 and 0.100), *Computer System Design* (0.079 and 0.207), *State Government Administration* (0.068 and 0.136) and *Road Freight Transport* (0.032 and 0.074), where the elasticities reported in the brackets refer to the ED and ECF parameters respectively.

Lastly, it is of interest to note that the only sectors to remain unaffected by the inclusion of the ECF are the mining industries and medical services. In the former case of the mining sector, all estimates remained insignificant while in the latter case of medical services, urbanization economies maintain their dominance over explaining spatial wage disparities.

Table 4.3.1: Sydney Results Using SLA Data (ED weighted by linear distance and ECF)

Ind ID	Industry Name	ε_U	S.E.	ε_{ECF}	S.E.	P-Value (F)	Adjusted R^2	N
70	Oil and Gas Extraction	N/A	N/A	N/A	N/A	N/A	N/A	N/A
80	Metal Ore Mining	0.228	0.444	-0.338	0.479	0.242473	0.251	21
101	Exploration	0.105	0.173	0.471	0.367	0.313414	0.189	23
109	Other mining support services	N/A	N/A	N/A	N/A	N/A	N/A	N/A
135	Clothing and Footwear Product Manufacturing	***0.188	0.069	-0.076	0.074	5.74E-06	0.490	62
184	Pharmaceutical and Medicinal Product Manufacturing	*0.086	0.051	**0.063	0.030	4.62E-08	0.682	51
241	Professional and Scientific Equipment Manufacturing	0.019	0.033	**0.067	0.031	5.45E-13	0.777	59
242	Computer and Electronic Equipment Manufacturing	0.042	0.035	***0.116	0.023	6.87E-18	0.878	57
246	Specialized Machinery and Equipment Manufacturing	0.017	0.053	-0.081	0.057	5.27E-07	0.636	49
C00	Manufacturing, nfd	**0.068	0.025	0.073	0.044	1.40E-12	0.732	64
301	Residential Building Construction	0.063	0.050	0.092	0.066	6.65E-09	0.615	64
310	Heavy and Civil Engineering Construction	**0.070	0.027	***0.113	0.035	3.68E-25	0.919	64
320	Construction Services, nfd	0.061	0.087	-0.112	0.161	0.563618	-0.025	57
341	Specialized Industrial Material Wholesaling	***0.209	0.019	-0.066	0.051	1.72E-24	0.944	56
391	Motor Vehicle Retailing	-0.012	0.032	***0.180	0.044	5.84E-26	0.939	60
411	Supermarket and Grocery Stores	***-0.081	0.022	0.011	0.030	1.49E-39	**0.979**	64
425	Clothing, Footwear and Personal Accessory Retailing	**0.044	0.018	***0.085	0.020	2.21E-37	0.974	64
427	Pharmaceutical and Other Store-Based Retailing	-0.020	0.017	***-0.108	0.037	7.89E-37	**0.973**	64
440	Accommodation	***0.078	0.027	**0.068	0.032	5.44E-25	0.922	63
451	Cafes, Restaurants &	0.003	0.035	0.036	0.073	2.31E-24	0.913	64

	Takeaway Food							
452	Pubs, Taverns & Bars	0.013	0.032	0.069	0.047	2.97E-08	0.588	64
461	Road Freight Transport	***0.032	0.011	***0.074	0.015	2.05E-23	**0.905**	64
462	Road Passenger Transport	0.021	0.062	-0.025	0.066	0.033975	0.183	63
521	Water Transport Support Services	-0.128	0.082	0.053	0.091	5.64E-22	0.993	35
529	Other Transport Support Services	*0.092	0.047	0.021	0.031	1.83E-08	0.626	60
530	Warehousing and Storage Services	-0.073	0.046	***0.183	0.036	1.57E-12	**0.751**	61
541	Newspaper, Periodical, Book, & Directory Publishing	***0.117	0.039	***0.348	0.051	4.66E-17	**0.832**	63
551	Motion Picture & Video Activities	0.043	0.041	***0.348	0.061	9.44E-29	0.954	60
562	Television Broadcasting	0.183	0.113	0.052	0.099	0.002901	0.418	44
580	Telecommunications Services	0.030	0.024	***0.155	0.037	2.65E-14	0.774	64
591	ISPs & Web Search Portals	0.008	0.066	**0.206	0.079	2.09E-22	0.980	42
592	Data Processing, Web Hosting, & Electronic Information Storage Services	***0.288	0.053	0.015	0.035	4.08E-12	0.878	43
620	Finance, nfd	***0.122	0.026	***0.189	0.059	2.17E-13	**0.753**	64
622	Depository Financial	-0.021	0.043	***0.110	0.039	2.85E-11	0.696	64
624	Financial Asset Investing	***0.150	0.050	***0.471	0.084	7.73E-16	**0.860**	55
631	Life Insurance	0.050	0.076	0.444	0.327	2.92E-09	0.864	36
632	Health & General Insurance	0.016	0.027	***0.150	0.045	4.78E-28	**0.941**	63
641	Auxiliary Finance & Investment Services	0.051	0.027	0.091	0.040	3.59E-27	**0.933**	64
642	Auxiliary Insurance Services	*0.069	0.037	*0.188	0.110	1.26E-16	0.864	55
670	Property Operators & Real Estate Services, nfd	0.108	0.117	0.430	0.327	0.543298	-0.020	57
671	Property Operators	0.028	0.022	*0.127	0.070	3.37E-10	0.669	63
672	Real Estate Services	0.024	0.017	***0.275	0.043	6.08E-35	0.967	64
690	Prof, Sci & Tech Services, nfd	*0.098	0.054	***0.522	0.162	7.69E-11	0.683	64
691	Scientific Research Services	0.029	0.052	0.036	0.034	8.69E-31	0.970	57
692	Arch, Eng & Tech Services	***0.071	0.021	***0.154	0.024	1.83E-33	**0.963**	64
693	Legal & Accounting Services	0.019	0.031	***0.277	0.046	1.65E-11	**0.703**	64
694	Advertising Services	*0.070	0.037	***0.259	0.071	3.08E-27	**0.933**	64
695	Market Research & Stat Services	0.032	0.040	*0.065	0.037	8.83E-32	0.959	63
696	Management &	***0.070	0.020	***0.100	0.031	5.51E-43	**0.984**	64

	Consulting Services							
699	Other Prof, Sci & Tech Services	***0.224	0.073	0.011	0.126	0.020073	0.206	64
700	Computer System Design	***0.079	0.021	***0.207	0.030	1.18E-24	0.915	64
751	Central Government Administration	0.037	0.027	**0.078	0.032	1.90E-23	**0.922**	60
752	State Government Administration	***0.068	0.015	***0.136	0.136	1.14E-25	**0.923**	64
753	Local Government Administration	***0.046	0.017	-0.002	0.030	1.18E-30	0.954	63
754	Justice	0.027	0.045	-0.075	0.121	1.71E-09	0.839	39
771	Public Order and Safety Services	-0.010	0.015	***0.085	0.024	1.61E-15	0.799	64
810	Tertiary Education	***0.042	0.013	**0.033	0.016	8.79E-39	**0.977**	64
840	Hospitals	***0.065	0.021	0.032	0.026	7.56E-06	0.490	61
851	Medical Services	***0.055	0.015	0.011	0.031	1.66E-10	0.672	64
853	Allied Health Services	***0.074	0.019	-0.001	0.031	1.54E-07	0.557	64

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

In the Melbourne case, the effect of incorporating the ECF into the model specification was similar to that in Sydney. The amount of spatial variation in earnings explained, as indicated by the adjusted R-squared value, increased for 32 of the 60 industries and increased by more than 0.10 in 12 of them. These results are given below in Table 4.3.2 where, once again, R-squared values that experienced an increase by more than 0.10 are reported in bold. These marked improvements in explanatory power were shared with Sydney in the industries of *Computer and Electronic Equipment Manufacturing*, *Road Freight Transport*, *Financial Asset Investing*, *Auxiliary Finance and Investment Services*, *Legal and Accounting Services*, *Advertising Services*, *State Government Administration* and *Tertiary Education*. This would lead one to believe that the contribution of the ECF to predicting income levels is not driven by a spurious relationship but rather by an actual causal effect, as a number of other industry wage function estimates are mutually improved in both cities by its inclusion as well.

The number of significant urbanization parameters falls in this specification from 35 to 29 while the localization parameter is significant for 33 industries. In some of these cases, the shift of significance from the urbanization to localization parameter might be an indication of multicollinearity. In *Management and Consulting Services*, for instance, the significance of the effective density coefficient is lost by the inclusion of the ECF which itself becomes very significant, while only a very minor change in the adjusted R-squared value occurs. In Sydney, however, both parameters maintain their significance when estimated simultaneously in this industry. It remains unclear in this circumstance which form of agglomeration benefit appears to be more important. *ISPs and Web Search Portals*, on the other hand, is significant with respect to the urbanization parameter in Analysis 1-2 for both cities, yet in Analysis 1-3 both cities experience a transfer of this significance to the localization measure coefficient. As this shift of significance is also accompanied by a relatively strong improvement in the adjusted R-squared values, one might be more inclined to conclude that localization economies have greater importance in this industry.

After controlling for localization, the largest urbanization effects were estimated for Melbourne for *Construction Services, nfd* (0.355); *Life Insurance* (0.217); *Property Operators* (0.188); *Clothing and Footwear Product Manufacturing* (0.174); and *Computer System Design* (0.171). Three industries remain in this list from those reported for Melbourne in Analysis 1-2 while those that dropped out experienced a shift in the influence on productivity from effective density to the ECF. *Computer System and Design* was the only industry to remain in the top five while still reporting a significant ECF parameter. Many of the media-related; finance; and professional, scientific and technical service industries reported significant ECF coefficients that took away some of the influence from effective density. The table below gives a complete summary of these results for Melbourne.

Table 4.3.2: Melbourne Results Using SLA Data (ED weighted by linear distance and ECF)

Ind ID	Industry Name	ε_U	S.E.	ε_{ECF}	S.E.	P-Value (F)	Adjusted R^2	N
70	Oil and Gas Extraction	0.043	0.265	0.081	0.317	0.182128	0.434	20
80	Metal Ore Mining	*0.274	0.150	-0.001	0.110	0.025714	0.586	23
101	Exploration	-0.046	0.259	0.054	0.298	0.422701	0.040	27
109	Other mining support services	-0.266	0.370	*-0.750	0.344	0.135544	0.554	18
135	Clothing and Footwear Product Manufacturing	***0.174	0.033	0.036	0.042	1.39E-25	0.887	73
184	Pharmaceutical and Medicinal Product Manufacturing	*0.110	0.058	***0.068	0.050	7.19E-07	0.717	42
241	Professional and Scientific Equipment Manufacturing	0.061	0.075	-0.018	0.074	3.30E-05	0.432	64
242	Computer and Electronic	**-0.093	0.043	**0.103	0.040	5.98E-11	**0.707**	61

	Equipment Manufacturing							
246	Specialized Machinery and Equipment Manufacturing	0.063	0.073	***0.154	0.051	4.40E-06	0.510	60
C00	Manufacturing, nfd	**0.096	0.041	**0.099	0.044	2.12E-09	0.544	79
301	Residential Building Construction	***0.081	0.026	0.019	0.034	1.60E-27	0.881	79
310	Heavy and Civil Engineering Construction	***0.130	0.024	*0.065	0.036	9.49E-34	0.924	79
320	Construction Services, nfd	***0.355	0.102	0.178	0.129	2.91E-11	0.857	42
341	Specialized Industrial Material Wholesaling	0.074	0.083	0.059	0.049	0.001786	0.326	60
391	Motor Vehicle Retailing	***0.065	0.024	***0.118	0.029	9.52E-19	0.815	71
411	Supermarket and Grocery Stores	***-0.081	0.028	0.053	0.037	5.91E-29	0.893	79
425	Clothing, Footwear and Personal Accessory Retailing	0.004	0.025	**0.072	0.036	1.68E-29	0.897	79
427	Pharmaceutical and Other Store-Based Retailing	0.043	0.028	0.069	0.053	6.48E-29	0.893	79
440	Accommodation	**0.088	0.038	-0.024	0.047	3.06E-10	0.573	79
451	Cafes, Restaurants & Takeaway Food	0.010	0.041	*0.084	0.044	2.19E-14	0.689	79
452	Pubs, Taverns & Bars	0.014	0.023	*-0.105	0.056	2.37E-11	**0.614**	78
461	Road Freight Transport	***0.113	0.019	***0.131	0.017	1.95E-23	**0.840**	79
462	Road Passenger Transport	0.053	0.039	0.044	0.036	3.13E-19	0.793	77
521	Water Transport Support Services	0.026	0.043	***-0.170	0.057	3.08E-27	0.999	30
529	Other Transport Support Services	**0.114	0.056	***0.159	0.037	8.63E-19	0.837	67
530	Warehousing and Storage Services	0.110	0.066	0.075	0.075	0.000843	0.333	64
541	Newspaper, Periodical, Book, & Directory Publishing	***0.144	0.035	0.073	0.049	8.09E-21	0.835	73
551	Motion Picture & Video Activities	0.083	0.090	0.176	0.113	4.77E-11	0.676	66
562	Television Broadcasting	-0.031	0.061	***0.104	0.028	6.87E-29	0.996	39
580	Telecommunications Services	0.025	0.042	**0.089	0.036	1.54E-18	**0.791**	75
591	ISPs & Web Search Portals	0.103	0.101	**0.248	0.114	5.63E-29	0.987	45
592	Data Processing, Web Hosting, & Electronic Information Storage Services	***0.570	0.178	*0.410	0.211	0.130354	0.148	45
620	Finance, nfd	-0.095	0.082	***0.311	0.083	1.93E-15	0.745	73

622	Depository Financial	-0.030	0.050	0.090	0.075	2.21E-12	0.638	79
624	Financial Asset Investing	***0.148	0.036	***0.209	0.052	4.94E-15	**0.783**	65
631	Life Insurance	***0.217	0.054	0.072	0.076	7.47E-10	**0.910**	33
632	Health & General Insurance	0.005	0.029	***0.083	0.028	3.53E-27	0.908	71
641	Auxiliary Finance & Investment Services	*0.075	0.041	***0.227	0.040	2.05E-20	**0.806**	78
642	Auxiliary Insurance Services	0.040	0.075	0.066	0.097	0.003634	0.298	60
670	Property Operators & Real Estate Services, nfd	0.088	0.084	0.108	0.095	2.38E-12	0.764	57
671	Property Operators	***0.188	0.041	0.046	0.078	1.07E-17	0.782	74
672	Real Estate Services	0.063	0.038	**0.143	0.056	1.10E-09	0.554	79
690	Prof, Sci & Tech Services, nfd	***0.171	0.032	***-0.220	0.069	1.88E-54	0.992	67
691	Scientific Research Services	-0.048	0.057	***0.170	0.042	3.46E-11	**0.687**	65
692	Arch, Eng & Tech Services	0.027	0.041	***0.208	0.054	9.57E-15	0.697	79
693	Legal & Accounting Services	0.029	0.028	***0.191	0.029	1.09E-34	**0.929**	79
694	Advertising Services	0.085	0.060	***0.194	0.043	8.12E-29	**0.909**	74
695	Market Research & Stat Services	***0.153	0.054	0.032	0.063	7.74E-31	0.906	79
696	Management & Consulting Services	0.061	0.037	***0.175	0.049	6.38E-12	0.625	79
699	Other Prof, Sci & Tech Services	***0.167	0.048	0.100	0.069	5.08E-17	0.781	72
700	Computer System Design	***0.171	0.030	***0.227	0.039	7.27E-28	0.884	79
751	Central Government Administration	0.020	0.017	0.019	0.020	4.92E-23	0.884	68
752	State Government Administration	***0.123	0.029	*0.047	0.026	4.00E-33	0.932	75
753	Local Government Administration	***0.093	0.027	***0.119	0.032	5.96E-26	0.875	77
754	Justice	0.033	0.469	0.482	0.487	1.74E-69	0.055	27
771	Public Order and Safety Services	0.039	0.025	-0.028	0.034	1.42E-06	0.432	78
810	Tertiary Education	**0.046	0.018	***0.099	0.022	2.83E-11	0.611	78
840	Hospitals	***0.098	0.020	**0.041	0.020	8.07E-38	0.953	75
851	Medical Services	***0.082	0.020	0.020	0.037	3.40E-05	0.357	79
853	Allied Health Services	***0.105	0.017	0.025	0.030	2.59E-14	0.687	79

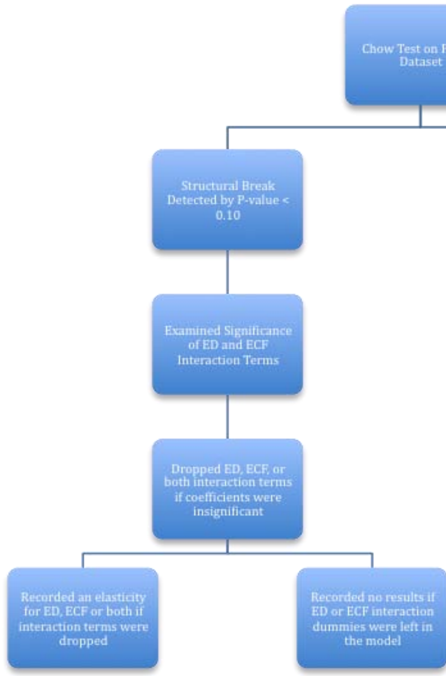
Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

Sydney and Melbourne Pooled Dataset Results

In order to determine whether the Sydney and Melbourne parameter estimates for the effective density and ECF variables are similar enough to deem their differences statistically insignificant, an approach similar to the pooled regressions in Analysis 1-2 was applied. Datasets were pooled, a city dummy variable was set to take on a value of zero for Melbourne and one for Sydney, and then interaction dummies were set by multiplying the city dummy by each control variable in the original model. Next, a Chow test was carried out by comparing the unrestricted to restricted models to test whether a structural break was likely to exist between the two city datasets. If a structural break was not detected at a significance level of 0.10 then the results of the restricted model were recorded, otherwise further investigation into the cause of the detected difference was carried out. This was achieved by examining the significance of the interaction terms.

If the parameter estimates on the ED and the ECF interaction dummies were insignificant at the 0.10 level then they were omitted and reduced models were estimated, concluding that their estimated differences were not reliable and significantly different enough to make it necessary for two separate coefficients to be maintained. If only one of the two coefficients on the interaction dummies was insignificant, the significant one was kept in the model while the other was omitted and a final model was estimated. Finally, if both interaction dummies were significant then the unrestricted model was maintained and no pooled parameter estimates were recorded. The flow diagram in Figure 4.6 illustrated below summarizes these steps. Robust standard errors were used in carrying out the Chow test if heteroskedasticity was present and GLS was used for the final model estimations if sample sizes were large enough to allow it.

Figure 4.3.2: Summary of steps in determining pooled city agglomeration elasticity estimates



The analysis produced statistically significant effective density parameter estimates for 30 industries and statistically significant results for the employment concentration factor for 33 industries. In 16 industries the parameter estimates were both significant. The largest significant ED estimates were for the industries *Other Professional, Scientific and Technical Services* (0.236), *Financial Asset Investing* (0.204), *Clothing and Footwear Product Manufacturing* (0.191), *Property Operators and Real Estate Services, nfd* (0.188), and *Market Research and Statistical Services* (0.174). The largest significant ECF parameter estimates were for the industries *Financial Asset Investing* (0.205), *Architectural, Engineering and Technical Services* (0.185), *Computer System Design* (0.176), *Advertising Services* (0.168) and *Finance, nfd* (0.160).

Despite the larger total sample size in this analysis compared to those for Sydney and Melbourne individually, the number of significant parameters for ED and ECF was roughly the same. With the pooled sample, one would have expected greater precision in the estimates as indicated by smaller standard errors. The likely reason that we did not find that the number of significant parameters increased is that the analysis generated 10 significant interaction dummies for each of ED and the ECF. This means that there are potentially 10 industries that could have significant ED and ECF coefficient estimates that simply were not reported because they differed between Sydney and Melbourne. *Data Processing, Web Hosting and Electronic Information Storage* is one such industry where in the pooled analysis no coefficients were reported because of significant interaction dummies, yet in the individual city analyses the effects of agglomeration were reported to be rather strong. It is likely that more industries have incompatible parameter estimates here in Analysis 1-3 compared to Analysis 1-2 because of potential multicollinearity issues generated by including controls for both industry urbanization and localization. While in this analysis several new industries reported at least one of the agglomeration controls to be statistically different between the two cities, there were also a few industries for which the inclusion of the ECF meant that the pooling of the city data resulted in pooled city estimates being able to be reported. These include the industries of *State Government Administration*, *Hospitals*, and *Medical Services*.

Once again, the pooling of the city data simply gives an opportunity to discover whether differences in agglomeration effects exist between Sydney and Melbourne and if not, then hopefully will provide us with more precise estimates because of the larger sample size. Where industry-specific effects from agglomeration are shown to differ between the two cities, it is difficult to judge whether this is truly the case or whether it is a result of statistical error that would not hold in repeated sampling. The addition of the ECF parameter could have created a multicollinearity issue in a number of industries that was affecting the estimates in Sydney and Melbourne differently and as such more differences between the two cities could have been detected than truly exist. This said, in the cities for which pooling allowed for shared ED and ECF parameters to be estimated we expect more efficiently estimated elasticities, while from the industry results that suggested pooling was not suitable we get some sense of differences that may exist between Sydney and Melbourne.

Table 4.3.3: Results for Sydney and Melbourne Using Pooled SLA Data (ED weighted by linear distance and ECF)

Ind ID	Industry Name	ε_U	S.E.	ε_{ECF}	S.E.	P-Value (F)	Adjusted R^2	N
70	Oil and Gas Extraction	0.048	0.108	0.142	0.102	9.38E-03	0.497	32
80	Metal Ore Mining	0.083	0.073	0.069	0.164	1.59E-31	0.992	44
101	Exploration	-0.075	0.125	0.330	0.363	0.647025	-0.053	50

109	Other mining support services+	Sig. Diff. Coeff.		Sig. Diff. Coeff.				30
135	Clothing and Footwear Product Manufacturing	***0.191	0.047	Sig. Diff. Coeff.		3.13E-18	0.403	135
184	Pharmaceutical and Medicinal Product Manufacturing	0.061	0.037	***0.057	0.015	5.44E-16	0.657	93
241	Professional and Scientific Equipment Manufacturing	*0.058	0.033	0.038	0.030	8.59E-22	0.721	123
242	Computer and Electronic Equipment Manufacturing	-0.031	0.038	**0.061	0.027	3.94E-55	0.926	118
246	Specialized Machinery and Equipment Manufacturing	***0.106	0.033	Sig. Diff. Coeff.		1.69E-32	0.873	109
C00	Manufacturing, nfd	***0.077	0.018	***0.092	0.024	3.22E-46	0.874	143
301	Residential Building Construction	*0.052	0.029	0.069	0.043	4.65E-28	0.674	143
310	Heavy and Civil Engineering Construction	Sig. Diff. Coeff.		***0.090	0.029	2.95E-61	0.933	143
320	Construction Services, nfd	0.198	0.062	0.192	0.123	2.77E-07	0.380	99
341	Specialized Industrial Material Wholesaling	***0.136	0.022	*0.050	0.027	2.70E-63	0.951	116
391	Motor Vehicle Retailing	***0.046	0.017	***0.123	0.023	1.01E-51	0.921	131
411	Supermarket and Grocery Stores	***-0.066	0.019	0.036	0.024	6.50E-86	0.974	143
425	Clothing, Footwear and Personal Accessory Retailing	0.007	0.020	***0.079	0.028	6.78E-50	0.853	143
427	Pharmaceutical and Other Store-Based Retailing	Sig. Diff. Coeff.		Sig. Diff. Coeff.		3.88e-46	0.723	143
440	Accommodation	***0.064	0.018	0.036	0.024	3.61E-53	0.907	142
451	Cafes, Restaurants & Takeaway Food	-0.019	0.020	Sig. Diff. Coeff.		8.03E-57	0.920	143
452	Pubs, Taverns & Bars	0.013	0.018	0.005	0.025	2.26E-26	0.656	142
461	Road Freight Transport	Sig. Diff. Coeff.		Sig. Diff. Coeff.		1.78E-47	0.886	143
462	Road Passenger Transport	0.036	0.026	**0.069	0.030	1.13E-27	0.741	140
521	Water Transport Support Services	0.090	0.057	0.023	0.066	3.86E-32	0.955	65
529	Other Transport Support Services	**0.095	0.041	Sig. Diff. Coeff.		1.98E-28	0.792	127
530	Warehousing and Storage Services	Sig. Diff. Coeff.		***0.099	0.027	4.66E-41	0.878	125
541	Newspaper, Periodical, Book, & Directory Publishing	***0.114	0.021	***0.175	0.027	1.64E-37	0.788	136

551	Motion Picture & Video Activities	**0.115	0.046	***0.180	0.051	2.02E-42	0.849	126
562	Television Broadcasting	Sig. Diff. Coeff.		**0.077	0.030	1.01E-99	0.998	83
580	Telecommunications Services	0.008	0.021	***0.107	0.026	6.16E-26	0.724	139
591	ISPs & Web Search Portals	0.050	0.044	**0.100	0.042	4.88E-77	0.994	87
592	Data Processing, Web Hosting, & Electronic Information Storage Services	Sig. Diff. Coeff.		Sig. Diff. Coeff.		5.02E-18	0.810	88
620	Finance, nfd	**0.059	0.029	***0.160	0.037	2.23E-82	0.961	137
622	Depository Financial	-0.024	0.023	***0.074	0.019	7.47E-52	0.863	143
624	Financial Asset Investing	***0.204	0.033	***0.205	0.040	7.19E-27	0.793	120
631	Life Insurance	0.034	0.042	***0.141	0.043	2.70E-18	0.811	69
632	Health & General Insurance	0.011	0.022	***0.105	0.023	5.39E-73	0.966	134
641	Auxiliary Finance & Investment Services	0.032	0.026	***0.154	0.025	1.19E-42	0.812	142
642	Auxiliary Insurance Services	***0.079	0.030	***0.078	0.033	3.16E-27	0.799	115
670	Property Operators & Real Estate Services, nfd	***0.188	0.054	0.056	0.082	3.46E-07	0.333	114
671	Property Operators	Sig. Diff. Coeff.		**0.079	0.038	4.41E-27	0.748	137
672	Real Estate Services	***0.039	0.013	***0.140	0.019	1.74E-60	0.927	143
690	Prof, Sci & Tech Services, nfd	***0.146	0.035	Sig. Diff. Coeff.		4.86E-63	0.953	131
691	Scientific Research Services	-0.039	0.038	***0.117	0.020	5.28E-55	0.947	122
692	Arch, Eng & Tech Services	***0.054	0.017	***0.185	0.025	1.42E-45	0.871	143
693	Legal & Accounting Services	0.012	0.011	***0.151	0.018	1.51E-57	0.889	143
694	Advertising Services	**0.071	0.033	***0.168	0.036	2.26E-33	0.746	138
695	Market Research & Stat Services	***0.174	0.034	***0.081	0.030	1.34E-58	0.925	142
696	Management & Consulting Services	***0.072	0.023	***0.123	0.026	1.12E-50	0.857	143
699	Other Prof, Sci & Tech Services	***0.236	0.036	-0.002	0.043	8.31E-21	0.593	136
700	Computer System Design	Sig. Diff. Coeff.		***0.176	0.016	1.58E-55	0.916	143
751	Central Government Administration	*0.028	0.014	Sig. Diff. Coeff.		1.36E-50	0.925	128
752	State Government Administration	***0.062	0.012	***0.089	0.021	2.44E-52	0.909	139
753	Local Government Administration	***0.069	0.015	***0.072	0.024	1.67E-52	0.908	140
754	Justice	0.037	0.130	-0.057	0.136	N/A	0.188	66

771	Public Order and Safety Services	Sig. Diff. Coeff.		Sig. Diff. Coeff.		4.76E-25	0.714	142
810	Tertiary Education	***0.041	0.011	Sig. Diff. Coeff.		1.33E-71	0.957	142
840	Hospitals	***0.077	0.011	***0.047	0.014	4.80E-138	0.998	136
851	Medical Services	***0.066	0.007	0.006	0.014	4.04E-32	0.777	143
853	Allied Health Services	***0.093	0.011	0.001	0.018	3.49E-29	0.749	143

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

4.3.2 Analysis 1-3 Conclusions

The debate over whether urbanization or local-industry concentration is more important to raising productivity levels is one well established in the literature on agglomeration economies. The econometric approach used in this analysis produced similarly conflicting results as experienced by other studies on the subject. Without directly observing the individual sources of agglomeration benefit, it is rather difficult to come to a clear understanding of which of the two forms of agglomeration economies are more important. The confounding effects of multicollinearity exacerbate this issue because local industry location can in many cases mimic overall industry location patterns and as such, OLS may not have enough information to identify their separate effects. It does not mean OLS fails, for if this is a consistent industry phenomenon with out-of-sample observations then predictions using the overall wage equation should not be affected – it is merely the ED and ECF coefficients that may not be accurately represented. This said, the reality is that within a single industry it may be the case that different sources of agglomeration benefit may arise because of different types of industry interactions. For instance, an industry may cluster because of labour-pooling benefits while the activities of knowledge exchange or service distribution may predominantly be across industry types. Furthermore the sources of agglomeration benefit may be particular to a firm and not necessarily suitable for generalization across an industry. There are many reasons why we may expect the results of an analysis including both urbanization and localization effects to be somewhat unclear.

What the results in this analysis have suggested is that urbanization and localization economies need not be at odds, as quite a number of industries experience significant estimated effects from both types. Moreover, broader industry sectors do not seem to be completely disposed to only one form of agglomeration externality. While convention postulates that manufacturing relies on benefits from localized industry clustering, the analysis here shows that benefit is also gleaned from overall urban scale. Media-related activities and financial services are shown to experience benefit from both, while generally showing more significant effects from industry localization. For professional, scientific and technical services it seems that the industry urbanization and localization effects are more or less balanced, while perhaps urbanization effects dominate in the health sector.

As for the likelihood that agglomeration benefits can be generalized across Sydney and Melbourne, for the most part the results suggest this can be done. While 10 industries showed incompatibilities between Sydney and Melbourne for each of the agglomeration parameters, this may not hold in repeated samples. Furthermore, if there were more observations to increase the power of the tests then there would be less of a chance of a Type II error being committed. This means avoiding false rejection of the null hypothesis of "no difference" in the agglomeration benefit between Sydney and Melbourne. Adhering to the results of this analysis, however, one might be wary of applying elasticity estimates derived from Sydney data to Melbourne projects for a select number of industries, and vice versa. Attempting to generalize these results to the rest of Australia may even be a riskier move, as Sydney and Melbourne are similar in size and age whereas other Australian capitals are of significantly smaller scale with potentially differing urban typologies and economic specialization.

What has been omitted from the Part 1 analyses is an investigation of any potential endogeneity bias that is affecting the agglomeration parameters results. While empirical works have typically shown the effects of endogeneity to be small, and even question the existence of an influence at all, it is still possible that the results in these analyses are being impacted by a reverse relationship where productivity reinforces agglomeration. This matter is explored in the next Part's analyses where the larger sample sizes allow tests of endogeneity to be made more reliably.

CHAPTER 5: Part 2 Analyses

5.0 Introduction to Part 2 Analyses

In Part 2 Analyses, the productivity impacts of agglomeration economies are estimated on a disaggregated industry basis using a much finer geographic unit of observation than in Part 1. The spatial unit of the Work Destination Zone (Melbourne) or the Travel Zone (Sydney) is the smallest spatial unit for which the Australian Bureau of Statistics provides place-of-work data. In addition to the benefit that its small size increases the observation counts, it solves the problem of using politically defined spatial boundaries. Employment zones are defined in each state and territory by their respective transport authorities and are typically used to analyse data on urban transport patterns, which are largely driven by the distribution of employment. Density measures constructed from employment data, as a result, are less susceptible to being diluted or misrepresented because of undeveloped geographic features or areas within a spatial unit.

Because of the finer geographic detail and perturbation practices of the ABS, maintaining an analysis using industry data at the 3-digit ANZSIC level would have been a risky endeavour because of randomization on small cell values. For this reason, Part 2 analyses here in Chapter 5 use the 2-digit level of industry aggregation and estimate wage functions for 30 industries while still covering a wide range of sectors. For this same reason, some of the categories on occupation type are collapsed to increase the number of observations in each geographic unit, thus preserving a level of robustness in the census data provided for these spatial units.

The utilization of Sydney's Travel Zones and Melbourne's Work Destination Zones increased the potential sample sizes to 2,690 and 2,083 respectively, which entailed the greater metropolitan areas (GMA) of both cities. The eventual industry-specific sample sizes were less than these figures because each employment zone did not contain employment in all of the industries selected for Part 2's analyses. Despite this, the smallest sample sizes to arise from the Sydney and Melbourne datasets were 163 and 194, respectively. In both cities, this was for the industry of Heritage Activities. At the other extreme, the largest respective sample sizes for Sydney and Melbourne were 1,863 and 1,603 – which again in the case of both cities applied to the industry of *Professional, scientific and technical services*.

The finer grain of geographic detail used in this section required a considerably different approach to handling the spatial and employment data from that in Part 1 Analyses. The data for this chapter's analyses were imported and stored in an open-source, object-relational database system called Postgres[21] which was coupled with a graphic user interface (GUI) called PGAdminIII.[22] These two programs were used in conjunction with a plug-in called PostGIS that spatially enabled Postgres to store and make calculations with spatial data. The functionalities of Postgres and PostGIS combined made for an extremely powerful tool for carrying out and storing the results of the complex calculations that merged spatial and employment data. In using this software to carry out all necessary tasks, a great number of SQL scripts had to be authored. The specifics of these scripts will be discussed more fully in the individual accounts of Part 2's analyses that follow.

Analysis 2-1 estimated agglomeration effects for 30 2-digit industry classifications using the employment zone data in a standard log-linear model specification. Additionally, it estimated the impact of adding quadratic terms, which essentially enable the elasticity estimates to vary with the level of effective density. Next, Analysis 2-2 re-estimated wage functions for the 30 industries while addressing the issue of endogeneity. Two different types of instruments were tested for their validity and effectiveness in controlling for the endogeneity issue in a 2SLS framework. Finally, an alternative approach to simultaneously accounting for urbanization and localization effects was carried out in Analysis 2-3, where the former form of agglomeration benefit was addressed by the effective density index and the latter by estimating industry-specific employment levels within concentric ring bands of each employment zone.

It should again be noted that, as with Part 1's analyses, the comprehensive model results are not reported here with the Part 2 analyses. Focus is given to the effects of agglomeration and the quality of the overall models, not to the significance of individual controls per se. If the complete model results are desired then they can be provided by contacting me.

---

[21] Available at http://www.postgresql.org/download/
[22] Available at http://www.pgadmin.org/download/

5.1.0 Analysis 2-1 Overview: Elasticity estimates using WDZ data

The estimating model utilized for this analysis was maintained from Analysis 1-2. It assumes that an estimated productivity effect from agglomeration (and any other control variable for that matter) is characterized by constant returns to scale because of the effective density parameter (and all other parameters as well) holding constant for all levels of effective density (and for all levels of the other controls). This model is illustrated here again for ease of reference.

$$\ln I_{k,m} = \alpha_k + \beta \ln U_m + \sum^{i} \delta_i \ln Occ_{i,k,m} + \sum^{j} \gamma_j \ln Edu_{j,k,m} + \psi \ln Exp_{k,m} + e_{k,m} \quad [5A]$$

What we can expect from estimating this model with a dataset compiled at the employment zone level is to have parameter estimates produced with smaller errors, thus seeing previously insignificant effects of agglomeration in some industries perhaps now become significant. Additionally, it is interesting to learn how using SLAs (which for the most part can be considered administrative units as they align closely with LGAs) may be having a confounding effect on the econometric estimations. While the larger sizes of SLAs can distort the understanding of urban form and the distribution of economic activity within them, the actual magnitude of this potential effect is at the moment unknown. Adjusting the SLA geographies for 'urbanized areas' would have bettered the situation, however the process was inexact and the geographic unit sizes remained rather large.

Following the industry-specific wage function estimations carried out in accordance with the econometric specification depicted above in equation 5A, quadratic terms were added to the industry models to allow the effects of agglomeration economies to vary with the level of effective density. This variation on the log-linear model is presented below.

$$\ln I_{k,m} = \alpha_k + \beta_{1k} \ln U_m + \beta_{2k} \ln U_m^2 + \sum^{i} \delta_i \ln Occ_{i,k,m} + \sum^{j} \gamma_j \ln Edu_{j,k,m} + \psi Exp_{k,m} + e_{k,m} \quad [5B]$$

The purpose of estimating this model was to identify whether industries were experiencing increasing, diminishing or constant returns to scale. The sign on the industry-specific parameter estimates, B_{k2}, indicates which of these happens to be the case. If the coefficient estimate is negative, this means that there are diminishing returns to scale; however if positive, increasing returns to scale are present. If the coefficient on the quadratic term is insignificant, regardless of the sign, interpretation is that the industry is experiencing constant returns to scale because the quadratic term can simply be dropped from the model. If this was the case then the model that was estimated reverted back to 5A.

Before moving on to the results, the method applied to estimating the effective densities of Sydney and Melbourne employment zones will be discussed.

Calculating Effective Density for Sydney and Melbourne Employment Zones

As mentioned in the overview of Part 2's analyses, the approach to calculating effective densities for Sydney and Melbourne differed significantly from the approach taken with SLA data. The process here involved the utilization of an open-source database software called PostgreSQL that was spatially enabled with an add-on called PostGIS. For ease of management, all employment data was imported into this software along with all spatial files. All the results of calculations carried out using Postgres and PostGIS were stored in the database as well. A substantial amount of SQL code had to be written to make all of this possible. While much of the code written to be executed in Postgres for importing and organizing the data was fairly standard, some of the code required a bit more time to develop and test for reliability. The majority of this code will be left for explanation in Appendix C but the core calculation of effective density will be discussed here.

When a spatial file, such as the ESRI Shapefile for Melbourne's Work Destination Zones, is being imported into Postrgres it is facilitated by using a program called 'shp2pgsql' which is included with the installation of PostGIS. The program writes a script that can subsequently be run to import a spatial file into a database. A command for this process had to be written in Terminal, which is a program that provides a line interface to control the foundations of the UNIX-based operating system on a Mac. This process is similar when using a Windows-based computer. When using shp2pgsql to convert an ESRI Shapefile file in preparation for import into Postgres, first the system has to be told where it can find the shp2pgsql binary to make this conversion from Shapefile to SQL script. This was done by entering the following command in Terminal.

export PATH=$PATH:/Library/PostgreSQL/8.4/bin [5C]

By doing this, the system would automatically know to search the directory "/Library/PostgreSQL/8.4/bin" for executable files. Next, navigation had to occur within Terminal to the folder where the Shapefile was located (See Appendix C for these commands). In this example where the Melbourne Work Destination Zone layer "tz2006_MGA55.shp" was being imported into PostgreSQL, command 5D was run in Terminal to create a script that could subsequently be run by Postgres. This script generated by shp2pgsql for use by Postgres is effectively what translated the spatial layer into a format that could be stored in the database.

shp2pgsql -I -s 28355 tz2006_MGA55.shp shp.melb_wdz > shp_melb_wdz.sql [5D]

This command illustrated above told the system to use the shp2pgsql binary (which it knew where to find because of running 5C) to index the data (-I) and write the spatial file into a SQL script titled shp_mel_wdz.sql while using projection 28355. Recalling the discussion in section 3.9 on data, a projection is a region-specific algorithm that enables spatial calculations using geometry data-types to take into account region-specific differences in terrain variability. With the installation of PostGIS is included a directory of projections along with their IDs. The projection type should usually be provided along with a spatial layer. When the SQL file was produced from running 5D in Terminal, it was subsequently executed in Terminal as well, where it was told to copy the spatial data into shp.melb_wdz, where 'shp' was the name of the schema and 'melb_wdz' was the name of the table to store the data in Postgres. Telling Postgres to run the script via Terminal was done by entering the following command:

psql -h localhost -p 5433 -d roman -U postgres -f shp_melb_tz.sql [5E] where the host (-h) was specified as 'localhost', the port (-p) was specified to 5433 which was the port on which Postgres was operating[23], the relevant database (-d) to receive the data was arbitrarily named 'roman', and the user (-U) given to access the database was titled 'postgres'. This command had to be run in Terminal from the folder in which the file (-f) 'shp_melb_tz.sql' was located.

The employment zone spatial files host a great deal of information about their contained spatial units, including their IDs, areas (in square metres in this case), perimeters (given in metres in this case), pertaining SLAs, and their spatial data (which is stored in the form of a 'geometry' data-type), in addition to other various types of information. To carry out the effective density calculation using the data contained within the table generated by running the command specified in 5C, additional data had to be added to the table. In this particular example, four new columns were created within the table 'shp.melb_wdz' that was storing the employment zone spatial data. These four columns were titled

1) "employment_no" (employment number)
2) "employment_density" (employment per square kilometre)
3) "distance_weighted_ed_sum" (the sum of all other WDZs weighted by their linear distance to the given unit) and
4) "total_ed" (which was the sum of columns 2 and 3)

Column 1 was populated from employment data on Melbourne's WDZs that was being stored in the database in a separate table. Populating this column was done by running the following script in Postgres:

```
UPDATE shp.melb_wdz t SET employment_no = e.num_employed FROM                [5F]
        import.melbourne _employment e WHERE
        substring(e.wdz_code FROM 7 FOR 4)::INTEGER = t.vicdznp06::INTEGER;
```

where the last line simply specified a way of pairing up the Work Destination Zone IDs in the spatial file with those in employment data table.

23 By default, Postgres installed to operate using post 5432.

Column 2 was populated by drawing on the data from the first column and the WDZ geometry data by running the following script:

```
UPDATE shp.melb_wdz t SET employment_density = t.no_employed /                    [5G]
        sqrt((st_area(st_transform(t.the_geom, 28355)) / 1000000) / pi());
```

where the second line calculated the area of a zone after transforming the spatial unit into the appropriate projection (which was unnecessary since it had already been stored in projection 28355 but stylistically embodied good practice) and changed the unit of measure from square metres to square kilometres.

Populating column 3 was the most complex and time-consuming piece of code to run in the calculation of effective density. It involved executing a 'self-join', or in other words the creation of an origin-distance matrix, on the spatial table and then eliminating duplicate entries such that the instance of zone 1 and 2, for example, would not be repeated with the instance of zone 2 and 1. This section of code took approximately 2 hours to run, whereas the others took only several seconds or minutes. It appears as below

```
UPDATE shp.melb_wdz t SET distance_weighted_ed_sum_error_test =                   [5H]
        temp.distance_weighted_ed_sum FROM (
SELECT t.gid AS tz_origin_gid,
         sum(t2.no_employed / (st_distance(st_centroid(st_transform(t.the_geom, 28355)),
        st_centroid(st_transform(t2.the_geom, 28355))) / 1000)) AS
        distance_weighted_ed_sum
FROM shp.melb_wdz t, shp.melb_wdz t2 -- this creates every combination of two WDZs
WHERE --t.gid < 10 AND -- for dev't, comment out this line after WHERE in production
        st_distance(st_centroid(st_transform(t.the_geom, 28355)),
        st_centroid(st_transform(t2.the_geom, 28355))) > 0 -- this eliminates self-joins
        AND t.is_duplicate = FALSE AND t2.is_duplicate = FALSE
GROUP BY t.gid ) AS temp
WHERE t.gid = temp.tz_origin_gid;
```

All comments in the above code that follow a double-hyphen (--) are ignored when run in PostgreSQL and simply provide a bit of clarity to sections of the code for future reference. On line 8, immediately after "WHERE", by deleting the double-hyphen before "t.gid < 10 AND" the script is restricted to run for only zones 1 through 10. It was written this way to enable one to run the script first on a reduced sample to review the results. In this way, one could determine whether the script was accomplishing its intended task without risking a 2-hour wait, only to realize that refinements had to be made.

The final step in the effective density calculation meant summing the results of columns 2 and 3 to give a measure consisting of an area's own density, plus the contributions to accessibility of all surrounding areas weighted by their linear distances. This script appears as follows:

```
UPDATE shp.melb_wdz SET total_ed = employment_density +                [5I]
        distance_weighted_ed_sum;
```

One fairly significant setback that occurred in the analyses using employment zone data was the realization that in the employment zone spatial files for both Sydney and Melbourne, there are a number of zones that have multiple entries while sharing one identification code. These are instances where islands are not physically part of mainland zones, yet fall under one given ID number. In Melbourne, for instance, there are 24 more polygons than there are unique zones with a total of 13 zones being affected.[24] The problem that this issue generated was that the script illustrated in 5H ended up allocating a given population figure that was intended for one zone to each of its multiple instances, which in turn generated multiple effective density values for these zones. This issue became evident when joining employment data together with the effective density indexes because some employment zones ended up having repeat observations.

This issue was resolved by identifying the repeat employment zone entries and creating a new column in the spatial tables titled 'is_duplicate'. The values in this column for all rows were given a Boolean default value of 'FALSE', except for those that were

[24] The zones to experience multiple instances in the Sydney TZ dataset are 212, 704, 719, 1871, 1871, 2869, 2966, 3010, 3047, 3088, 3101, 3107, 3144, 3151, 3171, 3350, 3378, 3385, 3514, 3516, 3614, and 3630. In the Melbourne WDZ dataset they are 1916, 1926, 1932, 1939, 2105, 2380, 2416, 2423, 2515, 2518, 2538, 2539, and 2540.

identified as duplicates. They were updated with the value 'TRUE'. Using the Melbourne dataset as an example, the identification of zones with repeat instances was determined by running the following query in PostgreSQL:

```
SELECT vicdznp06 from shp.melb_wdz GROUP BY vicdznp06 HAVING                [5J]
      COUNT(vicdznp06) > 1 ORDER BY vicdznp06;
```

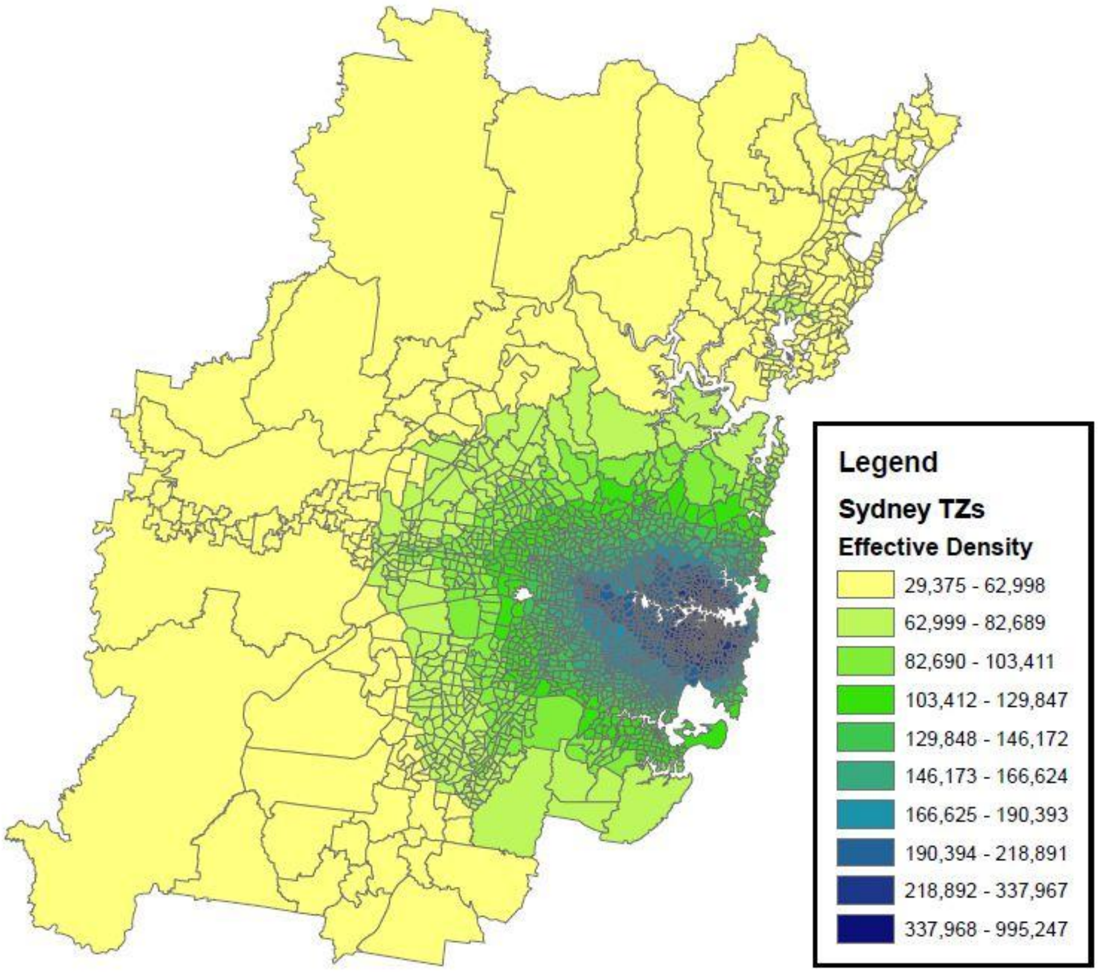
Updating the cells in the 'is_duplicate' column that corresponded to repeat entries was achieved by running the script indicated in 5K, where zone 1916 in Melbourne is being used as an example.

```
UPDATE shp.melb_wdz SET is_duplicate = TRUE WHERE vicdznp06 = 1916        [5K]
      AND gid != 2234;
```

The 'gid' is an arbitrary and unique value in a dataset, thus there was a way of identifying every zone regardless of instances where they may have shared an employment zone number. Duplicates were determined based on size, where the largest was assumed to be the actual zone to contain employment. It is possible that employment may be distributed among these zones; however, most repeat zones were far too small to logically contain any development. If employment is in fact distributed among multiple zones in some cases, there was no way of identifying this in the datasets. The most important outcome of this, however, was to designate one representative zone to be matched with the employment data otherwise repeat observations of different effective densities but shared control values would have disrupted the results of the analyses.

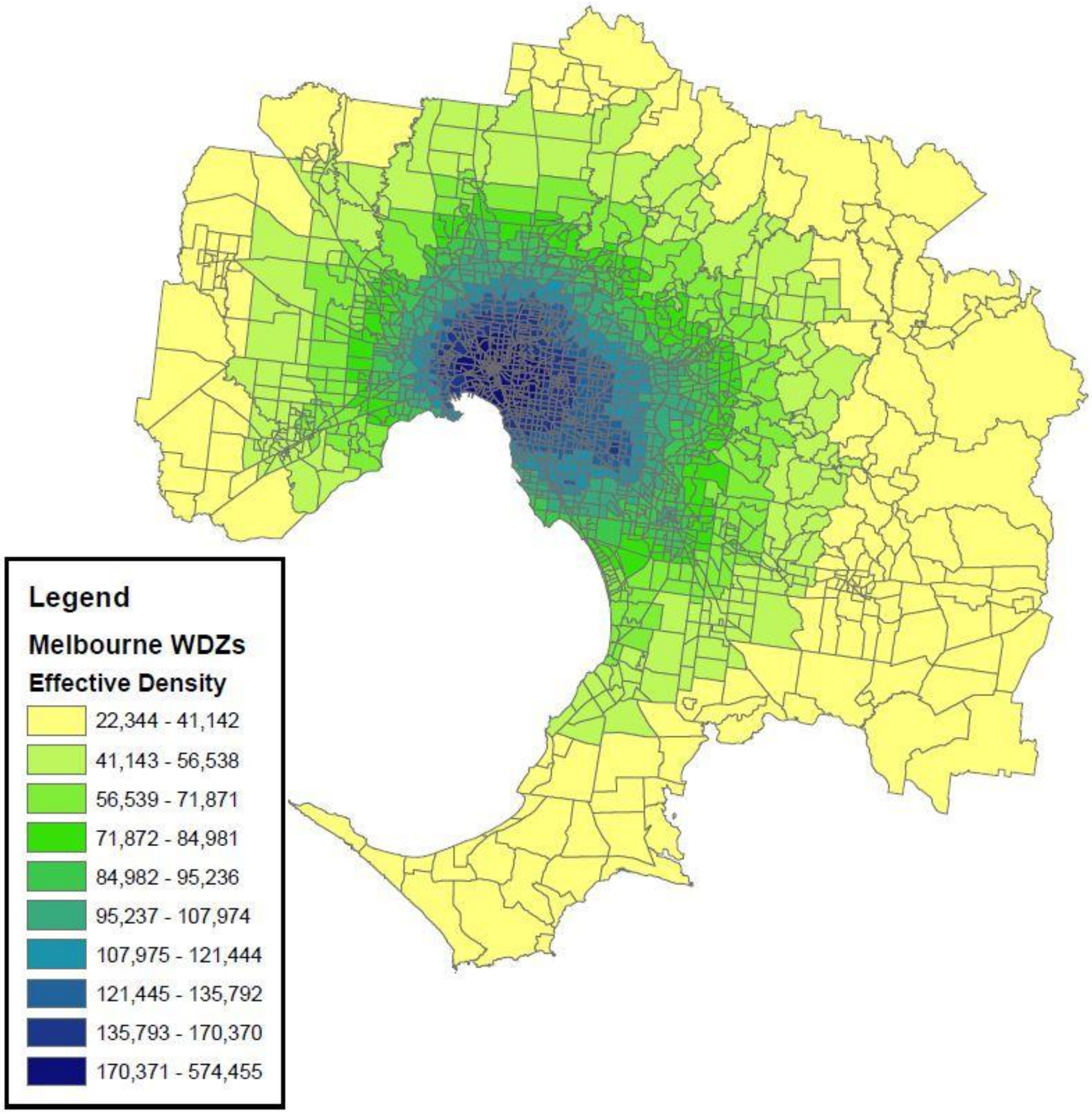
What has been described here was the general approach/method to calculating the effective density indexes for Sydney and Melbourne with proximity being measured by linear distance. Effective density was also calculated for Sydney by weighting proximity by travel-time, where mean journey time was calculated from the weighted average of peak a.m., daytime interpeak, and peak p.m. travel times provided for the 2006 base year to match the 2006 census data. As with the analyses on Sydney using SLA data, the version of the effective density index that employed travel-time as a proximity weight calculated own-area travel-time radii with a presumed average travel speed of 40 km/hr, as this information was not provided by the TDC. The computation of the travel-time weighted effective density index for Sydney was carried out in much the same manner as was described above for the linear distance version, with the exception that the issue of repeat zones did not apply. Further details of this code can be found in Appendix C as well. Thematic maps indicating effective density values across the Sydney and Melbourne Statistical Divisions appear below in Figures 5.1.1 and 5.1.2.

Figure 5.1.1: A Thematic Map of Effective Density Levels Across the Sydney SD[25]



[25] Note: The estimations of industry wage functions were carried out using Sydney SD's travel zones. These are displayed here in this thematic map. The effective density index values for Sydney, however, were calculated by including employment from two surrounding SD's outside of the Sydney SD, namely Hunter and Illawarra. This does not significantly influence the elasticity estimates produced but means that employment levels in Hunter and Illawarra influence Sydney SD effective density values displayed in the legend by increasing them all slightly. This is only important to keep in mind if comparing the effective density values to those reported for the Melbourne SD, which only considered employment within the immediate SD boundary.

Figure 5.1.2: A Thematic Map of Effective Density Levels across the Melbourne SD



With effective densities being calculated, matched up with the rest of the industry-specific wage function data in the database and exported to CSV format, the industry wage function regressions could be carried out. The ensuing approach to carrying out the regressions was the same as in Analysis 1-2, where first estimates were produced using OLS. If the Breusch-Pagan test for heteroskedasticity detected an inconsistent error variance in the residuals then the alternative regression model of generalized least squares (GLS) was used instead. A p-value of 0.05 was used as the critical value instead of 0.10 that was applied to the SLA analyses because the sample sizes were much larger when using employment zones. With a larger sample, one could expect greater precision in the tests and thus afford to lower the critical values below which the null hypothesis of 'no error variance' is rejected.

The next section will discuss the results of Analysis 2-1 as they pertain to Sydney's elasticity of productivity estimates measured with respect to linear effective density and travel-time effective density and linear effective density in Melbourne. First to be discussed are the Sydney regression results where effective density had been calculated by weighting employment by linear distance. Therein will be the identification of fundamental differences between using SLA and employment zone data. Next to be discussed are the results of the Sydney regressions where effective density was calculated by weighting employment by weighted-average travel-time, comparing the magnitudes with the linear ED results to see if the trend of larger travel-time estimates still holds. Following this will be a review of the Melbourne results and their comparison to the Sydney results. Finally, an account will be given of the outcomes of adding quadratic terms to the equations to allow productivity elasticities measured with respect to effective density to vary with levels of agglomeration.

5.1.1 Analysis 2-1 Results

Sydney and Melbourne Results

Immediately evident when reviewing the Sydney results is the increased precision of the elasticity estimates. Only five of the 30 industries came out insignificant in this analysis and the majority of the significant results had p-values below 0.01 (as indicated by the industries with three asterisks by their elasticity estimates in Table 5.1.1). The industries for which effective density parameter estimates came out insignificant were *Food Product*

Manufacturing (0.020), *Primary Metal and Metal Product Manufacturing* (0.020), *Fabricated Metal Product Manufacturing* (0.018), *Public Order, Safety and Regulatory Services* (0.000), and *Heritage Activities* (0.054). The magnitudes of their error terms were in line with the rest of the industries analysed, thus one could attribute their insignificance primarily to their very low point estimates rather than to their potentially large variation. The industries to display the largest elasticity estimates were *Creative and Performing Arts* (0.244), *Publishing (except internet and music publishing)* (0.218), *Motion Picture and Sound Recording Activities* (0.217), *Basic Chemical and Chemical Product Manufacturing* (0.198), and *Textile, Leather, Clothing and Footwear Manufacturing* (0.180).

In general, the magnitudes of the estimates did not deviate much from the SLA results. The 2-digit level of industry aggregation made it difficult to make direct comparisons but aggregating the related 3-digit industries and averaging them proved fairly effective. Aggregating industry codes 241, 242, and 246 and averaging their point elasticity estimates from Analysis 1-2 gives a value of 0.053. The higher-level aggregate of these industries here in Analysis 2-1, namely industry 24, had an elasticity estimate of 0.060, which is not far off. Similarly, in Analysis 1-2 the medical industries had elasticity estimates ranging from 0.085 to 0.108 while the results here estimated them at around 0.083. In contrast, the aggregation of 3-digit financial industries produced averages that differed quite substantially from the 2-digit estimates produced with employment zone data. The mean for industries 620, 622 and 624 is 0.115; for 631 and 632 it is 0.035; and for industries 641 and 642 the mean is 0.062. Here in Analysis 2-1 the results for these industry-classification aggregates are 0.172 for industry 62, 0.146 for industry 63, and 0.164 for industry 64. These constitute deviations in an order of magnitude of roughly 0.06 for *Finance*, 0.11 for *Insurance* and 0.10 for *Auxiliary Finance and Insurance Services*. Similarly, the mean elasticity from all 3-digit *Professional, Scientific and Technical Services* analysed is 0.093, whereas the 2-digit aggregate of these is estimated at 0.162.

The difficulty of direct comparison between the results of Analyses 1-2 and 2-1 is partially because the 3-digit level aggregates do not account for all related industries, but also because the averages given are not weighted and thus unequal employment shares will be affecting the results to some extent. Taking this into account, however, the cross-analysis comparison still suggests that using SLA level data is likely to be having some effect on the magnitudes of the parameters being estimated. The only 3-digit industry missing from the

Professional, Scientific and Technical Services aggregate is *Veterinary Services*, which is unlikely to account for a 0.07-point reduction in the 2-digit parameter estimate. The most notable difference in estimated industry elasticities is in the retail sector. In Analysis 1-2, the result for grocery retail was negative and significant with an estimate of -0.063, yet here in Analysis 2-1 it was positive and significant with an estimate of 0.087. Similarly, other retail activities in Analysis 1-2 had elasticities estimated close to zero while here the estimate for industry 42, *Other Store-Based Retailing*, in Sydney was 0.138. One could hypothesize that this disparity exists because of the perturbation of the datasets or general level of error in the sampling of the data; however, the results are very similar between Sydney and Melbourne in the two Analyses, which suggests that the cause of disparity is due to spatial unit size.

It is difficult to ascertain the reason(s) why the size differences between the two geographic units caused substantially different results between the two analyses for some, but not all, industries. Since the ABS data are averaged over the spatial units, variations in employment density within SLAs are much more prone to being "smoothed out" and this could result in a reduction in the size of the productivity gradient being estimated. This does not, however, necessarily explain how there could be a reversal in the sign by moving to using smaller spatial units. Too much emphasis should not be placed on trying to rationalize this, however, as it detracts from the fact that the majority of results meet expectations.

Moving away from comparing the results to those using SLAs, the relationship between knowledge-intensity and the magnitude of the effective density parameter estimate seems to hold very well. The manufacturing sector was still characterized by much variability in industry-specific estimates, thus the average of the results is fairly low at 0.085, which is consistent with the existing literature. The average of the aggregate of media and professional service activities is the highest at 0.182 while financial sector services are close behind with an average 0.161. Mining services and activities were omitted from this analysis because of the rather poor results that were estimated when using SLAs. These were replaced by an investigation of several new industries, notably *Social Assistance Services*, *Heritage Activities*, and *Creative and Performing Arts Activities*. They were chosen because of the importance that these types of industries and activities play in creating and maintaining social capital and quality of living.

Much of the focus of these empirical works on agglomeration has historically concerned itself with manufacturing industries and now more recently, professional and financial services. It is interesting to see that *Creative and Performing Arts Activities* have the highest elasticity estimate of all industries analysed, with *Social Assistance Services* having a rather large estimate as well at 0.154. It is rather intuitive that these two industries experience a substantial amount of productivity uplift from the density of activity for they would both thrive in places of good public transport access and where population density is rather large. The former industry may also benefit from the presence of a mass of restaurants, bars, and other like amenities offered by more urbanized areas, as patrons of the arts are likely to round out their evenings with other social activities. *Heritage Activities*, on the other hand, was among those industries with insignificant coefficient estimates. This could possibly be explained because certain sub-classifications of this industry, such as parks and gardens, do not operate for a profit thus their impetus to 'survive' as a firm, company or organization is not equivalent to that of for-profit activities. Moreover, industries such as these are not mobile or able to choose their locations readily.

The relatively high precision in the estimates using employment zones were accompanied by an overall reduction in the adjusted R-squared values of the regressions. This was expected because the provision of data in the form of averages smooths out variations in the data. At the employment zone level, the finer geographic scale means that the overall sample has a greater degree of variability than when using SLAs. The anticipated result of this is a reduction in the model's ability to explain variations in income. In contrast to this, the F-values produced by this analysis were far greater than those produced in Analysis 1-2. The F-value is generally more important to consider than the adjusted R-squared value, as it is a measure of a model's overall level of significance – the smaller the F-value, the more significant the cumulative effects of the variables in the model. The result of more significant model estimates as indicated by smaller F-values was most likely related to the larger sample sizes producing more precise parameter estimates. All of these industry regression results for Sydney can be found below in Table 5.1.1.

Table 5.1.1: Sydney Regression Results Using TZ Spatial Units (ED weighted by linear distance)

Ind ID	Industry Name	ε_U	S.E.	P-Value (F)	Adjusted R^2	N
11	Food Product Manufacturing	0.020	0.036	3.42E-56	0.311	740
13	Textile, Leather, Clothing and Footwear Manufacturing	***0.180	0.042	3.39E-27	0.209	588
18	Basic Chemical and Chemical Product Manufacturing	***0.198	0.036	1.06E-29	0.371	326
21	Primary Metal and Metal Product Manufacturing	0.020	0.040	1.20E-23	0.299	341
22	Fabricated Metal Product Manufacturing	0.018	0.030	7.33E-09	0.112	392
23	Transport Equipment Manufacturing	***0.111	0.035	8.16E-36	0.373	386
24	Machinery and Equipment Manufacturing	**0.060	0.024	3.99E-40	0.262	658
30	Building Construction	***0.166	0.014	3.10E-61	0.165	1654
31	Heavy and Civil Engineering Construction	***0.160	0.027	2.11E-30	0.283	453
33	Basic Material Wholesaling	***0.100	0.028	1.15E-38	0.234	721
34	Machinery and Equipment Wholesaling	***0.075	0.024	2.14E-29	0.183	733
41	Food Retailing	***0.087	0.017	1.30E-42	0.147	1327
42	Other Store-Based Retailing	***0.138	0.013	4.68E-79	0.203	1687
44	Accommodation	***0.112	0.016	4.28E-34	0.208	732
45	Food and Beverage Services	***0.087	0.011	7.20E-134	0.314	1694
54	Publishing (except Internet and Music Publishing)	***0.218	0.029	1.62E-24	0.186	603
55	Motion Picture and Sound Recording Activities	***0.217	0.041	2.51E-58	0.488	430
59	Internet Service Providers, Web Search Portals, and Data Processing Services	***0.132	0.033	1.19E-12	0.256	220
62	Finance	***0.172	0.014	2.13E-96	0.378	979
63	Insurance	***0.146	0.016	4.11E-50	0.399	487
64	Auxiliary Finance and Insurance Services	***0.164	0.016	1.89E-53	0.247	925
67	Property Operators and Real Estate Services	***0.129	0.015	1.24E-61	0.233	1136
69	Professional, Scientific and Technical Services (except Computer System Design and Related Services)	***0.162	0.011	1.80E-155	0.327	1863
75	Public Administration	***0.159	0.014	7.66E-77	0.302	1036
77	Public Order, Safety and Regulatory Services	0.000	0.017	1.91E-26	0.140	882
84	Hospitals	***0.085	0.022	1.25E-25	0.242	471
85	Medical and Other Health Care Services	***0.081	0.013	1.85E-52	0.154	1538
87	Social Assistance Services	***0.154	0.015	1.60E-56	0.158	1601
89	Heritage Activities	0.054	0.041	5.70E-05	0.126	194
90	Creative and Performing Arts Activities	***0.244	0.041	1.61E-41	0.315	546

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

When the industry regressions were carried out with the travel-time specification of the effective density index, only four industries generated insignificant parameter estimates rather than five when estimated with the linear specification. The precise industries that were estimated with insignificant parameters remained the same in both contexts, where proximity was weighted by linear distance and where it was weighted by travel time, except that *Heritage Activities* were significant in the latter case. This arose more because there was a slightly higher elasticity estimate rather than from a reduction in the estimated standard error. Differences in adjusted R-squared values show no distinguishable trend of one model explaining more of the variation in earnings than the other, and in most cases the values are nearly the same.

The situation of larger elasticity estimates being generated when effective density is measured by using travel-time as a proximity weight still holds here when using travel zones in Sydney. In the case of every industry analysed, the model specification with travel-time effective density produced a larger elasticity estimate than when linear effective density was used. Both the largest and the smallest differences occurred within the manufacturing sector, where it was a magnitude of 0.060 for *Textile, Leather, Clothing and Footwear Manufacturing* and 0.005 for *Fabricated Metal Product Manufacturing*. As argued by Graham (2006), the lower elasticity estimates generated by using linear effective density to represent accessibility is an indication that transport constraints are likely to be present. The main constraint that he refers to is road traffic congestion; however, it could also be a result of restricted access because of geological or geographic impediments such as impassable terrain or the presence of an inlet or river. The differences between the linear and travel-time effective density results do not seem to show any particular pattern, thus it is unlikely that one could postulate from this that some industries are experiencing worse effects from accessibility constraints than others. The rather ubiquitous and unsystematic nature of the differences between estimates could be a result of access restrictions because of the presence of Port Jackson and its numerous tributaries and/or a general level of citywide congestion that limits accessibility for all industries. For comparison, the travel-time effective density results are shown below in Table 5.2.1.

Table 5.1.2: Sydney Regression Results Using TZ Spatial Units (ED weighted by travel-time)

Ind ID	Industry Name	ε_U	S.E.	P-Value (F)	Adjusted R^2	N
11	Food Product Manufacturing	0.035	0.047	9.22E-57	0.314	740
13	Textile, Leather, Clothing and Footwear Manufacturing	***0.240	0.054	5.81E-28	0.213	588
18	Basic Chemical and Chemical Product Manufacturing	***0.224	0.048	1.00E-27	0.352	326
21	Primary Metal and Metal Product Manufacturing	0.039	0.047	2.66E-23	0.296	341
22	Fabricated Metal Product Manufacturing	0.024	0.039	2.27E-08	0.106	392
23	Transport Equipment Manufacturing	***0.149	0.044	2.00E-32	0.346	386
24	Machinery and Equipment Manufacturing	***0.078	0.030	1.84E-46	0.295	658
30	Building Construction	***0.210	0.018	3.59E-61	0.165	1654
31	Heavy and Civil Engineering Construction	***0.207	0.034	1.05E-30	0.288	454
33	Basic Material Wholesaling	***0.131	0.035	7.26E-39	0.235	721
34	Machinery and Equipment Wholesaling	***0.098	0.030	4.11E-29	0.181	733
41	Food Retailing	***0.117	0.022	9.89E-44	0.150	1327
42	Other Store-Based Retailing	***0.177	0.017	1.26E-78	0.202	1687
44	Accommodation	***0.133	0.020	3.21E-33	0.203	732
45	Food and Beverage Services	***0.111	0.014	2.00E-139	0.324	1694
54	Publishing (except Internet and Music Publishing)	***0.260	0.036	1.05E-24	0.188	603
55	Motion Picture and Sound Recording Activities	***0.243	0.053	1.20E-60	0.501	430
59	Internet Service Providers, Web Search Portals, and Data Processing Services	***0.166	0.042	2.13E-12	0.252	220
62	Finance	***0.181	0.018	1.81E-01	0.331	979
63	Insurance	***0.174	0.020	1.34E-49	0.396	487
64	Auxiliary Finance and Insurance Services	***0.194	0.018	4.07E-52	0.242	925
67	Property Operators and Real Estate Services	***0.151	0.018	6.75E-62	0.233	1136
69	Professional, Scientific and Technical Services (except Computer System Design and Related Services)	***0.191	0.012	7.40E-153	0.323	1863
75	Public Administration	***0.179	0.016	3.18E-01	0.318	1036
77	Public Order, Safety and Regulatory Services	0.008	0.021	2.51E-27	0.144	882
84	Hospitals	***0.134	0.025	5.45E-28	0.260	471
85	Medical and Other Health Care Services	***0.100	0.016	2.84E-52	0.153	1538
87	Social Assistance Services	***0.194	0.018	1.07E-56	0.159	1601
89	Heritage Activities	***0.077	0.045	8.10E-05	0.122	194
90	Creative and Performing Arts Activities	***0.282	0.051	1.63E-21	0.192	546

Note: * = Significant at 0.10   ** = Significant at 0.05   *** = Significant at 0.01

The results for Melbourne with respect to the linear specification of effective density are given below in Table 5.1.3. They show only two industries having insignificant coefficients, as opposed to five in Sydney. These two industries were *Primary Metal and Metal Product Manufacturing* and *Public Order, Safety and Regulatory Services*. Both of these came out insignificant in Sydney as well. The three industries insignificant in Sydney that came out significant in Melbourne were *Food Product Manufacturing*, *Fabricated Metal Product Manufacturing* and *Heritage Activities*, which were estimated to have substantially larger coefficients in the latter case. Comparing elasticity averages across aggregated, broader industry sectors between the two cities shows the impacts of agglomeration to be larger in Melbourne for manufacturing (0.114 compared to 0.085), retail (0.131 compared to 0.105), medical (0.108 compared to 0.083) and social services and activities (0.116 compared to 0.113) – though only marginally so in the latter case. Agglomeration impacts seem larger in Sydney, however, for the construction (0.163 compared to 0.153) and financial services (0.161 compared to 0.153) sectors. The average across media and professional, scientific and technical services industries is equal between the cities with an elasticity estimate of 0.182.

The largest elasticities to be estimated for Melbourne were in the industries *Motion Picture and Sound Recording Activities* (0.200); *Basic Material Wholesaling* (0.194); *Professional, Scientific and Technical Services* (0.190); *Public Administration* (0.181) and *Auxiliary Finance and Insurance Services* (0.176). The industries to return the lowest estimates were *Public Order, Safety and Regulatory* Services (0.016); *Primary Metal Product Manufacturing* (0.018); *Food Retailing* (0.041); *Medical and Other Health Care* Services (0.091) and *Food Product Manufacturing* (0.096). While most of these industries do not correspond to those returning the largest and smallest estimates produced for Sydney, the effects of agglomeration economies influence the broad-industry sectors in virtually the same order of magnitude. The effects were greatest for the aggregate of media and professional, scientific and technical services (0.182) and least for the manufacturing sector (0.114). The financial sector experienced the second largest effect with a broad industry average elasticity of 0.156.

In comparing the results between Analysis 1-2 and Analysis 2-1 for Melbourne, a similar pattern emerged that was evident in the comparison of the Sydney analyses. The most dramatic difference for Melbourne was also the disparity between the elasticity estimates for retail. Analysis 1-2 estimated a coefficient of -0.104 for *Supermarkets and Grocery Stores* but *Food Retailing* in Analysis 2-1 had an estimate of 0.041 – again showing a sign reversal. *Other Store-Based Retailing* had a mean parameter estimate of 0.030 in Analysis 1-2 while the estimate in this analysis was 0.148. The industries of *Professional, Scientific and Technical Services*, *Public Administration* and to a lesser degree, *Financial Services*, were also estimated to have greater agglomeration effects using the Work Destination Zone dataset. There appeared, however, to be very little change in the elasticity estimates for the broader sectors of media services and health services. Once again, it should be noted that one cannot rely too heavily on these comparisons because not all industry subcategories within broader sectors had wage functions estimated in Analysis 1-2 and the aggregate-industry averages were not weighted by employment shares. Despite this, there seems to be reason to believe that elasticity estimates are slightly larger for some sectors when using work destination zone data, as was the case for Sydney.

Lastly, making use of the smaller spatial units lowered the adjusted R-squared values and raised the F-values of the industry regressions for Melbourne as well. As with the Sydney results, this was to be expected as the finer geographic scale and lessened effect from the averaging of employment data within spatial units would leave more variation in the data needing explanation. Additionally, the increased sample size would improve the precision of the parameter estimates and subsequently the overall significance of the models.

Table 5.1.3: Melbourne Regression Results Using WDZ Spatial Units (ED weighted by linear distance)

Ind ID	Industry Name	ε_U	S.E.	P-Value (F)	Adjusted R^2	N
11	Food Product Manufacturing	***0.096	0.035	5.50E-03	0.362	815
13	Textile, Leather, Clothing and Footwear Manufacturing	***0.120	0.034	5.00E-04	0.181	666
18	Basic Chemical and Chemical Product Manufacturing	**0.121	0.049	1.46E-02	0.357	315
21	Primary Metal and Metal Product Manufacturing	0.018	0.046	6.94E-01	0.140	358
22	Fabricated Metal Product Manufacturing	***0.139	0.039	4.00E-04	0.087	417
23	Transport Equipment Manufacturing	***0.138	0.034	5.79E-05	0.156	486
24	Machinery and Equipment Manufacturing	***0.124	0.027	6.99E-06	0.165	655
30	Building Construction	***0.174	0.014	7.66E-36	0.238	1484
31	Heavy and Civil Engineering Construction	***0.131	0.031	2.84E-05	0.164	457
33	Basic Material Wholesaling	***0.194	0.027	8.37E-13	0.301	767
34	Machinery and Equipment Wholesaling	***0.154	0.024	4.82E-10	0.246	677
41	Food Retailing	**0.041	0.020	4.10E-02	0.202	1202
42	Other Store-Based Retailing	***0.148	0.017	3.31E-18	0.183	1523
44	Accommodation	***0.155	0.023	5.48E-11	0.158	590
45	Food and Beverage Services	***0.118	0.016	1.88E-13	0.227	1503
54	Publishing (except Internet and Music Publishing)	***0.169	0.028	2.36E-09	0.141	440
55	Motion Picture and Sound Recording Activities	***0.200	0.053	2.00E-04	0.410	336
59	Internet Service Providers, Web Search Portals, and Data Processing Services	***0.170	0.048	5.00E-04	0.197	200
62	Finance	***0.133	0.019	1.76E-11	0.686	772
63	Insurance	***0.161	0.023	1.04E-11	0.374	397
64	Auxiliary Finance and Insurance Services	***0.176	0.017	1.93E-23	0.300	772
67	Property Operators and Real Estate Services	***0.135	0.019	4.59E-12	0.150	871
69	Professional, Scientific and Technical Services (except Computer System Design and Related Services)	***0.190	0.015	1.07E-34	0.243	1603
75	Public Administration	***0.181	0.014	7.16E-35	0.299	984
77	Public Order, Safety and Regulatory Services	0.016	0.020	4.16E-01	0.621	675
84	Hospitals	***0.125	0.024	2.81E-07	0.205	509
85	Medical and Other Health Care Services	***0.091	0.014	1.36E-10	0.152	1352
87	Social Assistance Services	***0.113	0.018	3.03E-10	0.184	1353
89	Heritage Activities	***0.111	0.046	1.76E-02	0.414	163
90	Creative and Performing Arts Activities	***0.154	0.047	1.10E-03	0.094	476

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

Sydney and Melbourne Results with Added Quadratic Terms

The inclusion of a quadratic term for U, being U^2, allowed a log-linear wage function model specification to estimate an elasticity of productivity that was non-constant.

As such, it essentially allowed the relationship between productivity and effective density to bend. There are many instances where one can conceive of a situation where non-constant returns to agglomeration are characteristic of some industries. Not only can dense locations be prone to congestion, but land rents are also affected by their accessibility, which can make these locations unattractive to some industries after a certain point. Many high-tech industries, for instance, are understood to prefer medium density locations that benefit from their proximity to other dense centres such as the CBD while still experiencing only moderate rents. Similarly, too much competition in the densest locations may lower the value of firm output.

The coefficients on variables for which quadratic terms are estimated cannot be interpreted directly, but rather have to be evaluated at some given level of input. Recalling equation 5B above in the overview of this analysis, taking the first-order derivative with respect to U no longer leaves the coefficient β as a result, but rather the following:

$$\frac{\partial \ln I_{k,m}}{\partial \ln U_m} = \hat{\beta}_{1k} + (2)\hat{\beta}_{2k} \ln U_m \qquad [5L]$$

This resulting equation, when evaluated at a given level of effective density (U_m), returns an elasticity estimate that is predicated and conditional on the level of U_m. The equation can be evaluated for all levels of effective density present in a dataset, thus providing a tailored elasticity estimate for marginal increases in effective density in each employment zone. There is some level of risk involved in such an exercise, however, as predictions will not be equally reliable at all levels of U. This effect is apparent in the variance formula for the predicted values of income indicated below, extended to the situation where a first order derivative has been taken to produce an elasticity estimate.

$$\varepsilon = \frac{\partial \ln I_{k,m}}{\partial \ln U_m} = \hat{\beta}_{1k} + (2)\hat{\beta}_{2k} \ln U_m \qquad \text{let } k = (2)\ln U_m \qquad [5M]$$

$$Var(\varepsilon) = Var\left[\hat{\beta}_{1k} + k\hat{\beta}_{2k}\right] \qquad [5N]$$

$$Var(e) = Var(\hat{\beta}_{1k}) + k^2 Var(\hat{\beta}_{2k}) + (2)kCov(\hat{\beta}_{1k}, \hat{\beta}_{2k}) \quad [5O]$$

The variance of an elasticity estimate is a function of not only the variances of the individual parameters, but also of the covariance between them. In the case of a linear term and its quadratic, the covariance will be high by design. This is especially the case with effective density, as it cannot take on a negative value, thus both its linear term and its square will always be positive. In this situation, the quadratic specification can only take the shape of half a parabola and correlation will always be strongly positive. Examining 5O, one can see that the value k, defined above in 5M, appears twice: once in its linear form and once in its squared form. As k increases with values of U, so will the estimated variance of the predicted elasticity values. In this sense, variance is scaled up and down by the values taken on by effective density. In fact, the variance of an estimated elasticity will be at its smallest when $U = 1$, which will never be the case as effective density values tend to be quite large. This is something to keep in mind when using the non-constant elasticity formula to predict elasticities in highly dense locations.

Quadratic specifications, such as the one that appears here, are most commonly evaluated at mean values of x, or in this case $\overline{U}$. This does not provide a mean elasticity value, but merely an elasticity estimate evaluated when U_m is at its mean. To keep in alignment with convention, all reported elasticities in this section have been evaluated using mean values of effective density. To specify further, mean effective densities were calculated on an industry-specific basis as industries vary in terms of the employment zones in which they are located. As such, the mean effective density values are particular to each industry.

Table 5.1.4 presented below summarizes the outputs from the Sydney regressions with the linear specification of effective density and its squared term. β_1 represents the coefficient on the former and β_2 the coefficient on the latter, while ε_U denotes the estimated elasticity of productivity with respect to effective density. The elasticity estimates do not, and should not, deviate much from the former specification since each elasticity was evaluated at each industry's mean level of effective density. A standard error is not given with the elasticity estimates because they are derived from functions including estimates of β_1 and β_2, thus their significance depends on the joint significance of the two parameters. As evident by the very few industries with significant coefficients (as indicated by the asterisks), first inspection may lead one to infer that effective density no longer is a useful explanatory variable of industry income levels. In fact, only 11 of the 30 industries returned with any significant effective density parameter at all. This, however, can readily be explained because the squared value of ED will naturally be highly correlated with ED by construction. In such a situation, OLS can have troubles identifying the separate contributions of $\ln U$ and $(\ln U)^2$ and thus report them both as being insignificant. This does not cause concern for predicting levels of income, however, because it could never be the case that the two terms could not be correlated. To test whether both variables were in fact insignificant and the effects of agglomeration were negligible on these industries, a test of joint-significance was carried out.

The test of joint-significance was similar to the Chow test carried out in the pooled industry test of Analyses 2 and 3, insofar as it entailed carrying out an F-test and comparing two model specifications – one restricted and one unrestricted. The restricted equation was one that set each parameter in question equal to zero[26], namely β_1 and β_2. The null hypothesis of this test was that there was no difference between the two specifications. With a confidence level set at 0.10, the null was rejected in favour of the alternative hypothesis if the test's p-value was less than this critical value, leaving a 10% chance of incorrectly accepting the significance of the restricted model.

Carrying out this process on all the industry wage functions in Sydney led to the rejection of the null for 26 of the 30 industries, all of which were significant in the constant-elasticity model specification except for *Heritage Activities*. Apart from this industry, the only aspects of these results that differed from those previously stated with constant elasticities were that these reported elasticities that differed slightly. Along with this were minor changes in the F-values and adjusted R-squared values. The differences arose because in this specification any line plotted through the data-points would be allowed to bend slightly to accommodate any curvature in the trend of income with respect to changes in effective density. As aforementioned, however, confidence intervals around elasticity estimates derived from evaluating 5M will vary in accordance to the values specified for effective density.

[26] This can be done in an econometrics software package by inputting the linear restrictions as b[ED] = 0 and b[ED_sq] = 0, where the terms in the brackets are the column names attributed to the relevant variables.

When examining the sign of the coefficients on β_2, 14 of the industries were shown to be negative, suggesting that they may be experiencing diminishing returns to agglomeration. This means that there are positive effects of agglomeration for these industries, since none of their resulting elasticities are negative, but the magnitude of the benefit lessens with the scale of economic density. The remaining 16 industries with positive signs on their estimates of β_2 would suggest that they are experiencing increasing returns to scale. In other words, the benefits of agglomeration increase with economic scale in these industries. Considering the significance of these parameters on $(\ln U)^2$, however, could involve omitting this variable in the case of most industries. The ones that returned a negative and significant $(\ln U)^2$ coefficient were *Basic Chemical and Chemical Product Manufacturing*; *Transport Equipment Manufacturing*; *Accommodation* and *Public Order, Safety and Regulatory Services*, although the latter industry did not show the two parameters to be jointly significant. Industries that returned a positive and significant coefficient on $(\ln U)^2$ were *Food and Beverage Services*; *Finance*; *Professional, Scientific and Technical Services*; *Public Administration*; *Hospitals*; and *Social Assistance Services*.

Because for many of the results, the significance of the two parameters was confounded to some extent – a likely result of multicollinearity – it was preferable to keep to the constant elasticity specification in circumstances where the parameter estimate of the square of effective density was shown to not significantly add to the explanatory power of the model. While there may have been some reason to keep insignificant squared-term parameters in the model, as in the circumstances where overall adjusted R-squared values and F-values were improved and joint-significance was confirmed, the error terms resulting from the covariance of two insignificant and highly correlated variables would have been large. Following this, one could conclude from the results that the 18 industries to not report significant coefficients on the square of effective density were characterized by constant returns to scale (i.e. non-varying elasticity). Table 5.1.4 below reports on all of the industry results for Sydney where the industries reported in grey text did not have jointly significant U_m and U_m^2 parameters. As noted above, these industries aligned with those in Table 5.1.1 that received insignificant effective density coefficients except for *Heritage Activities*, which became significant with the addition of a squared effective density term that was jointly significant with the untransformed effective density variable.

Table 5.1.4: Sydney Regression Results with Added Squared ED Term (ED weighted by linear distance)

Ind ID	Industry Name	ε_U	β_1	S.E.	β_2	S.E.	P-Value (F)	Adjusted R^2	N
11	Food Product Manufacturing	0.023	-1.151	0.747	0.049	0.031	3.78E-55	0.309	740
13	Textile, Leather, Clothing and Footwear Manufacturing	0.177	-0.873	1.016	0.044	0.042	1.83E-28	0.220	588
18	Basic Chemical and Chemical Product Manufacturing	0.172	**2.269	1.076	*-0.087	0.045	1.62E-24	0.324	326
21	Primary Metal and Metal Product Manufacturing	0.026	1.241	0.941	-0.051	0.039	7.32E-23	0.296	341
22	Fabricated Metal Product Manufacturing	0.011	-1.007	0.743	0.043	0.031	6.05E-08	0.104	392
23	Transport Equipment Manufacturing	0.095	**2.662	1.107	**-0.108	0.047	1.01E-39	0.407	386
24	Machinery and Equipment Manufacturing	0.063	-0.781	0.624	0.035	0.026	2.96E-37	0.249	658
30	Building Construction	0.167	0.015	0.317	0.006	0.013	3.97E-57	0.157	1654
31	Heavy and Civil Engineering Construction	0.173	0.863	0.624	-0.029	0.026	1.11E-29	0.285	453
33	Basic Material Wholesaling	0.103	-0.338	0.547	0.019	0.023	8.33E-44	0.263	721
34	Machinery and Equipment Wholesaling	0.088	0.540	0.675	-0.019	0.028	4.91E-30	0.189	733
41	Food Retailing	0.087	0.053	0.382	0.001	0.016	3.15E-42	0.147	1327
42	Other Store-Based Retailing	0.137	0.192	0.277	-0.002	0.011	2.68E-77	0.200	1687
44	Accommodation	0.121	**0.698	0.341	*-0.024	0.014	1.46E-34	0.213	732
45	Food and Beverage Services	0.065	***-0.867	0.237	***0.039	0.010	5.00E-141	0.329	1694
54	Publishing (except Internet and Music Publishing)	0.252	*1.300	0.678	-0.043	0.027	2.13E-26	0.202	603
55	Motion Picture and Sound Recording Activities	0.216	1.341	0.963	-0.046	0.039	4.21E-59	0.496	430
59	Internet Service Providers, Web Search Portals, and Data Processing Services	0.130	0.209	0.853	-0.003	0.034	6.65E-12	0.249	220
62	Finance	0.149	*-0.593	0.311	**0.031	0.012	4.00E-104	0.403	979
63	Insurance	0.137	-0.082	0.537	0.009	0.021	4.37E-50	0.403	487
64	Auxiliary Finance and Insurance Services	0.155	-0.197	0.407	0.015	0.016	2.66E-56	0.260	925
67	Property Operators and Real Estate Services	0.121	-0.266	0.311	0.016	0.013	3.63E-64	0.243	1136
69	Professional, Scientific and Technical Services (except Computer System Design and Related Services)	0.139	**-0.486	0.218	***0.026	0.009	8.70E-171	0.354	1863
75	Public Administration	0.127	**-0.657	0.330	**0.033	0.013	5.17E-95	0.359	1036
77	Public Order, Safety and Regulatory Services	0.024	*0.848	0.477	*-0.034	0.019	1.62E-25	0.138	882
84	Hospitals	0.107	***-1.603	0.515	***0.072	0.022	1.89E-29	0.275	471
85	Medical and Other Health Care Services	0.087	0.422	0.300	-0.014	0.012	8.92E-52	0.154	1538
87	Social Assistance Services	0.144	***-0.843	0.324	***0.041	0.013	1.91E-58	0.165	1601
89	Heritage Activities	0.081	0.957	0.964	-0.036	0.040	5.22E-06	0.157	194
90	Creative and Performing Arts Activities	0.220	0.585	0.906	-0.015	0.037	3.95E-54	0.390	546

Note (1): * = Significant at 0.10   ** = Significant at 0.05   *** = Significant at 0.01

Note (2): Industries where the ED parameters are not jointly significant are indicated by their grey text. All industries reported in black text have jointly significant parameters at the 0.05 level.

The Melbourne results for industry regression models specified with the quadratic term of effective density are presented below in Table 5.1.5. A similar situation arose here as in Sydney where many industries that previously reported significant effects from agglomeration did not once $(\ln U)^2$ was added to the equation. The results were such that 12 industries reported significant squared effective density terms. Only one significant negative coefficient on the square of effective density was estimated from all of the industries, which was for *Internet Service Providers, Web Search Portals and Data Processing Services*. Industries to report positive and significant quadratic coefficients were *Basic Chemical and Chemical Product Manufacturing*; *Fabricated Metal Product Manufacturing*; *Building Construction*; *Basic Material Wholesaling*; *Food and Beverage Services*; *Finance*; *Insurance*; *Professional, Scientific and Technical Services*; *Hospitals*, *Medical and Other Health Care Services*; and *Social Assistance Services*. Six of these industries appeared in the list of industries that were estimated with positive and significant coefficients on the squared term of effective density in Sydney. It is difficult to say whether the industries that recurred here did so by mere chance, or if they affirm the presence of increasing returns to scale in these industries. Similarly, it is difficult to say whether the number of industries that did not share the same sign on the coefficient of effective density squared between the two cities, yet showed the terms to be significant, is an indication of incorrectly conferring their significance because of estimation error or because of differences inherent in the industries' activities in the two cities. Nevertheless, the results suggest that increasing returns to agglomeration are being experienced by 11 industries in Melbourne as indicated by the presence of positive and significant coefficients on the square of effective density. Decreasing returns are being experienced by one industry while constant returns to agglomeration characterize the 18 remaining.

A test of joint-significance for the effective density variable and its squared term revealed that all the industries that reported significant effects from effective density in Table 5.1.3 also did so with the addition of ED squared in Table 5.1.5, except for *Food Retailing* which became insignificant. In the case of this industry, a slight reduction in the adjusted R-squared value, a slightly larger F-value and the insignificance of the squared effective density parameter all suggest that the former model specification that excluded the squared-term was better. As such it can be maintained that the effects of agglomeration in Melbourne are positively influencing this industry.

The magnitude of the elasticity estimates were similar to those produced by the models without the squared terms, but in general the results are somewhat smaller with their inclusion. Considering only the industries for which the square of effective density was found to significantly improve the model, the largest elasticity decrease was for the industry of *Insurance*, with an estimated elasticity difference of -0.065. The largest increase was for *Internet Service Providers, Web Search Portals and Data Processing Services* with a difference of 0.045. Overall, adjusted R-squared values were improved in 20 of the 30 industries by this specification and the overall significance of the industry models was improved in 16 of the 30 industries. The preferred results for Sydney and Melbourne, however, were selected on the basis that constant elasticities would be reported unless the squared effective density parameter was shown to significantly contribute to a model's significance. This final tabulation of preferred results is available for reference in Appendices D and E.

Table 5.1.5: Melbourne Regression Results with Added Squared ED Term (ED weighted by linear distance)

Ind ID	Industry Name	ε_U	β_1	S.E.	β_2	S.E.	P-Value (F)	Adjusted R^2	N
11	Food Product Manufacturing	0.098	1.318	0.943	-0.053	0.041	8.21E-76	0.366	815
13	Textile, Leather, Clothing and Footwear Manufacturing	0.132	0.528	0.901	-0.017	0.038	1.31E-25	0.180	666
18	Basic Chemical and Chemical Product Manufacturing	0.071	*-2.447	1.253	**0.108	0.053	3.36E-27	0.361	315
21	Primary Metal and Metal Product Manufacturing	-0.024	-0.986	1.323	0.042	0.057	5.02E-09	0.128	358
22	Fabricated Metal Product Manufacturing	0.140	*-1.690	1.012	*0.080	0.044	1.34E-08	0.106	417
23	Transport Equipment Manufacturing	0.152	-0.012	0.839	0.007	0.036	1.18E-15	0.157	486
24	Machinery and Equipment Manufacturing	0.128	0.461	0.741	-0.014	0.031	9.55E-23	0.165	655
30	Building Construction	0.153	-0.304	0.288	*0.020	0.012	1.70E-109	0.300	1484
31	Heavy and Civil Engineering Construction	0.125	-0.360	0.665	0.021	0.028	1.01E-17	0.185	457
33	Basic Material Wholesaling	0.181	-0.607	0.465	*0.034	0.020	1.63E-76	0.387	767
34	Machinery and Equipment Wholesaling	0.136	-0.012	0.543	0.006	0.022	6.62E-39	0.252	677
41	Food Retailing	0.034	0.060	0.422	-0.001	0.018	5.72E-54	0.199	1202
42	Other Store-Based Retailing	0.147	0.070	0.376	0.003	0.016	8.82E-63	0.183	1523
44	Accommodation	0.155	0.179	0.554	-0.001	0.023	3.52E-19	0.156	590
45	Food and Beverage Services	0.058	***-1.849	0.353	***0.083	0.015	1.19E-85	0.242	1503
54	Publishing (except Internet and Music Publishing)	0.168	0.072	0.629	0.004	0.027	4.14E-12	0.137	440
55	Motion Picture and Sound Recording Activities	0.177	-0.306	1.187	0.020	0.050	1.89E-35	0.413	336

59	Internet Service Providers, Web Search Portals, and Data Processing Services	0.219	*2.879	1.561	*-0.111	0.063	1.06E-09	0.228	200
62	Finance	0.111	-0.664	0.460	*0.033	0.019	1.70E-225	0.751	772
63	Insurance	0.095	**-1.669	0.744	**0.075	0.030	2.25E-28	0.308	397
64	Auxiliary Finance and Insurance Services	0.165	-0.172	0.490	0.014	0.020	8.67E-58	0.310	772
67	Property Operators and Real Estate Services	0.129	-0.097	0.452	0.010	0.019	2.46E-31	0.167	871
69	Professional, Scientific and Technical Services (except Computer System Design and Related Services)	0.163	*-0.513	0.270	***0.029	0.011	1.10E-124	0.312	1603
75	Public Administration	0.160	-0.396	0.366	0.024	0.015	1.48E-75	0.313	984
77	Public Order, Safety and Regulatory Services	0.019	-0.137	0.507	0.007	0.021	3.00E-136	0.622	675
84	Hospitals	0.113	*-1.055	0.599	**0.050	0.025	1.52E-23	0.213	509
85	Medical and Other Health Care Services	0.065	*-0.613	0.323	**0.029	0.013	4.77E-50	0.168	1352
87	Social Assistance Services	0.106	-0.700	0.432	*0.035	0.019	5.59E-55	0.182	1353
89	Heritage Activities	0.085	-1.165	1.010	0.053	0.043	7.43E-12	0.323	163
90	Creative and Performing Arts Activities	0.154	-0.429	1.070	0.025	0.046	4.93E-08	0.085	476

Note (1): * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

Note (2): Industries where the ED parameters are not jointly significant are indicated by their grey text. All industries reported in black text have jointly significant parameters at the 0.05 level.

5.1.2 Analysis 2-1 Conclusions

In this section, industry-specific agglomeration elasticities were estimated with datasets based on the spatial units of the Travel Zone in Sydney and the Work Destination Zone in Melbourne. The use of the more finely geographically disaggregated datasets for the city wage function estimates greatly improved the precision of the industry-specific effective density parameters estimated. This was reflected in that only five industries in Sydney and two industries in Melbourne returned insignificant productivity elasticities with respect to the linear specification of effective. In addition to the greater efficiency in the parameter estimates, another effect of using smaller spatial units was that the industry-specific models were able to explain less of the variation in income as reflected in the reduction of the adjusted R-squared values. The most likely explanation for this is that since the census data for income are average over spatial units, the use of SLAs would have a much greater normalizing effect than when using employment zones. This means there is less variance in the SLA needing to be explained. The more precise parameter estimations with the use of employment zone data, however, meant that the overall significance of the models was improved as indicated by smaller F-values.

While the aggregation of industry classifications between Analyses 1-2 and 2-1 differed, which made direct comparison between their results an imprecise endeavour, general observation and computing broad-sector averages from the 3-digit results suggests that using a dataset with finer geographic detail affects results for some industries more than for others. This was affirmed by the same trends emerging in the Melbourne results as in the Sydney results. Using employment zones seems to have had a marginal effect on the estimates for medical-sector industries for instance, while the greatest effects seemed to have been on the elasticity estimates for retail. The general magnitudes of the elasticities, however, remained in concordance with existing studies insofar as manufacturing, construction, professional and financial services are concerned. The same cannot be said for retail, medical, government administration and food services because to my knowledge these industries have not had the effects of agglomeration on their productivity investigated in existing publications. Their elasticity estimates do, however, correspond to theory and meet expectations.

The effect of including a squared-term of the effective density variable was investigated to gain some understanding of whether the effects of agglomeration become more pronounced, less pronounced, or remain constant with increasing levels of effective density. These effects are synonymous with terms of increasing returns, diminishing returns, and constant returns to scale respectively in economic theory. Because the squared term was derived from another variable in the model, namely effective density, the effect of multicollinearity was strong and very evident. This resulted in both parameters being estimated with large errors and a lack of statistical significance of effective density in many industries. While joint-significance tests for the two variables confirmed their relevance in explaining spatial variations in income in the industries where insignificant results cast doubt, it was concluded that allowing the effects of agglomeration to vary with effective density did not contribute to the explanatory power of the models in a significant way. This most likely means that if the elasticities do vary, then they are likely to do so only in a minor way such that the existence of a variable effect cannot be proven beyond a statistically predetermined level of doubt. This was the case for 20 industries in Sydney and 18 industries in Melbourne, for which constant returns to scale were concluded. The industry to show the most significant impact from the inclusion of the square-term for both Sydney and Melbourne was *Food and Beverage Services*. The point estimate evaluated at a mean level of effective density was approximately 0.06, which suggests a rather small but significant productivity effect from agglomeration that is logical for this sector. Further, this relationship is estimated to increase in a pronounced way with increases in effective density: evident in clustering of these activities into food courts and restaurant strips in dense areas. Complete tables of the preferred results, assembled from the results in this section's analyses are provided in Appendices D and E.

Up to this point, none of the models specified have addressed the issue of endogeneity, that is, the concern that the possible presence of a two-way causal relationship between agglomeration and productivity may be having a biasing effect on the elasticities being estimated. This will be considered next in the following section, Analysis 2-2.

5.2.0 Analysis 2-2 Overview: Addressing endogeneity using WDZ data

The presence of endogeneity in a model specification means that one or more of the independent/explanatory variables in a model is being determined by factors within and as such is not exogenous. Endogeneity can also arise in the circumstance that an omitted variable is having a concurrent effect on the dependent and an independent variable. Thus, endogeneity can be the result of reverse-causality (simultaneity) or omitted variable bias in cross-sectional analyses. The situation here is that effective density may not be exogenous, meaning that its values are possibly being determined from influences outside of the system. More specifically, this effect is hypothesized to come from the productivity measure. If this holds then dense locations not only increase productivity for incumbent firms, but productive areas also act as attractive places for firms to locate and as such a feedback mechanism would be operating from income back to effective density. In such a circumstance, the result would be that effective density will be correlated with the error term of the model and the expected value of the error will no longer be zero, thus creating a bias. This would generate inconsistent effective density parameter estimates, as asymptotically their deviations from the true values would not converge on zero, but some other value. The direction of this bias would be in the direction of the sign on the covariance of the dependent and endogenous explanatory variables.

The issue of simultaneity bias is most commonly addressed by using two-stage least squares (2SLS). This procedure first involves estimating the suspected endogenous variable in an auxiliary regression as a function of an instrumental variable and all other independent variables in an original model specification. If the original specification is the one given in 5A, then the reduced form model would appear as indicated below.

$$\ln U_m = \alpha_{2k} + \beta_2 Inst + \sum^{i} \delta_{2i} \ln Occ_{i,k,m} + \sum^{j} \gamma_{2j} \ln Edu_{j,k,m} + \psi_2 \ln Exp_{k,m} + e_{2k,m} \qquad [5P]$$

where '$Inst$' represents the instrumental variable and the subscript '2' denotes a second set of coefficients that will differ from those estimated in 5A. A reduced form equation is one that expresses an endogenous variable as a function of all exogenous variables. In 2SLS the reduced form equation is used to estimate $\ln U_m$, which can be represented as $\ln \hat{U}_m$ and replace $\ln U_m$ on the right-hand side of the original structural equation. Thus industry-specific productivity is estimated by a two-stage process, hence the origin of the name 'two-stage least squares'. The model being estimated then becomes a variant of 5A and is as indicated below.

$$\ln I_{k,m} = \alpha_k + \beta_1 \ln \hat{U}_m + \sum^{i} \delta_i \ln Occ_{i,k,m} + \sum^{j} \gamma_j \ln Edu_{j,k,m} + \psi \ln Exp_{k,m} + e_{k,m} \qquad [5Q]$$

The challenge of such a procedure is to find an instrument that correlates strongly with effective density, but not with the error term. Long-lagged versions of the endogenous variable are most commonly used and the justification in this context, for instance, is that historic population numbers should be highly correlated with effective density through a legacy to current population patterns but not bear any present-day effect on productivity levels. Other instruments have been used in the literature as well, such as bedrock density (Rosenthal and Strange 2008) or the proximity to the coast or a railway line (Ciccone and Hall 1996). The most popular instrument is one based on historic population levels, but in this resides a challenge for an Australian analysis, as finding this sort of data lagged far enough back in time is very difficult or nearly impossible. Ciccone and Hall (1996) and Rice and Venables (2004) use population levels in the mid-19th century obtained from census data, yet the most temporally distant employment data that could be found in the Australian census data was from the 1970s. Even if this data was available, its use poses a challenge as allocating population numbers to spatial units that have had their boundaries changed significantly over time, or to spatial units that had not existed some number of years in the past, can be a complicated task.

An alternative to using these instruments, which is utilized here in the Australian industry regressions, is the geographic size of the units at which the datasets are being aggregated. The premise here is that the units have had their boundaries set at some point in time in the past and have remained relatively consistent over time. While this may not entirely hold true for Sydney and Melbourne's employment zones, the rationale for their use is still maintained.[27]

An additional specification of an instrumental variable that is tested in this section is the total road network density (local plus main roads) of Melbourne's work destination zones.[28] Here the premise is that a base level of road infrastructure is established and rooted in the past to service historical population levels, yet once areas become more developed it becomes rather difficult to increase the number of routes within them. An example of this is a central business district that might experience redevelopment in the form of higher or more ubiquitous high-rise development, yet the opportunities to expand road network capacity in the city do not increase much over time. Thus, we may see a strong correlation between road network density and effective density, but not necessarily with changing productivity levels. Using road network density as an instrumental variable means that the values that geographic units take on will not change with increasing road capacity in the form of adding lane kilometres, only with the addition of new routes, because adding lanes does not affect total route length. For this reason, in addition to the difficulties of adding road capacity to well-established areas, this measure should be fairly robust over time and

---

[27] In Melbourne, minor adjustments to zone boundaries may be made in periods leading up to new census years. Boundaries are set on the basis that a zone contains at least 100 employees and that the delineation of boundaries geographically makes sense in terms of physical features such as waterways that may restrict access. Furthermore, zone boundaries are also set to reflect the incumbency of differing land-uses. Other than these criteria that are unlikely to cause significant boundary changes over time, changes may also occur to extract more census information from particular employment centres of interest. These changes are more likely to be in the form of subdividing existing areas rather than moving boundaries, per se.

[28] In actuality, network density of main roads and network density of main roads plus local roads were both trialled. The former specification failed to show a strong correlation with effective density, largely because many zones had no main roads at all. For this reason, discussion is restricted to the latter specification.

consequently could perform well as an instrument. Furthermore, productive areas may increase public transport service levels as they growth, but this too would not be reflected in the road centreline data used for creating this instrument. This figure, road network density, was calculated in Postgres for each employment zone in Melbourne. The script to carry this out is discussed below.

Calculating Road Network Density in Postgres

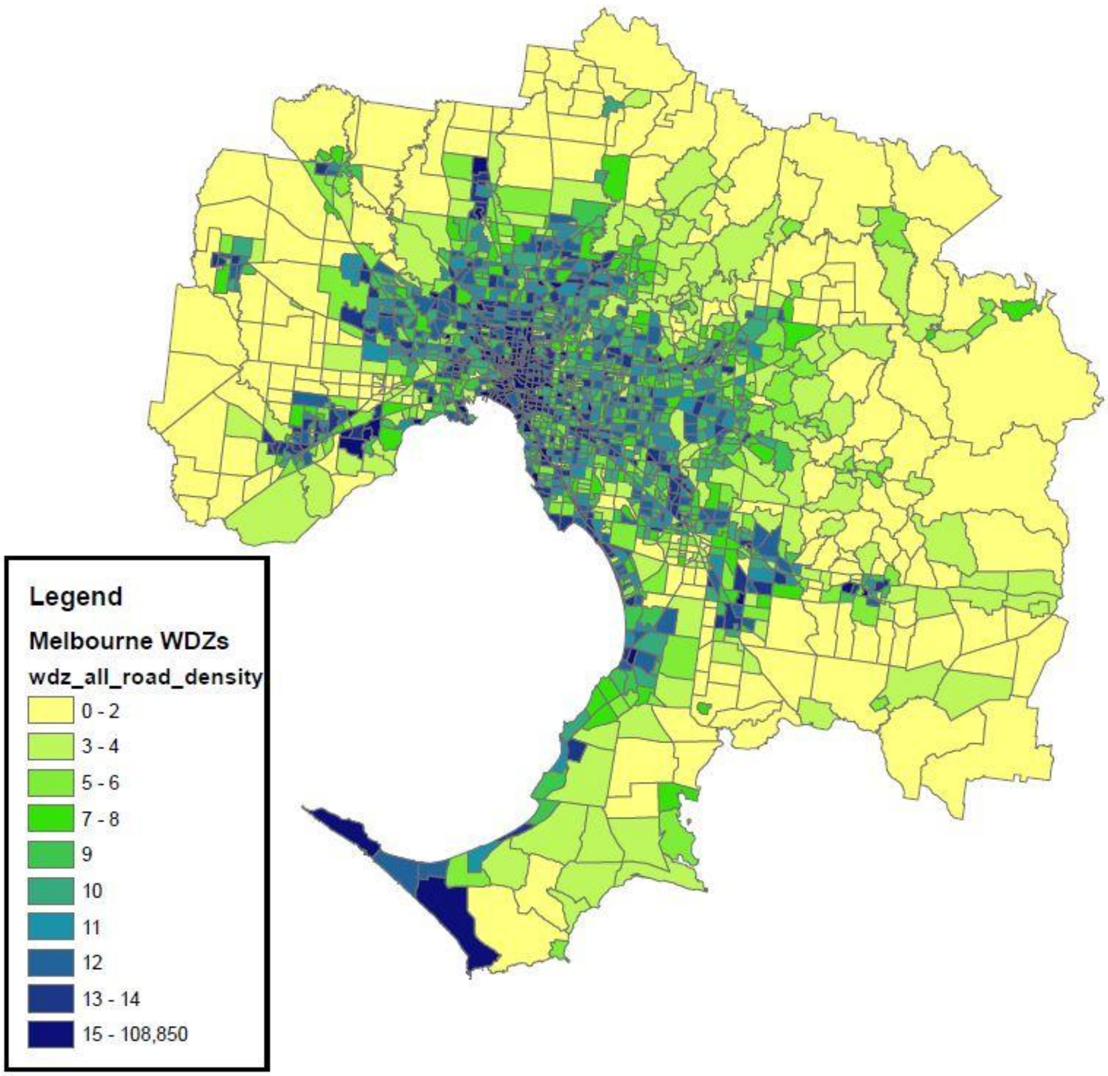
Importing the total road network spatial layer into Postgres for the city of Melbourne followed the same procedure as importing the employment zone layers described in Analysis 2-1 and as such will not be explained here. Once the layer was imported, a new column was added to the each of the Melbourne WDZ and Sydney TZ spatial tables titled 'total_all_road'. This provided a location for the results from a script to be stored that was written to calculate total road length within each employment zone. This script, written for Melbourne, is shown below.

```
UPDATE shp.melb_wdz AS wdz SET total_all_road = temp.roads_km FROM (          [5R]
        SELECT w.vicdznp06, sum(ST_Length(r.the_geom))/1000 AS roads_km
        FROM shp.melb_wdz AS w, shp.melb_all_roads AS r
        WHERE ST_Contains(w.the_geom, r.the_geom)
        GROUP BY w.vicdznp06
        ORDER BY w.vicdznp06 ASC
        ) AS temp
        WHERE wdz.vicdznp06 = temp.vicdznp06;
UPDATE shp.melb_wdz SET total_all_road = 0 WHERE total_all_road IS NULL;
```

The code was written to create a temporary table with a number of columns, including one titled 'roads_km' to temporarily store the data to be used to update the column 'total_all_road'. The central part of the entire code was the 'ST_Contains' function after the 'WHERE' condition that required that the length of a linestring geometry was only to be measured and counted insofar as it coincided with a particular employment zone. Once this script was executed and the 'total_all_road' column populated, an additional column was added to an employment zone spatial table to store the results of a final road-density calculation. In the case of Melbourne, the script for this final calculation appears below as equation/script 5S, followed by a thematic map of road network density indicated in Figure 5.2.1. As one can see by comparing Figures 5.1.2 and 5.2.1, road network density patterns mirror effective density patterns rather closely and as such, road network density is likely to correlate well with effective density and be relatively strong as an instrumental variable.

```
UPDATE shp.melb_wdz AS wdz SET wdz_all_road_density = wdz.total_all_road        [5S]
        / (ST_area(wdz.the_geom)/1000000);
```

Figure 5.2.1: A Thematic Map of Road Network Density in the Melbourne SD



These were the steps taken to prepare employment zone road network density as an IV to be applied in the 2SLS regressions. Preparing employment zone area as an IV was a significantly less involving process than the construction of this script. A simple script command using the 'ST_Area' function was used to calculate the areas of the employment zones and prepare them in a new column. Once the IVs had been prepared and exported along with the other wage and effective density data, industry-specific wage-functions using 2SLS were estimated.

When carried out in an econometric software package, conducting a 2SLS analysis typically results in the provision of outputs from two tests: the Hausman test and the Weak

Instrument test. The former is a test of whether 2SLS is even required, thus testing the null hypothesis of whether least squares (LS) estimates are consistent. More specifically, the Hausman test checks for a correlation between an explanatory variable and an error term, assuming a null hypothesis of $Ho:\text{cov}(U_m,e_m)=0$ and an alternative hypothesis of $H_1:\text{cov}(U_m,e_m)\neq 0$. A rejection of the null hypothesis suggests that LS estimates are inconsistent and endogeneity is present. Manually, this test can be carried out by first estimating a reduced form model and obtaining the residuals, then including the residuals in with the predicted values of a suspected endogenous variable in the estimation of the original model. If the residuals are significant, then endogeneity can be deemed to be present. The alternative is to carry out the test in an econometrics software package, which is likely to examine the differences between the LS and 2SLS estimates. In the circumstance that the null is true, the difference between the parameter estimates should converge on zero in large samples and in the case that the null is rejected, the difference should converge on some value other than zero. When the null is true, both analyses will produce consistent estimates but LS will be more efficient.

The Weak Instrument test also provides valuable output. If a weak instrument is chosen, its use in 2SLS can generate estimates with large biases and standard errors resulting in far worse estimates that those provided by LS. The test is carried out by first estimating a reduced model, where the endogenous variable U_m is once again specified as a function of the instrumental variable and all other exogenous variables. The Weak Instrument test examines the strength of the relationship between U_m and the instrument after the influence of all other variables has been accounted for. The null hypothesis of the coefficient on the instrumental variable being equal to zero must be 'soundly rejected' to prevent large biases from arising. The general rule of thumb applied to this criterion is that a weak instrument will take on an F-statistic of less than 10 or t-statistic of less than 3.3.[29] The results of a Weak Instrument test, when conducted in most econometrics software packages, are accompanied by criteria by which to select a tolerable level of bias in the 2SLS results. Figure 5.2.2 below gives an example of this output for the industry of *Food Product Manufacturing* in Sydney.

---

[29] See Stock and Watson (2003), *Introduction to Econometrics*, or Hill, Griffiths et al. (2007), *Principles of Econometrics* for further information on this issue.

Figure 5.2.2: An Example of a Weak Instrument Test Result for Food Product Manufacturing in Sydney

```
Weak instrument test -
  First-stage F-statistic (1, 731) = 1686.5
  Critical values for desired TSLS maximal size
  tests at a nominal 5% significance level:

     size       10%       15%       20%       25%
    value     16.38      8.96      6.66      5.53

  Maximal size is probably less than 10%
```

These criteria, developed by Stock and Yogo (2005), are provided because a weak instrument introduces a certain level of bias into the 2SLS results. The criteria enable the researcher to set his/her own level of tolerance for this bias, which arises because an instrument is an imperfect measure of an endogenous variable. The way these criteria are interpreted is that 'size' is the percentage of relative 2SLS bias and 'value' is a corresponding first-stage F-value required for the bias to be no greater than a selected acceptable amount. To specify further, these criteria are set at a 5% significance level, thus establishing that the criteria will hold in 95% of sample cases.

The challenge with instrumental variable estimation is that correlation between the endogenous variable and the instrument is a double-edged sword: high correlation is very desirable to avoid suffering the hazards of weak instruments and producing inconsistent and biased parameter estimates, but this also means that a strong instrument may not be uncorrelated with the error term in the model and thus instrumentation will have no benefit. Unfortunately, an investigation into the validity of instruments requires a greater number of instruments than there are endogenous variables. Even if this condition can be satisfied, a test of validity does not allow conclusions to be made on which instruments are valid, but simply tests the hypothesis that all instruments are valid with the assumption that at least one instrument is exogenous. A situation in which all instrumental variables are endogenous will not necessarily be detected. The situation that arises is that performing the Hausman test to discover if a variable in question is in fact endogenous, or causing an endogeneity problem, will only be effective if the instrument is valid, but testing for instrument validity is only possible when a surplus of IVs are available and at least one is known to be exogenous. If a Sargan over-identification test for instrument validity does not reject the null that the instruments are valid, this does not necessarily mean that this is actually the case. Hence, the importance of knowing that at least one IV is valid to make the test trustworthy. Essentially, the validity test is only effective at determining the validity of IVs additional to the one already established to be valid. In the case of this analysis where road network density and employment zone area are used as IVs, if neither are valid then this cannot be detected in a Sargan over-identification test and if only one is invalid then it cannot be confirmed which one. In the latter circumstance, determining which instrument is invalid becomes an endeavour that is sensitive to the interpretations of the Hausman and Weak Instrument tests. These issues will be explored further during the discussion of the results of the investigation into whether endogeneity is in fact affecting the results generated by Analysis 2-1.

5.2.1 Analysis 2-2 Results

Sydney and Melbourne Results

In this section, the results are reviewed for instrumental variable estimations of industry-specific wage functions for the cities of Sydney and Melbourne. The reason for carrying these out is to investigate which industries, if any, may be having their productivity elasticities biased by the effects of a potential endogeneity issue resulting from the reverse-causality between a location's effective density and level of productivity. For the city of Melbourne, two-stage least squares (2SLS) was applied with robust standard errors to estimate industry-specific wage functions while separately utilizing Work Destination Zone area and total road network density as well as their log transformations as instrumental variables. Thus, four 2SLS regressions were run on each of the 30 industries in Melbourne. For Sydney, 2SLS is carried out while utilizing only travel-zone area and its log transformation as instruments, thus the results of two 2SLS regressions are reported for each industry. The reason for trialling fewer instruments in the case of Sydney is that main roads and local roads centreline data were only obtained for the city of Melbourne. The complete sets of results for these regressions can be viewed in Appendices F and G for Sydney and Melbourne respectively.

Instrumented industry-specific wage functions were first estimated for the city of Melbourne because the availability of road-network spatial data enabled the extra instrument to be administered in the 2SLS regressions and thus their comparative performance could be evaluated. As such, the Melbourne results will be discussed first, followed by an account of those for Sydney.

Of the four instrument specifications used in Melbourne, by far the strongest performer in every industry was the natural log of Work Destination Zone area, as indicated by the large F-statistics on the Weak Instrument tests, which ranged from 206 to 1382 – all in excess of the minimum recommended threshold of 10. Interestingly, the one to correlate the least, again in the instance of every industry, was the untransformed geographic area of work destination zones that had F-values ranging from 4 to 78. In fact, this IV specification was the only one to report values below the critical value of 10, which it did in 13 of the 30 industries analysed. Consistently across all industries but one, the second strongest IV was the untransformed version of road-network density with F-statistics ranging from 71 to 547, with log-transformed road-network density placed third with values ranging from 55 to 381.

Reviewing the results for the Hausman test for exogeneity for the effective density variable, evidence of an endogeneity bias was detected in 14 industries when using the IV of log-transformed road network density, 10 industries using untransformed road network density, only 5 industries using log-transformed WDZ area, and 6 industries using untransformed WDZ area. With the high correlation between effective density and log-transformed WDZ area reflected in the Weak Instrument test results and the low detection rate of endogeneity bias, it possible that its use as an IV has low validity. The critical value used for the Hausman test was 0.10, which means that we would expect the null to be incorrectly rejected for 3 out of the 30 industries if endogeneity was not affecting GLS estimates. This number is not far off the actual number detected, being 5, which suggests that endogeneity may not be systematically present in the industry wage function estimations but arise because of estimation error. Table 5.2.1 shown below offers a summary of the results where endogeneity was detected by Hausman test p-values below 0.10.

Table 5.2.1: Checklist for Detection of Reverse-Causality for the 4 IV Specifications

	Instrumental Variable			
Industry Name	Ln(Road Density)	Road Density	Ln(WDZ Area)	WDZ Area
11 Food Product Manufacturing	✓	✓	✓	
13 Textile, Leather, Clothing and Footwear Manufacturing	✓		✓	
18 Basic Chemical and Chemical Product Manufacturing	✓	✓		
21 Primary Metal and Metal Product Manufacturing	✓	✓		
22 Fabricated Metal Product Manufacturing				
23 Transport Equipment Manufacturing	✓	✓		✓
24 Machinery and Equipment Manufacturing				
30 Building Construction	✓			
31 Heavy and Civil Engineering Construction				
33 Basic Material Wholesaling				
34 Machinery and Equipment Wholesaling	✓	✓		
41 Food Retailing				
42 Other Store-Based Retailing	✓	✓		
44 Accommodation	✓			
45 Food and Beverage Services			✓	✓
54 Publishing (except Internet and Music Publishing)				
55 Motion Picture and Sound Recording Activities				✓
59 Internet Service Providers, Web Search Portals, and Data Processing Services			✓	✓
62 Finance				
63 Insurance				
64 Auxiliary Finance and Insurance Services	✓			
67 Property Operators and Real Estate Services				
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)	✓	✓		✓
75 Public Administration	✓			
77 Public Order, Safety and Regulatory Services				
84 Hospitals	✓	✓		✓
85 Medical and Other Health Care Services				
87 Social Assistance Services		✓		
89 Heritage Activities				
90 Creative and Performing Arts Activities	✓	✓	✓	

There does not seem to be a particular trend evident in the results, such as one that would enable us to say endogeneity is picked up by the log of WDZ area only in instances when it is also picked up by road network density, yet perhaps with a lower frequency. This is not the case, as the former IV in some cases detects endogeneity when road network density does not. While it is quite obvious that road network density and the log of WDZ area perform the best as IVs in terms of the strength of their correlation with effective density, it is difficult to say which is more valid. The Sargan over-identification test for instrument validity was conducted on all the industries for Melbourne by running 2SLS with robust standard errors and by specifying the log of WDZ and untransformed road network density as IVs. This complete set of results is reported in Appendix F.

To reiterate, the Sargan over-identification test is only useful to judge instrument validity if at least one of the IVs is known to be valid – the exogeneity of an instrument in general cannot be tested. Thus if the test results do not reject the null of all instruments being valid then it does not necessarily mean that this is the case, as the reality could in fact be that both are invalid. Extremely useful in this investigation into endogeneity is that the 2SLS regressions were carried out on a multitude of industries. In the first-stage regression of every industry the vector of IV values is unchanging, thus their predictions of effective density only differ by the values of the other exogenous control variables. Considering that for an instrument to be strong it must be highly correlated with the suspect endogenous variable, it is likely that the IV will be making a majority contribution to predicting the levels of effective density in this first stage. Thus, we can use the occurrences of endogeneity being detected by the Hausman test as strong evidence of an endogeneity bias being present and the instruments being useful for its detection. This is because an IV that is strongly correlated with effective density and in fact endogenous itself can be expected to not reveal an endogeneity bias in the case of any of the 30 industries.

With this in mind, in instances where neither of the favoured IV specifications resulted in the detection of an endogeneity bias separately and combined in an over identified model could not reject the null of at least one being an invalid instrument, we can use their effectiveness in detecting an endogeneity bias in other industries as reasonable proof that the instruments were in fact valid. Such was the situation for the industries of *Fabricated Metal Product Manufacturing*; *Machinery and Equipment Manufacturing*; *Building Construction*; *Heavy and Civil Engineering Construction; Basic Material Wholesaling*; *Food Retailing*; *Accommodation*; *Publishing (except Internet and music publishing)*; *Motion Picture and Sound Recording Activities*; *Finance*; *Insurance*; *Auxiliary Finance and Insurance Services*; *Property Operators and Real Estate Services*; *Public Administration*; *Public Order, Safety and Regulatory Services*; *Medical and Other Health Care Services*; and *Heritage Activities*. In the wage-function estimations for these industries, endogeneity was not detected by the Hausman test for either log WDZ area or untransformed road network density while the null of the Sargan over-identification test for each industry was not rejected. While the outcome of the latter test does not necessarily mean that both instruments are valid (as the result could potentially be the same if both instruments were invalid), the detection of endogeneity by the IVs in a number of other industries suggests the reasonableness of the inference of their validity and that of the LS parameter estimates.

An alternative outcome for the 2SLS results with respect to the two preferred IVs, which occurred for a number of industries, was where they both produced results that showed strong evidence of endogeneity. This situation arose only in the industries of *Food Product Manufacturing* and *Creative and Performing Arts*. In the second industry, the validity test suggested that both instruments were valid while the test in the case of the first industry suggested both instruments to be invalid. This latter outcome countered expectations since both IVs detected a presence of endogeneity. The magnitudes of their elasticity point estimates, however, differed vastly as did the significance levels attributed to the conclusions of their Hausman tests.

A final possible outcome for the two 2SLS regressions, which occurred far more frequently than the one just mentioned, was that one IV could detect the presence of endogeneity while the other could not. Such was the situation in the industries of *Textile, Leather, Clothing and Footwear Manufacturing*; *Basic Chemical and Chemical Product Manufacturing*; *Primary Metal and Metal Product Manufacturing*; *Transport Equipment Manufacturing*; *Machinery and Equipment Wholesaling*; *Other Store-based Retailing*; *Food and Beverage Services*; *ISPs, Web Search Portals and Data Processing Services*; *Professional, Scientific and Technical Services (except computer system design and related services)*; *Hospitals*; and *Social Assistance Services*. In eight of these 11 industries the results of the 2SLS regressions with road network density as an IV indicated the presence of endogeneity while the log of WDZ did not. In the remaining three industries, the opposite is the case. In just under half of these 11 industries, the Sargan over-identification test was rejected at the 0.10 level concluding that at least one instrument was invalid, which was likely because of the differing conclusions from the Hausman tests. In the remaining cases where the validity of both instruments was maintained, the point estimates of the effective density coefficients tended to be similar between the two IV specifications.

Table 5.2.2, shown below, reports the results from the 2SLS regressions for Melbourne with respect to the two preferred IV specifications along with the point estimates generated from the regressions in Analysis 2-1. Next to the point estimates provided by instrumentation are reported the amounts by which the 2SLS estimates differed from the GLS estimates. A positive sign on this means the direction of the detected endogeneity bias is upwards, with the opposite being the case when a negative arises. Where the IV point estimates are reported in bold, the Hausman test generated p-values below 0.10 and thus the null hypotheses of the GLS estimators being consistent were rejected. In the results where IV estimates are not reported in bold, endogeneity was not detected with a reasonable level of confidence and thus the preferred estimate would be the one provided by GLS.

Table 5.2.2: A Comparison of GLS and IV Parameter Estimates of Effective Density in Melbourne

	GLS Results	IV Results				
Industry Name	Effective Density	Road Density	Est. Bias	Ln(WDZ Area)	Est. Bias	Validity p-value
11 Food Product Manufacturing	*** 0.096	*** -0.1318**	0.227	**0.009**	0.087	0.032
13 Textile, Leather, Clothing and Footwear Manufacturing	*** 0.119	0.071	0.119	**0.077**	0.042	0.006
18 Basic Chemical and Chemical Product Manufacturing	** 0.121	**** -0.191**	0.312	0.015	0.106	0.030
21 Primary Metal and Metal Product Manufacturing	0.018	**0.099**	-0.081	0.032	-0.014	0.151
22 Fabricated Metal Product Manufacturing	*** 0.138	0.138	0.000	*** 0.184	-0.046	0.564
23 Transport Equipment Manufacturing	*** 0.138	**-0.023**	0.161	*** 0.177	-0.039	0.005
24 Machinery and Equipment Manufacturing	*** 0.123	0.029	0.094	0.073	0.050	0.439
30 Building Construction	*** 0.174	*** 0.129	0.045	*** 0.149	0.025	0.541
31 Heavy and Civil Engineering Construction	*** 0.131	0.058	0.073	0.053	0.078	0.944
33 Basic Material Wholesaling	*** 0.194	*** 0.189	0.005	*** 0.201	-0.007	0.834
34 Machinery and Equipment Wholesaling	*** 0.153	**0.069**	0.084	*** 0.136	0.017	0.257
41 Food Retailing	** 0.041	0.016	0.025	* 0.070	-0.029	0.311
42 Other Store-Based Retailing	*** 0.147	***** 0.116**	0.031	*** 0.217	-0.070	0.015
44 Accommodation	*** 0.154	*** 0.227	-0.073	*** 0.143	0.011	0.134
45 Food and Beverage Services	*** 0.117	** 0.088	0.029	**0.042**	0.075	0.251
54 Publishing (except Internet and Music Publishing)	*** 0.169	0.076	0.093	*** 0.137	0.032	0.404
55 Motion Picture and Sound Recording Activities	*** 0.200	0.129	0.071	* 0.124	0.076	0.999
59 Internet Service Providers, Web Search Portals, and Data Processing Services	*** 0.169	0.141	0.028	*** 0.133**	0.036	0.998
62 Finance	*** 0.132	*** 0.211	-0.079	*** 0.153	-0.021	0.332
63 Insurance	*** 0.160	0.055	0.105	*** 0.174	-0.014	0.168
64 Auxiliary Finance and Insurance Services	*** 0.175	*** 0.133	0.042	*** 0.196	-0.021	0.257
67 Property Operators and Real Estate	*** 0.134	***** 0.194**	-0.060	*** 0.162	-0.028	0.517

Services						
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)	*** 0.190	** 0.093	0.097	*** 0.193	-0.003	0.006
75 Public Administration	*** 0.181	** 0.120	0.061	*** 0.144	0.037	0.634
77 Public Order, Safety and Regulatory Services	0.016	0.034	-0.018	* 0.054	-0.038	0.691
84 Hospitals	*** 0.125	***** 0.229**	-0.104	*** 0.156	-0.031	0.268
85 Medical and Other Health Care Services	*** 0.091	** 0.085	0.006	*** 0.080	0.011	0.903
87 Social Assistance Services	*** 0.113	***** 0.184**	-0.071	*** 0.142	-0.029	0.327
89 Heritage Activities	*** 0.111	0.064	0.047	** 0.129	-0.018	0.561
90 Creative and Performing Arts Activities	*** 0.154	**0.085**	0.069	**** 0.164**	-0.010	0.659

Note (1): * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01
Note (2): Elasticities in bold indicate results where the Hausman test detected an endogeneity bias at the 0.10 level.

In the case of the 2SLS estimates produced for the industries in Sydney, instrumentation was only carried out using the log and untransformed versions of the Travel Zone geographic areas. The discussion here of the results for these instrumented regressions will be restricted to those generated by the former specification because, as in Melbourne, untransformed TZ area performed poorly as an IV with F-statistics of values less than 10 reported by the Weak Instrument test for 10 of the 30 industries. F-statistics from the untransformed IV specification of TZ area ranged from 330 to 3088 over all industries in the analysis. The complete set of results can be viewed in Appendix G.

2SLS regressions run with log-transformed TZ area detected endogeneity present in five industries, including *Textile, Leather, Clothing and Footwear Manufacturing*; *Motion Picture and Sound Recording Activities*; *Professional, Scientific and Technical Services (except computer system design and related services)*; *Public Administration*; and *Social Assistance Services*. This number equalled the number detected in Melbourne with the same IV; however, only one industry overlapped between the two cities – that of *Textile, Clothing and Footwear Manufacturing*. In three of the five industries, IV estimates suggested that LS overestimated the impacts of effective density on productivity while underestimated them in two industries (although in one case the difference in the elasticity estimates was merely 0.002).

A Sargan over-identification test could not be carried out on the Sydney industry data because more than one instrument could not be obtained. With a p-value for the Hausman test set at 0.10, here too, in Sydney, one would have expected to incorrectly reject the null in roughly three industries by inferring the existence of endogeneity. Having only the one type of instrument, it is difficult to comment on its relative validity. Log-transformed TZ area may indeed be a valid instrument for Sydney but in the absence of an endogeneity bias, the LS specification would remain superior. On the other hand, it may be an invalid instrument that is correlated with the error term in the wage functions, thus leaving endogeneity to go undetected in most of the industries in the analysis. This cannot be proven but relies on the quality of reasoning behind the instrument's use and can be aided by the inference derived from having a second IV in the Melbourne analyses. The results of the 2SLS regressions for Sydney are given below in Table 5.2.3 in the same fashion as the Melbourne results in Table 5.2.2. Analysis 2-1 results are provided along with those generated by 2SLS to compare the magnitudes of the coefficients estimated and get an idea of the approximate level of bias, assuming that log-transformed TZ is valid as an instrument. Being reported in bold once again indicates 2SLS elasticity estimates in which endogeneity was detected. Lastly, a p-value for a Sargan over-identification test is not given because TZ area was the only IV possessed and it was not included in 2SLS analyses in combination with its logged form.

Table 5.2.3: A Comparison of GLS and IV Parameter Estimates of Effective Density in Sydney

	GLS Results	IV Results	
Industry Name	Effective Density	Ln(WDZ Area)	Est. Bias
11 Food Product Manufacturing	0.020	0.039	-0.019
13 Textile, Leather, Clothing and Footwear Manufacturing	*** 0.180	**** 0.122**	0.058
18 Basic Chemical and Chemical Product Manufacturing	*** 0.198	* 0.130	0.068
21 Primary Metal and Metal Product Manufacturing	0.020	** 0.141	-0.121
22 Fabricated Metal Product Manufacturing	0.018	0.088	-0.070
23 Transport Equipment Manufacturing	*** 0.111	** 0.136	-0.025
24 Machinery and Equipment Manufacturing	*** 0.060	*** 0.148	-0.088
30 Building Construction	*** 0.166	*** 0.168	-0.002
31 Heavy and Civil Engineering Construction	*** 0.160	*** 0.164	-0.324
33 Basic Material Wholesaling	*** 0.100	*** 0.149	-0.249
34 Machinery and Equipment Wholesaling	*** 0.075	* 0.076	-0.001
41 Food Retailing	*** 0.087	*** 0.107	-0.020
42 Other Store-Based Retailing	*** 0.138	*** 0.108	0.030
44 Accommodation	*** 0.112	*** 0.151	-0.039
45 Food and Beverage Services	*** 0.087	*** 0.061	0.026
54 Publishing (except Internet and Music Publishing)	*** 0.218	*** 0.309	-0.091
55 Motion Picture and Sound Recording Activities	*** 0.217	**** 0.129**	0.088
59 Internet Service Providers, Web Search Portals, and Data Processing Services	*** 0.132	0.104	0.028
62 Finance	*** 0.172	*** 0.187	-0.015
63 Insurance	*** 0.146	*** 0.131	0.015
64 Auxiliary Finance and Insurance Services	*** 0.164	*** 0.206	-0.042
67 Property Operators and Real Estate Services	*** 0.129	*** 0.143	-0.014
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)	*** 0.162	***** 0.129**	0.033
75 Public Administration	*** 0.159	***** 0.178**	-0.019
77 Public Order, Safety and Regulatory Services	0.000	0.020	-0.020
84 Hospitals	*** 0.085	*** 0.133	-0.048
85 Medical and Other Health Care Services	*** 0.081	*** 0.117	-0.036
87 Social Assistance Services	*** 0.154	***** 0.189**	-0.035
89 Heritage Activities	0.054	0.085	-0.031
90 Creative and Performing Arts Activities	*** 0.244	*** 0.266	-0.022

Note (1): * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01
Note (2): Elasticities in bold indicate results where the Hausman test detected an endogeneity bias at the 0.10 level.

5.2.2 Analysis 2-2 Conclusions

In this section's analyses, instrumental variable methods were applied with the intention of uncovering whether reverse-causality was influencing the GLS estimates in

Analysis 2-1 and if this was the case, getting parameter estimates that would be corrected for the induced level of bias. Two instrumental variables along with their log transformations were utilized for the Melbourne 2SLS regressions, giving a total of four IV specifications, and one instrumental variable along with its log transformation was utilized for the Sydney 2SLS regressions. Complications arose in this process because a Hausman test of exogeneity is only effective if the IVs applied are known to be valid, that is, that they are orthogonal to the error term in the original model specification. This was further complicated by inability to ascertain the validity of an IV. Only the hypothesis that all instruments in an over-identified model were valid could be tested, which still relied on the assumption that at least one IV was known to be valid.

The validity of the instruments and their transformations was loosely implied in endogeneity being detected in at least some industries and was further bolstered by the reporting of very strong correlations with effective density as indicated by large F-statistics in the Weak Instrument tests. This cannot be used as conclusive evidence, however, because it is unknown if the industries to reveal an endogeneity bias will continue to do so in repeated samples. There is little reason to believe that a detected bias must be systematic and as such asymptotically converge on a specific value that is always positive or always negative. A bias could merely be random. If this is the case, IV estimation may perform better or worse in some samples and the luxury of repeated sampling was not available. Having two cities to enable results comparisons was beneficial, but not a sure way of validating endogeneity bias as intra-industry differences could exist between Sydney and Melbourne.

The two strongest IV specifications were the log of WDZ area and untransformed road network density. The former performed better in terms of its strength of correlation with effective density; however, its application was accompanied by the less frequent detection of an endogeneity bias, which raises questions about its validity as an IV. Road density as an IV, on the other hand, detected many more cases of an endogeneity bias but produced a number of questionable parameter estimates in the Melbourne industry regressions that in turn led one to question their consistency. Conservatively, what can said about the GLS estimates is that endogeneity is likely to be affecting the parameter estimates for at least some industries; however, while some industries are more suspect than others the evidence is not overwhelming. Furthermore, there is a fair bit of uncertainty around the magnitude of the bias that this generates. If a preferred set of 2SLS results did need selection it would be those models that had been instrumented with the log of WDZ area, keeping in mind that the GLS estimates should be maintained in the cases where the null of the Hausman test could not be rejected as they would, on average, prove to be more efficient.

It is recommended that further investigation take place into suitable instrumental variables that could replace those applied here in the 2SLS estimations. If data on long-lagged population numbers cannot be obtained and easily allocated into current-day spatial boundaries, another IV that could prove promising would be a measure of sewerage capacity. A common exogenous factor in determining density levels and the locations for redevelopment is the location and excess capacity of water and sewerage infrastructure. The decisions of waterworks authorities on where to locate infrastructure are influenced by topography as relative inclines and declines can affect costs associated with pumping sewage and potable water. The location of excess capacity is commonly a determining criterion for the intensity of development in a given area as extending this capacity elsewhere can be costly, especially when capacity exists in current areas and intensification can aid in capital cost recovery. Using a metric based on this infrastructure as an instrument could be effective as long as a strong feedback effect does not arise; such as if infrastructure investment would follow productivity increases around cities. Unfortunately, theory dictates from a perspective that assumes that ideal conditions are present, which is infrequently the case in reality. Agglomeration economies do not have a standardized approach to instrument variables models.

5.3.0 Analysis 2-3 Overview: Re-estimating localization and urbanization effects using WDZ data

Here in Analysis 2-3, the simultaneous estimation of localization and urbanization effects on employment productivity are reinvestigated, this time applying more conventional methods to controlling for localization. The effects from urbanization are still represented by the effective density index while localization effects are calculated by estimating employment numbers present within concentric ring bangs emanating from the centroid of a given origin location. This method follows that used by Graham (2007b) who applies proximity bands of width 0-1 km, 1-5 km, 5-10 km, and 10-25 km and Rosenthal and Strange (2008) who apply bands of width 0-5 miles, 5-25 miles, 25-50 miles, and 50-100 miles. The overall estimating equation then can be expressed as illustrated below.

$$\ln I_{k,m} = \alpha_k \beta \ln U_m + \sum^{i} \delta_i \ln Occ_{i,k,m} + \sum^{j} \gamma_j \ln Edu_{j,k,m} + \psi Exp_{k,m} + \sum^{b} \phi_b Band_{b,k,m} + e_{k,m} \quad [5T]$$

The employment population within each ring was estimated assuming that employment was evenly distributed within each work destination zone. Thus, ring employment could be estimated by summing the product of ring-zone overlap and industry-specific employment across all zones that intersected with a given ring as depicted in the formula below.

$$\theta_{s,b,l} = \sum_{n=1}^{n} \left[E_{sz} \cdot \left(A_{l,n}^{b} / A_{n}^{z} \right) \right] \quad [5U]$$

Here, $\theta_{s,l,b}$ designates the employment θ in industry sector s and band b emanating from the centroid of location l. Next, $E_{s,z}$ represents the employment in industry s and zone n, $A_{l,n}^{b}$ is the area of the band emanating from location l that overlaps with zone n, and A_{n}^{z} is the area of zone n.

Rosenthal and Strange (2008) discuss the downside of this approach as it gives rise to an errors-in-variables problem. The error included in this measurement will bias the estimated influences of agglomeration towards zero, as the assumption of uniformly distributed employment is somewhat outside of what can realistically be expected – that employment even within small spatial units is likely to cluster. They minimize the effects of this measurement error by restricting their sample to include units from which the first concentric ring touches at least two surrounding areas. Rather than reducing the sample sizes in this analysis to meet this criterion, ring widths were selected such that they would minimize the instances where fewer than two adjacent employment zones were met by the first band whilst maintaining ring widths of sufficiently small scale to provide useful insight. Fortuitously, employment zones in Sydney and Melbourne are on the whole rather small so this could be carried out with little issue; however, there was still quite a bit of variability in employment zone sizes when one progressed from inner-city to fringe areas. The smallest zone in Sydney, for example, had an approximated radius of 56 m, the largest had one of approximately 51km and the citywide mean was approximately 1.15kms. In Melbourne these figures were approximately 32 m, 63 km, and 1.79 km respectively. The largest, outermost areas were unlikely to even enter into the samples because the minimum criteria for the establishment of a destination zone by VicRoads is that it contains at least 100 jobs, meaning there is a low likelihood that many of the industries analysed will have employment in these areas, especially since many of them are tied to the efficiencies of inner city location.

The selected ring sizes for this section's analyses were of the widths 0-2.5 km, 2.5-7.5 km, 7.5-15 km, and 15-25 km. This set the rings at sizes slightly larger than those utilized by Graham (2007b) and rather smaller than those used by Rosenthal and Strange (2008). Selecting rings of this size translated to 2567 zones in Sydney and 2400 zones in Melbourne having radii that set them within the bounds of a first ring. As with the calculation of effective density, the ring band employment calculations were carried out in the spatially enabled database, Postgres. A walkthrough of the code written for this purpose is given next.

Calculating Industry-Specific Concentric Ring Band Employment for Sydney and Melbourne

Much of the initial set-up for the ring band employment calculations had already been done to undertake the effective density calculations and as such can be referred to in the discussion of Analysis 2-1. The only new table to be created for this purpose was one to store all the industry-specific employment estimates for each ring band/employment zone pairing. The script for this in the case of Melbourne was as shown below.

```
CREATE TABLE public.melbourne_industry_band_employment (                [5V]
        id SERIAL PRIMARY KEY,
        vicdznp06 BIGINT,
        ind11_band1 DOUBLE PRECISION,
        ind11_band2 DOUBLE PRECISION,
        ind11_band3 DOUBLE PRECISION,
        ind11_band4 DOUBLE PRECISION
        );
```

This located the new table in the 'public' schema and merely set up all the columns required to store the values generated for the first industry's employment band estimates. In this case it was for industry 11 – Food Product Manufacturing. The next script populated the two generic columns that included unique ID codes and the work destination zone numbers. Incorporated in this script was the condition that every WDZ code had to be unique and thus omit the duplicates that created the error in the effective density calculations as described in Analysis 2-1. This script appears below.

```
INSERT INTO public.melbourne_industry_band_employment (vicdznp06)       [5W]
        SELECT vicdznp06 FROM shp.melb_wdz
        WHERE is_duplicate = FALSE
        ORDER BY vicdznp06 ASC;
```

Prior to running the code to carry out the necessary calculations, some amendments had to be made to the WDZ spatial file table. The amendments included making four new columns and within them creating the new geometrics of every ring band for every WDZ in Melbourne and doing the same for the table storing the Sydney geometries. This only had to be carried out once because the band geometries did not vary across the industry types – it was just their employment numbers within the bands that would differ. The scripts to create these four new geometry columns appear below in formula/script 5X.

```
ALTER TABLE shp.melb_wdz ADD COLUMN band_1b GEOMETRY;                    [5X]
COMMENT ON COLUMN shp.melb_wdz.band_1b IS 'Buffer on WDZ centroid of
        radius 2.5km';
UPDATE shp.melb_wdz SET band_1b = ST_Buffer(ST_Centroid(the_geom), 2500);
ALTER TABLE shp.melb_wdz ADD COLUMN band_2b GEOMETRY;
COMMENT ON COLUMN shp.melb_wdz.band_1b IS 'Buffer on WDZ centroid of
        radius 2.5km - 7.5km';
UPDATE shp.melb_wdz SET band_2b =
        ST_Difference(ST_Buffer(ST_Centroid(the_geom), 7500),
        ST_Buffer(ST_Centroid(the_geom), 2500));
ALTER TABLE shp.melb_wdz ADD COLUMN band_3b GEOMETRY;
COMMENT ON COLUMN shp.melb_wdz.band_1b IS 'Buffer on WDZ centroid of
        radius 7.5km - 15km';
UPDATE shp.melb_wdz SET band_3b =
        ST_Difference(ST_Buffer(ST_Centroid(the_geom), 15000),
        ST_Buffer(ST_Centroid(the_geom), 7500));
ALTER TABLE shp.melb_wdz ADD COLUMN band_4b GEOMETRY;
COMMENT ON COLUMN shp.melb_wdz.band_1b IS 'Buffer on WDZ centroid of
        radius 15km - 25km';
UPDATE shp.melb_wdz SET band_4b =
        ST_Difference(ST_Buffer(ST_Centroid(the_geom), 25000),
        ST_Buffer(ST_Centroid(the_geom), 15000));
```

The above section of script was written to contain three components: An 'ALTER TABLE' command, a "COMMENT ON COLUMN" command and an "UPDATE" command. The first created a new designated column to store the band data, the second simply enabled a comment to be inserted for the user's reference so that the precise widths of the bands would not be mistaken, and finally the third produced the new geometries. It did so via the "ST_Difference" command, which would subtract one circle from another thus creating a band, except in the case of the first ring band that in fact was just a circle.

Once these scripts were run and the ring band geometries created, the ring data was merged with the industry-specific employment data to populate the tables created to store the employment band data. The code for this calculation was carried out in four components for each industry, where each segment carried out the necessary calculations for one band. The code for the first (innermost) band for industry 11 is given below.

```
UPDATE public.melbourne_industry_band_employment SET ind11_band1 =            [5Y]
        temp.ind11_band1 FROM (
        SELECT m.vicdznp06 AS source_wdz, SUM (
        inc."total employed" * (ST_Area(ST_Intersection(m2.the_geom, m.band_1b))
        / ST_Area(m2.the_geom))) AS ind11_band1
        FROM shp.melb_wdz AS m, shp.melb_wdz AS m2, import.cities_income AS inc
        WHERE m2.vicdznp06 = substring(inc.wdz_code FROM 7 FOR 4)::BIGINT
        AND inc.ind_id = 11 AND inc.city_id = 2 AND m2.is_duplicate = FALSE AND
        m.is_duplicate = FALSE GROUP BY m.vicdznp06
        ) AS temp
        WHERE source_wdz = public.melbourne_industry_band_employment.vicdznp06;
```

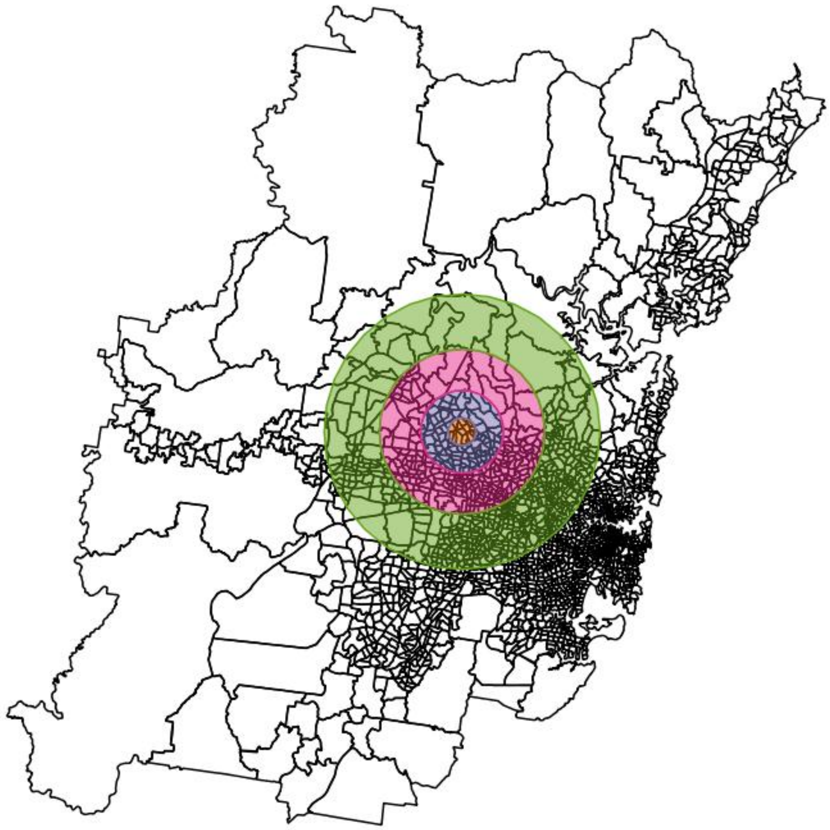
The code was written to create a temporary table, called "temp", to store data from a number of columns including the results of the band calculations. The results in the column named 'ind11_band1' were inserted into a designated table to store all the employment band data where the WDZ codes were appropriately matched. The crux of the code was that it had to be written to include the city spatial data twice, thus essentially creating an origin-destination matrix. The former was needed so that each employment zone in a city could draw on the spatial data of its respective inner-ring, while the latter was required to estimate the overlap between those rings and all possible zones that the rings intersect with. This could not be done by only drawing on the data from a spatial table once, hence the schema and table indicated by 'shp.melb_wdz' being specified and renamed twice: once as 'm' and again as 'm2'. The "GROUP BY" clause near the end of the code essentially told Postgres over which distinguishing characteristic the summation should occur, which was the WDZ identifier.

The SQL codes for the subsequent bands appeared the same as the one for an innermost band, simply the reference to 'band_1b' was replaced with a reference to 'band_2b', 'band_3b', or 'band_4b'. Running the scripts for all four bands for one industry was a laborious task for Postgres and would take up to eight hours. Repeating this for 30 industries in Melbourne and 30 industries in Sydney meant approximately 480 hours of run time to complete all the employment band calculations. Unfortunately, after running the calculations for Melbourne the widths of the bands were revised so the actual total run time was significantly greater than this. It took so long because each of the four bands emanating from each of the 2000-plus employment zones had to have the amount of overlap with every possible zone calculated, then had to have that amount divided by the respective WDZ area to get a share of area coverage, and then finally needed this share multiplied by an employment number. If carried out for one industry in the case of a city of 2,100 zones this meant nearly 18 million instances (4 bands x 2100 origin zones x 2100 destination zones) where a band was being associated with a zone during the calculation, before consideration was even given to the calculation of coverage share and the multiplication by a level of employment.

A different approach to these calculations was considered that conceivably would have carried them out with much greater efficiency. There was a part of the operation of the code that was shared among all industries and as such it was being repeated every time the four estimates of band employment were being recalculated for another industry. Part of this component is what was being referred to by the figure of 18 million associations. Neither the WDZ/band intersections, nor the shares of WDZs covered by the bands subsequently calculated, were unique to every industry; they were only unique to each city. This represented an opportunity for the shared section of code to be run and the results stored in Postgres so that separate industry-specific employment band scripts could be authored in such a way that they would simply draw on this data rather than recalculating it.

The approach adopted to take advantage of this fact is summarised in Appendix H. The issue that arose when running this script for Melbourne was that all of the free disk space was used up in executing the script to store this shared component. A little over 350 gigabytes that were available initially were subsequently filled up before the program crashed. The difference in the original method described above from the approach just described here is that the calculations in the former case were carried out piece-by-piece while using volatile memory to store what was needed, then only the grouped (or summed) results were actually stored permanently on the hard-drive. The approach elaborated on in Appendix H, on the other hand, attempted to store every possible combination of spatial unit interaction for every possible industry, which was generating over 130 million rows of data across a number of columns. In light of this, employment estimations for the proximity bands were calculated as initially described. The result was a series of four bands with industry-specific employment estimates being reported within them. The size of these bands is visually represented below in Figure 5.3.1 to give a better understanding of their scale.

Figure 5.3.1: An Example of Scale for One Set of Industry Employment Bands in Sydney



Once the industry employment band calculations were made, the resulting figures were exported along with the rest of the industry wage function data to CSV format and the analyses were carried out as with all the others in GRETL. The OLS estimator was used to estimate each industry's wage function for each city and where heteroskedasticity was present by detection with a Breusch-Pagan test – which happened to be the case for all industries in both cities – GLS was used instead. The following section discusses the outputs of these regressions first for the city of Sydney and then makes comparisons to the Melbourne results.

5.3.1 Analysis 2-3 Results

Sydney and Melbourne Results

As discussed in the overview above, a combination of the market potential index (effective density) and industry-specific employment within concentric ring bands was employed in an econometric model to simultaneously estimate the effects of urbanization and localization economies on industry-specific employment productivity. As in Analyses 2-1 and 2-2, wage functions are estimated for 30 industries across a range of sectors and are carried out using both Sydney and Melbourne employment data. Being able to compare the results between two large Australian capital cities not only allowed one city's results to be validated against the other's, but where differences arose one could find justification for city-specific elasticity estimates in light of differing levels of benefit from agglomeration.

In Sydney, the inclusion of industry-specific employment bands improved the model fit for 23 of the 30 industries when compared to the basic model in Analysis 2-1 without the quadratic terms. In most cases, the improvement was marginal, as only four industries had increases in adjusted R-squared values in excess of 0.05. The industry to benefit the most from the inclusion of the localization controls was *Heritage Activities* with an adjusted R-squared value improvement of 0.128, followed by *Public Administration* with an improvement of 0.080 and *Transport Equipment Manufacturing* with an improvement of 0.080. By far the industry to experience the greatest worsening of model fit because of the addition of the employment bands was *Creative and Performing Arts* with an adjusted R-squared value reduction of 0.154, which was then followed by *Basic Chemical and Chemical Product Manufacturing*, which experienced a reduction of 0.032 and *Professional*, *Scientific and Technical Services* with a reduction of 0.017.

With respect to the parameter estimates on the four employment band variables, expectations were that where localization economies were present, values would be largest and most significant for the inner-most band and progressively become smaller as one moved outward from an observed unit. Additionally, there was an expectation that effective density would in some cases experience a reduction in its coefficient's magnitude and lose some significance, indicating a shift to the importance of localization effects in generating positive externalities. These expectations were, for the most part, met by the model;

however, there were some instances where the results were somewhat perplexing and require some alternative explanations to be postulated. First to be discussed are the aspects of the results where expectations were met.

The inclusion of the employment bands generated results that meant only 15 industries experienced significant effects from urbanization, all of which were positive except for *Public Order, Safety and Regulatory Services* for which the effect came out negative. The first employment band of radius 0 km – 2.5 km was significant for 19 industries, the most of any of the agglomeration variables. Only one of the five industries from Analysis 2-1 continued to show no effects from agglomeration, which was *Fabricated Metal Product Manufacturing*, while the remaining four previously insignificant industry coefficients began reporting significant effects from agglomeration – a likely response to industry localization being a more meaningful determinant of productivity in these industries. All of these significant coefficients on the first band were positive, which supports this position. Beyond the first band, far fewer industries reported significant coefficients and perhaps not in the progressive fashion anticipated. The second employment band was significant in seven industries, the third employment band in eight and the fourth band in nine. This suggests that the effects of localization taper off rapidly and are rather geographically constrained, which is consistent with the existing literature.

The perplexing aspect of the employment band results is that a fair number of their significant coefficients were estimated to be negative. The second band was significant and negative in the instance of four industries and both the third and fourth bands were negative in the case of six industries. While expectations were that the effects of more distant own-industry concentrations would become negligible, the results reported for a number of industries suggest that the effect can actually be significant and negative for some industries. The same situation arises for Graham (2007b) when adopting a similar model on firm-level data on UK firms. He states that it is unclear from the data why own-industry employment density would have such an effect, but then follows with a couple of possible explanations. His first proposed explanation is that own-industry concentration could lead to fierce price competition that in turn leads to lower profits and value of output. An alternative reason he gives is that negative coefficients may be indicative of a lesser tendency in firms to concentrate, which would be characteristic of industries that service dispersed populations such as retail and energy distribution. Interestingly, in no industry was it the case that the innermost band was estimated to have a negative coefficient in Sydney, thus this latter explanation may be adapted with a slightly different justification given. It is possible that localization benefits are real and, because they exist over a small spatial scale, the impacts of competition may begin to dominate over the benefits of proximity when the market catchment extends further than the "reach" of the agglomeration benefit. Ascribing such ex post justifications, however, should not be pursued too far as they may detract from the results that do indeed meet expectations and all such justifications can only be speculative. The tolerance level for error in the model, alpha, also cannot be ruled out as an explanation in some cases as in repeated samples the tendency for outer bands to emerge as being significant and negative may not hold.

After controlling for the effects of localization, the largest productivity elasticities with respect to effective density were reported for the industries of *Social Assistance Services* (0.219), *Textile, Leather, Clothing and Footwear Manufacturing* (0.197), *Internet Service Providers, Web Search Portals and Data Processing Services* (0.163), *Hospitals* (0.162), and *Other Store-based Retailing* (0.146). *Finance* was a close sixth, with an estimated elasticity of 0.134. The strongest effects from localization, as indicated by the largest significant parameter estimates on the innermost band, were experienced by *Food Product Manufacturing* (0.098), *Motion Picture and Sound Recording Activities* (0.082), *Textile, Leather, Clothing and Footwear Manufacturing* (0.082), *Creative and Performing Arts Activities* (0.079), and *Basic Material Wholesaling* (0.069). The first industry mentioned, *Food Product Manufacturing*, was one of the industries to emerge insignificant in Analysis 2-1 and as such it was not surprising to see it at the top of the list for receiving the largest elasticity estimate with respect to industry localization. The complete set of results for Sydney can be viewed below in Table 5.3.1.

Table 5.3.1: Sydney Regression Parameter Results for ED and Employment Bands

Industry Name		ε_A	S.E.	Adj. R-Squared	F-Value	N
11 Food Product Manufacturing				0.342	8.81E-61	740
ED		-0.065	0.054			
Band 1	***	0.098	0.020			
Band 2		0.009	0.030			
Band 3		-0.057	0.037			
Band 4		0.018	0.025			
13 Textile, Leather, Clothing and Footwear Manufacturing				0.241	1.88E-30	588
ED	**	0.197	0.083			
Band 1	***	0.082	0.026			
Band 2	**	-0.063	0.032			
Band 3		-0.003	0.034			
Band 4		-0.032	0.033			
18 Basic Chemical and Chemical Product Manufacturing				0.338	1.33E-24	326
ED	**	0.129	0.058			
Band 1	*	0.032	0.016			
Band 2		0.017	0.019			
Band 3		-0.034	0.027			
Band 4		0.026	0.036			
21 Primary Metal and Metal Product Manufacturing				0.330	5.24E-25	341
ED		-0.029	0.044			
Band 1	***	0.066	0.018			
Band 2		-0.029	0.026			
Band 3		0.002	0.028			
Band 4		0.032	0.023			
22 Fabricated Metal Product Manufacturing				0.109	1.09E-07	392
ED		0.026	0.042			
Band 1		0.037	0.013			
Band 2		-0.050	0.018			
Band 3		0.004	0.020			
Band 4		0.012	0.017			
23 Transport Equipment Manufacturing				0.453	2.05E-44	386
ED		-0.047	0.057			
Band 1	***	0.053	0.017			
Band 2		0.024	0.022			
Band 3	*	-0.040	0.022			
Band 4	***	0.085	0.024			
24 Machinery and Equipment Manufacturing				0.260	8.03E-38	658
ED		0.028	0.036			
Band 1	***	0.035	0.010			
Band 2		0.015	0.013			
Band 3	**	-0.043	0.019			
Band 4		0.022	0.017			
30 Building Construction				0.181	1.68E-65	1654
ED	**	0.070	0.034			
Band 1	***	0.061	0.016			

Band 2		0.026	0.020			
Band 3	*	-0.034	0.020			
Band 4		-0.023	0.016			
31 Heavy and Civil Engineering Construction				0.283	4.15E-28	454
ED		0.068	0.045			
Band 1		0.009	0.022			
Band 2	***	0.078	0.030			
Band 3		0.002	0.039			
Band 4		-0.056	0.034			
33 Basic Material Wholesaling				0.254	1.40E-40	721
ED		0.045	0.049			
Band 1	***	0.069	0.019			
Band 2	**	-0.057	0.027			
Band 3		0.028	0.031			
Band 4		0.005	0.028			
34 Machinery and Equipment Wholesaling				0.220	1.25E-34	733
ED		0.002	0.039			
Band 1	***	0.055	0.011			
Band 2		0.018	0.018			
Band 3	**	-0.057	0.024			
Band 4	*	0.039	0.020			
41 Food Retailing				0.163	4.86E-46	1327
ED	**	0.069	0.032			
Band 1	*	0.023	0.013			
Band 2		-0.018	0.020			
Band 3	**	0.042	0.018			
Band 4	**	-0.043	0.017			
42 Other Store-Based Retailing				0.213	6.36E-81	1687
ED	***	0.146	0.025			
Band 1		-0.005	0.010			
Band 2		0.009	0.013			
Band 3		-0.005	0.015			
Band 4		-0.012	0.015			
44 Accommodation				0.216	1.03E-33	732
ED		0.063	0.041			
Band 1		0.016	0.016			
Band 2		0.009	0.014			
Band 3		0.004	0.017			
Band 4	*	0.027	0.016			
45 Food and Beverage Services				0.331	4.20E-140	1694
ED	***	0.125	0.025			
Band 1		0.002	0.012			
Band 2		0.004	0.014			
Band 3	**	-0.039	0.016			
Band 4		0.014	0.014			
54 Publishing (except Internet and Music Publishing)				0.209	2.41E-26	603
ED		0.000	0.070			
Band 1	***	0.059	0.022			
Band 2	*	0.034	0.018			
Band 3		0.012	0.026			
Band 4		-0.024	0.025			

55 Motion Picture and Sound Recording Activities				0.528903	3.17E-63	430
ED		-0.101	0.092			
Band 1	***	0.082	0.030			
Band 2		0.027	0.035			
Band 3		0.057	0.036			
Band 4		-0.022	0.036			
59 Internet Service Providers, Web Search Portals, and Data Processing Services				0.271	3.66E-12	220
ED	**	0.163	0.082			
Band 1		-0.035	0.044			
Band 2		0.007	0.032			
Band 3		0.014	0.023			
Band 4	**	-0.078	0.030			
62 Finance				0.398	3.40E-100	979
ED	***	0.134	0.036			
Band 1		0.014	0.011			
Band 2		0.004	0.009			
Band 3		-0.008	0.011			
Band 4	***	-0.030	0.009			
63 Insurance				0.440	7.73E-55	487
ED		0.050	0.040			
Band 1	***	0.036	0.012			
Band 2		-0.005	0.009			
Band 3		0.005	0.012			
Band 4		-0.018	0.011			
64 Auxiliary Finance and Insurance Services				0.285	2.74E-61	925
ED		-0.047	0.044			
Band 1	***	0.057	0.012			
Band 2	***	0.029	0.008			
Band 3		-0.004	0.009			
Band 4		-0.011	0.009			
67 Property Operators and Real Estate Services				0.241	5.06E-62	1136
ED	***	0.115	0.037			
Band 1		0.007	0.016			
Band 2		0.016	0.015			
Band 3		-0.019	0.018			
Band 4		-0.006	0.017			
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)				0.311	1.50E-142	1863
ED	**	0.059	0.027			
Band 1	***	0.036	0.009			
Band 2		0.005	0.007			
Band 3		0.003	0.008			
Band 4	*	-0.012	0.007			
75 Public Administration				0.383	5.60E-101	1036
ED	***	0.101	0.028			
Band 1	***	0.031	0.011			
Band 2		-0.012	0.010			

Band 3		-0.007	0.014			
Band 4	*	-0.031	0.017			
77 Public Order, Safety and Regulatory Services				0.153	2.17E-27	882
ED	***	-0.099	0.035			
Band 1	***	0.044	0.013			
Band 2		0.000	0.015			
Band 3		-0.011	0.020			
Band 4	**	-0.043	0.021			
84 Hospitals				0.314	2.00E-33	471
ED	***	0.162	0.045			
Band 1	***	0.027	0.009			
Band 2	*	-0.020	0.012			
Band 3	**	-0.026	0.013			
Band 4		-0.010	0.014			
85 Medical and Other Health Care Services				0.151	2.26E-49	1538
ED	*	0.051	0.027			
Band 1		0.012	0.010			
Band 2		0.007	0.013			
Band 3		-0.001	0.014			
Band 4		-0.009	0.013			
87 Social Assistance Services				0.173	4.49E-60	1601
ED	***	0.219	0.030			
Band 1		0.011	0.014			
Band 2		-0.027	0.019			
Band 3		-0.030	0.019			
Band 4		-0.009	0.016			
89 Heritage Activities				0.253	1.61E-09	194
ED		-0.035	0.107			
Band 1		0.026	0.032			
Band 2	*	-0.035	0.019			
Band 3	***	0.092	0.021			
Band 4		-0.004	0.019			
90 Creative and Performing Arts Activities				0.161	1.41E-17	546
ED		0.025	0.108			
Band 1	***	0.079	0.045			
Band 2		0.008	0.041			
Band 3		0.011	0.037			
Band 4		0.037	0.039			

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

The inclusion of local-industry employment bands in the wage equations for industries in Melbourne had a very similar impact on the fit of the models as it had in Sydney: explanatory power was improved for 22 industries and worsened for eight. The type of impact on industry-specific adjusted R-squared values (whether they were improved of worsened), however, did not align too closely between the two capital cities. The only industry to have its adjusted R-squared value lowered in both cities was *Creative and Performing Arts Activities*. This left 16 industry regressions mutually improved by the inclusion of the employment bands and 13 differing in the effect they experienced. The magnitude of improvement remained minor in Melbourne as it had in Sydney, except in a few industries. *Motion Picture and Sound Recording Activities* experienced the greatest R-squared value improvement of 0.143, followed by *Internet Service Providers, Web Search Portals and Data Processing Services* with an improvement of 0.117, *Building Construction* with an improvement of 0.108 and *Fabricated Metal Product Manufacturing* with an improvement of 0.082. The remaining industries experiencing improvements in model fit had their adjusted R-squared values increased by less than 0.050. As for the worsening of model fit, the greatest impact was on *Heritage Activities* with an adjusted R-squared reduction of 0.058 and the remaining seven industries experiencing reductions of less than 0.05.

After controlling for local-industry concentration in Melbourne, there were 16 industries in which effective density was estimated to have a positive and significant impact on productivity and two in which its effects were estimated to be significant yet negative. Of these the strongest positive effects were estimated for *Heritage Activities* (0.204), *Fabricated Metal Product Manufacturing* (0.191), *Transport Equipment Manufacturing* (0.164), *Food and Beverage Services* (0.159), and then tied at fifth were *Finance* (0.157) and *Machinery and Equipment Manufacturing* (0.157). None of these industries appeared among those with the largest significant effective density parameter estimates in the Sydney regressions. While many of the industries shared estimates of similar magnitude between the two cities, a fair number reported them to be vastly different. Industry 59 – *Internet Service Providers, Web Search Portals and Data Processing Services* – was one example of such a case where the Sydney and Melbourne results appeared contradictory. Industry 62 – *Finance* – was another to show differing results between the two capital cities. In Sydney, this latter industry showed rather strong effects from effective density with a significant coefficient estimate of 0.134 and no significant localization effects within the inner three employment bands. In Melbourne, on the other hand, the effects on this industry of effective density were estimated to be slight and significant yet negative with an estimated elasticity of -0.029 and a positive innermost band elasticity estimate of 0.053. The situation for the industry of *Insurance* for the two cities was the opposite: it was localization that was reported to be the dominant form of agglomeration to affect productivity in Sydney and urbanization in Melbourne. It is unlikely that the financial sectors would operate so differently in the two cities as to create such starkly different estimates, thus it is most likely due to high levels of multicollinearity between localization and urbanization patterns that some contrasting results between the two cities such as these were observed. In addition to this, the inclusion of industry localization controls improved the model fit for the industry of *Finance* by less than 0.02 in both cities and actually worsened the fit for Melbourne in the industry of *Insurance*, so maintaining the results for these industries from the models in Analysis 2-1 may be a preferred alternative.

The other industry to report a significant negative coefficient on effective density in Melbourne was *Food Product Manufacturing*. As was the case with this industry in Sydney, it displayed no significant effects from agglomeration in Analysis 2-1 but here in Analysis 2-3 with the inclusion of localization controls it became clear that own-industry scale is a valuable determinant of productivity in this industry. A total of 21 industries from those analysed reported strong effects from localization as indicated by significant coefficient estimates on their innermost employment band, all of which came out positive. The industries to experience the greatest estimated benefit were *Motion Picture and Sound Recording Activities* (0.122), *Food Product Manufacturing* (0.120), *Internet Service Providers, Web Search Portals and Data Processing Services* (0.082), *Building Construction* and Basic *Material Wholesaling* (0.069). Three of these held in the case of Sydney as well. The effects of progressively more distance bands on employment productivity were dramatically less than the innermost band, as was the case in Sydney. This gives additional validation to the claim that localization economies dissipate quickly from a source. Only six industry regressions were estimated with significant second bands, three with significant third bands, and six with significant outermost bands. As in Sydney, the somewhat unforeseen issue arose that most of these significant outer ring parameter estimates had negative signs. Among the significant parameters in the outer bands, only one was positive in each of Band 2, Band 3 and Band 4. As aforementioned, however, this situation was not exclusive to this analysis but has occurred in other published empirical work. The complete list of industry results appears below in Table 5.3.2.

Table 5.3.2: Melbourne Regression Parameter Results for ED and Employment Bands

Industry Name		ε_A	S.E.	Adj. R-Squared	F-Value	N
11 Food Product Manufacturing				0.409	7.26E-86	815
ED	***	-0.171	0.054			
Band 1	***	0.120	0.019			
Band 2		0.018	0.029			
Band 3		0.029	0.032			
Band 4		-0.049	0.037			
13 Textile, Leather, Clothing and Footwear Manufacturing				0.190	3.69E-26	666
ED		-0.037	0.058			
Band 1	***	0.055	0.022			
Band 2		0.016	0.035			
Band 3		0.008	0.037			
Band 4		-0.012	0.035			
18 Basic Chemical and Chemical Product Manufacturing				0.367449	1.80E-26	315
ED	**	0.116	0.055			
Band 1	**	0.034	0.016			
Band 2		0.009	0.014			
Band 3		-0.068	0.019			
Band 4		-0.007	0.034			
21 Primary Metal and Metal Product Manufacturing				0.143	1.97E-09	358
ED		0.032	0.067			
Band 1	***	0.063	0.020			
Band 2	*	-0.045	0.026			
Band 3		-0.015	0.028			
Band 4		0.030	0.032			
22 Fabricated Metal Product Manufacturing				0.169	1.38E-13	417
ED	***	0.191	0.056			
Band 1	***	0.037	0.014			
Band 2	**	0.044	0.021			
Band 3		0.002	0.022			
Band 4		-0.001	0.022			
23 Transport Equipment Manufacturing				0.166	9.46E-16	486
ED	***	0.164	0.051			
Band 1	*	0.023	0.012			
Band 2		0.007	0.015			
Band 3		-0.018	0.015			
Band 4		-0.014	0.019			
24 Machinery and Equipment Manufacturing				0.308	1.59E-46	655
ED	***	0.157	0.035			
Band 1	**	0.030	0.013			
Band 2	***	-0.054	0.019			
Band 3		0.026	0.025			
Band 4		-0.032	0.031			
30 Building Construction				0.345	1.40E-	1484

					128	
ED	***	0.082	0.030			
Band 1	***	0.070	0.017			
Band 2		-0.011	0.020			
Band 3		-0.011	0.021			
Band 4		0.030	0.022			
31 Heavy and Civil Engineering Construction				0.181	2.88E-16	457
ED		0.006	0.062			
Band 1	***	0.062	0.022			
Band 2		0.037	0.032			
Band 3		-0.015	0.031			
Band 4		-0.046	0.043			
33 Basic Material Wholesaling				0.297	1.22E-52	767
ED	***	0.117	0.039			
Band 1	***	0.069	0.015			
Band 2		0.017	0.023			
Band 3		0.042	0.030			
Band 4	**	0.056	0.028			
34 Machinery and Equipment Wholesaling				0.266	5.11E-40	677
ED	***	0.132	0.032			
Band 1	***	0.044	0.016			
Band 2	***	-0.062	0.022			
Band 3		0.021	0.023			
Band 4		0.008	0.034			
41 Food Retailing				0.216	1.04E-57	1202
ED		0.037	0.034			
Band 1		0.002	0.016			
Band 2		-0.005	0.018			
Band 3	*	0.035	0.019			
Band 4	**	-0.065	0.027			
42 Other Store-Based Retailing				0.183	5.24E-61	1523
ED	***	0.097	0.033			
Band 1		0.024	0.012			
Band 2		0.018	0.014			
Band 3		-0.024	0.016			
Band 4		0.002	0.021			
44 Accommodation				0.187	1.79E-22	590
ED		-0.012	0.069			
Band 1	**	0.067	0.026			
Band 2		0.001	0.020			
Band 3		0.023	0.024			
Band 4		0.010	0.023			
45 Food and Beverage Services				0.270	1.07E-95	1503
ED	***	0.159	0.037			
Band 1		-0.002	0.016			
Band 2		0.007	0.016			
Band 3	**	-0.035	0.016			
Band 4	***	-0.053	0.019			
54 Publishing (except Internet and Music Publishing)				0.132	8.89E-11	440
ED		0.072	0.084			
Band 1		0.033	0.023			

Band 2		0.008	0.020			
Band 3		-0.009	0.020			
Band 4		-0.011	0.021			
55 Motion Picture and Sound Recording Activities				0.552809	1.55E-52	336
ED		-0.132	0.110			
Band 1	***	0.122	0.042			
Band 2		0.064	0.044			
Band 3	*	-0.074	0.039			
Band 4		0.012	0.044			
59 Internet Service Providers, Web Search Portals, and Data Processing Services				0.313	3.50E-13	200
ED		-0.122	0.114			
Band 1	**	0.082	0.041			
Band 2		0.037	0.032			
Band 3		-0.003	0.035			
Band 4		-0.005	0.036			
62 Finance				0.705	2.20E-197	783
ED	***	-0.029	0.010			
Band 1	***	0.053	0.008			
Band 2		-0.005	0.010			
Band 3		-0.003	0.012			
Band 4		-0.011	0.013			
63 Insurance				0.348	3.64E-32	397
ED	***	0.157	0.056			
Band 1		-0.001	0.020			
Band 2		0.008	0.015			
Band 3		-0.002	0.014			
Band 4	***	-0.051	0.016			
64 Auxiliary Finance and Insurance Services				0.282	9.86E-50	772
ED	*	0.099	0.055			
Band 1		0.022	0.018			
Band 2		0.001	0.013			
Band 3		0.002	0.013			
Band 4		-0.020	0.015			
67 Property Operators and Real Estate Services				0.168	4.08E-30	871
ED		0.031	0.052			
Band 1	*	0.041	0.022			
Band 2		0.009	0.019			
Band 3		0.016	0.023			
Band 4		-0.032	0.025			
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)				0.289	3.10E-111	1603
ED		0.030	0.038			
Band 1	***	0.059	0.011			
Band 2	**	-0.018	0.008			
Band 3		0.007	0.008			
Band 4	*	-0.015	0.009			
75 Public Administration				0.343	9.24E-83	984

ED	**	0.076	0.034			
Band 1	***	0.041	0.010			
Band 2		0.010	0.010			
Band 3		-0.016	0.011			
Band 4		-0.018	0.013			
77 Public Order, Safety and Regulatory Services				0.635	1.10E-138	675
ED		-0.034	0.056			
Band 1	*	0.031	0.018			
Band 2		-0.014	0.018			
Band 3		-0.007	0.017			
Band 4		-0.038	0.022			
84 Hospitals				0.168	9.04E-17	509
ED	**	0.112	0.051			
Band 1	*	0.021	0.011			
Band 2		-0.002	0.013			
Band 3		-0.016	0.015			
Band 4		-0.012	0.021			
85 Medical and Other Health Care Services				0.170	3.13E-49	1352
ED	***	0.102	0.031			
Band 1		0.001	0.013			
Band 2		-0.002	0.014			
Band 3		-0.007	0.015			
Band 4	***	-0.043	0.016			
87 Social Assistance Services				0.182	2.67E-53	1353
ED	***	0.154	0.037			
Band 1	*	0.030	0.017			
Band 2	***	-0.043	0.015			
Band 3		-0.007	0.016			
Band 4		0.010	0.023			
89 Heritage Activities				0.357	1.94E-12	163
ED	*	0.204	0.105			
Band 1		-0.042	0.039			
Band 2		-0.004	0.029			
Band 3		-0.052	0.034			
Band 4		0.020	0.034			
90 Creative and Performing Arts Activities				0.087	1.46E-07	476
ED		-0.135	0.138			
Band 1		0.060	0.058			
Band 2		0.079	0.051			
Band 3		-0.062	0.053			
Band 4		0.018	0.041			

Note: * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

5.3.2 Analysis 2-3 Conclusions

In this analysis, the separate agglomeration effects of urbanization and localization economies were estimated by simultaneously including a market potential index to represent the former and industry-specific employment numbers within four concentric ring bands to represent the latter. The radii of the ring bands were 0 km to 2.5 km for the innermost ring, 2.5 km to 7.5 km for the second ring, 7.5 km to 15 km for the third ring and 15 km to 25 km for the outermost ring.

The inclusion of employment bands improved the fit of the wage-function models in 23 industries in Sydney and 22 industries in Melbourne; however, since the industries to experience an improvement and worsening of fit did not align too closely between the two cities, cross-city comparison did not prove to be an effective way of validating industry findings, nor did it support justification for generalizing the findings in this analysis to other cities. This could be because localization and urbanization effects in a number of given industries vary between the two cities because of differences in industry structure and/or specialization, or merely because of a multicollinearity issue obscuring distinction between the effects of the two variables. This leads to two fundamental limitations in the task of jointly estimating localization and urbanization effects.

The first is that because the actual processes generating the agglomeration benefits were not directly observed and measured, the effective density and ring band employment estimates had to act as proxies for the benefits that urban centres and industrial clusters are said to provide. This is the standard limitation in all empirical works on agglomeration economies but it extends to the second, which was that of identification and distinction between the two types of agglomeration economies in instances where the two variables may have been highly correlated. Localized industry patterns may mimic urbanization patterns closely and if this happens then the individual coefficients of the agglomeration proxies may be inaccurate. We do not directly observe the reasons why firms locate where they do, we observe only where they locate and this calls into question the results of models that attempt to separate out a localization and urbanization effect.

A number of industries such as those in the medical sector, non-grocery retailing, food and beverage services, and social assistance services seemed rather unaffected by the inclusion of employment band data in both cities, maintaining urbanization to be the dominant agglomeration force at work. A couple of others such as *Building Construction* and *Public Administration* showed both types of economies to be significant in both cities. Then a number of other industries such as *Primary Metal Product Manufacturing* and *Motion Picture and Sound Recording Activities* showed a complete shift of a productivity effect to the localization variables, but the majority of the results between the Sydney and Melbourne industry wage function estimates reported conflicting results in terms of which agglomeration force was a more relevant determinant of productivity.

The only industries to show a responsiveness to agglomeration in both cities in Analysis 2-3 that did not in Analysis 2-1 were *Primary Metal and Metal Product Manufacturing* and *Public Order, Regulatory and Safety Services*; otherwise the results of Analysis 2-1 seemed adequate at capturing agglomeration effects without raising the issue as to whether the added localised industry controls were providing reliable parameter estimates. In light of this, if one set of results were to be selected for application in transport infrastructure valuation assessments from the analyses then those from Analysis 2-1 would be recommended. The next chapter will discuss this further.

Chapter 6: Applying the Results, Conclusions and Further Work

6.1 Introduction

Up to this point, this thesis has covered matters ranging from the importance of cities and how they evolve, to transport and how it shapes them, to theoretical and empirical accounts of agglomeration economies. Most importantly, it has reported on the analyses undertaken for the purposes of this research. This final chapter concludes this work by first revisiting the research questions laid out in Section 1.3. These questions are restated below in Section 6.2 where they are followed by a progress check, accounting for the extent to which these questions have now been answered. They are followed by Section 6.3, which reviews the analyses of Chapters 4 and 5 for the purpose of selecting a set of elasticities most suitable for use in Australia. Section 6.4 discusses how these elasticity estimates can be applied and some of the limitations created by current transport network models. Section 6.5 then discusses how planning for more compact and well-connected centres can merge the productivity benefits of agglomeration with desirable planning outcomes such as sustainability and liveability, achieving a so-called 'sustainability multiplier' effect. Finally, Section 6.6 offers suggestions for future work.

6.2 Revisiting the Research Questions

Recalling Section 1.3, the purpose of this thesis has been to inform responses to five main research questions. These questions are restated below.

1. Can useful agglomeration elasticities be determined for Australian cities? This fundamental question not only asks whether there is a measurable productivity effect from economic density but also whether data sources exist to enable such analyses to occur whilst producing reliable results.
2. Are elasticity estimates amenable to being generalized across Australian cities? As there is a desire for agglomeration externalities to be incorporated into planning

efforts across the nation, it is useful to know if generalizing results from one city across others makes sense.

3. Are elasticity estimates robust to changes in the geographic scale of analysis? As the complexity of analyses increases with more detailed datasets, it is useful to understand the impacts of employing different geographic scales.
4. How can these elasticities best be applied in an Australian context? This question aims to gain some understanding of the ability of current transport modelling systems in Australia to apply these elasticities to infrastructure projects.
5. What are the broader implications of agglomeration economies for urban planning and infrastructure? This question entails taking a step back from the essentially econometric endeavour and determining the significance of the subject for cities and planning.

The first question, which has been central to this thesis, was explored in depth in Chapters 4 and 5. The short answer is that returns to employment density and accessibility are in fact evident in Australian cities and across a wide range of industries. While data-quality issues will almost always be present in any analysis of this type, initially there was a sense that data collection practices in Australia would limit the efficacy of an analysis of agglomeration effects. The results, however, show that it is possible to make estimates of agglomeration elasticities with existing datasets that meet expectations, are comparable to international studies, and can be reliably used in the appraisal of transport infrastructure and city planning strategies. The preferred sets of elasticity estimates are discussed below in Section 6.3.

Questions 2 and 3 can now also be answered. The former, which questions whether elasticity estimates generated for one Australian city can be applied in another, was explored in every one of the six analyses. The first analysis did so by comparing all Australian capital cities while the rest focussed on Sydney and Melbourne. What one could infer from comparing the results on the separate cities is that while many of the industries analysed show very similar effects from agglomeration, there is evidence that they may differ for at least some. Where possible, it would be best to apply elasticity estimates to cities from which data contributing to the estimates are sourced. The latter question, inquiring how geographic scale of analysis units affects elasticity estimates, can be answered by comparing the results between Chapters 4 and 5, though not without some limitation. Section 5.1.1 explored this in some detail. For a number of industries, elasticity estimates generated by using SLA units are similar to those estimated using the much smaller unit of the Travel or Work Destination Zone. For some other industries, primarily within the retail sector, differences are quite pronounced. The fact that these rather large differences exist between Part 1 and Part 2 Analyses for both Sydney and Melbourne tells us that they are unlikely to exist due to mere estimation error or slightly differing industry aggregation levels. The answer most likely resides in the fact that larger geographic units 'wash out' much of the variation in density and earnings. While the differing industry aggregation levels and model specifications limit the reliability of direct comparisons between Part 1 and Part 2 results, some of these obvious and large deviations are still strong indicators of the effects of spatial scale.

Questions 4 and 5 are somewhat subordinate to the previous three, but still highly relevant as they take a step back from what is most prominently an econometric endeavour and focus on the application of the findings. The former initiates an inquiry into how once productivity elasticity estimates are made with respect to industry agglomeration they can be applied. This question requires a two-part response as the application of agglomeration effects partly relies on the state of the art of transport network models and partly on an accounting approach for converting elasticities to an economic benefit. These matters are covered below in Section 6.4. The latter question, Question 5, is addressed below in Section 6.5 where the implications of agglomeration economies for sustainability are discussed.

6.3 The Preferred Elasticity Estimates

In the process of carrying out the empirical work in this thesis, the complexity of each subsequent analysis was increased while maintaining the general theoretical model and the overall approach to estimating agglomeration effects. The fundamental changes between the analyses were the scope of industry classification, the geographic scale of the observations in the samples, and the treatment of agglomeration externalities (i.e. how and whether or not localization effects were accounted for). Each of these posed a number of benefits and shortfalls. These will be weighed here in the process of selecting a preferred set of results to apply to appraisal of transport infrastructure projects.

The analyses in Part 1 used data on 3-digit industry classifications. This level of industrial disaggregation is rather well refined and allowed separate agglomeration impacts to be estimated for a number of industries in which differences were expected to be material. Sound reasoning exists for moving away from single, composite-industry measures, as by observing the spatial distribution of different industries it is evident that the benefits of agglomeration are not equal among them all. This becomes evident when considering the estimates that were made across industry types, but also within broader industry categories. A meaningful level of detail is lost for some industries when aggregated upwards, but there is more to consider.

Recalling the discussion in the overview of the Part 2 Analyses, the ABS consultants gave strong warnings over maintaining the use of 3-digit ANZSIC codes when choosing the finer geographic scale of work destination zones for analysis. The perturbation practices of the ABS in the context of requesting small spatial units with low employment counts will likely produce datasets with a significant amount of error. The potential gain, however, of using city datasets with larger sample sizes made possible by opting for the use of Work Destination Zones for generating more efficient and precisely measured parameter estimates, was deemed of greater value than maintaining a fine level of industry detail. This benefit became evident in the results derived from Analysis 2-1 where the vast majority of the industries reported significant agglomeration effects. In light of the considerations given thus far, the preferred results are narrowed down to those produced in Part 2's analyses.

Within Part 2, a number of variations were made on the econometric specification of the wage-function model initially proposed in Analysis 1-2. After first estimating a linear model that only considered agglomeration benefits from urbanization economies in Analysis 2-1, a model was then specified that let agglomeration effects vary with the level of effective density. While providing elasticity estimates that were more flexible, the confounding effects of multicollinearity sourced from interacting effective density with itself and the magnitude of error around the evaluated parameter estimates were likely to offset the benefits.

Next the topic of endogeneity was explored in the industry-specific estimates. In this analysis, some useful insights were gleaned into whether an endogeneity bias was present in some of the industry elasticity estimates and into the effectiveness of the instrumental variables that were trialled. The results, however, were somewhat inconclusive and in the case of the instrument that performed the best, namely the natural log of Work Destination Zone Area, the parameter estimates provided were not too far from the originals.

In the final analysis of Part 2, the effects of localization and urbanization were simultaneously estimated. This was done while applying different employment concentration measures to each effect to limit the potential issue invoked by multicollinearity. There were a few industries for which the model fit was improved by a fair amount when adding the industry-specific controls, but in most cases the benefit was rather trivial. Moreover, the frequent high correlation between employment band data and effective density had the general effect of reducing the efficiency of the parameter estimates. A final point to make is that while the insights from simultaneously including localization and urbanization parameters in a model may be useful from an economic development perspective, travel demand models are limited enough, as they are to accommodate the effects of agglomeration in appraisal. Including another form of an agglomeration effect would be well beyond the capacities of such models to include their differing effects in CBA analyses.

The overall recommendation would be to apply the results generated by Analysis 2-1 and laid out in Appendices D and E for application in transport infrastructure projects. The reason for this is that these results are estimated with relatively high levels of efficiency, reflected in nearly every industry reporting a highly significant effective density parameter estimate. The flexibility of allowing for industries to have variable elasticities (if the quadratic terms were significant and the fit of the models were improved) merely adds to the reliability of the results. These results are also of comparable magnitude to those produced by UK analyses (which gives a degree of validation) and the nature of the agglomeration elasticity variable applied makes them readily available for use in project appraisals. Having determined the preferred set of results, the following section will discuss how exactly they can be used to appraise the agglomeration benefit generated by transport infrastructure projects.

6.4 From Elasticities to Quantifying Investment Value

The treatment of the effects of changes in effective density on urban productivity can be addressed at a number of different scales. As such, there are very simplistic and also very complex ways of modelling travel and density patterns across a city, depending on the influence of time. This 'time' dimension is crucial because cities are dynamic entities in which population size and development patterns are being determined jointly by a myriad of forces. A change in any one of these forces can cause a shift in employment and transport patterns over time that reverberate through a city and consequently can also affect the effective density values of geographic units.

The factors that affect effective density specifically are distance (or travel time) and density. Cervero (2003) identifies two different 'path models', a short-term and a long-term, that affect travel patterns and travel times. In the short-term, increasing road network supply creates a benefit of increasing travel speeds, which in turn lower the cost of travel and cause a growth in demand on the improved route. Some of these gains in traffic are *generative* and others are *redistributive*. The former comprises shifts in the modes of travel, the release of suppressed trips, and distance changes while the latter is composed of route and schedule changes. In the first case, total vehicle kilometres travelled are increased and in the second case they are not. Then there exists another body of longer term impacts from transport infrastructure investments. These include induced development, which includes built environment shifts and land-use shifts; behavioural shifts, which include levels of car ownership and transit usage; and also long-term induced demand effects that are merely a continuation of the short-term impacts. All of these short and long-term effects influence the density of given locations (via induced development and land-use shifts) and travel times (via all effects mentioned above).

As one can see, projecting the impact of infrastructure projects on future effective density levels is not a simple task. Not only do most travel demand forecasting models not include most of the induced travel demand effects (DeCorla-Souza and Cohen 1999) but most travel demand elasticities are very case-specific and thus generalizing them to other projects can be ill advised (Heanue 1997). Broadly, what this suggests is that the accuracy of estimation around future effective density levels resulting from a transport improvement is only as good as the systems and methods in place to quantify them. What is very consistent in the literature is that traffic demand forecasting models are a long way from achieving a high level of refinement.

At the end of the UK Department for Transport's policy document on the wider economic benefits of transport investments they discuss some of the tools available for application to predicting future impacts of transport network improvements. They identify *Regional Economic Models* (REMs) that produce employment and output forecasts by industry and region but treat the transport sector quite superficially; *Spatial Computable General Equilibrium* (SCGE) *models* that give a great deal of detail to industrial and geographic disaggregation but again treat transport very simplistically; and *Land Use / Transport Interaction* (LUTI) *models* which perhaps perform the best as they directly model the impact of transport on land use; however, they are geographically limited and generally in their early stages of development. What this means is that although some "off the shelf" software packages are available (that typically use American data), cities and regions have a fair bit of developmental work to do and many years of data collection to calibrate reliable models.

As an Australian example, projects in Sydney commonly use the EMME2 platform for creating transport demand models. While the details and capabilities of this sophisticated modelling software are very flexible, thus allowing it to be tailored to the needs of the region in which it is being put to use, it is still plagued by a fair number of limitations. For one, these models are typically designed and built for a specific purpose and task, frequently requiring them to be updated and amended. In most cases, induced travel demand effects are included in a minimalistic way and effects on land-use are generally overlooked altogether. Alchin (personal communication, 21 April 2011) gives some examples of the criticism of the travel demand model as applied in Sydney. For one, he mentions that public transport is poorly integrated into the model because it disregards its capacity and service levels. A more general issue that he points out, which is more of a commentary on the use of the model rather than on the model's capabilities, is that in many instances the model benefits of a transport project are based on the assumption that certain other projects will receive funding and be constructed at some point in time in the future. These assumptions can frequently be made without their funding being sourced and approval granted.

One example of where the benefits of agglomeration were applied to a project in Australia is in the economic assessment of the Brisbane Cross-River Rail project. While the treatment of agglomeration economies in the appraisal was very advanced for the Australian context, they were still incorporated in a relatively simplistic manner. Again as with models in Sydney, the Brisbane Strategic Travel Demand Model includes the effects of induced travel demand to a limited extent by considering only trip redistribution and mode shift changes – interactions of transport with land-use are not accounted for. In this sense the model assumes land-use to be static and thus the location of industry-specific employment was unaltered in the computation of effective density values over the years following the project's completion in appraisal. Apart from the abilities of the Brisbane Strategic Model to predict travel-times and hence accessibility changes over time, the appraisal of agglomeration impacts from the cross-river rail project utilized elasticities at the highest level of industry aggregation and from sources where the estimates had been generated internationally (Oaten, personal communication, 16 May 2011). An obvious improvement here would be to use more detailed elasticities and ones derived from analyses using local data.

Being a new component of the appraisal of transport infrastructure projects in Australia, there is a lack of guidance and absence of precedence for how agglomeration economies can and should be quantified. Now, any treatment of them at all constitutes an improvement on convention; however, there is an opportunity for an organization such as the Bureau of Infrastructure, Transport and Regional Economics (BITRE) in Australia to define a consistent framework. The elasticities in this thesis are, to my knowledge, the most detailed advanced and possibly the only ones produced at all using data on Australian cities. In their current state, they can be used in the quantification of agglomeration benefits with the current state-of-the-art in travel demand models. Improvements on the travel demand and land-use models in Australia as vital components of generating the inputs for estimating the impacts on effective density would improve how accurately agglomeration effects are represented in appraisal; however, detailed recommendations for these models are beyond the scope of this thesis.

Once effective density indices have been estimated for the years following a new transport infrastructure project's completion, applying the elasticity estimates to estimate economic gain from a project becomes rather straightforward and the UK Department for Transport's document on WEBs shows us how. They propose the following formula:

$$A_j = \sum_{i,j}\left[\left(\frac{\Delta ED_j}{ED_j}\cdot ElP_{ij}\right)\cdot GDP_{ij}\cdot E_{ij}\right]$$

where i and j denote an industry sector and geographic location, respectively, ED represents effective density, ElP is the elasticity of productivity, GDP represents per capita gross domestic product, and E represents the number employed. As the elasticity formula is

$$\varepsilon_P = \frac{\Delta P}{\Delta ED}\cdot\frac{ED}{P}$$

the inclusion of the $\frac{\Delta ED_j}{ED_j}$ term in the formula above merely cancels out the effective density component of the previous formula and simply leaves us with a percentage change in productivity. Thus the formula eventually becomes the following:

$$A_j = \sum_{i,j}\left[\frac{\Delta P_{ij}}{P_{ij}}\cdot GDP_{ij}\cdot E_{ij}\right]$$

How the influence of a transport network improvement on travel-time affects ED depends on the way it was initially calculated when estimating the industry wage or production functions. In the case of effective density calculated by weighting surrounding area employment by travel-time the solution is easy: any change in travel speeds will automatically be picked up in the new effective density calculations with new network travel-time estimates. If linear distances are used as weights in the initial effective density calculations for estimating productivity elasticities, the situation becomes slightly more complicated. This is because changes in travel-times are not reflected in the index, only changes in employment levels in given locations. Graham (personal communication, 13 July 2009) explains that he prefers a linear specification of effective density because it avoids the potential of double counting since in conventional appraisal the effects of travel-time savings are already included. The more conservative elasticity estimates derived from a linear effective density measure can be used in the appraisal of transport infrastructure that affects travel times if an approximate change in linear distance can be estimated to reflect the improvements of travel speeds. This is a matter to be addressed by the operators of transport network models.

6.5 The Implications for Planning and Urban Sustainability

Taking a step back from discussion of the specific applications of industry productivity elasticities estimated with respect to levels of agglomeration, there are broader implications to appreciate at a city-scale. The benefits of agglomeration arise when the need for transportation is minimized and opportunities for economic interaction are made viable by high levels of accessibility. Travel-time constancy budgets suggest that reductions in time spent in commuting will be offset by accessing further destinations, making more trips, or by substituting private vehicle transport for more active forms of travel. In any of these cases, benefits accrue to the commuter and to the city as a whole. I would argue that, while road investment cannot be abandoned altogether, producing more compact cities through higher densities and increased public transport levels would have a positive effect on productivity levels as a result of the externalities of agglomeration. The polycentric city is one such compact city typology that seeks to maximize scale and density in a number of centres that are well connected by public transport. This might be especially beneficial for cities that have had a history of sprawl and now need to consolidate and densify in order to accommodate population growth while reducing resource consumption, environmental impact, and travel demand.

A polycentric urban form, reflective of the settlement patters observed in many European countries, means that within centres there are many destination opportunities within walking and cycling distance or perhaps made accessible by a short journey by tram or light rail. Space in this sense is optimized with less of it being taken up by private vehicle infrastructure such as roads and parking spaces. Opportunities for residential development within these centres further reduce the demand for transport infrastructure to carry workers into these employment centres. Concurrently, while a mix of uses not only reduces the need to leave the centre to access certain amenities and goods, it also stimulates the existence of urbanization economies touted by authors such as Jane Jacobs. The density of centres then makes public transport, especially rail, more viable. As elucidated by Newman and Kenworthy (1999), there is evidence from international data of the existence of an activity intensity threshold of around 35 persons plus jobs per hectare beyond which public transport service levels begin to dramatically increase because of the economies of urban scale. Rail has the advantage of being able to carry high volumes of people to other centres quickly and reliably whilst demand levels for the service can be managed by increasing service frequency or by adding carriages. Effective density levels (which have been shown to strongly influence productivity levels across a wide range of industry sectors) can be raised significantly by planning strategies that focus on building and improving centres, as centres are by definition rather dense. Moreover, the high levels of accessibility that they enable by making rail more viable and being efficient with the use of space play a key role in raising accessibility levels to surrounding areas.

The benefits of a polycentric urban form, well serviced by rail infrastructure, are not limited to agglomeration economies. Accompanying these benefits is a variety of others, all made possible by mindful planning of dense, mixed-use centres. Firms such as Gehl Architects[30], based in Copenhagen, are in great demand because of their work in urban design, believing that a focus on 'human scale' in urban centres can make them more liveable and sustainable places to live. The social capital and amenity made possible in cities is unsurpassable by alternative, sprawling suburban forms. Trubka, Newman et al. (2010a; 2010b; 2010c) focus on the economic benefits of dense, urban environments as they accrue because of cost savings in infrastructure, transportation, transport-related greenhouse gas emissions, and healthcare provision with an added benefit from an activity-induced health-related productivity improvement. The health component of urban planning is relatively new to the rhetoric of planning policy, yet it is a powerful one as bountiful evidence exists that individuals are more likely to achieve minimum physical activity health requirements in denser and more walkable environments (Active Living Research 2005) and suffer from significantly less chronic illness (Sturm and Cohen 2004).[31] The benefits of density are

---

[30] Gehl Architects have carried out projects in numerous major cities around the world such as Melbourne, Sydney, Perth, San Francisco, New York, Seattle, Vancouver, Toronto, London, Stockholm, and Prague to name several, where they emphasize alternative modes of transport, sustainability, and liveability in their recommendations and designs.

[31] Studies abound that discuss the environmental correlates of physical activity levels and urban form. See Frank, Saelens et al. (2007), Ewing (2005), Saelens (2003), and Stokols, Grzywacz et al. (2003) as merely a few examples for further discussion of the subject matter.

echoed by Wenban-Smith (2009) in his paper on the economies of scale in the costs of infrastructure provision, where he investigates the effects of urban spatial structure on water distribution costs in the UK. He argues that the benefits of agglomeration for infrastructure cost efficiencies are often taken for granted and finds density – and to a lesser extent, scale – to have a real and sizeable benefit. There is also the argument that brown-field redevelopment and the curtailment of sprawl by establishing greenbelts are vital for protecting valuable rural land and environmental amenity (Williams 2000). As one can see, planning for a polycentric urban form that is well connected by public transport and designed with the 'human scale' in mind addresses a suite of issues affecting the triple-bottom line in decision-making – the economic, social, and environmental facets of sustainability.

6.6 Recommendations for Further Work

The investigation into the industry-specific elasticity of productivity estimates carried out with respect to the density and accessibility of employment in this thesis has established an effective method for providing inputs into transport infrastructure appraisals. A logical next step would be to extend the industry coverage of the approach outlined in this thesis to include a greater number of industries. In this body of work, 30 2-digit industries had wage-functions estimated while the total number of industries at this level of aggregation nearly trebles this. Furthermore, one could consider investigating the agglomeration effects more comprehensively at the 3-digit level or even on a more disaggregated basis by using 4-digit ANZSIC data. When industries are treated on a more disaggregated basis, assessments of agglomeration economies can be made in greater detail. There is a trade-off, however, that must be considered between the benefit and increased effort of conducting assessments with highly disaggregated industry elasticities. Moreover, the statistical robustness of these estimates may begin to suffer as lower cell-counts become more unreliable with data that has been perturbed.

Apart for extending the industrial coverage of analysis, future works could look into other potential data sources. This would be particularly advisable if such data sources as the Longitudinal Business Survey (LBS) become geographically referenced in the future. LBS data would enable panel data techniques to be applied as well as a firm-level data to be used rather than drawing on census data for small-area units that has been reported as averages. It may also be that income data provided by the Australian Tax Office at the place-of-usual-residence may one day be provided by place-of-work, which would also be an improvement on the income measure used in this piece of research. Unless a valuable source of data had been overlooked in the process of researching this thesis, it is unlikely that a significantly better dataset currently exists in Australia, thus improving on the current estimates would likely require waiting for reform to occur in the current data collection practices and conventions. This reform primarily needs to be in the form of collecting and offering spatially referenced, small area data.

Another way in which to improve on the work embodied in this thesis is to investigate the effects of endogeneity further. The efforts made in this thesis produced results that were somewhat inconclusive. A lack of historical data in Australia for which collection practices and data formats have been held constant over time made finding an "ideal" instrument a challenge, but the search for suitable instruments in this thesis was by no means exhaustive. As such there may be reason to seek out and trial other instruments and other combinations of instruments that could provide better results. This said, given that the existing literature seems to be of the same mind that the endogeneity bias happens to be small, even if it is having an effect on the results here then it should not be of great concern.

A more novel extension of this work might seek to identify synergies among specific clusters of industries. The benefits of this, however, may be limited because as the microfoundations of agglomeration economies are unobserved, synergies across specific industry-types may simply be sporadic spatially and temporally. This means that inter-industry interactions may only occur intermittently between firms on a project-specific basis, or only occur in specific firms in a specific area of a city. The latter point refers to the rather uninformative nature of industry classifications, given that not all firms in the same industry provide the same product or service. A lawyer servicing victims of motor-vehicle accidents, for instance, will likely experience a different agglomeration benefit from one who deals primarily in corporate law. Identifying these differences in existing datasets, however, is simply not feasible.

In researching this thesis topic, it has also become evident that agglomeration measures generally fail to account for different modes of travel. Effective density, for instance, in the literature measures proximity by private vehicle travel-time or linear distance but does not account for bus or rail travel-times and the differing speeds of these modes. This is a matter worth considering, which can be evidenced by the example of the opening of the Southern Rail line in Perth in 2007. Commuters travelling from Mandurah to the city after the opening of the new line were able to make the journey in 48 minutes by train; by car the average speed was still an hour and ten minutes and by bus it was an average of an hour and thirty minutes (Newman 2011). Access by train in this context means something very different from access by car, and shares in transport modes can vary significantly from city to city – especially when comparisons are extended internationally. The resulting impact on elasticity estimates by using agglomeration measures that take transit mode share into account is unknown but may be worth investigating. This has been captured by some of Graham's analyses that used the generalized cost of travel in place of Euclidean distance in the effective density index. This, however, requires the outputs of a transport model which could not be accessed for this thesis research but could potentially be sourced for future works on the subject.

Another possible extension to this work might involve experimenting with different functional forms or model specifications. RESET tests could be conducted to determine whether the models specified could be improved. As the models applied in this research generally increased in comprehensiveness with each subsequent analysis, it might also be of interest to see how certain diagnostic tests (such as the RESET test) are affected.

The recommendations here are just some of the possible extensions and improvements that can potentially be made on the findings of this research. Economists may also find new theoretical frameworks and algebraic models by which to relate accessibility in cities to labour productivity; however, the purpose of this research has not been precisely this. This thesis has taken the econometric examination of agglomeration economies in Australian cities to its conclusions by applying the most relevant models to the best existing datasets that could be identified.

REFERENCES (Works Cited)

Active Living Research. (2005). *Designing for Active Transportation*. San Diego, CA: Active Living Research.

Andersson, F., Burgess, S., & Lane, J. (2007). Cities, Matching and the Productivity Gains of Agglomeration. *Journal of Urban Economics, 61*, 112-128.

Artis, M., Miguelez, E., & Moreno, R. (2009). *Assessing Agglomeration Economies in a Spatial Framework with Endogenous Regressors* (SERC Working Paper 23). London: London School of Economics.

Australian Bureau of Statistics. (2011, 03 May). Population Clock. Retrieved 03 May, 2011, from http://www.abs.gov.au/ausstats/abs@.nsf/web+pages/population+clock?opendocument

Australian Bureau of Statistics. (2009, 20 Aug). Household Income and Income Distribution. Retrieved 02 Nov, 2011, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6523.02007-08?OpenDocument

Baumgardner, J. (1988). The Division of Labour, Local Markets, and Worker Organization. *Journal of Political Economy, 96(3)*, 509-527.

Bhatt, K., Higgins, T., & Bert, J. (2008). *Lessons Learned from International Experience in Congestion Pricing*. Bethesda, Maryland: Report prepared by K.T. Analytics Inc for the U.S. Department of Transportation, Federal Highway Administration.

Boarnet, M., & Chalermpong, S. (2001). New Highways, Housing Prices and Urban Development: A case study of tolls roads in Orange County. *Transportation Research Record, 1706*, 1-8.

Bonsall, P. (1996). Can Induced Traffic be Measured by Surveys? *Transportation, 23*, 17-34.

Boshuizen, J., Geurts, P., & Van Der Veen, A. (2009). Regional Social Networks As Conduits for Knowledge Spillovers: Explaining performance of high-tech firms. *Tijdschrift voor Economische en Sociale Geografie, 100*(2), 183-197.

Brandt, A., Hahn, C., Kratke, S., & Kiese, M. (2008). Metropolitan Regions in the Knowledge Economy: Network analysis as a strategic information tool. *Tijdschrift voor Economische en Sociale Geografie, 100*(2), 235-249.

Burdett, R., & Sudjic, D. (2008). *The Endless City*. London: Phaidon Press.

Burger, M., Oort, F. V., Frenken, K., & Knaap, B. V. D. (2009). Networks and Economic Agglomerations: Introduction to the Special Issue. *Tijdschrift voor Economische en Sociale Geografie, 100*(2), 139-144.

Burgess, R., & Venables, A. (2004). *Toward a Microeconomics of Growth* (World Bank Policy Research Working Paper 3257). Washington: The World Bank.

Buxton, M. (2006). *Urban Form and Urban Efficiency*. Canberra: 2006 Australian State of the Environment Committee, Department of the Environment and Heritage.

Cervero, R. (2002). Induced Travel Demand: Research design, empirical evidence, and normative policies. *Journal of Planning Literature, 17*(1), 3-20.

Cervero, R. (2003). Road Expansion, Urban Growth, and Induced Travel: A path analysis. *Journal of the American Planning Association, 69*(2), 145-163.

Chinitz, B. (1961). Contrasts in Agglomeration: New York and Pittsburgh. *American Economic Review, 51*(2), 279-289.
Christaller, W. (1933). *Die Zentralen Orte in Suddeutschland* (C.W. Baskin (1966), Trans.). Translated to Central Places in Southern Germany, Englewood Cliffs, NJ: Prentice-Hall.
Ciccone, A. (2000). Agglomeration Effects in Europe. *European Economic Review, 46*, 213-227.
Ciccone, A., & Hall, R. (1996). Productivity and the Density of Economic Activity. *The American Economic Review, 86*(1), 54-70.
City Mayors Foundation. (2010a). The Largest Cities in the World by Land Area, Population and Density. Retrieved 02 May, 2011, from http://www.citymayors.com/statistics/largest-cities-population-125.html
City Mayors Foundation. (2010b). The Largest Cities in the World by Land Area, Population and Density. Retrieved 02 May, 2011, from http://www.citymayors.com/statistics/largest-cities-density-125.html
Coe, N. (1998). Exploring Uneven Development in Producer Service Sectors: Detailed evidence from the computer service industry in Britain. *Environment and Planning A, 30*, 2041-2068.
Combes, P.-P., Duranton, G., Diego Puga, & Roux, S. (2009). *The Productivity Advantages of Large Cities: Distinguishing agglomeration from firm selection* (SERC Discussion Paper 353). London: London School of Economics.
Combes, P.-P., Duranton, G., & Gobillon, L. (2008). Spatial Wage Disparities: Sorting matters! *Journal of Urban Economics, 63*, 723-742.
Coombes, M., & Raybould, S. (2001). Public Policy and Population Distribution: Developing appropriate indicators of settlement patterns. *Environment and Planning C: Government and Policy, 19*, 223-248.
Dantas, A., Krumdiek, S., & Page, S. (2010). Urban Form and Long-Term Fuel Supply Decline: A method to investigate the peak oil risks to essential activities. *Transportation Research Part A: Policy and Practice, 44*, 306-322.
DeCorla-Souza, P. (2000). Induced Highway Travel: Transportation policy implications for congested metropolitan areas. *Transportation Quarterly, 54*(2), 13-30.
DeCorla-Souza, P., & Cohen, H. (1999). Estimating Induced Travel for Evaluation of Metropolitan Highway Expansion. *Transportation, 26*, 249-262.
Department of Climate Change and Energy Efficiency. (2011). National Targets. Retrieved July 12, 2011, from http://www.climatechange.gov.au/government/reduce/national-targets.aspx
Desrochers, P., & Leppala, S. (2010). Rethinking "Jacobs Spillovers," or How Diverse Cities Actually Make Individuals More Creative and Economically Successful. In S. Goldsmith & L. Elizabeth (Eds.), *What We See: Advancing the observations of Jane Jacobs*. Oakland: New Village Press.
Dodson, J., & Sipe, N. (2008). *Shocking the Suburbs: Oil vulnerability in the Australian city*. Sydney: UNSW Press.
Duranton, G., & Overman, H. (2005). Testing for Localization Using Micro-Geographic Data. *The Review of Economic Studies, 72*(4), 1077-1106.
Duranton, G., & Puga, D. (2001). Nursery Cities: Urban diversity, process innovation, and the life-cyle of products. *The American Economic Review, 91*(5), 1454-1477.
Duranton, G., & Puga, D. (2005). From Sectoral to Functional Urban Specialisation. *Journal of Urban Economics, 57*(2), 343-373.

Duranton, G., & Turner, M. (2009). *The Fundamental Law of Road Congestion: Evidence from US cities* (SERC Discussion Paper 30). London: London School of Economics.
Ellison, G., Glaeser, E., & Kerr, W. (2007). *What Causes Industry Agglomeration? Evidence from coagglomeration patterns* (NBER Working Paper 13068). Cambridge, MA: National Bureau of Economic Research.
Eriksson, R., Lindgren, U., & Malmberg, G. (2008). Agglomeration Mobility: Effects of localisation, urbanisation, and scale of job changes. *Environment and Planning A, 40*, 2419-2434.
Florida, R. (2002). *The Rise of the Creative Class*. Cambridge, MA: Basic Books.
Frank, L., Saelens, B., Powell, L. & Chapman J. (2007). Stepping Towards Causation: Do built environments or neighborhood and travel preferences explain phsycial activity, driving and obesity? *Social Science and Medicine*, 65, 1898-1914.
Fujita, M., Krugman, P., & Mori, T. (1999). On the Evolution of Heirarchical Urban Systems. *European Economic Review, 43*, 209-251.
Fujita, M., Krugman, P., & Venables, A. (1999). *The Spatial Economy*. Cambridge, Massachusetts: The MIT Press.
Fujita, M., & Mori, T. (2005). Transport Development and the Evolution of Economic Geography. *Portuguese Economic Journal, 4*, 129:156.
Fujita, M., & Tabuchi, T. (1997). Regional Growth in Postwar Japan. *Regional Science and Urban Economics, 27*(643-670).
Fujita, M., & Thisse, J.-F. (2002). *Economics of Agglomeration: Cities, industrial location, and regional growth*. Cambridge: Cambridge University Press.
Fulton, L., Meszler, D., Noland, R., & Thomas, J. (2000). A Statistical Analysis of Induced Travel Effects in the U.S. Mid-Atlantic Region. *Journal of Transportation and Statistics, 3*(1), 1-14.
Glaeser, E. (1998). Are Cities Dying? *The Journal of Economic Perspectives, 12*(2), 139-160.
Glaeser, E. (2000). The Future of Urban Research: Nonmarket interactions. *Bookings-Wharton Papers on Urban Affairs 2000*, 101-138.
Glaeser, E. (2011). *Triumph of the City*. London: Macmillan.
Glaeser, E., & Gottlieb, J. (2009). The Wealth of Cities: Agglomeration economies and spatial equilibrium in the United States (NBER Working Paper 14806). Cambridge, MA: National Bureau of Economic Research.
Glaeser, E., Kallal, H., Scheinkman, J., & Shleifer, A. (1992). Growth in Cities. *The Journal of Political Economy, 100*(6), 1126-1152.
Glaeser, E., & Mare, D. (2001). Cities and Skills. *Journal of Labour Economics, 19*(2), 316-342.
Glaeser, E. & Saiz, A. (2004). The Rise of the Skilled City. *Brookings-Wharton Papers on Urban Affairs, 2004*, 47-105.
Glaeser, E., Scheinkman, J., & Shleifer, A. (1995). Economic Growth in a Cross-Section of Cities. *Journal of Monetary Economics, 36*(1), 117-143.
Gleeson, B., & Low, N. (2000). *Australian Urban Planning: New challenges, new agendas*. Sydney: Allen and Unwin.
Goodman, R., & Coote, M. (2007). Sustainable Urban Form and the Shopping Centre: And investigation of activity centres in Melbourne's growth areas. *Urban Policy and Research, 25*(1), 39-61.
Goodwin, P. (1996). Empirical evidence on induced traffic: A review and synthesis. *Transportation, 23*, 35-54.

Goodwin, P., Hass-Klau, C., & Cairns, S. (1997). *Evidence on the Effects of Road Capactiy Reduction on Traffic Levels*. London: Report prepared for the Department of the Environment, Transport and the Regions (DETR).
Government of New South Wales, DOP. (2010). *Metropolitan Plan for Sydney 2036*. Sydney, Deparment of Planning.
Government of Victoria, DOI. (2002). *Melbourne 2030*. Melbourne, Department of Infrastructure.
Government of Western Australia, WAPC. (2009). *Directions 2031 and Beyond*. Perth, Western Australian Planning Commission.
Graham, D. (2005). *Transport Investment, Agglomeration and Urban Productivity* (CTS Working Paper). London: Imperial College of London.
Graham, D. (2006). Variable Returns to Agglomeration and the Effect of Road Traffic Congestion. *Journal of Urban Economics, 62*, 103-120.
Graham, D. (2007a). *Agglomeration Economies and Transport Investment* (Discussion Paper 2007-11). London: Joint Transport Research Centre.
Graham, D. (2007b). Identifying Urbanisation and Localisation Externalities in Manufacturing and Service Industries. *Papers in Regional Science, 88*(1), 63-84.
Graham, D. (2007c). Agglomeration, Productivity and Transport Investment. *Journal of Transport Economics and Policy, 41*(3), 317-343.
Graham, D., Gibbons, S., & Martin, R. (2009). Transport Investment and the Distance Decay of Agglomeration Benefits (CTS Working Paper). London: Imperial College of London.
Green Building Council Australia. (2011). *Evolution: A year in green building 2011*. Sydney, NSW: Green Building Council Australia.
Hansen, M., Gillen, D., Dobbins, A., Huang, Y., & Puvathingal, M. (1993). *The Air Quality Impacts of Urban Highway Capacity Expansion: Traffic generation and land-use change*. Berkeley: University of California, Institute of Transportation Studies.
Hansen, W. (1959). How Accessibility Shapes Land Use. *Journal of the American Planning Association, 25*(2), 73-76.
Harris, G. (1954). The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers, 64*, 315-348.
Heanue, K. (1997). *Highway Capactiy and Induced Travel: Issues, evidence and implications*. Paper presented at the 1997 Annual Meeting of the Transport Research Board, Washington, DC.
Henderson, J. V. (1974). The Sizes and Types of Cities. *The American Economic Review, 64*(4), 640-656.
Henderson, J. V. (2003). Marshall's Scale Economies. *Journal of Urban Economics, 53*(2003), 1-28.
Henderson, V. (1986). Efficiency of Resource Usage and City Size. *Journal of Urban Economics, 19*, 47-70.
Hill,C., Griffiths, W., & Lim, G. (2007). *Princples of Econometrics*. New York: John Wiley and Sons Inc.
Holmes, T. (1999). Localization of Industry and Vertical Disintegration. *The Review of Economics and Statistics, 81(2)*, 314-325.
Huggins, R., Izushi, H., Davies, W., & Shougui, L. (2008). *World Knowledge Competitiveness Index 2008*. Cardiff: University of Wales Institute, Centre for International Competitiveness.

Infrastructure Australia. (2009). *National Infrastructure Priorities: Infrastructure for an economically, socially and environmentally sustainable future*. Canberra: Major Cities Unit.
Infrastructure Australia. (2010). *State of Australian Cities 2010*. Retrieved from http://www.infrastructureaustralia.gov.au/files/MCU_SOAC.pdf.
Jacobs, J. (1969). *The Economy of Cities*. New York: Random House.
Jara-Diaz, S. (2000). Allocation and Valuation of Travel Time Savings. In D. Hensher & K. Button (Eds.), *Handbook of Transport Modelling* (pp. 304-319). Amsterdam: Pergamon.
Johansson, B., & Quigley, J. (2004). Agglomeration and Networks in Spatial Economies. *Papers in Regional Science, 83*, 1-13.
Karlsson, C., & Johansson, B. (2004). *Towards a Dynamic Theory for the Spatial Knowledge Economy* (CESIS Working Paper 20). Stockholm: Royal Institute of Technology.
Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy, 99*(3), 483-499.
Krugman, P. (1993). First Nature, Second Nature and Metropolitan Location. *Journal of Regional Science, 33(2)*, 129-144.
Lampard. (1955). The History of Cities in the Economically Advanced Areas. *Economic Development and Cultural Change, 3*(2), 81-136.
Lee, D., Klein, L., & Camus, G. (1999). Induced Traffic and Induced Demand. *Transportation Research Record: Journal of the Transportation Research Board, 1659*, 68-75.
Losch, A. (1940). *The Economics of Location* (F. Jena, Trans.). Fischer Jena: English Translation (1954). New Haven, CT: Yale University Press.
Luk, J., & Chung, E. (1997). *Induced Demand and Road Investment: An initial appraisal* (Research Report 299). Vermont South, VIC: ARRB Transport Research.
Lyons, D. (2000). Embeddedness, Milieu and Innovation Among High-Technology Firms: A Richardson, Texas case study. *Environment and Planning A, 32*, 891-908.
Malmberg, A., & Maskell, P. (2001, 27 February). *The Elusive Concept of Localization Economies.* Paper presented at the AAG Annual Conference, New York.
Marshall, A. (1890). *Principles of Economics*. London: Macmillan (8th ed. published 1920).
Melitz, M., & Ottaviano, G. (2008). Market Size, Trade and Productivity *Review of Economic Studies, 71*(1), 295-316.
Melo, P., & Graham, D. (2009). *Agglomeration Economies and Labour Productivity: Evidence from longitudinal worker data for GB's Travel-to-Work areas* (SERC Discussion Paper 31). London: London School of Economics.
Melo, P., Graham, D., & Noland, R. (2009). A Meta-Analysis of Estimates of Urban Agglomeration Economies. *Regional Science and Urban Economics, 39*, 332-342.
Metz, D. (2008). The Myth of Travel Time Savings. *Transport Reviews, 28*(3), 321-336.
Meyer, J. (2000). *The Role of Industrial and Post-Industrial Cities in Economic Development*. Boston: Harvard University.
Montgomery, J. (2007). *The New Wealth of Cities: City dynamics and the Fifth Wave*. Hampshire: Ashgate Publishing Limited.
Moomaw, R. (1985). Firm Location and City Size: Reduced productivity advantages as a factor in the decline of manufacturing in urban areas. *Journal of Urban Economics, 17*, 73-89.
Mumford, L. (1961). *The City in History*. Great Britain: Secker and Warburg.
Nakamura, R. (1985). Agglomeration Economies in Urban Manufacturing Industries: A Case of Japanese Cities. *Journal of Urban Economics, 17*, 108-124.
Neutze, G. (1977). *Urban Development in Australia*. Hornsby, NSW: George Allen and Unwin Australia Pty Ltd.

Newman, P. (2011, 6th April). *The Perth Rail Transformation: Some political lessons learned.* Paper presented at the Light Rail 2011 Conference, Sydney, NSW.
Newman, P., & Jennings, I. (2008). *Cities as Sustainable Ecosystems*. Washington, D.C.: Island Press.
Newman, P., & Kenworthy, J. (1981). A Review of Urban Density Models: Towards a resolution of the conflict between populace and planner. *Human Ecology, 9*(3), 269-303.
Newman, P., & Kenworthy, J. (1991). Cities and Automobile Dependence. *Viewpoint, 8*(3), 170-173.
Newman, P., & Kenworthy, J. (1999). *Sustainability and Cities*. Washington, DC: Island Press.
Noland, R., & Cowart, W. (2000). Analysis of Metropolitan Highway Capacity and the Growth in Vehicles Miles of Travel. *Transportation, 27*(4), 363-390.
Noland, R., & Lem, L. (2002). A Review of the Evidence for Induced Travel and Changes in Transportation and Environmental Policy in the US and the UK. *Transportation Research Part D: Transport and Environment, 7*(1), 1-26.
O'Sullivan, A. (1996). *Urban Economics*. Chicago: Irwin.
Pearce, D., & Barbier, E. (2000). *Blueprint for a Sustainable Economy*. London: Earthscan Publications Ltd.
Population Reference Bureau. (2009). 2009 World Population Data Sheet. Washington, DC: Population Reference Bureau.
Puga, D. (2010). The Magnitude and Causes of Agglomeration Economies. *Journal of Regional Science, 50(1)*, 203-219.
Rice, P., & Venables, A. (2004). *Spatial Determinants of Productivity: Analysis for the regions of Great Britain* (CEP Discussion Paper 642). London: London School of Economics and Social Science.
Romer, P. (1990). Endogeneous Technological Change. *The Journal of Political Economy, 98*(5), S71-S102.
Rosenthal, S., & Strange, W. (2003). Geography, Industrial Organization, and Agglomeration. *The Review of Economics and Statistics, 85*(2), 377-393.
Rosenthal, S., & Strange, W. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4). Amsterdam: Elsevier.
Rosenthal, S., & Strange, W. (2008). The Attenuation of Human Capital Spillovers. *Journal of Urban Economics, 64*(2008), 373-389.
Ruiter, E., Loudon, W., Kern, C., & Bell, D. (1980). *The Vehicle-Miles of Travel/Urban Highway Supply Relationship*. Washington, DC: Transportation Research Board.
Saelens, B. (2003). Environmental Correlates of Walking and Cycling: Findings from the transportation, urban design and planning literatures. *Annals of Behavioural Medivine, 25(2)*, 80-91.
Saxenian, A. (1996). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
Schafer, A. (2000). Regularities in Travel Demand. *Journal of Transportation and Statistics, 3*(3), 1-31.
Scott, A. (1998). *Regions and the World Economy*. Oxford: Oxford University Press.
Scott, A., & Storper, M. (2003). Regions, Globalization, Development. *Regional Studies, 37*(579-593).
Searle, G., & Pritchard, B. (2005). Industry Clusters and Sydney's ITT Sector: Northern Sydney as 'Australia's Silicon Valley'? *Australian Geographer, 36*(2), 145-169.

Sheppard, D. (Producer). (2009, 20/01/2011) World Oil Demand Growth to be Led by Asia - IEA. *Reuters*. Retrieved from http://in.reuters.com/article/idINIndia-43841720091110?pageNumber=1

Short, J. (1991). *Imagined Country: Environment, culture and society*. New York: Routledge.

Simmie, J., Carpenter, J. Chadwick, A., Martin, R., & Wood, P. (2006). *State of English Cities: The competitive economic performance of english cities*. London: Department for Communities and Local Government.

Song, S. (1996). Some Tests of Alternative Accessibility Measures: A Population Density Approach. *Land Economics, 72*(4), 474-482.

Standing Advisory Committee on Trunk Road Assessment (SACTRA). (1994). *Trunk Roads and the Generation of Traffic*. London: UK Department of Transport.

Stretton, H. (1971). *Ideas for Australian Cities*. Melbourne: Georgian House.

Stigler, G. (1951). The Division of Labour is Limited by the Extent of the Market. *Journal of Political Economy, 59(3)*, 185-193.

Stock, J., & Watson, M. (2003). *Introduction to Econometrics*. Boston: Addison-Wesley.

Stock, J., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. Andrews & J. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in honour of Thomas Rothenberg*. New York: Cambridge University Press.

Stockols, D. (2003). Increasing the Health Promotive Capactiy of Human Environments. *American Journal of Health Promotion, 18*, 4-13.

Sturm, R., & Cohen, D. (2004). Suburban Sprawl and Physical and Mental Health. *Journal of the Royal Institute of Public Health, 118*, 448-496.

Teece, D. (1998). Capturing Value from Knowledge Assets: The New Economy, Markets for Know-How and Intangible Assets. *California Management Review, 40*(3), 55-79.

Trubka, R., Newman, P., & Bilsborough, D. (2010a). The Costs of Urban Sprawl - Infrastructure and Transportation. *Environment Design Guide, GEN83*, 1-6.

Trubka, R., Newman, P., & Bilsborough, D. (2010b). The Costs of Urban Sprawl - Predicting Transport Greenhouse Gases From Urban Form. *Environment Design Guide, GEN84*, 1-16.

Trubka, R., Newman, P., & Bilsborough, D. (2010c). The Costs of Urban Sprawl - Physical Activity Links to Healthcare Costs and Productivity. *Environment Design Guide, GEN85*, 1-13.

UK Department for Transport. (2006a). *Transport Statistics Bulletin: National Travel Survey: 2005*. London: Department for Transport.

UK Department for Transport. (2006b). *Transport, Wider Economic Benefits and Impacts of GDP*. London: Department for Transport,.

Venables, A. (2003). *Evaluating Urban Transport Improvements: Cost-benefit analysis in the presence of agglomeration and income taxation*: Paper prepared for the UK Department for Transport, June 2007.

Vogiatzis, N., Karuppannan, S., & Sivam, A. (2009). *Employment Distribution and Sustainable Urban Structure*. Paper presented at the State of Australian Cities 2009.

Voith, R. (1993). Changing Capitalization of CBD-Oriented Transportation Systems: Evidence from Philidelphia. *Journal of Urban Economics, 33*(3), 361-376.

von Thunen, J.H. (1826). *Der Isolierte Staat in Beziehung auf landwrirtschaft und Nationalekonomie*. Hamburg.

Warner, S. (1972). *The Urban Wilderness: A history of the American city*. New York: Harper and Row.

Wenban-Smith, H. (2009). *Spatial Costs in a Monocentric City (And Implications for Agglomeration)* (Discussion Paper 32). London: SERC.

Williams, K. (2000). Does Intensifying Cities Make Them More Sustainable? In K. Williams, M. Jenks & E. Burton (Eds.), *Achieving Sustainable Urban Form*. London: Routledge.

Zahavi, Y., & Talvitie, A. (1980). Regularities in Travel Time and Money Expenditures. *Transportation Research Record*(750), 13-19.

APPENDICES

Appendix A: Explaining How Endogeneity Can Bias Parameter Estimates

To explain the implications of a bias arising from simultaneous equations, or the presence of reverse causality, take for instance the subject of this thesis in which economic density positively affects wages. There is some reason to believe that the opposite may also be true – that productive locations may create a gravitas for firms to locate. If this is the case, we can assume that two equations are being determined at the same time. Equation 1a is a simple model specification where a level of output in a given location (Y_i) is determined by an intercept term (β_o), the location's degree of economic concentration (A_i), and an error term (e_i). Concurrently, a location's economic density (A_i) might be determined by a separate intercept term (β_1), the level of productivity in the area (Y_i), and an error term of its own (μ_i). This is expressed in equation 1b.

1a) $Y_i = \beta_o + \gamma A_i + e_i$ 1b) $A_i = \beta_1 + \alpha Y_i + \mu_i$

If we substitute the right-hand side of 1b in for A_i in 1a, we get the equation expressed in 2a. Similarly, if we substitute the right-hand side of 1a in for Y_i in 1b, we get what is expressed in equation 2b.

2a) $Y_i = \beta_o + \gamma\left(\beta_1 + \alpha Y_i + \mu_i\right) + e_i$

2b) $A_i = \beta_1 + \alpha\left(\beta_o + \gamma A_i + e_i\right) + \mu_i$

Equations 3a and 3b simply restate equations 2a and 2b, respectively, but are reorganized to isolate for the error terms.

3a) $Y_i = \beta_o + \gamma\left(\beta_1 + \alpha Y_i\right) + \left(\gamma \mu_i + e_i\right)$

3b) $A_i = \beta_1 + \alpha(\beta_o + \gamma A_i) + (\alpha e_i + \mu_i)$

The issue of simultaneity bias arises because if both expressions 1a and 1b are true, then the coefficients γ and α cannot equal zero, leaving $\gamma(\beta_1 + \alpha Y_i)$ to be correlated with the e_i component of the error term in 3a and $\alpha(\beta_o + \gamma A_i)$ to be correlated with the μ_i part of the error term in 3b. This creates an endogeneity issue where one of the right-hand side variables is being determined from within the system. We can say that the values of Y_i and A_i are being jointly determined and that there is a feedback mechanism working between them – hence the term "reverse-causality" being ascribed to the situation.

In this basic example, unbiased parameters cannot be estimated unless at least one of the equations is 'identified', which means that there must be at least one exogenous variable that appears in one equation but not in the other. If this can be satisfied, then two-stage least squares (2SLS) can be used to estimate the unbiased parameter γ in equation 1a. The ultimate purpose of 2SLS is to use one or more instruments to predict values of A_i, represented as $\hat{A}_i$, and substitute them in for A_i in equation 1a. If the instrument is not correlated with the disturbance term in the equation for Y_i, then it will produce an estimate of A_i that is uncorrelated with the error term and produce an unbiased estimate of Y_i. In other words, a suitable instrument for use in 2SLS should be strongly correlated with the part of A_i that is not correlated with Y_i. This criterion is typically achieved by using a lagged version of the endogenous variable A_i under the assumption that its historic levels will be associated with its currents levels, but not have any impact on present day productivity.

Appendix B: Proof for Log-Log Models Generating Elasticities as Coefficients

The generic formula for calculating an elasticity, which is a ratio of the percentage change in a y variable given a percentage change in a x variable, is as follows:

$$\varepsilon = \frac{\Delta x}{\Delta y} \cdot \frac{y}{x} = \frac{\frac{\Delta x}{x}}{\frac{\Delta y}{y}} = \frac{\frac{x_2 - x_1}{x_1}}{\frac{y_2 - y_1}{y_1}}$$

or

$$\varepsilon = \frac{d \ln y}{d \ln x} = \frac{dy}{dx} \cdot \frac{x}{y}$$

Proof:

$$\frac{\partial \ln y}{\partial \ln x} = \frac{d \ln y}{dy} \cdot \frac{dy}{dx} \Big/ \frac{d \ln x}{dx}$$

$$= \frac{1}{y} \cdot \frac{dy}{dx} \Big/ \frac{1}{x}$$

$$= \frac{x}{y} \cdot \frac{dy}{dx} = \varepsilon$$

Appendix C: A Summary of SQL Code Used for Part 2 Analyses

Part 2 Analyses, as described in Chapter 5, utilized significantly larger datasets than those in Part 1 and as such needed more powerful software suitable to the task of handling them. The software used was called PostreSQL, which was coupled with the PostGIS add-on to allow spatial files to be stored in the database and spatial calculations to be carried out with them. Code, written in the process of utilizing the spatially enabled database, is called SQL script and in the circumstance of this work, scripts were authored and executed using PGAdminIII, which is a graphic user interface (GUI) linked to the Postgres database management system (DBMS). In this appendix, a number of script commands that were authored for setting up tables and importing data will be explained; however, the commands documented here will by no means be a comprehensive account of all the code that was actually written. Datasets and tables were constantly 'tweaked' to satisfy conditions that had to be met for merging data from different tables. Quite often these actions were done retroactively, as the limitations of the formats of the data were not made evident until scripts combining sources had to be authored. This appendix will proceed by giving examples of the code for a number of fundamental actions; however, for comprehensive detail of commands and functions one should refer to Postgres documentation site (http://www.postgresql.org/docs/).

Importing Data:

The data provided by the ABS was in the .txt format, which does not impose a limit on the number of columns that can be contained in a file. The database was constructed such that one table strictly contained the city employment zone codes and city along with arbitrarily set primary key IDs. Subsequent tables storing the ABS data were organized such that each of income, occupation figures, education figures, and mean age had their data stored in their own table. The initial table was established by importing all the data from the 'mean age' file and then dropping the columns of mean age, industry and total employment so that only the employment zone codes and industry names remained. The following lines of script accomplished this.

```
DROP TABLE IF EXISTS import.cities_wdzs CASCADE; -- drop the table to start from zero
CREATE TABLE import.cities_wdzs (id serial PRIMARY KEY, industry TEXT, wdz_code
        TEXT, "total employed" integer, "mean age" integer); --create new, empty table
COMMENT ON TABLE import.cities_wdzs IS 'List of Sydney and Melbourne wdzs along with
        city names'; -- comment on the table
COPY import.cities_wdzs (industry, wdz_code, "total employed", "mean age") FROM
        '/Users/rotru/files 2/import/2006 IND06 Mean Age.txt';
ALTER TABLE import.cities_wdzs DROP COLUMN "total employed"; --ALTER TABLE
        import.cities_wdzs DROP COLUMN "mean age";
ALTER TABLE import.cities_wdzs ADD COLUMN city_name TEXT;
UPDATE import.cities_wdzs AS e SET city_name = 'melbourne' WHERE substring(e.wdz_code
        FROM 1 FOR 1)::TEXT = 2::TEXT;
UPDATE import.cities_wdzs AS e SET city_name = 'sydney' WHERE substring(e.wdz_code
        FROM 1 FOR 1)::TEXT = 1::TEXT;
DELETE FROM import.cities_wdzs WHERE id > 5183;
ALTER TABLE import.cities_wdzs DROP COLUMN industry;
```

The command that actually imports the data is the ‘COPY’ command, where public access to where the file is stored must be enabled, otherwise Postgres will not be able retrieve the data. Using the table on industry-specific mean income as an example, appearing below is how data for the other data-types was imported.

```
DROP TABLE IF EXISTS import.cities_income CASCADE; -- drop the table to start from zero
CREATE TABLE import.cities_income (id serial PRIMARY KEY, industry TEXT, wdz_code
        TEXT, "total employed" INTEGER, "mean income" INTEGER); --create new, empty
        table
COMMENT ON TABLE import.cities_income IS 'Mean income by industry for sydney and
        melbourne WDZs'; -- comment on the table
COPY import.cities_income (industry, wdz_code, "total employed", "mean income") FROM
        '/Users/rotru/files 2/import/2006 IND06 Mean Income.txt'; --import income data from
        text file
```

Exporting the Data for subsequent Analysis in GRETL:

Data on each industry was monotonically shifted (but only if the data type was expressed as a share), log-transformed, compiled and exported into separate .csv files to keep the data uncluttered. All this was accomplished in one long script for each city/industry combination. One such script appears below for the industry of *Food Product Manufacturing* in the city of Melbourne. Exporting data for a different industry or for another city meant changing the references after the 'WHERE' command from '11 Food Product Manufacturing' and 'melbourne' to a different desired industry or city, respectively.

```
COPY (
    SELECT substring(wdz.wdz_code FROM 7 FOR 4),
    ln(inc."mean income") AS ln_income,
    ln(ed.total_ed_2) AS ln_ed,
    (ln(ed.total_ed_2))^2 AS ln_ed_sq,
    ln(1 + occ."share of managers and professionals") AS ln_occ_1,
    ln(1 + occ."share of tech trade and labour workers") AS ln_occ_2,
    ln(1 + occ."share of community personal service and sales workers") AS ln_occ_3,
    ln(1 + edu."share of tertiary") AS ln_edu_1,
    ln(1 + edu."share of non-tertiary") AS ln_edu_2,
    ln(1 + edu."share of no education") AS ln_edu_3,
    ln(age."mean age") AS ln_age,
    (ln(age."mean age"))^2 AS ln_age_sq
    FROM import.cities_wdzs AS wdz
    LEFT JOIN import.cities_income AS inc
    ON substring(wdz.wdz_code FROM 7 FOR 4)::INTEGER =
    substring(inc.wdz_code FROM 7 FOR 4)::INTEGER
    LEFT JOIN shp.melb_wdz AS ed
    ON substring(wdz.wdz_code FROM 7 FOR 4)::INTEGER =
    ed.vicdznp06::INTEGER LEFT JOIN import.cities_occupation AS occ
    ON substring(wdz.wdz_code FROM 7 FOR 4)::INTEGER =
    substring(occ.wdz_code FROM 7 FOR 4)::INTEGER
```

```
LEFT JOIN import.cities_education AS edu
ON substring(wdz.wdz_code FROM 7 FOR 4)::INTEGER =
substring(edu.wdz_code FROM 7 FOR 4)::INTEGER
LEFT JOIN import.cities_age AS age
ON substring(wdz.wdz_code FROM 7 FOR 4)::INTEGER =
substring(age.wdz_code FROM 7 FOR 4)::INTEGER
WHERE wdz.city_name = 'melbourne'
AND inc.city_name = 'melbourne'
AND occ.city_name = 'melbourne'
AND edu.city_name = 'melbourne'
AND age.city_name = 'melbourne'
AND inc.industry = '11 Food Product Manufacturing'::TEXT
AND occ.industry = '11 Food Product Manufacturing'::TEXT
AND edu.industry = '11 Food Product Manufacturing'::TEXT
AND age.industry = '11 Food Product Manufacturing'::TEXT
AND inc."mean income" > 0
AND age."mean age" > 0
AND ed.total_ed_2 > 0
AND occ.share_sum > 0
AND edu.share_sum > 0
AND ed.is_duplicate = FALSE
ORDER BY substring(wdz.wdz_code FROM 7 FOR 4) ASC )
TO '/Users/romantrubka/Desktop/Melbourne Data – Linear
ED/ln_melbourne_industry11.csv'
WITH NULL AS '0' CSV HEADER;
```

The numerous 'AND' subquery expressions, giving the condition that the various control variables must be greater than zero, are one example of a 'tweak' that was required on the code as a result of the perturbation practices of the ABS. Since cell values were randomized, the number of observations comprising the means in the various data tables varied and in some cases equaled zero. This would have caused a major issue in the data-sets because if income, for instance, was randomized such that a zone had zero employees in a given industry yet a positive number of employees in the other controls, then mean income would be reported as being zero while data would exist reporting on magnitudes of the other controls. This was one of the drawbacks of using ABS datasets and. The only remedy was to prevent the zones that reported zero employment in any of the control variable types from being exported.

Calculating Travel-Time Effective Density for Sydney:

The SQL code calculating effective density values that were weighted by Euclidean distance was presented in the overview of Analysis 2-1. The set-up for this calculation required a table to be prepared for storing the Travel Zone ID codes, the effective density values and the values from calculations that estimated travel-times for traversing the radii of individual Travel Zones. Also needing calculation were the weighted average travel times as the TDC travel-time data were provided separately for a.m. peak, daytime interpeak, and p.m. peak periods. Description of these scripts will not be given here, only of the script that carried out the travel-time weighted effective density calculation. This script appears is given below.

```
UPDATE import.sydney_tt_ed AS ed
    SET surrounding_density = temp.surrounding_density FROM (
    SELECT ed.wdz06_id, tt.origin,
    sum(tt.destination_employment / tt.weighted_avg_traveltime) AS
    surrounding_density
    FROM import.sydney_tt_ed AS ed
    LEFT JOIN import.traveltime_peakam_ip_peakpm AS tt
    ON ed.wdz06_id::INTEGER = tt.origin::INTEGER
    GROUP BY ed.wdz06_id, tt.origin) AS temp,
    import.traveltime_peakam_ip_peakpm AS tt
    WHERE ed.wdz06_id::INTEGER = temp.origin::INTEGER;
```

This component only calculates surrounding area contributions to effective density. The values of own-area employment in each WDZ that had been divided by their estimated travel-time radii were then added to the outputs of this script.

Appendix D: Sydney Preferred Industry Results from Linear and Quadratic Model Specification

Industry	Elasticity		Ln(ED)	S.E.	P-Value		Ln(ED)^2	S.E.	P-Value	Adjusted R-Squared	P-Value (F)	Sample Size	Mean Ln(ED)	Returns to ED
11 Food Product Manufacturing	0.020		0.020	0.036	0.573		-	-	-	0.311	3.42E-56	740	11.929	Constant
13 Textile, Leather, Clothing and Footwear Manufacturing	0.180	***	0.180	0.042	0.000		-	-	-	0.209	3.39E-27	588	11.992	Constant
18 Basic Chemical and Chemical Product Manufacturing	0.172	**	2.269	1.076	0.036	*	-0.087	0.045	0.051	0.324	1.62E-24	326	12.016	Diminishing
21 Primary Metal and Metal Product Manufacturing	0.020		0.020	0.040	0.624		-	-	-	0.299	1.20E-23	341	11.835	Constant
22 Fabricated Metal Product Manufacturing	0.018		0.018	0.030	0.536		-	-	-	0.112	7.33E-09	392	11.845	Constant
23 Transport Equipment Manufacturing	0.095	**	2.662	1.107	0.017	**	-0.108	0.047	0.022	0.407	1.01E-39	386	11.92	Diminishing
24 Machinery and Equipment Manufacturing	0.060	**	0.060	0.024	0.015		-	-	-	0.262	3.99E-40	658	11.997	Constant
30 Building Construction	0.166	***	0.166	0.014	0.000		-	-	-	0.165	3.10E-61	1654	11.945	Constant
31 Heavy and Civil Engineering Construction	0.160	***	0.160	0.027	0.000		-	-	-	0.283	2.11E-30	454	11.892	Constant
33 Basic Material Wholesaling	0.100	***	0.100	0.028	0.000		-	-	-	0.234	1.15E-38	721	11.901	Constant
34 Machinery and Equipment Wholesaling	0.075	***	0.075	0.024	0.002		-	-	-	0.183	2.14E-29	733	12.058	Constant
41 Food Retailing	0.087	***	0.087	0.017	0.000		-	-	-	0.147	1.30E-42	1327	11.953	Constant
42 Other Store-Based Retailing	0.138	***	0.138	0.013	0.000		-	-	-	0.203	4.68E-79	1687	11.973	Constant
44 Accommodation	0.121	**	0.698	0.341	0.041	*	-0.024	0.014	0.087	0.213	1.46E-34	732	12.146	Diminishing
45 Food and Beverage Services	0.065	***	-0.867	0.237	0.000	***	0.039	0.010	0.000	0.329	5.00E-141	1694	12.001	Increasing
54 Publishing (except Internet and Music Publishing)	0.218	***	0.218	0.029	0.000		-	-	-	0.186	1.62E-24	603	12.172	Constant
55 Motion Picture and Sound Recording Activities	0.217	***	0.217	0.041	0.000		-	-	-	0.488	2.51E-58	430	12.233	Constant
59 Internet Service Providers, Web Search Portals, and Data Processing Services	0.132	***	0.132	0.033	0.000		-	-	-	0.256	1.19E-12	220	12.46	Constant
62 Finance	0.149	*	-0.593	0.311	0.057	**	0.031	0.012	0.014	0.403	4.00E-104	979	12.137	Increasing
63 Insurance	0.146	***	0.146	0.016	0.000		-	-	-	0.399	4.11E-50	487	12.271	Constant
64 Auxiliary Finance and Insurance Services	0.164	***	0.016	0.016	0.000		-	-	-	0.247	1.89E-53	925	12.136	Constant
67 Property Operators and Real Estate Services	0.129	***	0.129	0.015	0.000		-	-	-	0.233	1.24E-61	1136	12.06	Constant
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)	0.139	**	-0.486	0.218	0.026	***	0.026	0.009	0.003	0.354	8.70E-171	1863	11.973	Increasing
75 Public Administration	0.127	**	-0.657	0.330	0.047	**	0.033	0.013	0.014	0.359	5.17E-95	1036	12.063	Increasing
77 Public Order, Safety and Regulatory Services	0.024	*	0.848	0.477	0.076	*	-0.034	0.019	0.077	0.138	1.62E-25	882	12.089	Diminishing
84 Hospitals	0.107	***	-1.603	0.515	0.002	***	0.072	0.022	0.001	0.275	1.89E-29	471	11.956	Increasing
85 Medical and Other Health Care Services	0.081	***	0.081	0.013	0.000		-	-	-	0.154	1.85E-52	1538	11.983	Constant
87 Social Assistance Services	0.144	***	-0.843	0.324	0.009	***	0.041	0.013	0.002	0.165	1.91E-58	1601	11.916	Increasing
89 Heritage Activities	0.054		0.054	0.041	0.186		-	-	-	0.126	5.70E-05	194	12.205	Constant
90 Creative and Performing Arts Activities	0.244	***	0.244	0.041	0.000		-	-	-	0.315	1.61E-41	546	12.144	Constant

Note: The full model specification (with ED-sq) was first run and then the linear restriction of whether the coefficient on Ln(ED)^2 = 0 was tested. If this restriction was rejected, the squared term was omitted and a reduced model was estimated.

Appendix E: Melbourne Preferred Industry Results from Linear and Quadratic Model Specifications

Industry	Elasticity		Ln(ED)	S.E.	P-Value		Ln(ED)^2	S.E.	P-Value	Adjusted R-Squared	P-Value (F)	Sample Size	Mean Ln(ED)	Returns to ED
11 Food Product Manufacturing	0.096	***	0.096	0.035	0.006		-	-	-	0.362	2.34E-75	815	11.527	Constant
13 Textile, Leather, Clothing and Footwear Manufacturing	0.120	***	0.120	0.034	0.001		-	-	-	0.181	3.08E-26	666	11.624	Constant
18 Basic Chemical and Chemical Product Manufacturing	0.071	*	-2.447	1.253	0.052	**	0.108	0.053	0.043	0.361	3.36E-27	315	11.630	Increasing
21 Primary Metal and Metal Product Manufacturing	0.018		0.018	0.046	0.694		-	-	-	0.140	2.82E-10	358	11.446	Constant
22 Fabricated Metal Product Manufacturing	0.140	*	-1.690	1.012	0.096	*	0.080	0.044	0.069	0.106	1.34E-08	417	11.446	Increasing
23 Transport Equipment Manufacturing	0.138	***	0.138	0.034	0.000		-	-	-	0.156	6.50E-16	486	11.516	Constant
24 Machinery and Equipment Manufacturing	0.124	***	0.124	0.027	0.000		-	-	-	0.165	3.39E-23	655	11.608	Constant
30 Building Construction	0.153		-0.304	0.288	0.292	*	0.020	0.012	0.098	0.300	1.70E-109	1484	11.488	Increasing
31 Heavy and Civil Engineering Construction	0.131	***	0.131	0.031	0.000		-	-	-	0.164	9.19E-16	457	11.496	Constant
33 Basic Material Wholesaling	0.181		-0.607	0.465	0.192	*	0.034	0.020	0.088	0.387	1.63E-76	767	11.457	Increasing
34 Machinery and Equipment Wholesaling	0.154	***	0.154	0.024	0.000		-	-	-	0.246	1.81E-38	677	11.669	Constant
41 Food Retailing	0.041	**	0.041	0.020	0.041		-	-	-	0.202	1.93E-55	1202	11.543	Constant
42 Other Store-Based Retailing	0.148	***	0.148	0.017	0.000		-	-	-	0.183	3.34E-63	1523	11.540	Constant
44 Accommodation	0.155	***	0.155	0.023	0.000		-	-	-	0.158	7.26E-20	590	11.644	Constant
45 Food and Beverage Services	0.057	***	-1.849	0.353	0.000	***	0.083	0.015	0.000	0.242	1.19E-85	1503	11.536	Increasing
54 Publishing (except Internet and Music Publishing)	0.169	***	0.169	0.028	0.000		-	-	-	0.141	7.91E-13	440	11.772	Constant
55 Motion Picture and Sound Recording Activities	0.200	***	0.200	0.053	0.000		-	-	-	0.410	1.11E-35	336	11.814	Constant
59 Internet Service Providers, Web Search Portals, and Data Processing Services	0.219	*	2.879	1.561	0.067	*	-0.111	0.063	0.082	0.228	1.06E-09	200	11.992	Decreasing
62 Finance	0.111		-0.664	0.460	0.149	*	0.033	0.019	0.080	0.751	1.70E-225	772	11.671	Increasing
63 Insurance	0.095	**	-1.669	0.744	0.026	**	0.075	0.030	0.014	0.308	2.25E-28	398	11.776	Increasing
64 Auxiliary Finance and Insurance Services	0.176	***	0.176	0.017	0.000		-	-	-	0.300	3.50E-56	772	11.673	Constant
67 Property Operators and Real Estate Services	0.135	***	0.135	0.019	0.000		-	-	-	0.150	3.50E-28	871	11.632	Constant
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)	0.163	*	-0.513	0.270	0.057	***	0.029	0.011	0.009	0.312	1.10E-124	1603	11.533	Increasing
75 Public Administration	0.181	***	0.181	0.014	0.000		-	-	-	0.299	8.82E-72	984	11.584	Constant
77 Public Order, Safety and Regulatory Services	0.016		0.016	0.020	0.416		-	-	-	0.621	1.10E-136	675	11.666	Constant
84 Hospitals	0.113	*	-1.055	0.599	0.079	**	0.050	0.025	0.050	0.213	1.52E-23	509	11.658	Increasing
85 Medical and Other Health Care Services	0.065	*	-0.613	0.323	0.058	**	0.029	0.013	0.030	0.168	4.77E-50	1352	11.589	Increasing
87 Social Assistance Services	0.106		-0.700	0.432	0.106	*	0.035	0.019	0.061	0.182	5.59E-55	1353	11.549	Increasing
89 Heritage Activities	0.111	**	0.111	0.046	0.018		-	-	-	0.414	7.17E-17	163	11.706	Constant
90 Creative and Performing Arts Activities	0.154	***	0.154	0.047	0.001		-	-	-	0.081	7.01E-08	476	11.733	Constant

Note: The full model specification (with ED-sq) was first run and then the linear restriction of whether the coefficient on Ln(ED)^2 = 0 was tested. If this restriction was rejected, the squared term was omitted and a reduced model was estimated.

Appendix F: Sydney 2SLS Regression Results (with Robust Standard Errors)

Industry		Elasticity	S.E.	P-Value	Adj. R-Squared	F-Value	N	Hausman P-Value	Weak Instrument Test F-Stat
11 Food Product Manufacturing									
Ln(WDZ Area) Instrument		0.039	0.048	4.14E-01	0.375	5.83E-62	740	9.67E-01	1194.17
WDZ Area Instrument		-0.133	0.067	0.048	0.358	2.14E-56	740	1.34E-01	15.3
13 Textile, Leather, Clothing and Footwear Manufacturing									
Ln(WDZ Area) Instrument	**	0.122	0.062	0.049	0.157	2.75E-22	588	**5.53E-02**	704.2
WDZ Area Instrument		0.000	0.078	1.000	0.137	5.34E-21	588	2.37E-01	8.8
18 Basic Chemical and Chemical Product Manufacturing									
Ln(WDZ Area) Instrument	*	0.130	0.074	0.080	0.225	4.23E-17	326	1.82E-01	369.9
WDZ Area Instrument		0.029	0.126	0.815	0.205	3.51E-15	326	2.95E-01	14.7
21 Primary Metal and Metal Product Manufacturing									
Ln(WDZ Area) Instrument	**	0.141	0.072	0.048	0.232	4.37E-18	341	9.74E-01	342.7
WDZ Area Instrument	***	0.503	0.150	0.001	0.166	4.88E-16	341	**7.02E-03**	28.6
22 Fabricated Metal Product Manufacturing									
Ln(WDZ Area) Instrument		0.088	0.068	0.195	0.097	9.29E-09	392	8.77E-01	443.8
WDZ Area Instrument		0.266	0.230	0.248	0.072	3.34E-09	392	2.12E-01	6.2
23 Transport Equipment Manufacturing									
Ln(WDZ Area) Instrument	**	0.136	0.057	0.017	0.219	4.74E-20	386	3.70E-01	424.0
WDZ Area Instrument	**	0.279	0.115	0.016	0.196	7.02E-20	386	1.90E-01	12.4
24 Machinery and Equipment Manufacturing									
Ln(WDZ Area) Instrument	***	0.148	0.045	0.001	0.162	1.95E-19	658	1.71E-01	1018.3
WDZ Area Instrument	*	0.174	0.095	0.067	0.160	1.17E-18	658	5.59E-01	19.0
30 Building Construction									
Ln(WDZ Area) Instrument	***	0.168	0.019	0.000	0.152	3.06E-48	1654	7.17E-01	2372.6
WDZ Area Instrument	*	0.132	0.076	0.082	0.149	2.33E-29	1654	5.70E-01	15.9
31 Heavy and Civil Engineering Construction									
Ln(WDZ Area) Instrument	***	0.164	0.038	0.000	0.235	1.53E-25	454	7.47E-01	789.2
WDZ Area Instrument		-0.061	0.101	0.547	0.154	1.83E-20	454	**1.91E-02**	14.5
33 Basic Material Wholesaling									
Ln(WDZ Area) Instrument	***	0.149	0.037	0.000	0.164	5.62E-25	721	8.42E-01	1048.6
WDZ Area Instrument	*	0.139	0.074	0.059	0.164	6.24E-21	721	9.59E-01	44.5
34 Machinery and Equipment Wholesaling									
Ln(WDZ Area) Instrument	*	0.076	0.046	0.098	0.147	2.46E-22	733	5.97E-01	1423.6
WDZ Area Instrument	*	0.342	0.193	0.077	0.105	9.64E-02	733	**9.64E-02**	4.6
41 Food Retailing									
Ln(WDZ Area) Instrument	***	0.107	0.030	0.000	0.176	2.22E-28	1327	7.95E-01	1839.1
WDZ Area Instrument	*	0.240	0.133	0.071	0.159	3.73E-25	1327	1.71E-01	8.8
42 Other Store-Based Retailing									
Ln(WDZ Area) Instrument	***	0.108	0.024	0.000	0.187	1.28E-37	1687	1.39E-01	2708.1
WDZ Area Instrument		0.104	0.139	0.454	0.187	8.09E-35	1687	7.95E-01	18.7
44 Accommodation									
Ln(WDZ Area) Instrument	***	0.151	0.025	0.000	0.140	2.43E-15	732	4.01E-01	1428.4
WDZ Area Instrument	*	0.178	0.107	0.098	0.138	1.31E-09	732	7.27E-01	13.2
45 Food and Beverage Services									
Ln(WDZ Area) Instrument	***	0.061	0.016	0.000	0.310	3.70E-101	1694	1.36E-01	3087.8
WDZ Area Instrument		0.027	0.054	0.608	0.306	4.60E-89	1694	4.85E-01	11.0
54 Publishing (except Internet and Music Publishing)									
Ln(WDZ Area) Instrument	***	0.309	0.040	0.000	0.258	6.43E-23	603	4.04E-01	1187.2
WDZ Area Instrument	***	0.524	0.086	0.000	0.237	4.15E-21	603	1.82E-01	4.5
55 Motion Picture and Sound Recording Activities									
Ln(WDZ Area) Instrument	**	0.129	0.053	0.014	0.311	1.89E-41	430	**9.49E-02**	668.1
WDZ Area Instrument		0.124	0.086	0.148	0.310	1.52E-39	430	7.61E-01	23.2
59 Internet Service Providers, Web Search Portals, and Data Processing Services									
Ln(WDZ Area) Instrument		0.104	0.064	0.101	0.153	5.12E-07	220	1.40E-01	928.0
WDZ Area Instrument		0.045	0.066	0.501	0.143	3.96E-06	220	5.31E-01	4.3
62 Finance									
Ln(WDZ Area) Instrument	***	0.187	0.023	0.000	0.198	1.41E-40	979	8.66E-01	2089.6
WDZ Area Instrument	*	0.202	0.116	0.081	0.197	8.11E-02	979	9.24E-01	3.9
63 Insurance									
Ln(WDZ Area) Instrument	***	0.131	0.030	0.000	0.281	7.93E-37	487	1.48E-01	1371.3
WDZ Area Instrument	*	0.129	0.071	0.068	0.281	2.62E-33	487	8.08E-01	11.7
64 Auxiliary Finance and Insurance Services									
Ln(WDZ Area) Instrument	***	0.206	0.030	0.000	0.178	4.41E-29	925	9.17E-01	1989.9
WDZ Area Instrument		0.177	0.121	0.145	0.177	3.39E-16	925	7.93E-01	13.5
67 Property Operators and Real Estate Services									
Ln(WDZ Area) Instrument	***	0.143	0.025	0.000	0.156	2.77E-37	1136	7.15E-01	2111.2
WDZ Area Instrument	***	0.221	0.062	0.000	0.149	1.29E-27	1136	3.96E-01	11.0
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)									
Ln(WDZ Area) Instrument	***	0.129	0.019	0.000	0.208	6.74E-87	1863	**1.25E-03**	2592.7
WDZ Area Instrument	*	0.160	0.091	0.077	0.210	6.79E-61	1863	9.97E-01	6.5
75 Public Administration									
Ln(WDZ Area) Instrument	***	0.178	0.023	0.000	0.192	1.01E-41	1036	**3.32E-02**	2197.5
WDZ Area Instrument		0.161	0.125	0.199	0.193	1.99E-25	1036	9.24E-01	20.9
77 Public Order, Safety and Regulatory Services									
Ln(WDZ Area) Instrument		0.020	0.025	0.421	0.120	7.14E-19	882	8.34E-01	1854.8
WDZ Area Instrument	*	0.111	0.060	0.067	0.105	9.90E-21	882	4.20E-01	4.5
84 Hospitals									
Ln(WDZ Area) Instrument	***	0.133	0.036	0.000	0.252	5.92E-23	471	8.60E-01	329.8
WDZ Area Instrument	***	0.268	0.045	0.000	0.230	4.60E-20	471	3.15E-01	22.1
85 Medical and Other Health Care Services									
Ln(WDZ Area) Instrument	***	0.117	0.021	0.000	0.182	2.75E-39	1538	2.35E-01	2358.1
WDZ Area Instrument		0.247	0.165	0.135	0.160	3.44E-32	1538	**4.57E-02**	13.1
87 Social Assistance Services									
Ln(WDZ Area) Instrument	***	0.189	0.024	0.000	0.154	1.37E-36	1601	**8.88E-02**	1864.3
WDZ Area Instrument		0.120	0.140	0.394	0.150	9.94E-17	1601	4.70E-01	17.5
89 Heritage Activities									
Ln(WDZ Area) Instrument		0.085	0.052	0.106	0.109	9.27E-03	194	8.63E-01	417.5
WDZ Area Instrument		-0.024	0.071	0.740	0.086	1.49E-02	194	5.06E-01	9.2
90 Creative and Performing Arts Activities									
Ln(WDZ Area) Instrument	***	0.266	0.061	0.000	0.068	2.98E-08	546	4.67E-01	922.1
WDZ Area Instrument		0.399	0.248	0.108	0.064	1.65E-06	546	2.78E-01	30.5

Note (1): * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01

Note (2): Hausmann P-values in bold indicate regressions in which endogeneity was detected at a significance level of 0.10

Appendix G: Melbourne 2SLS Regression Results (with Robust Standard Errors)

Industry		Elasticity	S.E.	P-Value	Adj. R-Squared	F-Value	N	Hausman P-Value	Weak Instrument F-Stat	Sargan validity test p-values
11 Food Product Manufacturing										
Ln(Road Network Density) Instrument		-0.118	0.075	0.114	0.340	2.23E-63	815	**1.07E-03**	163.2	
Road Network Density Instrument	*	-0.131	0.073	0.071	0.337	9.59E-64	815	**1.48E-04**	301.2	0.032
Ln(WDZ Area) Instrument		0.009	0.051	0.861	0.362	6.84E-63	815	**2.39E-02**	596.5	
WDZ Area Instrument		0.053	0.065	0.414	0.364	4.83E-63	815	6.11E-01	78.2	
13 Textile, Leather, Clothing and Footwear Manufacturing										
Ln(Road Network Density) Instrument		-0.021	0.081	0.796	0.128	8.08E-18	666	**2.24E-02**	206.2	
Road Network Density Instrument		0.071	0.068	0.298	0.143	4.12E-18	666	2.98E-01	296.1	0.006
Ln(WDZ Area) Instrument		0.077	0.062	0.217	0.143	3.02E-18	666	**6.72E-02**	746.7	
WDZ Area Instrument		0.161	0.119	0.175	0.145	3.62E-18	666	7.70E-01	46.6	
18 Basic Chemical and Chemical Product Manufacturing										
Ln(Road Network Density) Instrument	**	-0.228	0.112	0.042	0.146	7.16E-31	315	**2.57E-03**	77.6	
Road Network Density Instrument	**	-0.191	0.094	0.042	0.158	9.91E-33	315	**1.77E-03**	114.7	0.030
Ln(WDZ Area) Instrument		0.015	0.076	0.844	0.201	1.73E-26	315	5.79E-01	206.4	
WDZ Area Instrument		-0.132	0.152	0.385	0.175	1.28E-27	315	2.54E-01	14.1	
21 Primary Metal and Metal Product Manufacturing										
Ln(Road Network Density) Instrument		-0.151	0.107	0.160	0.118	1.80E-08	358	**4.08E-02**	63.3	
Road Network Density Instrument		0.099	0.099	0.314	0.132	1.51E-08	358	**8.27E-02**	104.8	0.151
Ln(WDZ Area) Instrument		0.032	0.075	0.672	0.150	6.02E-09	358	9.37E-01	267.1	
WDZ Area Instrument		0.004	0.072	0.961	0.149	3.87E-09	358	7.88E-01	4.7	
22 Fabricated Metal Product Manufacturing										
Ln(Road Network Density) Instrument		0.097	0.093	0.294	0.114	1.51E-07	417	1.81E-01	95.2	
Road Network Density Instrument		0.138	0.087	0.112	0.121	5.27E-08	417	3.95E-01	118.9	0.564
Ln(WDZ Area) Instrument	***	0.184	0.067	0.006	0.125	1.45E-08	417	7.47E-01	338.0	
WDZ Area Instrument	***	0.212	0.074	0.004	0.125	1.22E-10	417	9.04E-01	6.4	
23 Transport Equipment Manufacturing										
Ln(Road Network Density) Instrument		-0.033	0.095	0.726	0.102	1.63E-09	486	**6.41E-03**	103.8	
Road Network Density Instrument		-0.023	0.083	0.780	0.106	2.13E-09	486	**2.95E-03**	199.7	0.005
Ln(WDZ Area) Instrument	***	0.177	0.066	0.007	0.142	1.05E-11	486	6.92E-01	375.8	
WDZ Area Instrument		-0.067	0.079	0.393	0.088	4.75E-09	486	**4.96E-02**	8.1	
24 Machinery and Equipment Manufacturing										
Ln(Road Network Density) Instrument		0.031	0.077	0.687	0.102	2.12E-09	655	3.57E-01	184.1	
Road Network Density Instrument		0.029	0.062	0.644	0.102	2.84E-09	655	2.84E-01	263.7	0.439
Ln(WDZ Area) Instrument		0.073	0.049	0.133	0.105	2.99E-09	655	7.12E-01	737.1	
WDZ Area Instrument		0.161	0.123	0.191	0.100	3.23E-09	655	2.89E-01	18.7	
30 Building Construction										
Ln(Road Network Density) Instrument	**	0.094	0.042	0.025	0.134	4.80E-21	1484	**1.76E-02**	380.5	
Road Network Density Instrument	***	0.129	0.035	0.000	0.140	1.40E-23	1484	1.66E-01	547.3	0.541
Ln(WDZ Area) Instrument	***	0.149	0.028	0.000	0.142	1.01E-27	1484	1.90E-01	1382.3	
WDZ Area Instrument	*	0.126	0.070	0.070	0.140	3.38E-20	1484	3.21E-01	19.2	
31 Heavy and Civil Engineering Construction										
Ln(Road Network Density) Instrument		0.046	0.073	0.531	0.170	8.16E-14	457	3.37E-01	154.7	
Road Network Density Instrument		0.058	0.070	0.402	0.172	6.15E-14	457	4.22E-01	199.0	0.944
Ln(WDZ Area) Instrument		0.053	0.056	0.339	0.171	1.16E-15	457	1.06E-01	497.1	
WDZ Area Instrument		0.129	0.104	0.216	0.174	5.12E-14	457	8.63E-01	4.2	
33 Basic Material Wholesaling										
Ln(Road Network Density) Instrument	**	0.168	0.071	0.018	0.174	2.33E-29	767	5.72E-01	244.6	
Road Network Density Instrument	***	0.189	0.068	0.005	0.175	4.31E-31	767	8.17E-01	315.3	0.834
Ln(WDZ Area) Instrument	***	0.201	0.052	0.000	0.175	6.03E-36	767	9.97E-01	715.9	
WDZ Area Instrument	**	0.206	0.090	0.021	0.175	7.40E-30	767	9.57E-01	14.0	
34 Machinery and Equipment Wholesaling										
Ln(Road Network Density) Instrument		0.033	0.064	0.604	0.093	2.02E-27	677	**3.46E-02**	169.0	
Road Network Density Instrument		0.069	0.058	0.232	0.102	7.14E-28	677	**8.20E-02**	274.9	0.257
Ln(WDZ Area) Instrument	***	0.136	0.044	0.002	0.110	2.37E-28	677	3.75E-01	847.5	
WDZ Area Instrument		0.103	0.092	0.262	0.107	2.23E-26	677	4.59E-01	34.4	
41 Food Retailing										
Ln(Road Network Density) Instrument		-0.013	0.064	0.844	0.187	3.58E-29	1202	1.17E-01	192.2	
Road Network Density Instrument		0.016	0.058	0.778	0.191	5.64E-29	1202	2.56E-01	342.5	0.311
Ln(WDZ Area) Instrument	*	0.070	0.036	0.053	0.193	1.34E-31	1202	9.59E-01	917.6	
WDZ Area Instrument	*	0.111	0.063	0.075	0.192	4.04E-29	1202	5.64E-01	14.2	
42 Other Store-Based Retailing										
Ln(Road Network Density) Instrument	*	0.085	0.052	0.100	0.169	1.11E-27	1523	**7.79E-03**	230.1	
Road Network Density Instrument	***	0.116	0.040	0.004	0.176	3.13E-29	1523	**2.86E-02**	425.3	0.015
Ln(WDZ Area) Instrument	***	0.217	0.033	0.000	0.183	7.87E-37	1523	3.88E-01	1007.2	
WDZ Area Instrument	***	0.241	0.088	0.006	0.182	5.15E-31	1523	4.63E-01	11.3	
44 Accommodation										
Ln(Road Network Density) Instrument	***	0.260	0.063	0.000	0.106	1.17E-13	590	**9.61E-02**	229.9	
Road Network Density Instrument	***	0.227	0.052	0.000	0.108	4.14E-14	590	2.85E-01	235.8	0.134
Ln(WDZ Area) Instrument	***	0.143	0.044	0.001	0.110	9.80E-15	590	2.86E-01	1222.5	
WDZ Area Instrument	**	0.189	0.080	0.018	0.110	7.28E-12	590	8.36E-01	10.4	
45 Food and Beverage Services										
Ln(Road Network Density) Instrument		0.036	0.054	0.512	0.218	3.70E-44	1503	2.05E-01	288.5	
Road Network Density Instrument	**	0.088	0.041	0.031	0.221	8.88E-46	1503	9.89E-01	473.2	0.251
Ln(WDZ Area) Instrument		0.042	0.034	0.215	0.219	1.15E-44	1503	**3.02E-02**	1156.7	
WDZ Area Instrument		-0.033	0.106	0.752	0.206	9.93E-44	1503	**1.90E-02**	31.7	

Appendix G Continued…

54 Publishing (except Internet and Music Publishing)										
Ln(Road Network Density) Instrument		0.056	0.069	0.413	0.095	0.000428	440	1.06E-01	148.2	
Road Network Density Instrument		0.076	0.063	0.228	0.099	0.000262	440	1.68E-01	175.5	0.404
Ln(WDZ Area) Instrument	***	0.137	0.049	0.005	0.106	3.95E-06	440	3.41E-01	701.6	
WDZ Area Instrument	**	0.203	0.080	0.011	0.106	0.000037	440	7.80E-01	5.6	
55 Motion Picture and Sound Recording Activities										
Ln(Road Network Density) Instrument		0.114	0.144	0.428	0.246	2.07E-17	336	4.70E-01	57.9	
Road Network Density Instrument		0.129	0.111	0.244	0.247	5.06E-18	336	5.07E-01	129.3	0.999
Ln(WDZ Area) Instrument	*	0.124	0.069	0.073	0.246	5.17E-17	336	1.41E-01	490.3	
WDZ Area Instrument		-0.024	0.066	0.714	0.223	9.53E-16	336	**5.03E-02**	28.5	
59 Internet Service Providers, Web Search Portals, and Data Processing Services										
Ln(Road Network Density) Instrument		0.104	0.103	0.314	0.113	3.57E-03	200	2.15E-01	82.6	
Road Network Density Instrument		0.141	0.089	0.112	0.121	1.97E-03	200	3.62E-01	70.8	0.998
Ln(WDZ Area) Instrument	*	0.133	0.078	0.088	0.120	4.80E-04	200	**2.04E-02**	472.8	
WDZ Area Instrument		-0.007	0.137	0.958	0.073	6.39E-03	200	**4.97E-02**	30.9	
62 Finance										
Ln(Road Network Density) Instrument	***	0.224	0.081	0.006	0.176	3.10E-120	772	2.21E-01	135.0	
Road Network Density Instrument	***	0.211	0.058	0.000	0.177	1.20E-125	772	2.67E-01	256.7	0.332
Ln(WDZ Area) Instrument	***	0.153	0.039	0.000	0.180	3.90E-121	772	9.61E-01	661.1	
WDZ Area Instrument	**	0.180	0.083	0.030	0.180	6.80E-114	772	8.00E-01	4.6	
63 Insurance										
Ln(Road Network Density) Instrument		-0.005	0.082	0.947	0.135	5.84E-17	398	1.43E-01	55.2	
Road Network Density Instrument		0.055	0.067	0.411	0.150	3.39E-17	398	3.80E-01	115.2	0.168
Ln(WDZ Area) Instrument	***	0.174	0.054	0.001	0.153	4.51E-18	398	1.16E-01	581.3	
WDZ Area Instrument		0.075	0.107	0.482	0.153	2.04E-17	398	6.25E-01	9.5	
64 Auxiliary Finance and Insurance Services										
Ln(Road Network Density) Instrument	*	0.101	0.058	0.084	0.150	1.47E-46	772	**9.51E-02**	117.3	
Road Network Density Instrument	***	0.133	0.045	0.003	0.157	4.62E-48	772	1.95E-01	279.3	0.257
Ln(WDZ Area) Instrument	***	0.196	0.037	0.000	0.161	2.21E-52	772	9.42E-01	757.5	
WDZ Area Instrument	**	0.117	0.048	0.015	0.154	7.30E-49	772	2.53E-01	9.1	
67 Property Operators and Real Estate Services										
Ln(Road Network Density) Instrument	***	0.193	0.063	0.002	0.102	1.38E-15	871	3.94E-01	174.6	
Road Network Density Instrument	***	0.194	0.050	0.000	0.101	5.66E-17	871	3.37E-01	320.0	0.517
Ln(WDZ Area) Instrument	***	0.162	0.037	0.000	0.103	1.99E-20	871	6.71E-01	826.9	
WDZ Area Instrument		0.089	0.055	0.104	0.099	5.24E-14	871	3.25E-01	12.1	
69 Professional, Scientific and Technical Services (except Computer System Design and Related Services)										
Ln(Road Network Density) Instrument		0.065	0.052	0.213	0.191	8.40E-32	1603	**1.94E-03**	259.9	
Road Network Density Instrument	**	0.093	0.041	0.023	0.199	9.15E-34	1603	**1.15E-02**	424.3	0.006
Ln(WDZ Area) Instrument	***	0.193	0.031	0.000	0.208	4.87E-41	1603	4.63E-01	1250.1	
WDZ Area Instrument	*	0.095	0.050	0.057	0.200	1.13E-31	1603	**8.33E-02**	15.3	
75 Public Administration										
Ln(Road Network Density) Instrument		0.066	0.061	0.281	0.238	1.85E-31	984	**8.57E-02**	131.0	
Road Network Density Instrument	**	0.120	0.048	0.012	0.246	1.13E-34	984	4.13E-01	294.7	0.634
Ln(WDZ Area) Instrument	***	0.144	0.033	0.000	0.248	1.01E-36	984	5.63E-01	853.2	
WDZ Area Instrument		0.063	0.047	0.179	0.238	2.75E-31	984	2.36E-01	6.0	
77 Public Order, Safety and Regulatory Services										
Ln(Road Network Density) Instrument		0.043	0.064	0.499	0.047	9.80E-115	675	7.91E-01	113.8	
Road Network Density Instrument		0.034	0.048	0.486	0.047	7.90E-115	675	9.17E-01	264.0	0.691
Ln(WDZ Area) Instrument	*	0.054	0.032	0.089	0.046	9.40E-116	675	3.33E-01	633.4	
WDZ Area Instrument		0.071	0.077	0.360	0.044	2.10E-114	675	5.83E-01	10.3	
84 Hospitals										
Ln(Road Network Density) Instrument	**	0.263	0.114	0.021	0.128	7.31E-13	509	**3.80E-02**	76.9	
Road Network Density Instrument	***	0.229	0.085	0.007	0.134	3.25E-13	509	**6.88E-02**	151.6	0.268
Ln(WDZ Area) Instrument	***	0.156	0.050	0.002	0.145	9.48E-14	509	1.99E-01	483.6	
WDZ Area Instrument	***	0.281	0.094	0.003	0.124	4.78E-13	509	**1.80E-02**	9.4	
85 Medical and Other Health Care Services										
Ln(Road Network Density) Instrument		0.055	0.054	0.307	0.163	3.33E-20	1352	4.23E-01	180.4	
Road Network Density Instrument	**	0.085	0.042	0.042	0.164	1.77E-21	1352	9.57E-01	360.6	0.903
Ln(WDZ Area) Instrument	***	0.080	0.030	0.007	0.164	1.41E-23	1352	7.22E-01	1043.0	
WDZ Area Instrument		0.018	0.046	0.688	0.157	2.50E-20	1352	2.21E-01	8.9	
87 Social Assistance Services										
Ln(Road Network Density) Instrument	***	0.185	0.060	0.002	0.173	3.75E-37	1353	1.35E-01	156.4	
Road Network Density Instrument	***	0.184	0.048	0.000	0.173	8.37E-39	1353	**9.75E-02**	323.8	0.327
Ln(WDZ Area) Instrument	***	0.142	0.031	0.000	0.177	2.39E-43	1353	2.76E-01	988.0	
WDZ Area Instrument	**	0.080	0.039	0.039	0.176	2.93E-37	1353	4.95E-01	7.8	
89 Heritage Activities										
Ln(Road Network Density) Instrument		0.052	0.076	0.494	0.214	2.68E-10	163	1.04E-01	208.9	
Road Network Density Instrument		0.064	0.068	0.345	0.217	5.93E-10	163	1.54E-01	168.9	0.561
Ln(WDZ Area) Instrument	**	0.129	0.059	0.027	0.227	1.19E-09	163	5.74E-01	411.2	
WDZ Area Instrument		-0.022	0.092	0.813	0.185	2.23E-09	163	2.44E-01	4.2	
90 Creative and Performing Arts Activities										
Ln(Road Network Density) Instrument		0.003	0.109	0.977	0.034	1.58E-03	476	**3.29E-03**	119.4	
Road Network Density Instrument		0.085	0.089	0.339	0.053	1.76E-03	476	**4.10E-02**	230.0	0.659
Ln(WDZ Area) Instrument	**	0.164	0.077	0.034	0.063	7.43E-04	476	**5.11E-02**	1044.9	
WDZ Area Instrument		0.131	0.125	0.296	0.060	1.73E-03	476	3.21E-01	16.6	

Note (1): * = Significant at 0.10 ** = Significant at 0.05 *** = Significant at 0.01
Note (2): Hausmann P-values in bold indicate regressions in which endogeneity was detected at a significance level of 0.10

Appendix H: An Alternative Method to Calculating Ring-Band Employment in Postgres

This appendix gives an alternate method to estimating industry-specific ring band employment, authored for the Melbourne industry data in particular. Some of the sections of code, such as H1, H2 and H3 are merely maintenance scripts to make the actual working code easier to write and interpret. Much of the data in its original form, for instance, includes data of the 'text' data type, which slows laborious calculations significantly compared to calculations referencing integer data. This slowing also occurs when substrings are referenced as opposed using complete reference values. Other maintenance scripts exist mainly to omit rows that include values such as 'totals', as they are of no use for our purposes.

The main function of this script is to calculate and store a set of values that are shared among all industry ring band calculations within a city. Section H4 of the code creates indices on the WDZ and ring band geometries that serve to speed up calculations that employ them. The main sections of the code, however, are H5 and H6. The former script sets up a table that can be subsequently populated by running the latter. H6 essentially breaks down the code given in Analysis 2-3, replicating the geometries where necessary and organizing everything in columns of unique values rather than matrices. H8 then creates indices on all the newly inserted geometry data to speed up subsequent calculations. The calculations estimating the area of each WDZ covered by each ring band, the share of each WDZ covered by each ring band, and the amount of industry-specific employment in each section where a WDZ and ring band overlap are all done in section H9 of the code. Finally, section H10 of the code sums up the employment within each ring for each industry and exports the values to CSV format.

This method to estimating industry-ring band employment would benefit the user by speeding up the entire employment band estimation process, as it would only need to be carried out once, but its shortfall is the massive amount of data that it generates and that requires a large amount of storage space. When executed in Postgres, the code filled up 350 gigabytes of storage space (the entire availability of free space) and then crashed. The returned error message and timing of the message is reported at the end of section H6. If hard-disk space was in abundance, this method would have been more efficient and simpler than the alternative that was used for generating the ring band employment figures in Analysis 2-3.

H1

```
--Find all wdz_codes containing any string plus "SD Total".
/*
SELECT * FROM import.cities_income WHERE wdz_code ~~ '%SD Total%';
--DELETE FROM import.cities_income WHERE wdz_code ~~ '%SD Total%';
*/
```

H2

```
--Create import.cities_income.city_id and set to different numbers for each city.
/*
ALTER TABLE import.cities_income DROP COLUMN city_id;
ALTER TABLE import.cities_income ADD COLUMN city_id Integer;
UPDATE import.cities_income set city_id = 2 WHERE city_name ~ 'melbourne';
UPDATE import.cities_income set city_id = 1 WHERE city_name ~ 'sydney';
*/
```

H3

```
--Create import.cities_income.ind_id and set to industries' prefix number
/*
SELECT DISTINCT substring(industry from 1 for 2)::INTEGER
        FROM import.cities_income
        ORDER BY substring ASC;
*/
/*
ALTER TABLE import.cities_income DROP COLUMN ind_id;
ALTER TABLE import.cities_income ADD COLUMN ind_id INTEGER;
UPDATE import.cities_income SET ind_id = substring(industry from 1 for 2)::INTEGER;
 */
```

H4

```
--Create indices for all geometry columns.
/*
DROP index shp.shp_melb_wdz_b1;
CREATE INDEX shp_melb_wdz_b1 ON shp.melb_wdz USING gist (band_1);
DROP index shp.shp_melb_wdz_b2;
CREATE INDEX shp_melb_wdz_b2 ON shp.melb_wdz USING gist (band_2);
DROP index shp.shp_melb_wdz_b3;
CREATE INDEX shp_melb_wdz_b3 ON shp.melb_wdz USING gist (band_3);
DROP index shp.shp_melb_wdz_b4;
CREATE INDEX shp_melb_wdz_b4 ON shp.melb_wdz USING gist (band_4);
*/
```

H5

```
--DROP TABLE public.melb;
/*
CREATE TABLE public.melb (gid SERIAL PRIMARY KEY,
        m1_gid BIGINT,
        m1_geom GEOMETRY,
        m2_gid bigint,  m2_geom GEOMETRY,
        m2_band1 GEOMETRY,
        m2_band2 GEOMETRY,
        m2_band3 GEOMETRY,
        m2_band4 GEOMETRY,
        ind_id BIGINT,
        no_employed BIGINT,
        area_m1_b1 DOUBLE PRECISION,
        area_m1_b2 DOUBLE PRECISION,
        area_m1_b3 DOUBLE PRECISION,
        area_m1_b4 DOUBLE PRECISION,
        share_m1b1_over_m1 DOUBLE PRECISION,
        share_m1b2_over_m1 DOUBLE PRECISION,
        share_m1b3_over_m1 DOUBLE PRECISION,
        share_m1b4_over_m1 DOUBLE PRECISION,
        emp_m1_b1 DOUBLE PRECISION,
        emp_m1_b2 DOUBLE PRECISION,
        emp_m1_b3 DOUBLE PRECISION,
        emp_m1_b4 DOUBLE PRECISION
 ); */
```

H6

```
/*
INSERT INTO melb (
        m1_gid,
        m1_geom,-- geometry,
        m2_gid,-- bigint,
        m2_geom,-- geometry,
        m2_band1,-- geometry,
        m2_band2,-- geometry,
        m2_band3,-- geometry,
        m2_band4,-- geometry,
        ind_id,-- bigint,
        no_employed
) SELECT temp.m1_gid,
        temp.m1_the_geom,
        temp.m2_gid,
        temp.m2_the_geom,
        temp.m2_band1,
        temp.m2_band2,
        temp.m2_band3,
        temp.m2_band4,
```

```
        inc.ind_id,
        inc."total employed"
FROM (
        SELECT
        m1.vicdznp06 AS vicdznp06,
        m1.gid AS m1_gid,
        m1.the_geom AS m1_the_geom,
        m2.gid AS m2_gid,
        m2.the_geom AS m2_the_geom,
        m2.band_1b AS m2_band1,
        m2.band_2b AS m2_band2,
        m2.band_3b AS m2_band3,
        m2.band_4b AS m2_band4
        FROM shp.melb_wdz AS m1, shp.melb_wdz AS m2) AS temp
        JOIN import.cities_income AS inc ON inc.wdz_id = temp.vicdznp06;
*/
```

Error received after running this stage: "HINT: Check free disk space." after running for 17,394,146ms.

H7

```
--Create a new column to store WDZ ID numbers as integers.
/*
ALTER TABLE import.cities_income ADD COLUMN wdz_id BIGINT;
UPDATE import.cities_income SET wdz_id = substring(wdz_code FROM 7 FOR
4)::BIGINT WHERE city_id =2;
*/
```

H8

```
--Create index on all primary-stage columns in public.melb (run after INSERT statement).
/*
--DROP INDEX public_melb_m1_gid;
CREATE INDEX public_melb_m1_gid ON public.melb USING btree (m1_gid);
--DROP INDEX public_melb_m1_geom;
CREATE INDEX public_melb_m1_geom ON public.melb USING gist (m1_geom);
--DROP INDEX public_melb_m2_gid; CREATE INDEX public_melb_m2_gid ON
public.melb USING btree (m2_gid);
--DROP INDEX public_melb_m2_geom;
CREATE INDEX public_melb_m2_geom ON public.melb USING gist (m2_geom);

--DROP INDEX public_melb_m2_band1;
CREATE INDEX public_melb_m2_band1 ON public.melb USING gist (m2_band1);
--DROP INDEX public_melb_m2_band2;
CREATE INDEX public_melb_m2_band2 ON public.melb USING gist (m2_band2);
--DROP INDEX public_melb_m2_band3;
CREATE INDEX public_melb_m2_band3 ON public.melb USING gist (m2_band3);
--DROP INDEX public_melb_m2_band4;
CREATE INDEX public_melb_m2_band4 ON public.melb USING gist (m2_band4);
```

```
--DROP INDEX public_melb_ind_id;
CREATE INDEX public_melb_ind_id ON public.melb USING btree (ind_id);
--DROP INDEX public_melb_no_employed;
CREATE INDEX public_melb_no_employed ON public.melb USING btree
(no_employed);
*/
```

H9

```
--Update melb area of intersection between m1.the_geom and bands m2.band(1-4).
/*
UPDATE public.melb SET
    area_m1_b1 = st_area(st_intersect(m1_geom, m2_band1)),
    area_m1_b2 = st_area(st_intersect(m1_geom, m2_band2)),
    area_m1_b3 = st_area(st_intersect(m1_geom, m2_band3)),
    area_m1_b4 = st_area(st_intersect(m1_geom, m2_band4));
UPDATE public.melb SET
    share_m1b1_over_m1 = area_m1_b1 / st_area(m1_geom),
    share_m1b2_over_m1 = area_m1_b2 / st_area(m1_geom),
    share_m1b3_over_m1 = area_m1_b3 / st_area(m1_geom),
    share_m1b4_over_m1 = area_m1_b4 / st_area(m1_geom);
UPDATE public.melb SET
    emp_m1_b1 = no_employed * share_m1b1_over_m1,
    emp_m1_b2 = no_employed * share_m1b2_over_m1,
    emp_m1_b3 = no_employed * share_m1b3_over_m1,
    emp_m1_b4 = no_employed * share_m1b4_over_m1;
*/
```

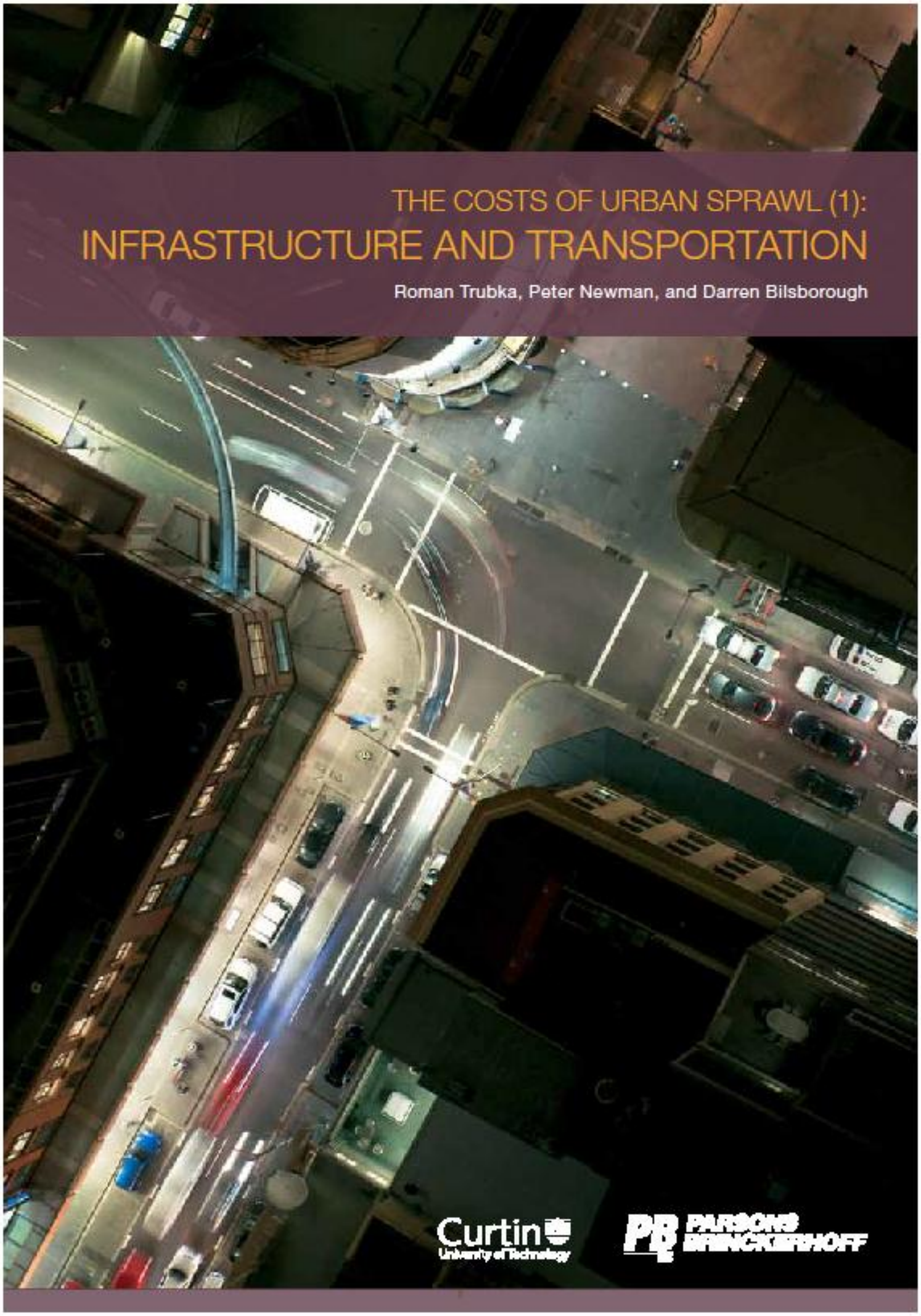
H10

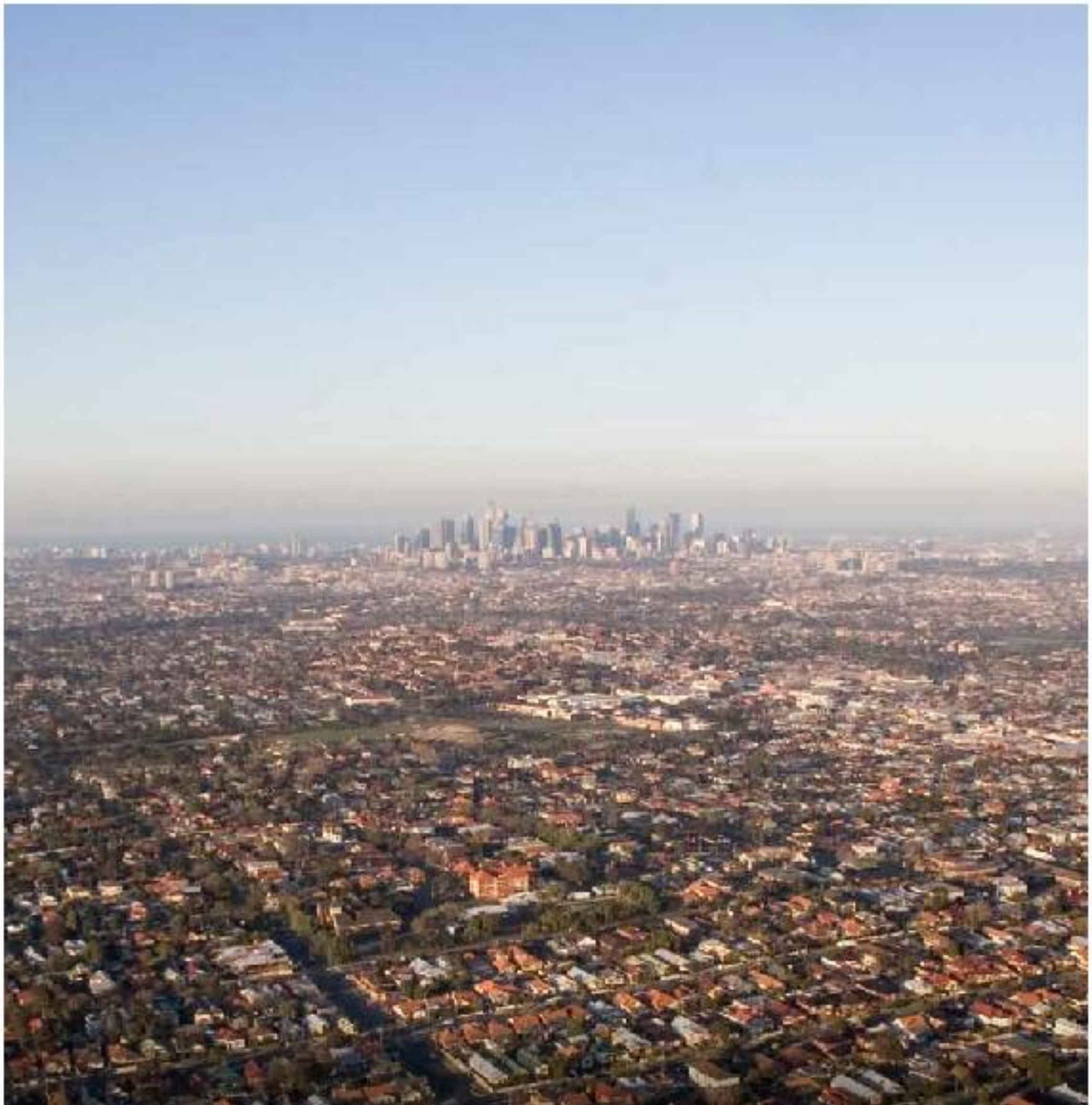
```
/* Example of how one might want to export the results to a CSV file, however one will
most likely wish to do this in conjunction with the other employment data to make a
complete dataset. */
/*
COPY
SELECT
    m1_gid,
    w.vicdznp06,
    ind_id,
    sum(emp_m1_b1),
    sum(emp_m1_b2),
    sum(emp_m1_b3),
    sum(emp_m1_b4)
FROM melb AS m JOIN shp.melb_wdz AS w ON m.m1_gid = w.vicdznp06
--WHERE ind_id = 1 --Optional clause to restrict the industries for which data are
exported
GROUP BY m1_gid, ind_id
TO ('employment_bands.csv') CSV HEADER;
*/
```

Appendix I: Previous Publications Supporting the Work in this Thesis

See Next Page

NOTE: The versions of the papers in this appendix are those that were published by the PB-CUSP Alliance. These papers have also been peer reviewed and published by the Environment Design Guide (EDG), which are the versions indicated in the reference section of this thesis.


THE COSTS OF URBAN SPRAWL (1):
INFRASTRUCTURE AND TRANSPORTATION
Roman Trubka, Peter Newman, and Darren Bilsborough



We have compared the economic impacts associated with urban redevelopment to that of fringe development. The findings are presented in three companion papers on costs associated with infrastructure and transportation (this paper), greenhouse gas emissions, and activity-related health. The savings in transport and infrastructure for 1000 dwellings are around $86 million up-front for infrastructure and $250 million for annualised transportation costs over 50 years.

Keywords: sprawl, redevelopment, greenfields, transport, infrastructure, planning

1.0 Introduction

Australian cities are in focus as the Prime Minister has announced a Federal Government involvement in urban planning to cope with expected population increases (BCA 2009). In the future if there is Federal money in urban infrastructure, transport and healthcare, then projects will need to meet certain criteria about costs, climate change, and health. These papers are designed to assess these implications by examining two alternative approaches to urban development: redevelopment in walkable, transit-oriented developments and fringe development in conventional low-density car dependent suburbs.

Redevelopment is based around present urban areas that are already well served by public transport but can also include new developments, so long as transit accessibility, walkability, and density are implemented in the planning and design process. Table 1 sets out the two development types according to some of their defining characteristics. More detailed data can be found in the Appendices attached to paper 2, based on Sydney and Melbourne local government areas.

Table 1: Defining criteria of the dichotomous urban forms

	Urban redevelopment	Fringe development
Daily per capita GHG emissions from transport	0 to 4 kg CO_2-e	8 up to 10 kg CO_2-e
Distance to CBD	less than10 km	more than 40 km
Activity intensity[1] (measured by population and jobs per ha)	more than 35 pop. + jobs/ha	less than 20 pop. + jobs/ha
Transit accessibility[2]	more than 80% with >15min service	less than 15% with >15min service

[1] Activity intensity is a measure of density that includes residential and commercial activity on urbanised spaces. As such, areas occupied by large natural bodies such as lakes, rivers and parklands are omitted from its calculation.

[2] Transit accessibility relates to the proportion of land within an urbanised area that is within 400 m of a full-service bus or tram, or within 800 m of a train station. 'Full-service' is defined as a route operating seven days a week with at least four services an hour on weekdays and Saturdays during the day and two services an hour on Sundays and holidays.

Source: data collected for this study

The two development types are based on the data collected for Sydney and Melbourne that included a local government area's distance to the CDB, estimated daily per capita greenhouse gas emissions, activity intensity and transit accessibility rating. The LGAs of the CBD, Port Phillip and Yarra for Melbourne and the CBD, South Sydney and Leichhardt in Sydney are examples of inner-city/core areas that rate well within the above-defined thresholds of the defining criteria. As for the criteria into which fringe developments would fall, LGAs to represent these in Melbourne would include the Yarra Ranges, Cardinia and Mornington Peninsula and among others, Penrith, Camden, Gosford, and the Blue Mountains in Sydney. Table 2 sets out some other transport data that characterises Melbourne's local area differences to show how much more walkable and transit oriented the core/inner area is. These are also wealthier areas as Table 2 indicates.

Table 2: Trips per day per person by area – Melbourne

	Core	Inner	Middle	Outer/fringe
Car	2.12	2.52	2.86	3.92
Transit	0.66	0.46	0.29	0.04
Walk/bike	2.62	1.61	1.08	0.81
Income >$70,000	12%	11%	10%	6%

Source: Kenworthy and Newman 2000

The research has examined the economic costs associated with these two modes of development, first assessing the physical planning costs associated with the different transport and infrastructure requirements, and then two new areas of public policy — greenhouse gas emissions and activity-related health costs. These are the subjects of increasing interest and their economic costs can be compared with the more traditional costs of physical planning.

2.0 Infrastructure costs

2.1 Background

The economic assessment of infrastructure costs associated with urban sprawl is not a recent concept. Such assessments have been done in Australia as early as the 1970s and numerous assessments have been done since[3]; however, the most recent studies in Australia that could be found were from 2001 and 2003 and simply capitalised the costs reported in previous assessments to then current values[4]. The challenge in interpreting the assessments is that infrastructure costs are so heavily dependent on area-specific factors. For instance, road costs among different prospective development areas may vary based on the necessity for major arterial roads, costs for sewerage and water infrastructure could vary immensely depending on terrain and soil conditions, and many infrastructure components will differ depending on the level and degree of excess capacity. It is also difficult to determine who bears the costs of new infrastructure developments because of constantly changing government-induced fees, taxes, policies, and building standards.

Despite the area-specific nature of calculating development costs, the evidence suggests that initial capital costs and operating costs of sprawling developments outweigh the costs associated with inner-city redevelopment. Perhaps the most significant infrastructure category to mark an economic difference in provisioning is road construction. In many cases it can make up 50% of the cost difference between the two iconic development forms (SGS 2003). The provision of water and sewerage infrastructure is another expensive infrastructure requirement. Markedly in these two categories, but in the others to some degree as well, inner-city redevelopment offers significant cost savings by either using excess capacity or requiring less of the service because of shorter distances and greater compactness.

2.2 Calculating the costs of infrastructure

The main source of data for determining the infrastructure costs of inner city and fringe developments was drawn from a paper prepared for the Western Australia Planning Commission in 2001. Environmental Resource Management Pty Ltd (ERM) compiled the report, titled *Future Perth,* to identify the economic cost differences between developments in inner, middle and fringe areas (ERM 2001). It reviewed the information produced by 22 studies across Australia, America, and Canada and sorted the cost findings into three different measures of urban form: inner, middle, and outer.

[3] Previous studies on costs of developments in inner and outer areas have been carried out in Australia by PG Pak Poy and Associates (1973), Voran Consultants (1991), Birrell (1991) and Newman et al. (1992) to name a few.

[4] See the studies titled Future Perth by ERM (2001) and Costs of Urban Sprawl by SGS (2003).

Cost indexing

The *Future Perth* report drew on studies that ranged between 1972 and 2000 but adjusted the reported costs to 1999 prices. The same would have been done for this study to inflate those reported values by a standard inflation rate to 2007 prices; however, since 2002 the prices of materials and labour in construction have increased disproportionately to the general consumer price and labour price indices (due largely to the impact of mining operations on the labour market). To account for this, infrastructure costs were inflated according the Australian Bureau of Statistics' (ABS) reported price indices for 1999 to 2007. Table 3 shows how some of these categorical costs have changed between those years.

Table 3: National labour and construction input price increases from June 1999 to June 2007

Labour wage increases	Jun-99	Jun-07	Index change	% Change
Electrical	83.1	114.8	31.7	38.1%
Gas	83.1	114.8	31.7	38.1%
Water	83.1	114.8	31.7	38.1%
Construction	83.8	115.7	31.9	38.1%
Transportation	86.8	110.8	24	27.6%
Government admin and def	84.4	113.4	29	34.4%
Health and community	85.9	113.5	27.6	32.1%
Property and business	83.3	111.4	28.1	33.7%
Education	83.8	113.5	29.7	35.4%
Price increases of supplies				
Weighted avg. of six cap cities	119.2	148.3	29.1	24.41%
Consumer Price Index				
CPI	122.3	157.5	35.2	28.78%

Source: ABS (2008a, 2008b, 2008c)

When consolidating and inflating the reported residential development costs reported in *Future Perth*, the appropriate price increase was matched to each category according to the type of industry it fell into and if it was likely to include a labour component, a materials component, or both.

Muncipal services

Table 4 displays the economic breakdown of inner city and urban fringe initial capital costs in 2007 prices, and represents the higher estimates reported by the studies surveyed by *Future Perth*. In the case of the inner-city provision of fire, ambulance, and police infrastructure, none of the *Future Perth* studies reported estimates. This was explained as being a likely result of excess capacity use. Cities typically have staffing ratios they maintain of police officers to residents, but these costs are covered incrementally and would likely appear as operating costs, not needing new investments in physical infrastructure. Municipal services price estimates were not provided for either of the two urban forms, but this was merely for the reason that none of the surveyed studies researched these costs. They will be higher in the fringe areas thus the results are conservative.

Table 4: Breakdown of initial capital costs for redevelopment versus fringe development infrastructure (1000 dwellings)

	Inner	Outer
Roads	$5,086,562	$30,378,881
Water and sewerage	$14,747,616	$22,377,459
Telecommunications	$2,576,106	$3,711,851
Electricity	$4,082,117	$9,696,505
Gas	$0	$3,690,843
Fire and ambulance	$0	$302,509
Police	$0	$388,416
Municipal services	Not reported	Not reported
Education	$3,895,458	$33,147,274
Health	$20,114,867	$32,347,327
Total	**$50,502,726**	**$136,041,065**

Source: Future Perth (2001))

Operating costs

The *Future Perth* study also reviewed the operating costs reported by the numerous studies, however, they were incomplete. The majority of studies in the report either did not research operating costs, reported costs only for certain infrastructure items, or only reported them for one type of urban form. Aggregating the costs did not give a comprehensive depiction of infrastructure operating costs associated with inner city and fringe developments and, therefore, they have not been included in this assessment; however, the one area for which operational costs are well known is transportation.

3.0 Transportation costs

3.1 Background

Transportation is a derived need, meaning that people typically travel for some purpose other than for the simple reason of travelling; yet Australian cities are reaching an expansiveness necessitating many residents to commit upwards of an hour or two daily for commuting purposes. The private, public, and external costs associated with the proliferation of roadways are substantial and have largely been driven by automobile dependence, a by-product of fringe development (Newman and Kenworthy 1999).

In many sprawling suburbs, predominantly in the United States, the private costs of transportation have led to home values dropping and in some cases to a point where homes have been boarded up and abandoned (Newman et al. 2009). This should be a signifier that there are limits to urban growth based on car dependence and that housing affordability does come at a transportation cost.

3.2 Associated costs of transportation

The transportation costs associated with inner-city and fringe development were drawn from a study by Newman and Kenworthy (1999), which together with infrastructure constituted part of an economic assessment of urban form. The estimated costs were calculated as functions of vehicle kilometres travelled and covered all of private, public and external costs. Table 5 displays a summary of the costs in 2007 prices, which constitute the recurring annual costs of a development of 1000 dwellings.

Table 5: Transportation costs for 1,000 inner-city and fringe dwellings
Prices shown are calculated for 2007

	Inner	Outer
Capital cost of cars	$2,990,802	$8,628,654
Fuel Costs	$1,203,925	$3,255,349
Other operating car Costs	$1,476,392	$4,259,675
Time costs (total)5	$6,158,348	$8,210,448
Private transport	$3,116,810	$8,210,448
Public transport	$3,041,538	$0
Walking and cycling	$0	$0
Road costs	$1,216,597	$3,508,806
Parking costs	$2,184,489	$7,709,869
Externalities (total)	$243,731	$703,250
Fatalities	$73,368	$211,693
Injuries	$23,627	$68,172
Property damage	$38,549	$111,228
Air pollution	$90,777	$261,925
Noise pollution	$17,409	$50,232
Transit costs (capital & opportunity)	$3,136,540	$470,481
Total	**$18,610,824**	**$36,746,532**

Source: Newman and Kenworthy (1999)

5 Public transport travel time costs are not allotted a value for fringe developments because like in the outermost suburbs of Sydney and Melbourne, the level of public transport service is low to non-existent. Travel time costs are not allotted to walking and cycling because the act may also be discretionary, or done for enjoyment, and little empirical evidence exists to quantify the disutility of active commuting modes.

Cars

The capital costs of cars are represented as annual depreciation figures. The increased wear of longer trip distances in outer city developments is why the operating and capital costs of cars appear much higher than in inner areas. The higher fuel costs are also a reflection of longer trip distances but are likely to also be underestimated in this account because of more recent spikes in oil prices.

Parking

Parking costs represent a significant expenditure that can frequently be overlooked. The higher cost associated with fringe development is due to more parking spaces being required for fringe than inner-city residents. In Perth, the parking requirement for inner areas is approximately four spaces per resident, while outer areas require approximately 10 (Newman and Kenworthy 1999). The level of parking provision is representative of the greater diversity of car trips and the greater proportion of trips requiring private transport that is characteristic of fringe areas.

Understanding future value

To account for the annual stream of costs associated with transportation, their present values were calculated over a period of 15 years as well as 50 years. The 15-year **annuity** (a term that means the annual costs would be recurring over a period of 15 years) was calculated as a reference point to numerous other economic assessments of development expenditures that tend to use a 15-year time period. The 50-year annuity was calculated for the purpose of this economic assessment to synchronise with the other sections. A **discount rate** (a term that accounts for the diminishing value of money in time and the risk of an investment) of 7% was used for all of the transportation-related costs as suggested by the U.S. Department of Transportation (1994). The high rate is used because empirical evidence suggests that immediate benefits are valued higher than future benefits, hence a lower present value of future costs. Discount rates are frequently debated, however, Tables 6 and 7 present the values with a 7% rate to stay consistent with a generally agreed upon value.

Table 6: 15-year present value

Total	**$256,843,265**	**$507,129,572**
Transport (7%)	$136,309,097	$226,100,382
Roads and parking (7%)	$30,976,806	$102,178,732
Externalities (7%)	$2,219,884	$6,504,143
Total	**$169,505,787**	**$334,783,257**

Table 7: 50-year present value

Total	**$169,505,787**	**$334,783,257**
Transport (7%)	$206,542,055	$342,598,098
Roads and parking (7%)	$46,937,535	$154,826,095
Externalities (7%)	$3,363,675	$9,705,379
Total	**$256,843,265**	**$507,129,572**

Source: Newman and Kenworthy (1999)

Discussion

Infrastructure, especially transport infrastructure, shapes cities. The past 50 years of urban development in Australia has been based around building new car-dependent suburbs on the urban fringe. Around the world this form of city building is now under serious reconsideration. The data in this paper show why this is the case — it is very costly. For each new block on the urban fringe compared to redevelopment there is an infrastructure subsidy from various levels of government of around $85,000.

Governments are concerned about affordability and justify this infrastructure investment as part of the subsidy that this generation provides for those who need to buy a house. There is also some evidence that Treasury officials are unaware of the extent of the subsidy as the provision of new greenfields infrastructure is automatic rather than a decision assessed by Treasury. On the other hand, redevelopment projects often need some up-front costs and are seen as a burden on the state. This is a reflection of the dominant framing of the issue of urban development as a greenfield issue alone (Schon and Rein 1995).

Infrastructure subsidies establish suburbs for at least 50 years before they need renewing (Lucy and Phillips 2006). It is a major decision that should be made based on the full costs involved. This study would suggest that if a city has land that can be redeveloped then there will be highly significant infrastructure cost savings associated with this compared to new development on the urban fringe.

Once established, there are many ongoing operational costs of both urban typologies but the most significant operational costs are associated with transport. Private and public costs are needed to ensure people travel more easily to and from these urban areas. The difference between redeveloped and greenfields development is around $18,000 per household per year spent on private and public transport operations. This needs to feed into debates about affordability as over 50 years this adds up to a difference of $251 million for 1000 dwellings, or $251,000 per household. A number of US studies are now showing that transport costs on the urban fringe are higher than mortgage costs (CTCNT 2006; Lipman 2006).

Thus infrastructure and transport costs are suggesting the need to focus on redevelopment rather than fringe development. As Australian cities have developed planning structures and funding structures that are oriented to greenfields development, the reversal of this to focus on redevelopment is not straightforward. The involvement of the Federal Government in infrastructure and planning is now an opportunity to reshape urban development around more cost-effective urban regeneration initiatives.

The next two companion papers assess what this could mean for costs associated with greenhouse gases and inactivity-related health impacts and productivity loss.

References

Australian Bureau of Statistics (ABS), 2008a, 6401.0 – *Consumer Price Index, Australia*, Mar 2008, viewed 14 May 2008, www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6401.0Mar%202008?OpenDocument

Australian Bureau of Statistics (ABS) 2008b, 6345.0 – *Labour Price Index, Australia*, Mar 2008, viewed 30 May 2008, www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6345.0Mar%202008?OpenDocument

Australian Bureau of Statistics (ABS) 2008c, 6427.0 – *Producer Price Indexes, Australia*, viewed 13 May 2008, www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6427.0Mar%202008?OpenDocument

Birrell, R 1991, *Infrastructure Costs on the Urban Fringe: Sydney and Melbourne Compared. Working Paper No. 8 in Background Papers on Urban and Regional Issues: Studies Prepared for the Office of Economic Planning Advisory Council*, Australian Government Publishing Service, Canberra.

Business Council of Australia (BCA) 2009, *Building a Big Australia: Future Planning Needs of Our Major Cities*, Prime Minister Kevin Rudd's address to the Business Council of Australia 2009 Annual Dinner, viewed 7 Dec 2009, www.bca.com.au/Content/101620.aspx

Center for TOD and Center for Neighborhood Technology (CTCNT) 2006, *The Affordability Index: A New Tool for Measuring the Affordability of Housing Choice*, Brookings Institution Metropolitan Policy Program, Washington DC

Environmental Resources Management (ERM) 2001, *Future Perth: Costs of Urban Form Working Paper No. 2*, (unpublished), report prepared for the Western Australian Planning Commission by ERM Australia Pty. Ltd., Perth, Australia.

Lakoff, G 2004, *Don't Think of an Elephant: Know your Values and Frame the Debate*, Chelsea Green Publishers, White River Junction, VT.

Lipman, B 2006, *A Heavy Load: The Combined Housing and Transportation Burdens of Working Families*, Center for Housing Policy, Washington, D.C.

Lucy, W & Phillips, D 2006, *Tomorrow's Cities, Tomorrow's Suburbs*. American Planning Association, Planners Press, Chicago.

Newman, P et al. 1992, *An Economic Impact Statement Comparing Fringe and Inner Developments*, in *Housing, Transport and Urban Form*, The National Housing Strategy.

Newman, P & Kenworthy, J 1999, *Sustainability and Cities: Overcoming Automobile Dependence*, Island Press, Washington.

PG Pak Poy and Associates 1973, *Inner Suburban – Outer Suburban: A comparison of Costs*, Publication No. 42, Australian Institute of Urban Studies.

Reserve Bank of Australia 2007, *Statement on Monetary Policy: November 2007*, viewed 23 April 2008, www.rba.gov.au/PublicationsAndResearch/StatementsOnMonetaryPolicy/index.html

Schon, D & Rein, M 1995, *Frame Reflection: Toward the Resolution of Intractable Controversies*, Lexington Books, Landham, MD.

SGS Economics & Planning 2003. *Costs of Urban Form Discussion Paper*, (unpublished), report prepared for Western Australian Planning Commission by SGS Economics and Planning Pty. Ltd., Melbourne, Australia.

Urbis JHD 2006, *Residential Development Cost Benchmarking Study*, (unpublished), prepared for The Residential Development Council of Australia by Urbis JHD, Perth, Australia.

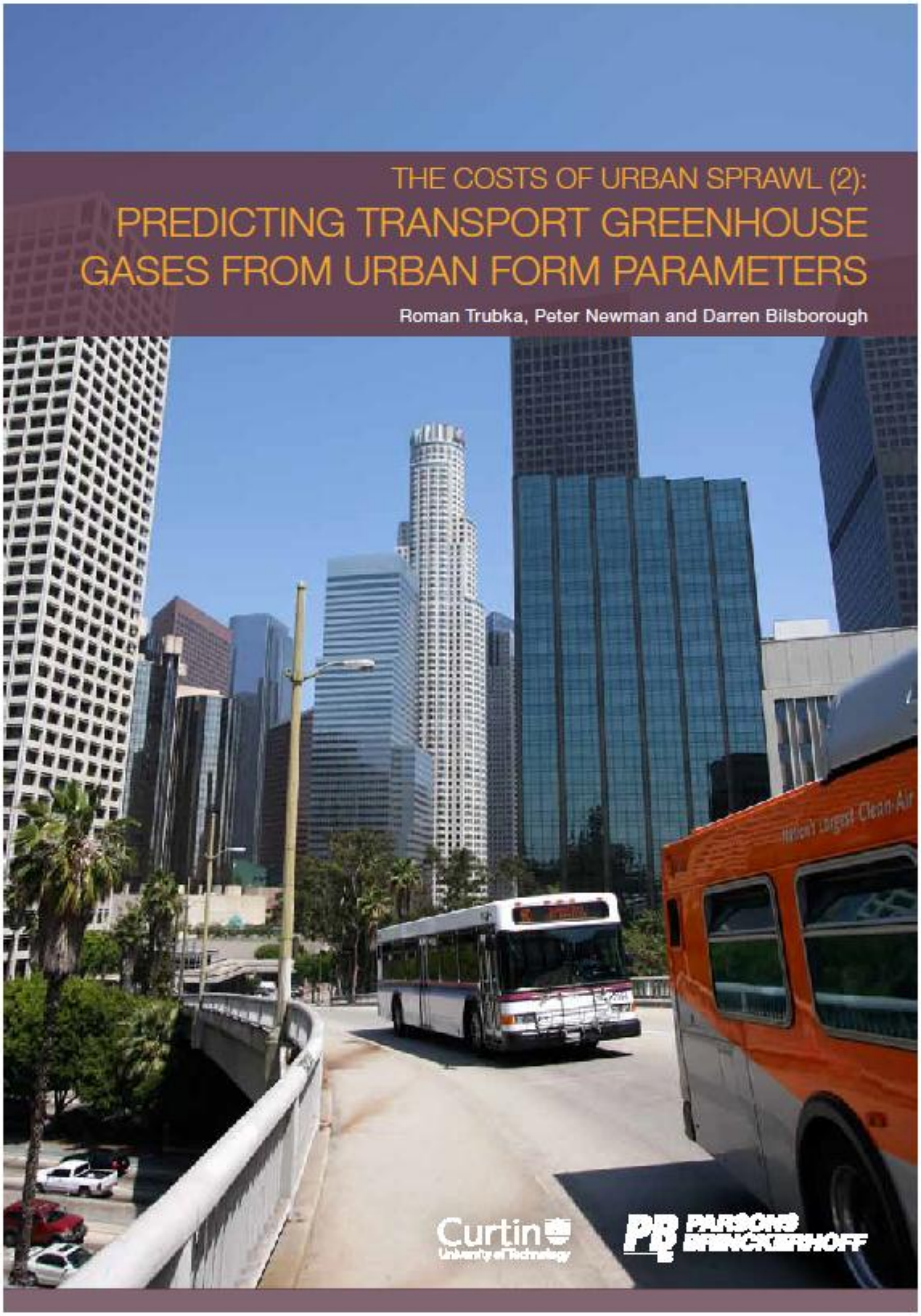
Voran Consultants 1991, *A comparison of Urban Consolidation and Fringe Development – Housing Costs and Benefits*, prepared for Homeswest and the Department of Health, Housing and Community Services, Western Australia.

Biography

Roman Trubka is a PhD candidate at the Curtin University Sustainability Policy (CUSP) Institute in Western Australia.

Peter Newman is a Professor of Sustainability and Director at the Curtin University Sustainability Policy (CUSP) Institute in Western Australia.

Darren Bilsborough is the Director of Sustainability at Parsons Brinckerhoff (PB) and an adjunct professor in sustainability at CUSP.


THE COSTS OF URBAN SPRAWL (2):
PREDICTING TRANSPORT GREENHOUSE GASES FROM URBAN FORM PARAMETERS
Roman Trubka, Peter Newman and Darren Bilsborough
Curtin University of Technology
PARSONS BRINCKERHOFF

This is a companion paper to a study assessing the comparative costs of urban redevelopment with the costs of greenfields development. The first paper showed that substantial costs would be saved in infrastructure and transport if redevelopment were the focus. This paper assesses how these different urban typologies perform with respect to greenhouse gases. The redevelopment option in Australian cities is around 4.4 tonnes less greenhouse gas intensive per household. The study shows how greenhouse gases can be calculated for any development based on simple physical planning parameters, such as the distance to the CBD (a reflection of distance travelled and a proxy for density) and transit accessibility. Although the actual costs of greenhouse gas are small the significance in this work is as governments will need to demonstrate how they are reducing climate change impacts, greenfields developments will find it hard to pass this fundamental criterion of assessment.

Keywords: sprawl, transportation, greenhouse gases, urban form, urban planning

1.0 Greenhouse gas issues

1.1 Background

The combustion of fossil fuels is leading to anthropogenic increases of CO_2 equivalents (CO_2-e) in the earth's atmosphere and, as a result, global warming has become a primary international concern. The United Nations Framework Convention on Climate Change initiated the Kyoto Protocol, an international treaty designed to limit global greenhouse gas emissions, in 1997. When Australia ratified this agreement in December 2007, it made a commitment to limit its greenhouse gas emissions to 108% of its 1990 levels in the 2008/12 period. Reducing the rate of growth in greenhouse gases is difficult, as the 'Business As Usual' (BAU) scenario suggests that Australia's greenhouse gas emissions would reach 124% of 1990 levels over the Kyoto period (AGDCC 2007). However, the world is now moving to reductions in greenhouse gases and hence Australia is committed to 5% reductions by 2020 from 1990 levels (or 15-30% if land-based carbon offsets are included). Many people want much stronger targets but even these are not going to be easy. Urban development priorities can make a big difference as to whether this will be possible to reach. In 2009, the Council of Australian Governments (COAG) chose to support the new Federal planning agenda that requires all infrastructure investment to demonstrate reductions in climate change impacts. All future development will need to demonstrate reductions in greenhouse gases by estimating their carbon content and implied usage patterns.

This paper will examine the greenhouse gas implications of two urban development typologies: urban redevelopment and greenfields development. Although there will be differences due to the greenhouse gas associated with buildings, this study will focus more on transport greenhouse gases. There is evidence that higher density buildings do save on GHGs over lower density buildings due to the shared insulating effect (Newton & Tucker 2007), though there is still some uncertainty about this as argued by Troy (1996). Rickwood (2009) in his empirical work estimates that semi-detached units and flats have 15% less energy use than detached housing units. Apart from this recent empirical work, there are plenty of thermal simulation studies suggesting apartments and terraces have significantly lower heating/cooling through a combination of less air volume and better insulation. Examples of the simulation work include the works of Miller and Ambrose (2005) and Newton et al. (2000).

There is considerable work going into the carbon implications of different built forms but often they neglect the transport aspects (e.g. the Bed Zed development in the UK only examines building energy in its methodology for calculating carbon neutral, see Lazarus 2003).

Thus this paper will examine the transport greenhouse implications of different urban forms in Australian cities and will try to set up a model for estimating these in any urban development.

1.2 Emissions from transport

Currently, transportation in Australia accounts for 14% of total GHG production and road transport is responsible for 88% of this figure or 12.3% of total GHG production (AGDCC 2007). Projections for this sector are based on demographic indicators, such as GDP and population forecasts, vehicle technology and the future travel behaviours of Australian residents. According to these variables, GHG emissions in this sector are expected to increase by 42% of the 1990 level during the Kyoto period, reaching 88 million tonnes (t) annually (AGDCC 2007).

1.3 Reducing emissions from transport

Approaches to reduce the effects of GHG emissions from transportation have been identified as either technological in nature or demand based. The limitations of technological solutions are that they can be expensive (e.g. hybrid vehicles), currently unviable or unavailable on a regional scale (e.g. electric vehicles and high-intensity recharge stations), or simply shift the GHG production to elsewhere in the supply chain (e.g. biofuels and hydrogen power), and are susceptible to the rebound effect (where increased efficiency leads to increased use and hence no/marginal net benefit).

Demand-based solutions reduce the need for private travel and, in some cases, remove the need for motorised travel altogether. This can be achieved through urban planning by bringing people closer to their desired destinations and making non-motorised modes of travel more attractive. Both solutions have their benefits and deficiencies depending on the timeframe of reference; however, they are not mutually exclusive. An opportunity exists for urban planning and technological development to combine and produce sustainable outcomes on a city scale.

1.4 Setting a price for GHG production

Electing a price to place on the generation of greenhouse gases for our purposes is not a clear and distinct matter, especially because carbon offsetting is not yet a mandatory responsibility in Australia. The emerging emissions trading scheme (ETS) in Australia only covers large industry polluters, although the Garnaut ETS was designed to include all emitters. Even with an ETS in place, however, the price of carbon is not set because only the quantity is capped.

Alternatively, it is possible to use a value of carbon equal to the price of a carbon offset, which in Australia can be anywhere within the range of A$8 to A$40. The variation in price can depend on a number of criteria, including the type of project being invested in (i.e. biosequestration projects versus technological investments), the level or type of assurance or accreditation, and the offset provider's business model (Ribon and Scott 2007). It is important to identify, however, that the price of a carbon offset has little to do with the actual cost to society of producing the carbon in the first place.

In 2002, the UK Government Economic Service presented a review of available literature on the estimated social costs of carbon. Within a range of 35–140£/tCO_2-e, the paper suggests an average social cost of 70£/tCO_2-e or roughly A$175 in 2000 prices (Clarkson & Deyes 2002). In 2007, prices this value would be more in the area of A$215. A social cost of carbon (SCC) can be defined as the cost of impacts (market and non-market) associated with the additional production of a unit of greenhouse gas emissions. It represents what society should be willing to pay today to avoid the carbon damages from emissions produced today and their contribution to world emissions in the future.

As a result of the GHG pricing uncertainty and variability, the economic valuation of an urban form's transport-induced greenhouse gas impact becomes equally susceptible to cost variation. Where carbon credits only reflect the value of traded emissions rights and not the social cost of carbon, they fail to reflect the incremental damage being done. On the other hand, if confronted with the choice to suffer a loss of A$215 or pay A$25, the rational choice would be to abate and this would continue to be the case until the marginal cost of abating would equal the social cost of the carbon. Pursuing this logic, the ensuing calculation in this paper will assume carbon to be valued at A$25 a tonne but will also show the implications of choosing the social cost of carbon at $215 a tonne.

Emissions related to city planning

2.1 The effect of urban form and transport provision factors

There are some simple observations that can be made about transport:

(1) The further that people live from their frequented travel destinations, the less likely they are to accomplish that travel through active means, whether it is by walking or cycling.

(2) Areas with poor or no transit services are reliant on private vehicle travel.

(3) The more people living or working in a particular area, the more viable public transport becomes to that place.

(4) The easier it is to walk or cycle in an area the more likely it is that they will do that.

(5) The closer that a development is to the CBD then all these factors come into play providing increased amenity and economy for less car dependent travel.

The importance of quantifying these relationships in a model is that they can then become part of any planning system.

To determine greenhouse gas emissions for the two alternative forms of development a predictive model had to be made. A subjective method does not reliably show the trend that develops between transport greenhouse gas emissions and various urban form parameters, nor can it report on the explanatory significance of any of the parameters of interest. A predictive model in this sense had to be developed where urban form characteristics could be proven on some account to be associated with transport emissions.

To establish a quantitative model for how urban form in Australian cities affects per capita CO_2-e, data and information were drawn from a study previously done at Murdoch University's ISTP department in Western Australia (Chandra 2006). The study estimated fuel use (and therefore GHG production) at the local government level from household travel survey data and average fuel consumption figures sourced from the Australia Greenhouse Office for Sydney, Perth, and Melbourne. The Perth data were very dated so they were not used in the conclusions. The amount of daily per capita private vehicle travel was regressed on several transport-influencing factors such as:

- density (people per hectare)
- jobs (jobs per hectare)
- activity intensity (people plus jobs per hectare)
- permeability (number of intersections per hectare)
- linear distance to CBD (kilometres)
- and transit accessibility (% of area with public transport services beyond a specified threshold).

2.2 Factors affecting emissions

A summary of the simple regression results of the best three explanatory parameters from the list above follows[1]:

Distance to the CBD: The distance to the CBD was the dominant factor in explaining daily per capita GHG emissions (explains 71% of the variance in Sydney and Melbourne). This single variable alone is a very powerful determinant of vehicle energy use and emissions production. From the data a formula was generated to model GHG emissions as a function of distance to a city centre's CBD and it proved to

[1] In some circumstances it may be beneficial to express how the various explanatory variables in a multiple regression model contribute individually to the overall explanatory power of an estimated equation. The individual variables' contribution to explaining the variance in GHG, however, will only sum up meaningfully if they are perfectly uncorrelated. With these variables, this is not the case as locations close to the CBD will likely be denser and characterised by having greater public transport servicing.

hold in all three cities: Perth, Sydney and Melbourne. The formula is $y = x/10 + 3$, where y represents daily per capita GHG emissions in kg CO_2-e and x represents the distance to the CBD in kilometres. The limitation of the formula is most evident in comparing its estimates against outliers, such as Blue Mountain and Mornington (ex-urban areas), where their actual emissions tend to be much higher than are predicted — a potential result of poor proximal amenity in outlying regions.

Activity intensity: Activity intensity (AI) is calculated by adding together the population (number of people) and the number of jobs per hectare for a given area. This explains 56% of the variance in Melbourne and 71% in Sydney. The difference in explanatory power is perhaps because Sydney has a larger variation in AI across its local government areas (LGAs). Sydney's CBD has an activity intensity of roughly 330 people/ha where as Melbourne's is closer to 100 people/ha.

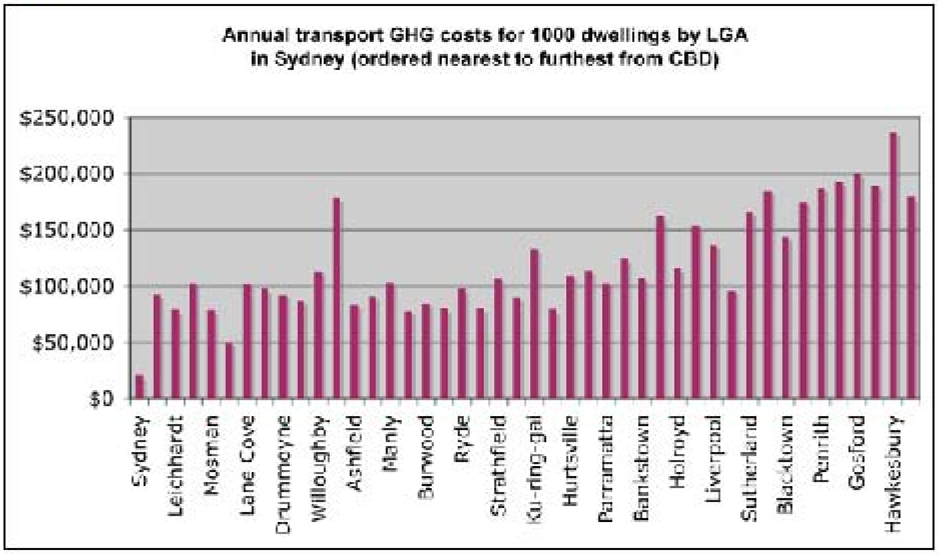
Public transit access: Public transit access has similar statistical significance to activity intensity as it explains 61% of the variance in Melbourne and 58% of the variance in Sydney. Public transit ties in closely with activity intensity because an efficient transit system needs a higher population level to support it. This parameter is defined as the proportion of an LGA that has a transit service frequency of greater than 15 minutes as well as service on weeknights and weekends.

2.3 Results

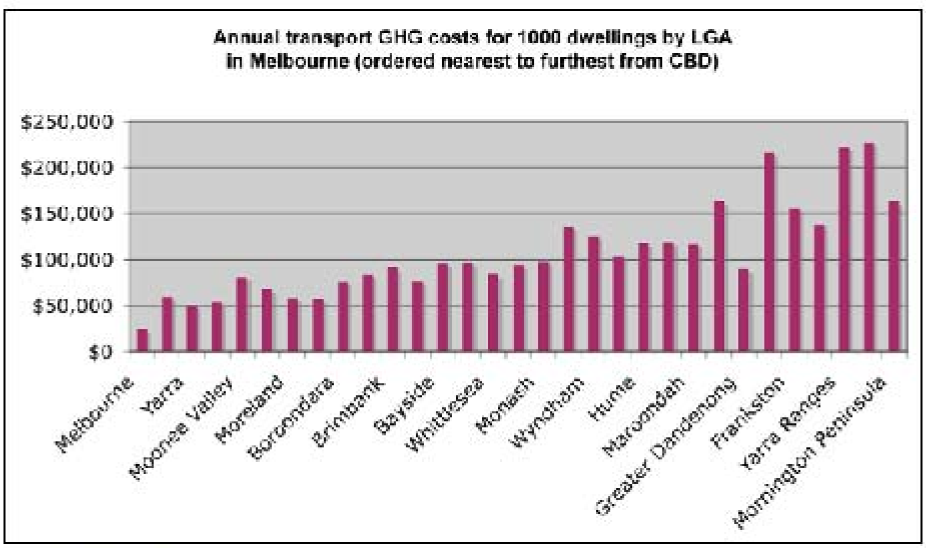
The data from Chandra's (2006) study were made available for this work. Before another model was attempted, a spreadsheet was made for each of the three cities with their local government areas ordered from nearest to furthest from their respective central business districts (see Appendices). For each local government area, data were entered for their respective actual daily per capita transport CO_2-e production. Another column was added calculating the daily transport CO_2-e emissions costs for 1000 dwellings and a final column calculating an annual figure for these values. The price attributed to the generation of 1 t of CO_2-e was hypothetically set at AU$25, a reasonable estimate given current carbon offset prices. The resulting information when translated into a graph appears as follows:

Figure 1: Annual transport GHG costs for 1000 Dwellings in Sydney and Melbourne
Shown for local government areas ordered nearest to furthers from the CBD

Sydney



Melbourne



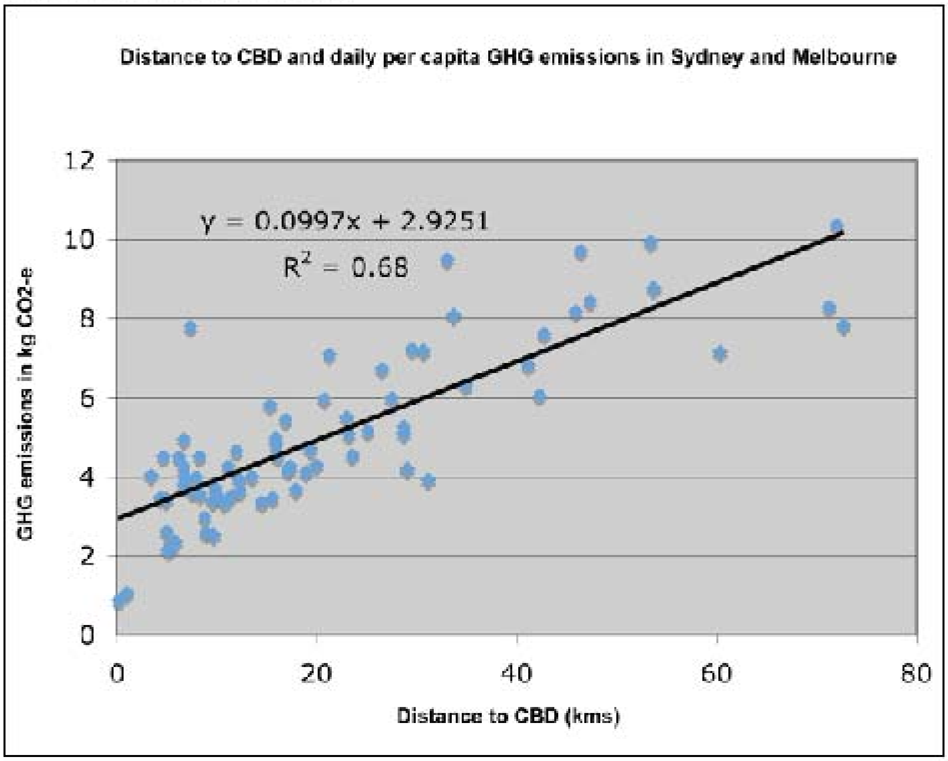
Source: Chandra (2006)

Basing an economic assessment on distance to CBD alone creates a fairly good prediction of per capita greenhouse gas emissions and therefore associated costs, but it was thought that working other variables into the model could improve its effectiveness. The data from Chandra's report was run again several times with different parameters. As distance to CBD, activity intensity, and transit access are the three variables of distinct interest, a simple linear regression was done for each of them with combined data for Sydney and Melbourne.

Distance to CBD

The relationship of 'distance to CBD' to daily per capita greenhouse gas emissions is quite clear in this linear regression. The R-squared value of 0.68 is quite high, implying that distance to CBD explains about 68% of the variance in greenhouse gases. There is, however, a fair bit of dispersion around the line. In particular it tends to highly overestimate per capita GHG emissions for CBD residents where the factors relating to the ease of walking become very obvious.

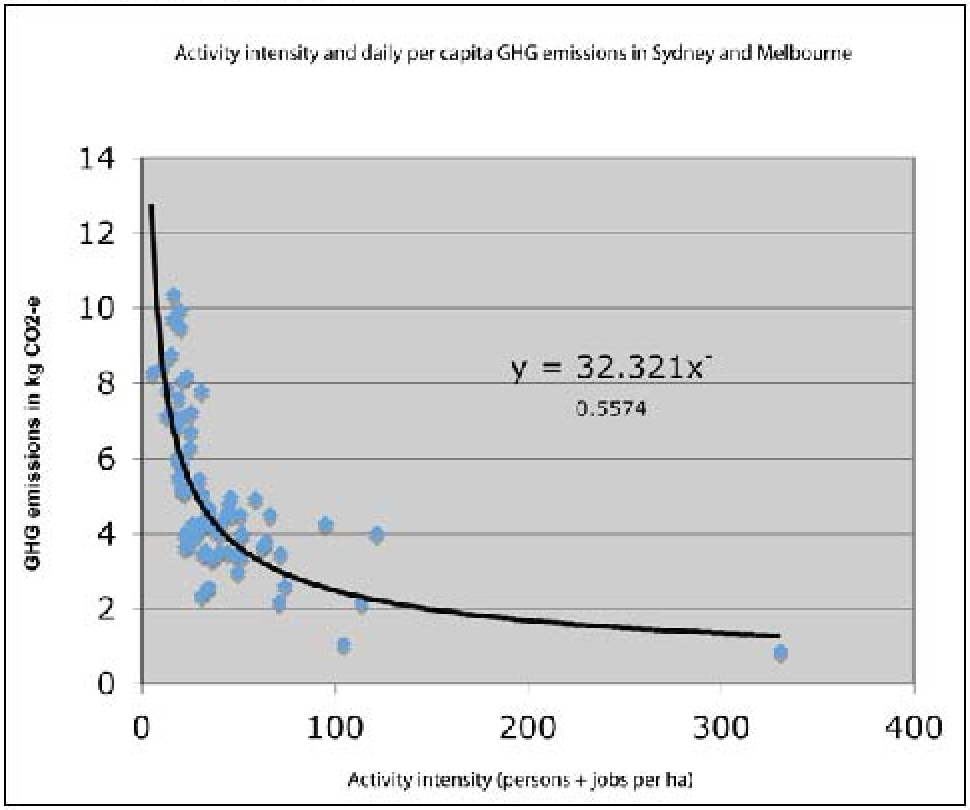
Figure 2: GHG emissions regressed on distance to CBD — Sydney and Melbourne
Daily per capita GHG emissions shown



Activity intensity

The exponential relationship of activity intensity to daily per capita GHG emissions is also strong in this example, but with an R-squared value of 0.6016 it is a less effective predictor than distance to CBD is alone. By observing the trend-line it is evident there is an activity intensity level of roughly 35 people and jobs per hectare above which GHG emissions are most dramatically reduced. This is consistent with findings from Newman and Kenworthy (1999, 2006).

Figure 3: GHG emissions regressed on activity Intensity — Sydney and Melbourne
Daily per capita GHG emissions shown

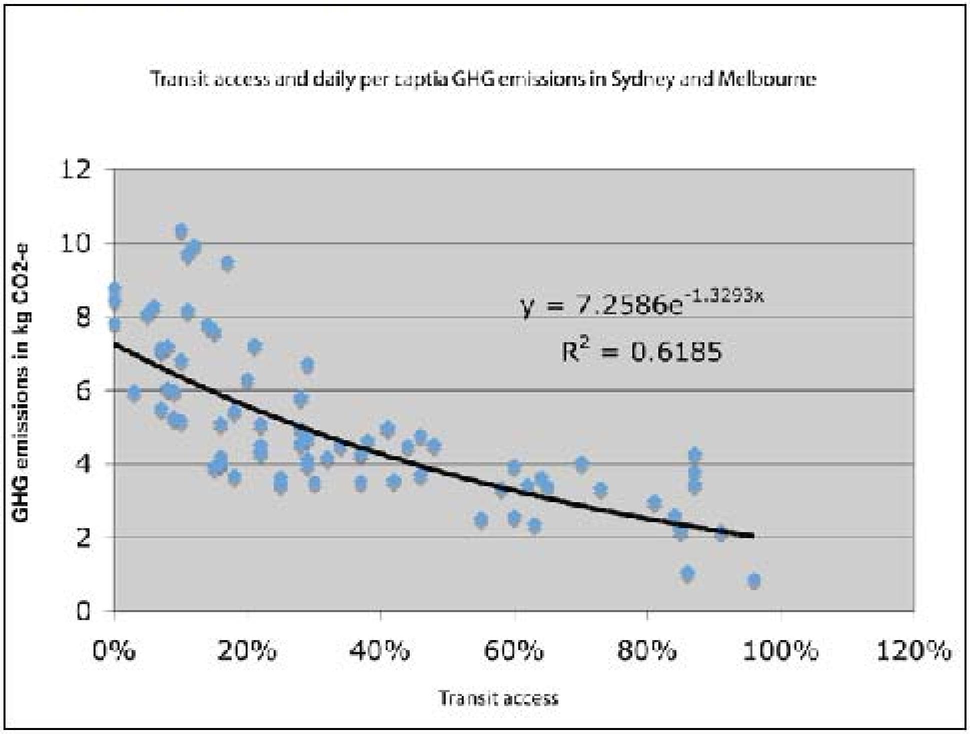


7

Transit access

The parameter of transit access (TA) also seems to be a fairly good predictor of CO_2-e emissions with a similar R-squared to activity intensity. The majority of the variance seems to be in areas of poor transit access. Again, alone it does not seem to be as strong a predictor of emissions as distance to CBD.

Figure 4: GHG emissions regressed on transit accessibility — Sydney and Melbourne
Daily per capita GHG emissions shown



Modelling emissions as a function of various parameters

Running a multiple regression analysis of the three parameters of distance to CBD, activity intensity and transit access, together resulted in the highest R-squared value of all the combinations (see figure 5). In this case, however, activity intensity as a parameter comes out highly insignificant with a p-value of -0.5503 and a very low coefficient of -0.002, so there is no use leaving it in the model.

Figure 5: Predicted against actual GHG emissions – Sydney and Melbourne
Predicted values include distance to CBD, activity intensity and transit accessibility parameters in their estimation

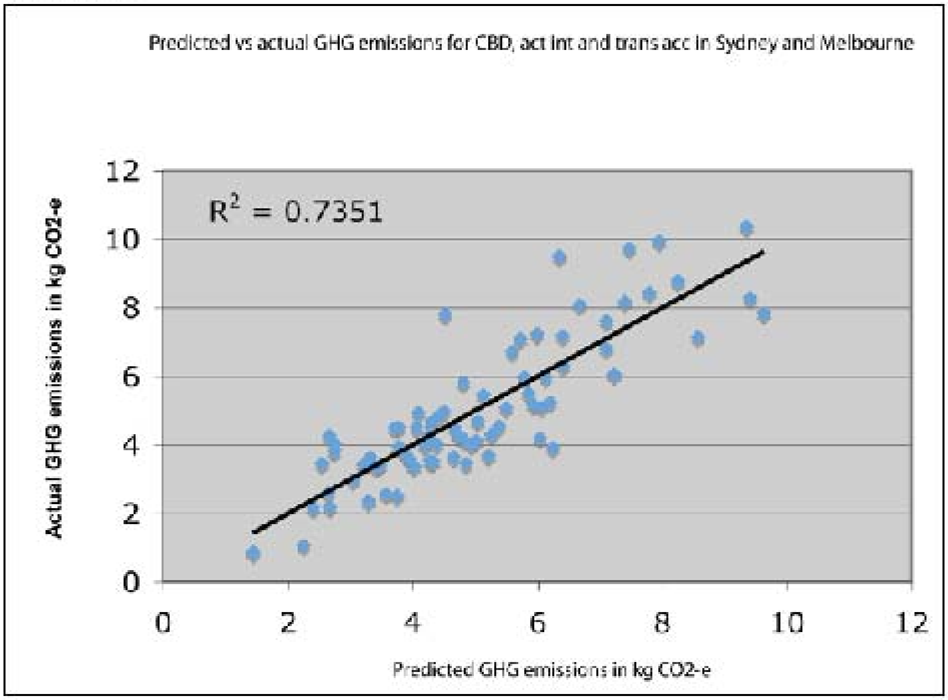


Table 5: Output results from regressing GHG on Distance to CBD, AI, and TA

	Coefficients	P-value
Intercept	4.38	1.44E-15
To CBD	0.07	7.30E-10
Activity intensity	-0.00	0.59
Transit access	-2.28	0.00

Regression statistics	
Multiple R	0.86
R Square	0.74
Adjusted R Square	0.72
Observations	75
Observations	75

The outcome of activity intensity becoming insignificant in the multiple regression is not surprising as it tends to be a surrogate of distance to CBD. In other words, it is characteristic of most modern cities that the inner cores would be high in number of residents and jobs and that this intensity would diminish as distance from the city centre increases. Activity intensity was withdrawn from the model and another attempt was made at modelling GHG emissions with just distance to CBD and transit accessibility (see figure 6).

Figure 6: Predicted against actual GHG emissions – Sydney and Melbourne
Predicted values include the distance to CBD and transit accessibility parameters in their estimation

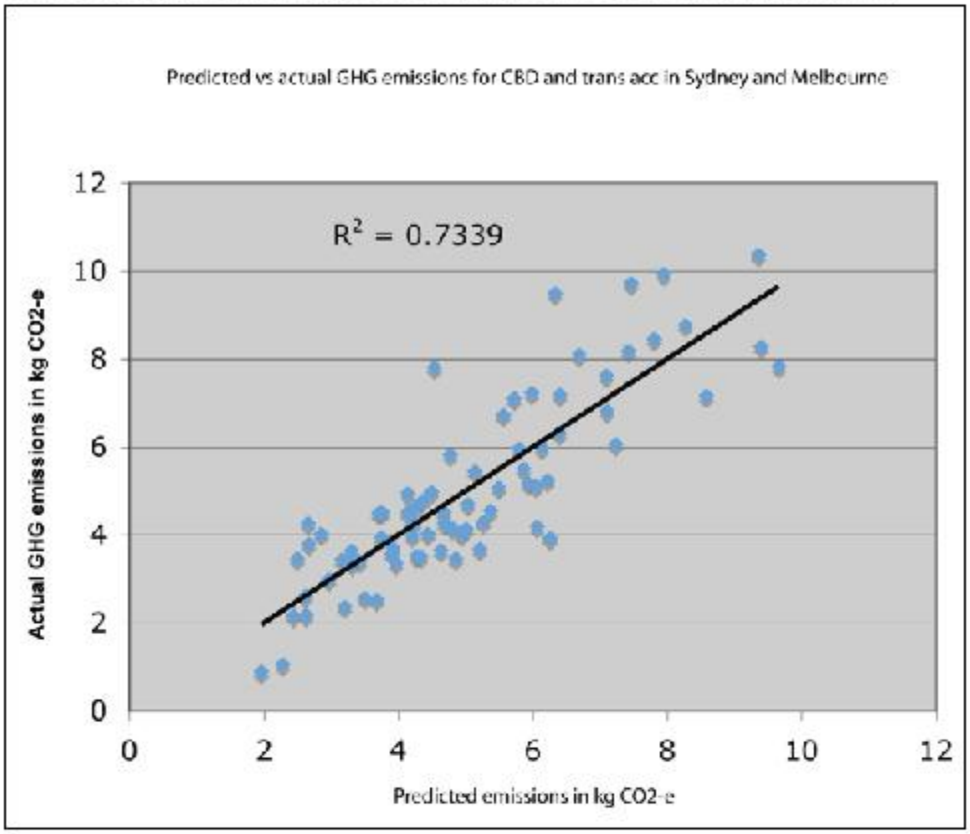


Table 6: Output results from regressing GHG on Distance to CBD and TA

	Coefficients	P-value	Regression statistics	
Intercept	4.35	9.80E-16	Multiple R	0.86
To CBD	0.07	4.97E-10	R Square	0.73
Transit access	-2.50	0.00	Adjusted R Square	0.73
			Observations	75
			Observations	75

The resulting analysis produced the equation **y = .073x – 2.5z + 4.35,** where y is the daily per capita GHG emissions, x is the distance to the CBD, and z is the level of transit service expressed as a percentage of the area covered. The results were significant at the 5% level and the equation generated accounts for 73.4% of the variance in CO_2-e.

It is worth considering any other possible factors, such as social or economic, which could be influencing this equation. In Australian cities income decreases with distance from the CBD and as higher income households are known to drive more (Luk and Hepburn 1993), then the physical parameters in the model could only be stronger in explaining transport greenhouse. Social factors, such as different ethnic and racial groups, are also not likely to explain the data as there are no strong groupings in Australian cities that appear to have any different travel patterns.

2.4 Calculating emissions costs for the alternative development typologies

The equation $y = 0.073x - 2.5z + 4.35$ is useful in estimating the daily per capita GHG emissions for any given area where the distance to the city centre and the transit accessibility is known. For the purpose of this economic assessment, however, it was used to estimate the GHG costs for the two dichotomous development types: inner-city type redevelopments and fringe developments. For this the equation had to be amended to accommodate the economic reporting requirements.

The desired end estimate of the equation was to predict the annual economic CO_2-e emissions impact for a development of 1,000 dwellings. By adjusting the equation to achieve this we get the following:

GHG gas cost as a function of distance to CBD and transit accessibility

$$Y = (365\ \text{days/yr})(\text{price/kg } CO_2\text{-e})\ (\text{No. of dwellings})(\text{Inhabitants/dwelling})(.073x - .25z + 4.35)$$

$$= (365)(0.025)(1000)(2.5)(.073x - .25z + 4.35)$$

$$= 22{,}812.5(.073x - .25z + 4.35)$$

where Y = annual cost   x = distance to CBD   z = transit accessibility

The next critical step was to choose values for the variables in the equation to represent the two opposing urban forms of interest. The furthest local government area from the city centre of Melbourne in the data was measured at just over 60 km and for Sydney it was nearly 73 km, while the transit accessibility for the most distant local government areas was 7% and 0% for the two cities respectively. Averages were taken of these respective figures to represent the inputs for estimating the annual cost of CO_2-e production in a fringe development. The figures for an inner-city type development we determined a little differently. Choosing the immediate centre would perhaps be an ambitious suggested location for a new development of 1,000 dwellings, although still very possible as demonstrated by Vancouver and envisioned in projects like North Port Quay and the Gateway project in Perth. Therefore a development within a 30 minute walking distance (3 km) from the city centre was chosen. For both Sydney and Perth this represents roughly a transit accessibility of 85%. It is feasible to imagine developments in various parts of the city that demonstrate the kind of central/inner-core urban form where 85% of the development has a transit service of better than 15 minutes. Taking these inputs into account, the estimated daily per capita GHG emissions and the annual emissions costs for 1,000 dwellings (at A$25/t) for the two opposing city forms were estimated as follows:

Calculating inner-city and outer/fringe development per capita daily GHG emissions

Daily emissions: Y(Outer)	= [.073(66.5) - .25(.035) + 4.35]
	= 9.2 kg CO_2-e per capita
Y(Inner)	= [.073(3.0) - .25(.85) + 4.35]
	= 4.4 kg CO_2-e per capita
Annual costs: Y(Outer)	= 22,812.5[.073(66.5) - .25(.035) + 4.35]
	= 22,812.5(9.19575)
	= A$209,778
Y(Inner)	= 22,812.5[.073(3.0) - .25(.85) + 4.35]
	= 22,812.5(4.3565)
	= A$99,383

where Y = annual cost x = distance to CBD z = transit accessibility

For the most part, the estimates are quite accurate against the actual data for the two cities; however, the formula tends to overestimate the per capita GHG emissions (and therefore costs) of inner-city-type areas while providing a more accurate estimate as distance to CBD increases. Data points from Sydney's and Melbourne's inner areas report daily per capita emissions of roughly 2.5 kg CO_2-e, which is roughly half of what the equation generates as a predicted value. This is likely due to the greater proportion of walking trips and higher level of amenity in their inner cores. This disparity for inner area estimates in the end adds a conservative property to the economic comparison between the two urban forms.

When applying the formula, the annual greenhouse gas emissions associated with the two iconic development types for a development of 1000 units is 8,400 t for the fringe and 4,000 t for the redevelopment (2,300 t if the actual average core/inner value is taken).

Next, the present value of the recurring greenhouse gas costs was calculated over 50 years at a 3% discount rate[2]. For fringe developments this equates to a greenhouse gas cost of A$5.40 million and for inner-city type developments this equates to A$2.56 million. This means a 1000 unit development over 50 years saves $2.84 million or if the social cost of carbon ($A215) were used then this would be $24.42 million.

[2] An annuity period of 50 years was chosen for the entire economic assessment of urban form under the assumption that 50 years is a reasonable average life expectancy of a residential development. Beyond 50 years, the decision to redevelop an area may be made once again.

3.0 Conclusion

If inner area-type development is preferred to fringe area-type development there will be a savings of around 4.4 t per household in transport greenhouse gases, which is around 20% of an average Australian household's contribution to climate change. This will mean a saving of around $2840 per household over 50 years (NPV) if carbon is just offset at around $25/t, or $$24,420 if it is to cover the social cost at $215/t. Compared to the magnitude of the costs associated with transport and infrastructure, GHG costs alone are unlikely to catalyse urban planning reform based on a carbon cost associated with offsets. The cost of climate change impacts are not indicative of a benefit that is aligned with the growing attention that all governments must give to reducing national emissions. The next phase of urban development is likely to require a simple assessment that asks what it is doing to reduce GHG. Our model can be used to show savings if there is evidence of redevelopment at higher densities with reasonable transit accessibility. A model or tool of this sort is in itself useful to planners who can then build accordingly to reduce greenhouse gas emissions by estimated amounts. The next paper in this series will examine the health and productivity aspects of more activity-oriented urban design.

References

Australian Government Department of Climate Change (AGDCC) 2007, *Tracking to the Kyoto Target: Australia's Greenhouse Emissions Trends 1990 to 2008-2012 and 2020*, Department of Climate Change, Canberra.

Chandra, L 2006, 'Modeling the impact of urban form and transport provision on transport-related greenhouse gas emissions', Masters Thesis, Murdoch University, Perth.

Clarkson, R & Deyes, K 2002, *Estimating the Social Cost of Carbon Emissions*, report by the Environment Protection Economics Division, DEFRA, London.

Garnaut, R 2008, *The Garnaut Climate Change Review*. Cambridge University Press, Port Melbourne.

Lazarus, N 2003, Beddington Zero (Fossil) *Energy Development: Toolkit for Carbon Neutral Developments*, Bioregional Development Group, London.

Luk, J & Hepburn, S 1993, *New Review of Australian Travel Demand Elasticities*, Australian Road Research Board.

Miller, A & Ambrose, M 2005, *Energy efficient multi-storey residential developments*, Conference on Sustainable Building South East Asia, Malaysia.

Newman, P & Kenworthy, J 1999, *Sustainability and Cities: Overcoming Automobile Dependence*, Island Press, Washington.

Newman, P, and Kenworthy, J 2006, 'Urban Design to Reduce Automobile Dependence', *Opolis*, 2(1), p.35-52.

Newton PW, Tucker SN & Ambrose MD 2000, 'Housing form, energy use and greenhouse gas emissions', in Williams K, Burton E & Jenks M. (eds), *Achieving sustainable urban form*, E & FN Spon: London, UK, pp. 74-83.

Newton, P & Tucker, S 2009, 'Hybrid Buildings: Towards zero carbon housing', *Architectural Science Review*, 52(2), in press.

Price, R, Thornton, S & Nelson, S 2007, *The Social Cost of Carbon and the Shadow Price of Carbon: What they are, and how to use them in economic appraisal in the UK*, Office of Public Sector Information, DEFRA, London.

Ribon, L & Scott, H 2007, *Carbon Offset Providers in Australia 2007*, (unpublished), Global Sustainability at RMIT University, Melbourne.

Stern, N H 2007, *The Economics of Climate Change: The Stern Review*, Cambridge University Press, Cambridge.

Troy, P 1996, *The Perils of Urban Consolidation*, The Federation Press, Leichhardt.

United Nations Framework Convention on Climate Change (UNFCCC), (n.d.), *An introduction to the Kyoto Protocol Compliance Mechanism*, viewed 16 May 2008, http://unfccc.int/kyoto_protocol/compliance/introduction/items/3024.php

Biography

Roman Trubka is a PhD candidate at the Curtin University Sustainability Policy (CUSP) Institute in Western Australia.

Peter Newman is a Professor of Sustainability and Director at the Curtin University Sustainability Policy (CUSP) Institute in Western Australia.

Darren Bilsborough is the Director of Sustainability at Parsons Brinckerhoff (PB) and an adjunct professor of sustainability at CUSP.

Appendix 1: GHG emissions costs for local government areas — Sydney

Assuming a Cost of A$25 per tonne CO_2-e

LGA	Distance from CBD	GHG emissions car CO_2-e	Car emissions per capita	GHG cost per day	GHG cost per year	Share of total cost	GHG savings per year
	km	kg/day	kg/capita/day	$/1000 dwellings/day	$/1000 dwellings/yr	%	$/1000 dwellings per year
Sydney (CC)	0.10	17,695	0.889	$55.56	$20,280.31	0.38%	$216,330.94
North Sydney	3.40	228,084	4.034	$252.13	$92,025.63	1.75%	$144,585.63
Leichhardt	4.30	216,510	3.467	$216.69	$79,090.94	1.50%	$157,520.31
Woollahra	4.60	229,842	4.516	$282.25	$103,021.25	1.95%	$133,590.00
Mosman	4.90	88,898	3.434	$214.63	$78,338.13	1.49%	$158,273.13
South Sydney	5.00	200,239	2.171	$135.69	$49,525.94	0.94%	$187,085.31
Lane Cove	6.20	138,528	4.504	$281.50	$102,747.50	1.95%	$133,863.75
Waverly	6.60	259,363	4.275	$267.19	$97,523.44	1.85%	$139,087.81
Drummoyne	6.70	132,690	4.024	$251.50	$91,797.50	1.74%	$144,813.75
Marrickville	6.70	279,022	3.800	$237.50	$86,687.50	1.64%	$149,923.75
Willoughby	6.70	293,790	4.950	$309.38	$112,921.88	2.14%	$123,689.38
Hunters Hill	7.30	99,129	7.810	$488.13	$178,165.63	3.38%	$58,445.63
Ashfield	7.50	143,874	3.643	$227.69	$83,105.94	1.58%	$153,505.31
Botany	7.90	143,281	3.991	$249.44	$91,044.69	1.73%	$145,566.56
Manly	8.20	170,071	4.525	$282.81	$103,226.56	1.96%	$133,384.69
Randwick	9.60	413,492	3.403	$212.69	$77,630.94	1.47%	$158,980.31
Burwood	9.80	111,118	3.709	$231.81	$84,611.56	1.60%	$151,999.69
Concord	10.00	94,393	3.513	$219.56	$80,140.31	1.52%	$156,470.94
Ryde	11.10	409,550	4.278	$267.38	$97,591.88	1.85%	$139,019.38
Rockdale	11.50	312,232	3.527	$220.44	$80,459.69	1.53%	$156,151.56
Strathfield	11.90	131,386	4.658	$291.13	$106,260.63	2.02%	$130,350.63
Canterbury	12.20	517,443	3.952	$247.00	$90,155.00	1.71%	$146,456.25
Ku-ring-gai	15.30	590,484	5.826	$364.13	$132,905.63	2.52%	$103,705.63
Auburn	15.50	195,499	3.468	$216.75	$79,113.75	1.50%	$157,497.50
Hurstville	15.90	338,012	4.785	$299.06	$109,157.81	2.07%	$127,453.44
Kogarah	15.90	250,806	4.982	$311.38	$113,651.88	2.16%	$122,959.38
Parramatta	16.10	652,438	4.515	$282.19	$102,998.44	1.95%	$133,612.81
Warringah	16.90	702,831	5.455	$340.94	$124,442.19	2.36%	$112,169.06
Bankstown	19.30	779,676	4.708	$294.25	$107,401.25	2.04%	$129,210.00
Pittwater	21.20	374,942	7.101	$443.81	$161,991.56	3.07%	$74,619.69

LGA	Distance from CBD	GHG emissions car CO_2-e	Car emissions per capita	GHG cost per day	GHG cost per year	Share of total cost	GHG savings per year
	km	kg/day	kg/capita/ day	$/1000 dwellings/ day	$/1000 dwellings/yr	%	$/1000 dwellings per year
Holroyd	23.20	436,193	5.086	$317.88	$116,024.38	2.20%	$120,586.88
Hornsby	26.50	982,347	6.730	$420.63	$153,528.13	2.91%	$83,083.13
Liverpool	27.50	924,301	5.991	$374.44	$136,669.69	2.59%	$99,941.56
Fairfield	29.00	764,935	4.204	$262.75	$95,903.75	1.82%	$140,707.50
Sutherland	29.50	1,471,140	7.244	$452.75	$165,253.75	3.13%	$71,357.50
Baulkham Hills	33.70	1,129,224	8.100	$506.25	$184,781.25	3.50%	$51,830.00
Blacktown	34.90	1,618,947	6.315	$394.69	$144,060.94	2.73%	$92,550.31
Campbelltown	42.70	1,111,737	7.622	$476.38	$173,876.88	3.30%	$62,734.38
Penrith	45.80	1,411,810	8.189	$511.81	$186,811.56	3.54%	$49,799.69
Camden	47.30	371,556	8.455	$528.44	$192,879.69	3.66%	$43,731.56
Gosford	53.70	1,358,030	8.781	$548.81	$200,316.56	3.80%	$36,294.69
Blue Mountains	71.20	616,266	8.292	$518.25	$189,161.25	3.59%	$47,450.00
Hawkesbury	72.00	633,436	10.372	$648.25	$236,611.25	4.49%	-
Wyong	72.70	1,026,623	7.846	$490.38	$178,986.88	3.39%	$57,624.38
Total		22,371,863			$5,272,881.25	100.00%	

Source: Chandra, 2006

Appendix 2: GHG emissions costs for local government areas — Melbourne

Assuming a cost of A$25 per tonne CO_2-e

LGA	Distance from CBD	GHG emissions car CO_2-e	Car emissions per capita	GHG cost per day	GHG cost per year	Share of total cost	GHG savings per year
	km	kg/day	kg/capita/ day	$/1000 dwellings/ day	$/1000 dwellings/yr	%	$/1000 dwellings per year
Melbourne	1.00	72,560	1.070	$66.88	$24,409.38	0.73%	$202,506.57
Port Phillip	4.97	209,989	2.620	$163.75	$59,768.75	1.78%	$167,147.19
Yarra	5.41	148,770	2.187	$136.69	$49,890.94	1.49%	$177,025.00
Maribyrnong	5.81	142,019	2.376	$148.50	$54,202.50	1.62%	$172,713.44

LGA	Distance from CBD	GHG emissions car CO_2-e	Car emissions per capita	GHG cost per day	GHG cost per year	Share of total cost	GHG savings per year
	km	kg/day	kg/capita/day	$/1000 dwellings/day	$/1000 dwellings/yr	%	$/1000 dwellings per year
Moonee Valley	8.31	378,122	3.563	$222.69	$81,280.94	2.43%	$145,635.00
Stonnington	8.75	261,513	2.992	$187.00	$68,255.00	2.04%	158,660.94
Moreland	8.87	339,095	2.581	$161.31	$58,879.06	1.76%	168,036.88
Darebin	9.63	314,038	2.536	$158.50	$57,852.50	1.73%	$169,063.44
Boroondara	10.72	503,250	3.350	$209.38	$76,421.88	2.28%	$150,494.07
Hobsons Bay	12.31	293,734	3.652	$228.25	$83,311.25	2.49%	$143,604.69
Brimbank	13.50	654,809	4.011	$250.69	$91,500.94	2.73%	$135,415.00
Glen Eira	14.49	397,819	3.367	$210.44	$76,809.69	2.29%	$150,106.25
Bayside	17.10	350,496	4.168	$260.50	$95,082.50	2.84%	$131,833.44
Banyule	17.28	488,445	4.276	$267.25	$97,546.25	2.91%	$129,369.69
Whittlesea	17.90	392,445	3.687	$230.44	$84,109.69	2.51%	$142,806.25
Whitehorse	18.90	582,401	4.138	$258.63	$94,398.13	2.82%	$132,517.82
Monash	19.93	675,837	4.307	$269.19	$98,253.44	2.93%	$128,662.50
Manningham	20.76	636,298	5.973	$373.31	$136,259.06	4.07%	$90,656.88
Wyndham	23.00	447,716	5.515	$344.69	$125,810.94	3.76%	$101,105.00
Kingston	23.56	584,181	4.558	$284.88	$103,979.38	3.10%	$122,936.57
Hume	25.11	654,918	5.175	$323.44	$118,054.69	3.52%	$108,861.25
Knox	28.69	745,187	5.257	$328.56	$119,925.31	3.58%	$106,990.63
Maroondah	28.73	493,494	5.116	$319.75	$116,708.75	3.48%	$110,207.19
Melton	30.70	350,424	7.201	$450.06	$164,272.81	4.90%	$62,643.13
Greater Dandenong	31.19	485,786	3.929	$245.56	$89,630.31	2.68%	$137,285.63
Nillumbik	33.00	453,164	9.501	$593.81	$216,741.56	6.47%	$10,174.38
Frankston	41.09	736,346	6.822	$426.38	$155,626.88	4.65%	$71,289.07
Casey	42.20	993,651	6.066	$379.13	$138,380.63	4.13%	$88,535.32
Yarra Ranges	46.40	1,032,522	9.731	$608.19	$221,988.44	6.63%	$4,927.50
Cardinia	53.40	128,669	9.947	$621.69	$226,915.94	6.77%	-
Mornington Peninsula	60.30	813,414	7.164	$447.75	$163,428.75	4.88%	$63,487.19
Total		**14,761,112**			**3,349,696.25**	**100.00%**	

Source: Chandra, 2006

Appendix 3: GHG emissions costs for local government areas — Perth

Assuming a Cost of A$25 per tonne CO_2-e

LGA	Distance from CBD	GHG emissions car CO_2-e	Car emissions per capita	GHG cost per day	GHG cost per year	Share of total cost	GHG savings per year
	km	kg/day	kg/capita/day	$/1000 dwellings/day	$/1000 dwellings/yr	%	$/1000 dwellings per year
Perth	1.40	36,047	2.749	$171.81	$62,711.56	1.59%	$292,159.69
Vincent	3.00	110,906	5.606	$350.38	$127,886.88	3.24%	$226,984.38
Subiaco	4.10	36,350	4.949	$309.31	$112,899.06	2.86%	$241,972.19
South Perth	4.80	196,886	6.035	$377.19	$137,673.44	3.49%	$217,197.81
Victoria Park	5.20	73,750	3.735	$233.44	$85,204.69	2.16%	$269,666.56
Claremont	7.80	30,064	4.924	$307.75	$112,328.75	2.85%	$242,542.50
Bayswater	8.10	362,442	5.914	$369.63	$134,913.13	3.42%	$219,958.13
Nedlands	8.10	82,524	4.351	$271.94	$99,257.19	2.52%	$255,614.06
Cambridge	8.30	189,199	5.661	$353.81	$129,141.56	3.27%	$225,729.69
Belmont	8.60	120,167	4.13	$258.13	$94,215.63	2.39%	$260,655.63
Stirling	9.50	981,428	5.64	$352.50	$128,662.50	3.26%	$226,208.75
Canning	9.80	213,977	5.55	$346.88	$126,609.38	3.21%	$228,261.88
Melville	9.80	462,567	6.418	$401.13	$146,410.63	3.71%	$208,460.63
Peppermint Grove	9.90	6,930	5.575	$348.44	$127,179.69	3.22%	$227,691.56
Bassendean	10.00	32,148	4.065	$254.06	$92,732.81	2.35%	$262,138.44
Cottesloe	10.30	37,436	6.958	$434.88	$158,729.38	4.02%	$196,141.88
Mosman Park	10.90	69,687	9.043	$565.19	$206,293.44	5.23%	$148,577.81
East Fremantle	12.50	29,641	5.209	$325.56	$118,830.31	3.01%	$236,040.94
Fremantle	14.50	95,283	4.481	$280.06	$102,222.81	2.59%	$252,648.44
Gosnells	16.70	356,948	6.687	$417.94	$152,547.19	3.87%	$202,324.06
Kalamunda	17.00	234,881	6.349	$396.81	$144,836.56	3.67%	$210,034.69
Swan	17.30	78,706	4.799	$299.94	$109,477.19	2.77%	$245,394.06
Cockburn	19.40	171,920	5.463	$341.44	$124,624.69	3.16%	$230,246.56
Joondalup	19.60	443,553	6.661	$416.31	$151,954.06	3.85%	$202,917.19
Armadale	25.50	279,105	7.267	$454.19	$165,778.44	4.20%	$189,092.81
Wanneroo	27.70	142,278	5.203	$325.19	$118,693.44	3.01%	$236,177.81
Kwinana	30.50	38,170	6.909	$431.81	$157,611.56	3.99%	$197,259.69
Mundaring	37.30	290,837	15.556	$972.25	$354,871.25	8.99%	-

LGA	Distance from CBD	GHG emissions car CO_2-e	Car emissions per capita	GHG cost per day	GHG cost per year	Share of total cost	GHG savings per year
	km	kg/day	kg/capita/ day	$/1000 dwellings/ day	$/1000 dwellings/yr	%	$/1000 dwellings per year
Rockingham	45.00	110,844	7.072	$442.00	$161,330.00	4.09%	$193,541.25
Total		5,314,674			$3,945,627.19	100.00%	

Source: Chandra, 2006

Appendix 4: Sydney and Melbourne data: daily per capita CO_2-e and modeled parameters

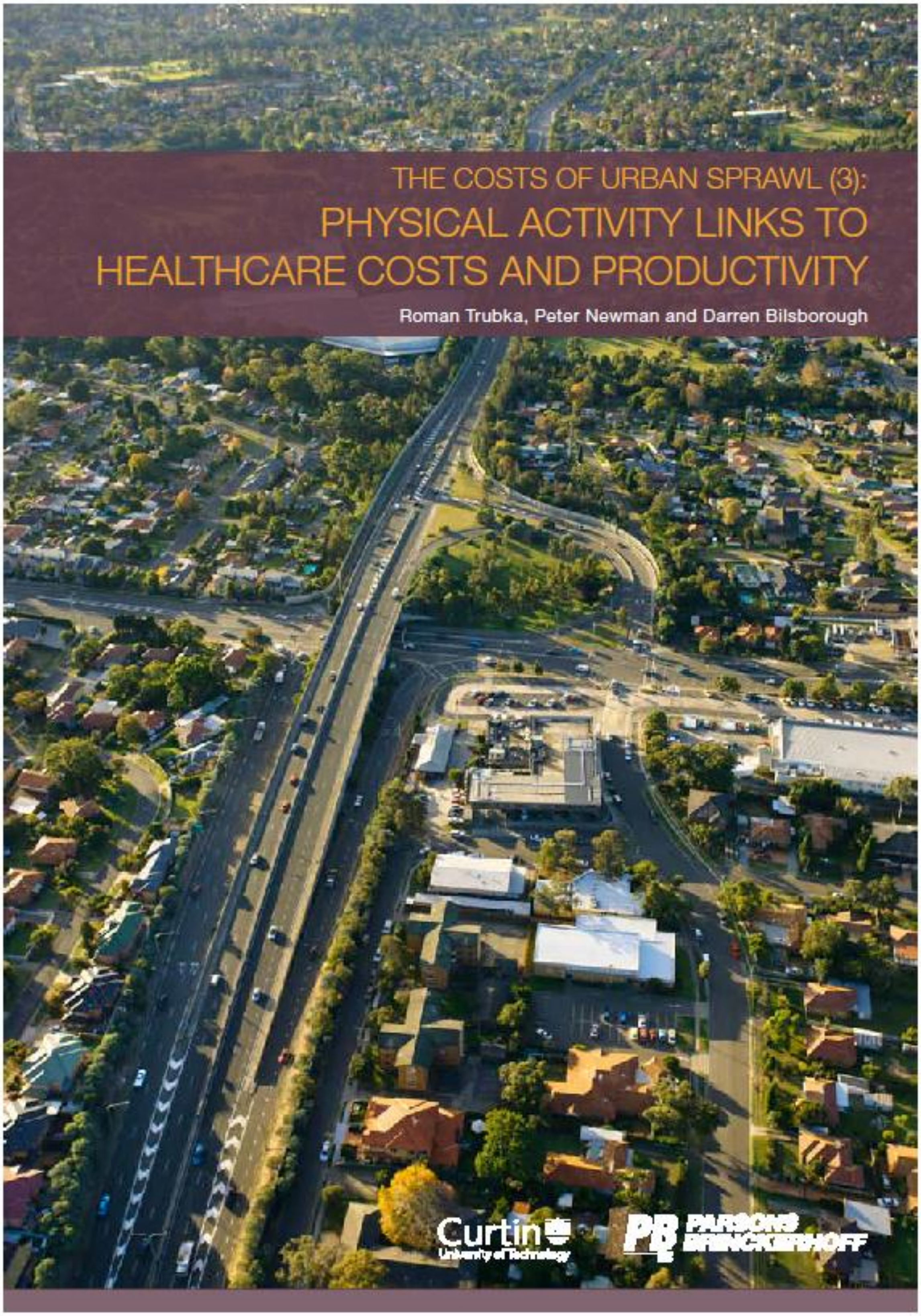
Suburb	GHG/capita	To CBD	Activity intensity	Transit access
Melbourne	1.070	1.00	104.491	86%
Port Phillip	2.620	4.97	74.37067	84%
Yarra	2.187	5.41	71.32699	85%
Maribyrnong	2.376	5.81	31.18083	63%
Moonee Valley	3.563	8.31	33.04913	42%
Stonnington	2.992	8.75	49.66764	81%
Moreland	2.581	8.87	34.94962	60%
Darebin	2.536	9.63	33.6832	55%
Boroondara	3.350	10.72	36.07256	73%
Hobsons Bay	3.652	12.31	22.11571	25%
Brimbank	4.011	13.50	22.11571	16%
Glen Eira	3.367	14.49	37.8036	58%
Bayside	4.168	17.10	30.5488	32%
Banyule	4.276	17.28	26.69309	37%
Whittlesea	3.687	17.90	24.69786	18%
Whitehorse	4.138	18.90	30.32901	29%
Monash	4.307	19.93	30.74724	22%
Manningham	5.973	20.76	18.33446	3%
Wyndham	5.515	23.00	19.19991	7%
Kingston	4.558	23.56	34.87298	28%
Hume	5.175	25.11	20.36164	10%
Knox	5.257	28.69	24.13857	9%
Maroondah	5.116	28.73	22.07024	16%

Suburb	GHG/capita	To CBD	Activity intensity	Transit access
Melton	7.201	30.70	15.19053	8%
Greater Dandenong	3.929	31.19	27.47892	15%
Nillumbik	9.501	33.00	20.12773	17%
Frankston	6.822	41.09	17.71661	10%
Casey	6.066	42.20	18.92039	8%
Yarra Ranges	9.731	46.40	16.0437	11%
Cardinia	9.947	53.40	19.82883	12%
Mornington Peninsula	7.164	60.30	13.20078	7%

Suburb	GHG/capita	To CBD	Activity intensity	Transit access
Sydney (CC)	0.889	0.10	330.6339	96%
North Sydney	4.034	3.40	121.8281	70%
Leichhardt	3.467	4.30	71.86342	87%
Woollahra	4.516	4.60	66.41607	22%
Mosman	3.434	4.90	48.7702	62%
South Sydney	2.171	5.00	113.8737	91%
Lane Cove	4.504	6.20	51.24671	44%
Waverly	4.275	6.60	95.10611	87%
Drummoyne	4.024	6.70	51.78653	16%
Marrickville	3.800	6.70	64.37259	87%
Willoughby	4.950	6.70	58.96366	28%
Hunters Hill	7.810	7.30	30.77002	14%
Ashfield	3.643	7.50	62.40785	64%
Botany	3.991	7.90	40.32756	29%
Manly	4.525	8.20	46.70056	48%
Randwick	3.403	9.60	52.13946	65%
Burwood	3.709	9.80	63.90592	46%
Concord	3.513	10.00	40.06209	30%
Ryde	4.278	11.10	40.98237	37%
Rockdale	3.527	11.50	44.24853	37%
Strathfield	4.658	11.90	34.91641	38%
Canterbury	3.952	12.20	51.74705	60%
Ku-ring-gai	5.826	15.30	20.88535	28%
Auburn	3.468	15.50	31.85958	25%
Hurstville	4.785	15.90	44.94621	46%
Kogarah	4.982	15.90	46.03009	41%
Parramatta	4.515	16.10	43.71826	34%
Warringah	5.455	16.90	29.96319	18%

Suburb	GHG/capita	To CBD	Activity intensity	Transit access
Bankstown	4.708	19.30	34.55737	29%
Pittwater	7.101	21.20	20.54202	7%
Holroyd	5.086	23.20	31.12426	22%
Hornsby	6.730	26.50	25.35693	29%
Liverpool	5.991	27.50	20.89103	9%
Fairfield	4.204	29.00	35.95417	16%
Sutherland	7.244	29.50	25.4291	21%
Baulkham Hills	8.100	33.70	21.09561	5%
Blacktown	6.315	34.90	24.84111	20%
Campbelltown	7.622	42.70	19.0944	15%
Penrith	8.189	45.80	23.51239	11%
Camden	8.455	47.30	10.79457	0%
Gosford	8.781	53.70	15.06618	0%
Blue Mountains	8.292	71.20	5.30491	6%
Hawkesbury	10.372	72.00	16.30905	10%
Wyong	7.846	72.70	13.44757	0%

Source: Chandra, 2006


THE COSTS OF URBAN SPRAWL (3):
PHYSICAL ACTIVITY LINKS TO HEALTHCARE COSTS AND PRODUCTIVITY
Roman Trubka, Peter Newman and Darren Bilsborough
Curtin
University of Technology
PB
PARSONS BRINCKERHOFF

The overall study has compared the costs of development between two alternative development paths: urban redevelopment and greenfield development on the urban fringe. The results of the research are presented in the form of three companion papers: (1) investigating the economic costs associated with infrastructure and transportation, (2) greenhouse gases, and (3) the health and productivity benefits of active-travel associated with the different urban forms due to levels of density, connectivity, and variety in amenity.

This paper shows that healthcare savings related to active forms of travel over a 50-year urban lifetime are quite small at $2.3 million for 1000 dwellings. But if these more walkable developments are pursued then the benefits to employment productivity are large — estimated at a present value of $34 million. This is a substantial benefit that is comparable in scale to the savings in transport and infrastructure, as well as the social costs of greenhouse gases, and should provide a critical input to urban planning decision-making.

Keywords: sprawl, redevelopment, urban planning, activity, health, productivity

1.0 Introduction

Australian cities are under focus as the Federal Government begins to invest in urban infrastructure and questions are raised about the fuel, greenhouse gas and health impacts of such investment. This paper examines two alternative approaches to urban development: redevelopment in walkable, transit-oriented developments and fringe development in conventional low-density, car-dependent suburbs. Redevelopment is based around present urban areas that are already well served by public transport but can also include new developments, so long as transit accessibility, walkability, and density are included in the planning and design process. The two development types are set out in detail in the first paper of the series.

The companion papers discuss the physical planning costs associated with the different transport and infrastructure requirements and then examine a new area of public policy — greenhouse gas emissions. This paper discusses activity-related health and productivity costs. These are the subjects of increasing interest and their economic costs can then be compared with the more traditional costs of physical planning.

Section 2 of this paper will discuss the health effects of sedentary living and walk through a calculation of the economic benefits of active-travel neighbourhoods on the healthcare system. *Section 3* will briefly review the situation of workplace fitness programs in western countries and bridge their employee productivity benefits with the potential benefits of active travel neighbourhoods in a basic economic appraisal. *Section 4* discusses some of the issues surrounding the calculation of both types of activity-related benefits. *Section 5* summarises the cumulative impacts of the findings from the paper and the companion papers and finally *Section 6* concludes.

2.0 Health-related costs

Recent years have set the stage for increased interest in the topic of urban form's influence on public health. The view that car dependency has led to the creation of obesogenic environments is now supported by a substantial number of studies and the case is being built for urban planning reform for active lifestyle improvements (Ewing et al. 2003; Frank et al. 2004). Nations have made estimates of healthcare costs as experienced by the burden of inactivity among their populations — for instance it is estimated that 1.5–3.0% of total direct healthcare costs are related to inactivity in developed countries (Oldridge 2008); however, the economic assessment associating the costs of illness, inactivity, and urban form to an urban planning mindset has yet to been done.

The allotting of more residential zones in greenfield areas is a further commitment to car dependency and inactive travel but has been the conventional model for residential growth since World War II[1] (Newman & Kenworthy 1999). The environments that we create, aesthetically and functionally, have profound consequences on our emotional connectivity to the people around us and to our physical settings, affecting both our quality of life and the manner in which we interact with the cities we live in (Frank & Engelke 2005; Stokols et al. 2003).

In addition to psychological effects, research has also been able to link aspects of the built environment directly to human activity patterns and travel choices for both non-discretionary travel and leisure (Frank & Engelke 2005). The purpose of this research is to economically quantify the health benefit of refocusing future development to inner city type areas where transit and active means of travel can make for a

[1] The term 'Walking Cities' is associated with traditional cities built before the era of the tram and train (from the late 19th century), which created the Transit City (with walking cities associated with station precincts). Auto Cities are said to have formed after the war as personalised transport by car became more widespread, allowing people to travel further and faster (Newman and Kenworthy 1999).

healthier population. We have defined an **active-travel** neighbourhood as one that is conducive to both cycling and walking, which in daily life activities could lead to most able bodied people engaging in at least 30 minutes of active travel per day.

2.1 Background

A growing body of evidence suggests that neighbourhoods characterised by low density, poor connectivity, and poor access to shops and services, are associated with low levels of walking. Moreover, sprawling areas of low walkability have been linked to obesity and numerous other chronic illnesses (Giles-Corti 2006; Sturm & Cohen 2004). Australia now has one of the most obese populations, ranking 21st in the world and 3rd among all English-speaking countries (Forbes 2007).

2.2 Obesity

Between 1980 and 2000, obesity levels increased among males (10%) and females (12%) in Australia as measured by a BMI (Body Mass Index) of equal to or greater than 30 kg/m2 (ANZOS n.d.). Over the same period, the portion of overweight Australians increased 17.5% and 18% for males and females respectively, measured by a BMI equal to or greater than 25 kg/m2 (ANZOS n.d.). Although not all obesity is attributable to urban form, one study has documented in its research that each additional hour spent in a car per day was associated with a 6% increase in the odds of being obese, while each additional kilometre walked per day was associated with a 4.8% reduction in the odds of being obese (Frank et al. 2006). This can have significant implications, as another study has shown that individuals in the least walkable environments drive the most per day (74 km), while those in the most walkable environments drive the least (43 km) (Frank et al. 2007a).

2.3 Other illnesses

Obesity is not the only link to urban form. Other illnesses and costs to the healthcare system arise from inactivity, such as falls, coronary heart disease, type 2 diabetes, depression, stroke, colon cancer, and breast cancer on which Econtech[2] reported a cost of $1.5 billion to the Australian healthcare system in 2007 (Econtech 2007). Furthermore, it is stated that about 54.2% of Australian adults were found to be insufficiently active according to National Physical Activity guidelines in the year 2000 (Econtech 2007). This suggests an opportunity to substantially reduce direct and indirect costs due to inactivity in Australia by facilitating incidental activity with urban planning. A joint American-Canadian study found that residents of more walkable environments are 2.4 times more likely to meet or exceed the recommended minimum levels of moderate activity than people in the most sprawling areas (Frank et al. 2005), further showing the potential for urban planning policy intervention for healthcare savings.

2.4 Active travel

Recent advances in research on urban form have researchers starting to identify and separately measure utilitarian active forms of travel from leisure forms of conveyance to acknowledge that sometimes both are accomplished in the same outing, yet also may be affected differently by certain aspects of urban form. In other words, they record and measure how urban form affects walking for the sake of transportation as distinct from walking for the sake of leisure, then cumulatively measure if there is a net gain or loss in active living in walkable neighbourhoods.

2.5 Measuring the factors

Multidimensional measures of urban sprawl are another advancement in urban planning research. They allow for a more objective measure of urban form to be made and thus explain the relationship to travel behaviour with more confidence. This is done by agglomerating urban features such as density, land-use-

[2] Econtech is an Australian consultancy contracted by Medibank Private, Australia's government-owned private health insurer.

mix, proximity, connectivity, and degree of centring to make indexes for experimental designs. All have been proven to be associated with increased active travel (see companion papers).

2.6 Making positive change

Newman and Kenworthy (1999) in their research found that 35 people and jobs per hectare was the threshold density for decreased auto dependence, and beyond that travel by car lessens and active travel and transit use begin to increase. This measure will depend on the degree of transit service provided and the walkability of the design. Other research, such as that by Sturm and Cohen, has closely linked overall physical health to urban density. They found that a difference in their sprawl index of 100 points, which would be the difference between Riverside in California, which is very sprawled, and Boston, Massachusetts, which is characterised by a low level of sprawl, was associated with 200 fewer chronic illnesses per 1,000 persons (Sturm and Cohen 2004).

All this implies that a step towards designing our cities around active transport instead of the automobile can have some profound effects on physical health and possibly mental health. As a result, an increase in discretional and non-discretional active transport could identifiably benefit social capital and public health while saving the healthcare system considerable money. This paper attempts to quantify this potential benefit for the Australian context.

2.7 Data and method for healthcare cost calculation

The calculation method selected for use in this research is best described as a *cost-of-illness approach* to economically appraising the health impacts of urban form. Other methods exist, such as a *years-of-life-lost* (YLL) *approach* or agreeing on a standard value of a 'statistical life;' however, the *cost-of-illness approach* worked best with the available information and allows for the fewest assumptions to be made.

In the process of the economic assessment, information was drawn upon from two separate areas of study, being the cost of inactivity in Australia and the variation in active travel among cities of differing urban form. The overall calculation is done in a series of parts, starting with a top-down approach to place a value on an hour of moderate-intensity activity per person. Next, the hourly per person savings estimate is attributed to an expected increase in activity levels characteristic of active travel neighbourhoods and finally, the healthcare savings for the development of a high-density, mixed-use development of 1,000 dwellings is calculated.

2.7.1 Identifying the value of physical activity in Australia

In 2007, Medibank Private contracted Econtech to produce a report on the direct inactivity costs of Australian adults. This value was estimated at $1.5 billion[3] and included the following seven illnesses: falls, coronary heart disease, type 2 diabetes, depression, stroke, colon cancer, and breast cancer (Econtech 2007). An adult was defined as anyone of the age 18 and over and the value represented the potential savings that could be achieved if more adults became sufficiently active. Their report quoted the 2000 *National Physical Activity Survey* in stating that 54.2% of Australia's adult population is not getting enough physical activity to remain healthy. Using this figure and assuming that the $1.5 billion estimate can be applied to the inactive portion of the adult population, an overall value of $2.8 billion[4] was estimated for the physical-activity related component of health for all Australian adults.

Indirect costs are more difficult to calculate because of the complexity of the assumptions required. Health Canada's *Economic Burden of Illness* (1993) assigns an overall ratio to its economic health assessments

[3] The cost of inactivity in Australia was last calculated in 1993–94 and valued at $377.4 million in 1993 dollars and did not include falls and causally related diseases. Accounting for inflation, falls, increased population, increased obesity levels, and increased inactivity levels would account for most of the difference between this value and the one produced by Econtech Pty. Ltd. for Medibank Private.

[4] Calculated as .542/(1-.542)=(1.5*10^9)/x and isolated for 'x' to determine the value of the current Australian adult population meeting sufficient activity levels and then adding 1.5*10^9 for the value attributed to inactivity.

that approximates indirect costs at 54.3% of the total cost of illness. This approach takes into account productivity losses due to mortality and short- and long-term disability. Using this ratio would estimate Australia's indirect cost of inactivity at $1.78 billion[5], the total cost of inactivity at $3.82 billion[6], and the total value of all Australian adults meeting recommended activity levels at $6.1 billion.

Australia's National Public Health Partnership (ANPHP) estimates that indirect costs would more than double direct costs but provides no numerical value (Bauman et al. 2002). To be conservative and simply say that indirect costs would amount to double the value of direct costs would produce an estimate of $3 billion. Using this estimate would translate into a total cost of $4.5 billion due to inactivity and a total value for all Australian adults of $8.3 billion.

For the purpose of this calculation, the figure of $6.1 billion for the total health value of activity among Australian adults is used in determining the estimate for the healthcare savings of active travel neighbourhoods as the ANPHP's indirect-cost estimate lacks specificity and rigor for our purposes.

2.7.2 Demographic information

The population of Australia is roughly 21 million people (ABS 2007). The cost of inactivity in Australia, however, was determined for the ages 18 and over. This had to be taken into account when calculating a value for each hour of moderate-intensity activity for the Australian adult population. According to data provided by the Australian Bureau of Statistics, 73.3%, or 15.4 million people, fall within this age group (ABS 2006).

Furthermore, there are an average of 2.5 people living in each household in 2003–04 (ABS 2007). With 73.3% of Australia's population over the age of 17, it was estimated that each household contains an average of 1.83 people within this age group. This figure is used in the overall calculation to determine the health-related savings in developing 1000 dwellings as inner city type developments.

2.7.3 Recommended minimum activity levels and associated savings

The National Physical Activity Guidelines for Australians recommends that people should engage in 30 minutes of moderate-intensity physical activity a day over at least five sessions per week to be considered physically active. These 2.5 hours per week can be met by walking 15 minutes to and from the bus during a standard workweek, or more generally by engaging in more active travel. This is the criterion on which the costs of inactivity are based. In the tool it is assumed that any increase in moderate activity is associated with a proportional decrease in health costs. In other words, it assumes that if the Australian adults that are insufficiently active begin to increase their activity levels by 50% of the required amount, then a cost reduction of 50% would be experienced. Furthermore it is assumed that if the entire adult population became sufficiently active according to National Physical Activity (NPA) guidelines, then the costs of inactivity would be averted.

By knowing the adult population of Australia[7], the minimum recommended activity levels[8], and the estimated value of those activity levels being met by all adults[9], a value of $3.02 was then determined for each hour that an individual engages in moderate physical activity[10]. The calculation does not account for varying proportions of inactive people by specific region or state. The usefulness of the economic impact estimate is in its versatility in calculating the value of active lifestyles in urban settings, not in making specific economic assessments of specific neighbourhoods or demographics.

[5] Calculated as .457/.543=(1.5*10^9)/x and isolated for "x".

[6] No estimate was available for the health costs due to inactivity for Australians 17 and under, nor could any objective studies be found linking their activity levels and health to urban form; therefore, they are not accounted for in the calculation.

[7] Adult population of Australia roughly 15.4 million (ages 18 and over).

[8] Minimum moderate-intensity activity recommended of 30 minutes a day, five sessions a week.

[9] Lower health-related savings estimate of $3.82 billion and upper estimate of $4.5 billion.

[10] Value in an hour of moderate-intensity activity per person = Total national savings potential/adult population of Australia/ recommended hours of moderate-intensity activity a year.

2.7.4 Estimated activity increase in active travel neighbourhoods

Keeping in mind that the goal of the tool was to monetise the benefits of developing an area that is well suited for active travel, it was important to quantify the health benefits that cycling and walking could have as people convert to them and away from car dependency; however, studies that objectively measure physical activity with objectively measured urban form have focused on walking for active transport. Cycling-specific data correlated with objectively measured urban form could not be found. Information on walking, conversely, was more readily available and the assumption had to be made that the two would vary proportionally as functions of an area's suitability for active travel.

Research by Active Living Research in the US has shown that residents of more walkable areas spend about 30 minutes more per week (20% of the recommended amount) on walking trips than residents in sprawling areas (Active Living Research 2005). Another study conducted by findings from SMARTRAQ in the US found that 19% more people (the difference between 37% and 18%) are likely to meet or exceed the recommended minimum activity quota of 2.5 hours a week (or 130 hours a year) in highly walkable areas than people in the most sprawling neighbourhoods (Frank et al 2005). Total potential health-related savings were then calculated using the logic that if 19% more of the total resident population meets the NPA's minimum recommended level of moderate activity a week in active travel neighbourhoods, then a 19% discount could be expected in inactivity-related health costs.

The annual difference in cumulative time spent walking between active travel and sprawling neighbourhoods was then calculated as 19% x 1000 dwellings x 1.8325 adults per dwelling x 130 hours per year per person, resulting in 45,263 hours. Note that this figure is not an estimated difference in total hours of activity between sprawling and walkable neighbourhoods; instead it is simply an estimated difference in minimum activity level hours of walking between the two types of developments. It does not include hours that exceed the minimum recommended levels, nor does it include time spent on other recreational or non-discretional forms of activity.

As mentioned before, similar data for cycling was not available so it had to be calculated a little differently. Socialdata Australia provides some data on travel mode distributions among various Western Australian suburbs. A weighted average of bicycle trips as a proportion of walking trips was calculated and found to be roughly 21% (Socialdata 2008). Assuming that cycling levels remain proportionate to walking levels and that their average trip duration is approximately the same (Newman & Kenworthy 1999), the increase in annual hours of cycling for transport was worked out to be 9,505 hours (21% of 45,263). Since the NPA guidelines do not distinguish between types of physical activity and simply recommend 'moderate-intensity activity,' it was also assumed that walking and cycling for transportation share the same level of benefit.

2.7.5 An economic impact estimate of healthcare costs

The estimated savings benefit due to increased physical activity levels in an active travel neighbourhood was calculated for a development of 1,000 dwellings:

Walking at	45,263 hours x $3.02/hr = $136,694
Cycling at	9,505 hours x $3.02/hr = $28,706.
Total	$164,399 a year in savings

Thus, the savings in public health due to an active travel neighbourhood of 1,000 dwellings is estimated to be $164,400 a year or $164 a dwelling.

2.7.6 Discounting

When discounting recurring savings such as these there are a few timelines that we could consider, such as using the turnover period for a development, the average life expectancy for an Australian, the average life expectancy of a development, or we could even discount the annual savings as perpetuities if making the assumption that the property will remain zoned for residential use indefinitely. The decision was made to use 50 years, which is considered the minimum duration that a residential building would be erected for. It is assumed that after 50 years the decision of if and how to redevelop the piece of land will be made once again.

In addition to deciding on the number of years over which to discount the annual savings, a discount rate of 3% was chosen to reflect Australia's average annual rate of inflation (Reserve Bank of Australia 2007). A higher discount rate could have been used; however, since the figure represents a savings benefit and not an investment with associated risk, 3% was considered suitable. Conversely, a lower rate could have been used or future figures adjusted if other technological or medical considerations could be foreseen, but the calculation assumes treatments for the associated illnesses will remain constant.

The final calculation after the considerations previously mentioned estimates the present value of the economic health benefits of an active-travel development of 1000 dwellings at $4,229,950. This figure reflects the incremental economic health savings of developing 1000 residential dwellings if deciding to redevelop inner city-type areas as active-travel neighbourhoods as opposed to further expanding into greenfield areas. This is a small figure compared with the transport and infrastructure costs over 50 years and even compared with the social costs of greenhouse emissions. Thus our attention turns to health-related productivity.

3.0 The activity — employment productivity link

While there exists one body of research investigating the link between urban form and activity levels and health, there exists another body of research exploring the impacts of physical activity on workplace productivity. As of yet, we are unaware of any studies that directly tie urban form characteristics to employment productivity via the physical activity link, but it does not require a stretch of the imagination to see that such a link exists. This section pursues this avenue of thought, arguing that active-travel neighbourhoods are likely to have a workplace productivity benefit that is distinct from any potential healthcare savings. The estimated economic impact on productivity will assume the same scenario as with the activity-related healthcare costs: that the impact is from a development of 1,000 dwellings with average occupancies of 1.83 adults over the age of 17 per household.

3.1 Background

The majority of empirical studies relating exercise to workplace productivity have been focused on workplace fitness and wellness programs. The rationale behind their introduction by firms is that if their employees are healthier, this may result in fewer sick days being taken, better productivity on the job (presenteeism), and better employee relationships (Pronk et al. 2004). Furthermore, firms can expect greater savings in medical insurance expenditure if their employees are healthier (Proper et al. 2002) and can simultaneously improve their corporate image (Aldana & Pronk 2001). In the US, the percentage of worksites offering health programs and facilities increased from 22% to 42% between 1985 and 1992 (Wattles & Harris 2003). In Canada, this number grew from 44% in 1996 to 64% in 2004 (Chenoweth 2007).

The majority of empirical works examining the physical activity/employment productivity relationship do so by employing an experimental design that involves measuring the effects of worksite wellness programs on productivity-augmenting issues such as absenteeism, presenteeism, stress levels, job satisfaction, and job turnover. This is typically accomplished by conducting longitudinal studies involving intervention and control groups where pre- and post-intervention physiological and mental criteria can be assessed. For the purpose of these experiments, data on absenteeism can with relative ease be retrieved from human resources departments within firms; however, measuring employee job performance tends to be more challenging. Studies typically turn to the World Health Organization's Health and Work Performance

Questionnaire (HPQ), which is a self-report instrument designed to help organisations estimate the employee health costs associated with absenteeism, job performance and work-related accidents and injuries (Kessler, R et al.). Although self-report surveys such as this are prone to bias, the HPQ has been reconciled with employer archives on employee performance and good concordance has been found.

Given the difficulties of quantifying the effects of physical activity on employee performance, it is not surprising that these programs have not been economically justified in any rigorous or detailed manner. Some of these difficulties include (but are not limited to) clearly understanding who is being affected by the programs and the respective values of their time (i.e. a CEO missing a day of work is likely to have greater financial repercussions than a lower-level office clerk); being able to predict the participation levels of programs; knowing whether the employees participating are the ones standing to benefit the most; designing programs well enough to keep participant interest long enough for health benefits to be experienced; understanding how improved health may benefit employees who work in teams differently from those who typically work alone; and understanding how embodied knowledge in a position may affect the flow of output in an organisation. Despite these measurement challenges, larger employers have been opting to provide these facilities and programs for altruistic purposes, to demonstrate good corporate citizenship, and to improve employee well being rather than solely for pecuniary purposes (Shephard 1992).

3.2 The opportunity for planning

The choice for firms to invest in workplace fitness programs and facilities may be one justified (or partially justified) on imprecise calculations, but for many smaller firms the whole concept is simply financially unrealistic. In Australia, small businesses (defined as employing fewer than 20 workers) represented 97% of private firms and nearly 50% of employment in 2000–2001 (ABS 2001). In the US, 50% of companies with over 750 employees offer health programs, yet this number drops to 38% in those employing between 250 and 749 employees and drops a further 5% in those employing fewer than 49 (Chenoweth 2007). If similar health benefits can be accomplished through urban design efforts, ones particularly geared towards increasing incidental travel by way of walking or cycling, it is logical to see how planning policy can thus influence the economic productivity of cities (inclusive of companies of all sizes). In this way, health policy and community-scale urban planning may currently be missing opportunities to effectively improve health and productivity simultaneously and across multiple sectors (Yancey et al 2007). This is of striking importance because once communities are built, it becomes extremely difficult to reconfigure them and many opportunities may be lost.

Berger et al. (2001) do well to put things into perspective when they argue that a firm's investment in workers extends beyond that of wages to include other things that affect performance and tenure, such as health. Furthermore, they argue that health can be viewed as a commodity that gets "used up" and that employees manage their time so that it gets produced and consumed. In this sense, just as worksite wellness programs can be viewed as investments in 'replenishing' employee health, a company's choice of location can be viewed as a similar type of investment if it is a comparatively healthier environment and better supports employee wellbeing. In their study of the effects of an employee fitness program on reduced absenteeism, Lechner and Vries (1997) conclude that a significant decline in sick days only occurs for those who participate in the programs at least once per week, which is the threshold that they used to distinguish between "high" and "low" participation. While elevating and maintaining employee participation in these programs could prove to be a challenge, an urban form that allows or even gives employees an incentive to travel by physically active modes would conceivably be much more effective by getting larger volumes of people more active and sustaining this activity level over longer periods of time.

3.3 The empirical productivity evidence

Keeping with some of the assumptions made in the calculation of the activity-related health care benefits and sourcing some empirically estimated productivity benefits from existing studies, we can make some rough estimates of how active-travel has an economic impact on employee productivity through urban form. For the sake of simplicity, we will focus on benefits as they relate to absenteeism and on-the-job productivity — the two most common empirically studied effects. Mills et al. (2007) found that after a 12-month intervention-control study on a multinational corporation, the intervention group benefited from 4.3 fewer absentee days and an on-the-job productivity increase of 10.4%. Similarly, Lechner et al. (1997) in their longitudinal pre-test – post-test study found that fitness program participants experienced a decline of 4.8 sick days with a sample consisting of employees in the police force, the chemical industry, and in banking. Furthermore, Sheppard (1992) provides a critical analysis of worksite fitness programs and generates a table summarising the results from a number of reviewed studies. **Table 1** displays several results from the Sheppard (1992) study, with results selected on the basis that absenteeism is reported in 'days' gained or lost and on-the-job productivity is reported as a proportional increase.

Table 1: Influence of employee fitness programs on productivity and absenteeism

Author	Effect type	Results	Company/occupation
Cox et al. (1981)	*On-the-job productivity*	2.7% gain over controls	Canada Life
Feigin et al. (1960)	*On-the-job productivity*	4% increase	Electrical Assembly
Health and Welfare Canada (1976)	*On-the-job productivity*	4% gain	Office workers
Kmuzoz (1975)	*On-the-job productivity*	4-10% higher productivity	Worker-athletes
Pravosudov (1978)	*On-the-job productivity*	2–5% to 10–15% gain	Industrial work
Zoltik et al. (1990)	*On-the-job productivity*	5.6% gain	Pentagon
Bertera (1990)	*Absenteeism*	0.5 days/yr decrease	Blue-collar chemical
Blair et al. (1986)	*Absenteeism*	1.25 days/yr decrease	Dallas School Board
Bowne et al. (1984)	*Absenteeism*	0.8 days/yr decrease	Prudential Assurance
Garson (1977)	*Absenteeism*	2.1 days/yr decrease	Metropolitan Life
Mealey (1979)	*Absenteeism*	1.4 days/yr decrease	Police
Montgomery and Byrne (1988)	*Absenteeism*	0.6 days/yr decrease	
Pravosudov (1978)	*Absenteeism*	4 days/yr decrease	Russian industries
Terborg (1986)	*Absenteeism*	2.8 days/yr decrease	Batelle Memorial Institute
Zoltick et al. (1990)	*Absenteeism*	0.2 days/yr decrease	Pentagon

Source: Shephard (1992)

3.4 An economic impact estimate on productivity

For the sake of this calculation, we will assume a 6.2% on-the-job productivity increase and for absenteeism we will assume that more active employees benefit from 2.1 fewer sick days, both of which represent averages of the surveyed results. We adopt the neoclassical assumption that employees are remunerated according to the marginal contribution of labour and thus use the Australian weekly earnings average of $1165.40[11] as the baseline level of productivity. Furthermore, corresponding with the activity-related healthcare savings we will make the calculation for a 1000-dwelling development, each inhabited by an average of 1.83 adults of the age 18 or over.

Keeping with these assumptions and making the necessary calculations suggests that the productivity benefit due to a reduction in absenteeism would accrue to an annual average of $489.47 per person, with an additional $3,468.23 in benefit due to improved presenteeism. For an active-travel development of 1000 dwellings where 19% more of the population meets their minimum physical activity requirements, these values surge to $170,420 and $1,207,550 a year respectively, with a total annual health benefit of $1,377,970. After discounting over 50-year time spans at rates of 3%, these values translate to $4,384,900 and $31,070,000 for absenteeism and presenteeism respectively, totalling $35,454,900 for the productivity-related health benefits. These are substantial cost savings.

3.5 Other sources of productivity-related benefits

The possible benefits suggested here are not trivial and are now possible to be considered and weighted along with other quantifiable costs and benefits of planning. The productivity effect of active travel urban form was calculated here as a function of a reduction in workdays lost due to illness, stress, or waning workplace satisfaction (absenteeism) and of the increased ability for employees to focus on tasks and maintain focus for longer periods of time (presenteeism or on-the-job productivity).

The effects of increased daily activity levels on productivity are not limited to these two benefits, however, outside of them the empirical evidence and causal links are inadequately explored for their inclusion in a quantitative economic estimate. Some of these additional productivity benefits may include improved employee relationships, which may enhance cooperation within group settings, and lower employee turnover, which may reduce hiring and training costs.

4.0 Discussion of health and productivity benefits

The process used to economically assess the health impacts of urban form used a 'cost of illness approach', while the productivity benefits were estimated by a bottom-up accounting method. What the data collection endeavour for these calculations revealed was the scarcity of objective data available, more particularly with an Australian context, and the complexity of the subject matters at hand.

A noteworthy limitation of the analysis is the assumption that 19% more of the population meets their minimum daily activity requirements in walkable as opposed to non-walkable neighbourhoods. The findings reported in the study by Frank et al. (2005), from which this figure was sourced, are only reported in relation to the minimum activity threshold. Thus, we have no way to account for increased activity levels among 1) the inactive population that may experience increased activity levels but not enough to surpass the threshold and 2) the sufficiently active population that may become increasingly active in a highly walkable neighbourhood. Also, since marginal activity increases are likely to benefit the insufficiently active more than the already sufficiently active, we could expect potentially greater financial benefit from increased

[11] $1165.40 represents the average weekly earning for full-time, non-managerial employees in 2009. This wage corresponds to an average of 41.1 hours of work per week and an hourly wage of $28.35.

activity among sedentary individuals. This kind of disaggregation in the effect of walkable neighbourhoods on activity levels among residents would allow a more detailed estimate and likely increase the estimated value as well. Unfortunately, to our knowledge such empirical work has not been done to date.

Also worth addressing is that critics of the view that urban form affects health and levels of activity commonly refer to neighbourhood selection as the reason for measured differences among communities of varying walkability. They argue that those living in walkable areas choose to do so because they desire an environment conducive to active transport and would be active regardless of where they lived. A study by Frank et al. (2007) does support this to some extent, revealing that environment strongly influences active travel among individuals who prefer more walkable neighbourhoods, but those who prefer car-dependant environments are affected to a lesser extent. A quasi-longitudinal study by Handy et al. (2005), on the other hand, found that over the period of one year after a move to a more walkable neighbourhood, travel behaviours began to change more significantly in their sample population. This may suggest that as people have become accustomed to car-dependent lifestyles, they can also readjust over time and adopt healthier transportation habits. There are a number of other considerations that could also be taken into account which suggest that in some instances the economic estimates could be undervalued. Sections 4.1 and 4.2 will pursue some of these considerations further.

4.1 International comparison

The studies reporting empirically measured differences between activity levels were based on US cities and these findings may or may not hold in other countries such as Australia for cultural and/or other reasons. If studies of similar design had been conducted in Australia, that information could have been used instead; however, the differences between the two countries, given the subject matter, are not vast (Newman and Kenworthy 1999). For Seattle and Perth, for instance, it is documented that cycling comprises 2.3% (Nelson and Scholar ca.2006) and 2.4% (Socialdata 2008) of total trips, respectively. The share of total trips for walking is roughly 7% in Seattle (Nelson and Scholar ca.2006) and 11% in Perth (Socialdata 2008). In this sense, using empirical evidence that has originated in the US is suitable for an economic assessment in an Australian setting, but this does not mean that there is no room for improvement from here.

Copenhagen[12], Denmark is one of many European cities that has shown that it is possible to design a city with active modes of travel in mind with a modal split of 27% cent for driving, 33% for transit, 36% for bicycling, and 5% for walking (Nelson and Scholar ca.2006). This is a reflection of policy intervention, cultural characteristics, urban design and urban planning. No estimates of the health benefits of this substantial increase in active travel have been found but if done would enable a perspective on the upper boundaries of this approach.

4.2 Other physical determinants of urban form

The objective measurements of urban form include factors such as density, land-use mix, connectivity, and proximity, but they do not measure some very important travel-related design considerations. Often the determinants of a neighbourhood's walkability, or orientation to supporting active modes of travel, extend beyond these measures to include a number of other factors. Some of these may include the quality or amount of sidewalk space and bicycle paths, the level of public transport service and the regulations around carrying bicycles on-board, the lighting and sense of safety for travel during darker or less populated hours, and the inclusion of natural landscaping and aesthetic value. The potential for even higher levels of active travel than was considered for this calculation is evident in many European examples, as outlined with Copenhagen.

[12] Albeit high in active transport modal share, Copenhagen has never experienced the type of modal splits characterised by sprawled, auto-dependent cities. It is not an example of a city that successfully underwent a dramatic transformation, but an example of one that has avoided the plight of car-dependent cities by careful planning and selection of transport infrastructure (see Newman and Kenworthy 1999).

Gehl Architects in Copenhagen have made some significant alterations and enhancements to cities around the world, such as Sydney, Melbourne, Cape Town, London and Zurich and in their own city of Copenhagen, enhancing their urban realms for active transportation[13]. If more studies take place in different cities and eventually include factors such as kilometres of bike lanes and widths of pedestrian paths in their measures of walkability, then maybe a larger health cost reduction percentage than the one used in the calculation could be substituted. These calculations only drew on the results of studies that suggest there is a relationship between urban sprawl and physical activity levels. Other factors in the mindset of urban planning, such as safety policy, bicycle schemes, education, economic incentives, traffic mitigation and infrastructure, could all have a huge impact on active travel levels and potentially, health and productivity; however, the role of urban form remains a fundamental factor.

The objective measurement of 19% more individuals meeting the recommended quota of moderate-intensity physical activity translates into considerable economic health-related and productivity-related savings from a conservative standpoint. In the future, other figures representing expected changes in active travel could be substituted into the calculation as more objective studies linking urban form to activity levels report new findings. The way forward would be to continue to measure urban form's impacts on physical activity levels with factors that make active transport safer, more efficient, and more enjoyable while making driving a more unattractive option. Similarly, it would be interesting to investigate how these factors specifically influence travel behaviour among individuals who are predisposed to auto-dependent lifestyles, for therein potentially lies the greatest benefit in activity increase.

5.0 A cumulative economic impact statement of alternate development patterns

This paper examined the economic impacts that urban form can have on healthcare costs and workplace productivity via its ability to facilitate active modes of travel. These calculations reveal that in addition to the traditional development costs associated with infrastructure and transportation (and more recently greenhouse gases), there is strong financial justification for including health and productivity costs in appraisal. As such, the case has strengthened for inner city and greyfield redevelopment in an urban form that favours active modes of transport.

The cumulative economic benefits of redeveloping inner-city type areas, merging the findings of this paper with its companion papers, are summarised below in **Table 2.**

[13] Gehl Architects has conducted Public Life studies in Melbourne (1993, 2004), Perth (1994), Copenhagen (numerous studies including 1996), London (2004), Zurich (2004), Stockholm (1990, 2005) and are carrying out studies in Sydney and New York (2007). This is not a comprehensive list. Further information can be found on their website: www.gehlarchitects.dk

Table 2: Estimated development costs for an urban redevelopment compared to a fringe development of 1,000 dwellings

	Inner	Outer	Difference
Infrastructure costs			
Roads	$5,086,560	$30,378,880	$25,292,320
Water and sewerage	$14,747,620	$22,377,460	$7,629,840
Telecommunications	$2,576,110	$3,711,850	$1,135,740
Electricity	$4,082,120	$9,696,510	$5,614,390
Gas		$3,690,840	$3,690,840
Fire and ambulance		$302,510	$302,510
Police		$388,420	$388,420
Education	$3,895,460	$33,147,270	$29,251,810
Health (hospitals etc.)	$20,114,870	$32,347,330	$12,232,460
Total infrastructure	**$50,502,740**	**$136,041,070**	**$85,538,330**
Transport costs			
Transport and travel time	$206,542,060	$342,598,100	$136,056,040
Roads and parking	$46,937,540	$154,826,100	$107,888,560
Externalities	$2,219,880	$9,705,380	$7,485,500
Total transport	**$255,699,480**	**$507,129,580**	**$251,430,100**
Greenhouse gas cost			
Offset cost ($25/t)	$2,500,000	$5,400,000	$2,900,000
Social cost ($215/t) — NOT INCLUDED IN TOTAL	$21,500,000	$46,440,000	$24,940,000
Total greenhouse	**$2,500,000**	**$5,400,000**	**$2,900,000**
Physical activity costs			
Healthcare cost		$4,229,950	$4,229,950
Productivity loss		$34,454,900	$34,454,900
Total activity costs		**$38,684,850**	**$38,684,850**
Total	**$308,702,220**	**$687,255,500**	**$378,553,280**

14 Transport costs are calculated as operating costs and thus are a function of vehicle kilometres traveled (VKT) and patronage

15 The cumulative economic impact statement does not include the social cost of carbon (SCC) from transport greenhouse gas emissions. It is assumed that abatement, being the more cost-effective option, would be preferred over sustaining the harm or global warming; however, without proper policy measures put in place the social costs will continue to surmount..

The costs figures in Table 2 display the differential cost streams associated with the two alternative development paths of inner city versus greenfield. Many of the estimates were made conservatively: infrastructure costs may vary depending on excess capacity levels and area-specific requirements; greenhouse gas costs depend on a price for carbon and will go up or down accordingly; and health savings depend on the types of mode-specific infrastructure put into place combined with incentive schemes and public education. It should also be noted that the infrastructure costs are up-front costs that require payment upon initial development. The transport, greenhouse gas, health costs and productivity losses are present values calculated over 50 years and could be considered as operating costs of the respective types of development. Also, as the health and productivity costs were calculated as foregone benefits by choosing to develop on the fringe, they appear as a cost in the 'Outer' column.

Nevertheless, the size of the different costs does provide some perspective on the overall costs of sprawl. Other studies (e.g. Costs of Sprawl — 2000) have indicated the substantial costs of infrastructure and transport, which we too have demonstrated and which should be of considerable concern to cities contemplating a future with plans to continue major greenfields development. However, we can now see that two of the newer parameters relating to urban development — the social costs of greenhouse and health-related productivity — add another substantial cost rationale to the value of redevelopment with more active transport modes built-in.

6.0 Conclusion

Despite the variations that can be expected from these types of calculations, the data provided indicate very substantial differences in costs between fringe and inner-city types of development. The dominant factors are infrastructure and transportation costs. The newer topics of greenhouse and health impacts are a little less yet are still substantial. They are important in policy decisions as they are part of a global and local governance system that will only want to see them reduced, not increased. The newly added productivity benefits are especially significant and comparable in scale to the infrastructure and transport savings.

The benefits to inner-city redevelopment compared to greenfields can be seen across a range of portfolios that go beyond the usual urban planning focus. Town planning decisions will influence Australia's ability to meet its greenhouse gas targets and the National Physical Activity Taskforce meeting its target of a 5% increase in efficiently active Australians. The synergies are numerous and as shown can be economically quantified.

The values of infrastructure, transportation, greenhouse gas, activity-related health and activity-related productivity for the different development types have been calculated from models and methods that can be used to predict values for any other planned development. This type of approach should become standard practice in evaluating different development projects.

References

Active Living Research 2005, *Designing for Active Transportation*, viewed 12 April 2008, www.activelivingresearch.org/alr/files/transportationrevised021105.pdf

Aldana, S & Pronk, N 2001, 'Health Promotion Programs, Modifiable Health Risks, and Employee Absenteeism', *Journal of Occupational and Environmental Health*, 43, pp. 36–46.

Australian and New Zealand Obesity Society (ANZOS) n.d., *Obesity in Australian Adults: Prevalence Data*, viewed April 24 2008, www.asso.org.au/freestyler/gui/files//factsheet_adults_prevalence.pdf

Australian Bureau of Statistics (ABS) 2009, 6302.0 *Average Weekly Earnings, Australia, Aug. 2009*, viewed 26 November 2009.

Australian Bureau of Statistics (ABS) 2007, 3101.0 *Australian Demographic Statistics*, Sept. 2007, viewed 11 April 2008.

Australian Bureau of Statistics (ABS) 2006, 2068.0 – 2006 *Census Tables of Population and Housing*, viewed 17 April 2008.

Australian Bureau of Statistics (ABS) 2004, 3236.0 *Household and Family Projections*, Australia, 2001 to 2026, viewed 14 April 2008

Australian Bureau of Statistics (ABS) 2001, 1321.0 *Small businesses in Australia*, 2001, viewed 26 November 2009.

Bauman, A, et al. 2002, *Getting Australia active: towards better practice for the promotion of physical activity*, National Public Health Partnership, Melbourne, Australia, March, 2002.

Berger, M et al. 2001, 'Alternative Valuations of Work Loss and Productivity'. *Journal of Occupational and Environmental Medicine*, 43, pp. 18–24.

Bray, R et al. 2005, *Report on Public Health and Urban Sprawl in Ontario: A review of the pertinent literature*, (unpublished), Environmental Health Committee, Ontario College of Family Physicians, viewed April 11 2008, www.cfpc.ca/local/files/Communications/Current%20Issues/Urban%20Sprawl-Jan-05.pdf

Cao, X et al. 2007, 'Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations model approach', *Transportation*, 34, pp. 535–556.

Cao, X et al. 2006, 'The influences of the built environment and residential self-selection on pedestrian behavior: evidence from Austin, TX', *Transportation*, 33:1, pp. 1–20.

Chenoweth, D 2007, *Worksite Health Promotion 2nd E*, Human Kinetics, Champaign.

Department of Planning and Infrastructure (DPI) (2008). *Travelsmart in Perth 2000-2005: Increasing Physical Activity*, (unpublished), Perth.

Econtech 2007, *The Cost of Physical Inactivity: What is the lack of participation in physical activity costing Australia?* Prepared by Econtech Pty. Ltd. For Medibank Private.

Ewing, R 2005, 'Building Environment to Promote Health', *Journal of Epidemiology and Community Health*, 000, pp. 1–2.

Ewing, R 2005, 'Can the Physical Environment Determine Physical Activity Levels?', *Exercise and Sports Sciences Reviews*, 33:2, pp. 69–75.

Ewing, R et al. 2003, 'Relationship between Urban Sprawl and Physical Activity, Obesity, and Morbidity', *American Journal of Health Promotion*, 18:1, p.47–57.

Frank, L et al. 2007a, 'Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and Obesity', *Social Science & Medicine*, 65, pp. 1898–1914.

Frank, L et al. 2007b, 'Urban form, travel time, and cost relationships with tour complexity and mode choice', *Transportation*, 35, pp. 37–54.

Frank, L et al. 2006, 'Many Pathways from Land Use to Health', *Journal of the American Planning Association*, 72:1, pp.75–87.

Frank, L et al. 2005, 'Linking Objectively Measured Physical Activity with Objectively Measured Urban Form', *American Journal of Preventative Medicine*. 28:2, pp. 117–125.

Frank, L. et al (2004). Obesity Relationships with Community Design, Physical Activity, and Time Spent in Cars. *American Journal of Preventative Medicine*, 27:2, pp. 87–96.

Frank, L & Engelke, P 2005, 'Multiple Impacts of the Build Environment on Public Health: Walkable places and the exposure to air pollution', *International Regional Science Review*, 28:2, pp. 193–216.

Giles-Corti, B 2006, *'The impact of urban form on public health'*, paper prepared for the 2006 Australian State of the Environment Committee, Department of the Environment and Heritage, Canberra, www.deh.gov.au/soe/2006/emerging/public-health/index.html

Handy, S et al 2005, 'Correlation of Causality between the built environment and travel behavior? Evidence from North California', *Transportation Research Part D*, 10, p. 427–444.

James, G et al. 2001, *Potential for Increased Walking Trips:* Proceedings of the 2001 Walking the 21st Century conference held in Perth, WA, viewed 27 April 2008, www.dpi.wa.gov.au/mediaFiles/walking_21centconf01bpaper_james.pdf

Kelly-Schwartz, AC et al. 2004, 'Is Sprawl Unhealthy? A Multilevel Analysis of the Relationship of Metropolitan Sprawl to the Health of Individuals', *Journal of Planning Education and Research*, 24, pp.184–196.

Lechner, L & de Vries, H 1997, 'Effects of an Employee Fitness Program on Reduced Absenteeism', *Journal of Occupational and Environmental Medicine*, 39:9, pp. 827–831.

Mills, P et al. 2007, 'Impact of a Health Promotion Program on Employee Health Risks and Work Productivity', *American Journal of Health Promotion*, 22:1, pp. 45–53.

Nelson, A, Scholar, V ca. 2006, *'Livable Copenhagen: The Design of a Bicycle City*, (unpublished), Center for Public Space Research, Copenhagen and the University of Washington, Seattle, viewed April 23 2008, www.sightline.org/research/sprawl/res_pubs/Livable_Copenhagen_reduced.pdf

National Research Council 2002, *'Costs of Sprawl – 2000*, National Academy Press, Washington.

Newman, P & Kenworthy, J 1999, *Sustainability and Cities: Overcoming Automobile Dependence*, Island Press, Washington.

Oldridge, N 2008, 'Economic Burden of Physical Inactivity: Healthcare costs associated with cardiovascular disease', *European Journal of Cardiovascular Prevention and Rehabilitation*, 15:2, pp.130–139.

Pronk, N et al. 2004, 'The Association Between Work Performance and Physical Activity, Cardiorespiratory Fitness, and Obesity', *Journal of Occupational and Environmental Health*, 46, 19–25.

Proper, K et al. 2002, 'Effectiveness of Physical Activity Programs at Worksites with Respect to Work-Related Outcomes', *Scandinavian Journal of Work, Environment and Health*, 28:2, pp. 75–84.

Public Health Agency of Canada (1993). *Economic Burden of Illness in Canada, 1993*, viewed 11 April 2008, www.phac-aspc.gc.ca/publicat/ebic-femc93/burd4_e.html

Reserve Bank of Australia 2007, *Statement on Monetary Policy: November 2007*, viewed 23 April 2008, www.rba.gov.au/PublicationsAndResearch/StatementsOnMonetaryPolicy/index.html

Saelens, B et al. 2003, 'Environmental Correlates of Walking and Cycling: Findings From the Transportation, Urban Design, and Planning Literatures', *Annals of Behavioural Medicine*, 25:2, pp. 80–91.

Shephard, R 1992, 'A Critical Analysis of Work-Site Fitness Programs and Their Postulated Economic Benefits', *Medicine and Science in Sports and Exercise,* 24:3, pp. 354–370.

Socialdata Australia 2008, *Mode Choice,* viewed 22 April 2008, www.socialdata.de/daten/vm_e.php

Stokols, D et al. 2003, 'Increasing the Health Promotive Capacity of Human Environments' *American Journal of Health Promotion,* 18, pp. 4–13.

Streib, L 8 February 2007, *World's Fattest Countries,* Forbes, accessible at www.forbes.com/2007/02/07/worlds-fattest-countries-forbeslife-cx_ls_0208worldfat_2.html

Sturm, R 2004, 'The Economics of Physical Activity: Societal trends and rationales for interventions', *American Journal of Preventative Medicine,* 27:3S, pp. 125–135.

Sturm, R, Cohen, DA 2004, 'Suburban Sprawl and Physical and Mental Health', *Journal of the Royal Institute of Public Health,* 118, p.448–496.

Trubka, R et al. 2008, *Assessing the Costs of Alternative Development Paths in Australian Cities.* A discussion paper prepared by PB-CUSP.

Wattles, M & Harris, C 2003, 'The Relationship Between Fitness Levels and Employee's Perceived Productivity, Job Satisfaction, and Absenteeism', *Journal of Exercise Physiology,* 6:1, pp. 24–32.

World Health Organization 2007, *Economic Assessment of Transport Infrastructure and Policies: Methodological guidance on the economic appraisal of health effects related to walking and cycling,* viewed 23 April 2008, www.euro.who.int/Document/E90944.pdf

Yancey, A et al. 2007, 'Creating a Robust Public Health Infrastructure for Physical Activity Promotion', *American Journal of Preventative Medicine,* 32:1, pp. 68–78.

Biography

Roman Trubka is a PhD candidate at the Curtin University Sustainability Policy (CUSP) Institute in Western Australia.

Peter Newman is a Professor of Sustainability and Director at the Curtin University Sustainability Policy (CUSP) Institute in Western Australia.

Darren Bilsborough is the Director of Sustainability at Parsons Brinckerhoff (PB) and an adjunct professor in sustainability at CUSP.