**School of Earth and Planetary Sciences**

**Department of Spatial Sciences**

# Deregulation and Competition: Comparison of Regional Aviation Markets in Western Australia

**Heng Zhou**

**0000-0002-8586-3053**

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

**March 2020**

## 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.

**Human Ethics** (For projects involving human participants/tissue, etc) The research presented and reported in this thesis was conducted in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2007). The proposed research study received human research ethics approval from the Curtin University Human Research Ethics Committee (EC00262). Approval Number is **HRE2017-0815**.

Date: 06/Mar/2020

**ABSTRACT**

In Australia, airlines are a key enabling industry, with a central role in the broader economy. In both leisure and work travel, the public rely on air transport for both domestic and international transit. Domestically, regional aviation services in Australia perform a crucial role in linking hub cities to remote, rural and regional areas. This is particularly true for Western Australia – the largest state with extremely large isolated areas that are difficult to reach by land. Regional aviation in Western Australia satisfies the air service demands of communities, mining, tourism, government access, and other business activities, which collectively provide significant contributions to the social and economic development of the state. However, regional air passenger movement in Western Australia has experienced a considerable decrease in recent years. Several other important aviation issues have been recognised by the government transport agencies. High regional airfares compared to the rest of Australia are a key problem that seriously hampers regional public movements and hinders economic growth. Some public transport air routes are experiencing the issue of unbalanced passenger demand on round trips. Another problem is that some regional airports are incapable of upgrading their infrastructure, and hence struggle to introduce larger aircraft. Although the State transport department has set up projects to invest in these regional airports, they require insight concerning issues such as reliable demand forecasts to guide the plan implementation to overcome impediments such as funding and commercial limitations.

While considerable research has been conducted into international and domestic air travel demand and aviation markets, limited attention has been given to the regional aviation market in Western Australia. In particular, not enough attention has been given to understand how air travel competes with other travel modes in the Western Australia context. Therefore, to address these gaps and assist policy-makers tackling these aviation issues, this study develops a systematic and robust methodology for forecasting regional air travel demand, exploring the regional aviation market and estimating travel mode and airline choice that could yield a more comprehensive understanding of this regional aviation market and competition.

Firstly, modified gravity models with Poisson pseudo-maximum likelihood estimators are developed to forecast bilateral air travel demand of regional airport-pairs, and to investigate the impact of different airport catchment area definitions on influencing demand estimation. The modelling results not only find that airfare, distance, population, tourism and mining sectors

can significantly affect air travel demand, but also indicate that the size/boundary defined for the catchment area of the airports can impact on the magnitude of factors and therefore affect the modelling results.

Then, the thesis applies exploratory data analysis to investigate the regional aviation market based on collected survey data about passengers and their demand preferences. This involves summarising and visualising the air travel survey data that provides a preliminary understanding of regional air passengers' characteristics. Then a mixture model-based market segmentation approach is used to identify and investigate existing and potential aviation markets for an in-depth insight into the regional aviation market.

This study integrates a semi-automatic method for generating a realistic and statistically efficient stated preference (SP) design by extending the widely-used Modified Federov Algorithm. Data for this SP survey were collected in regional Western Australia and analysed in a variety of ways. Therefore, in the next stage of this study, multinomial and nested logit models are developed to estimate travel mode and airline choice using the SP data. The results show that travel cost, journey time, service frequency and seat comfort are statistically significantly in representing travel mode and airline choice. Airport and non-airport respondents also have a different magnitude of sensitivity to these factors, business travellers are more time sensitive and less price sensitive compared to non-business travellers. However, preference heterogeneity may exist among the individuals in each of the specified groups. Therefore, to give a better characterisation of preferences across the community, a market segmentation approach based on latent class modelling is used to accommodate the preference heterogeneity across the respondents for further investigating travel mode and airline choice. This identifies two distinct market segments of respondents with different demographics, economics and trip characteristics and quantifies a different sensitivity magnitude to these factors.

This thesis develops a holistic methodology for investigating the regional aviation market and modelling travel mode and airline choice. The results, with intuitive interpretations, can help to guide the development of policy to tackle future aviation issues. Further research is recommended to identify other potential parameters or interactions that are uncertain but may also influence the aggregated travel demand and/or airline choices.

**ACKNOWLEDGEMENTS**

This PhD study has been a truly life-changing experience for me and I am grateful for the encouragement, support and guidance from many people, which made the completion of this thesis possible.

I would like to first express my sincere gratitude and thanks to my supervisor Associate Prof. Jianhong (Cecilia) Xia, my co-supervisor Associate Prof. Richard Norman and my associate supervisors Dr. Brett Hughes and Prof. Torbjorn Falkmer. Thank you to Associate Prof. Jianhong (Cecilia) Xia, for all the constant support of my PhD study and related research you have given me. Your guidance, patience and immense knowledge has helped me through all the time of the research and writing this thesis and related papers. Your meticulous guidance and encouragement has helped me overcome the difficulties during the research. I cannot image how my PhD study would have been without your great supervision and the regular meetings we had. My appreciation also goes to my co-supervisor Associate Prof. Richard Norman, for his time, guidance, expertise and contribution in strengthening the methodology development in this research and in improving this thesis and related publications. Great thanks also go to my associate supervisors Dr. Brett Hughes and Prof. Torbjorn Falkmer, for their professional guidance and the continuous support in this research. Thank you, Dr. Brett Hughes, as an associate supervisor and the industry partner (Department of Transport), you have not only provided great supervision and efforts to help me better completing this aviation research, but also provided a lot of important and useful information and advice, which significantly improved the quality of this thesis. Many thanks also to my associate supervisor Prof. Torbjorn Falkmer, for his professional guidance and the insightful comments to the research and this thesis.

A very special thank you to Mr. Gary McCarney, for his editing of my thesis, as well as his clear, rational and intuitive comments that have contributed to the explanation of the results.

I gratefully acknowledge the financial support received towards my PhD study from Curtin International Postgraduate Research Scholarships (CIPRS) and the Department of Transport Western Australia grant [grant number CTR-8890]. Thanks to the project industry partners Dr. Brett Hughes and Miss Gabi Nikolova from the Department of Transport for their continuative support and assistance to this project.

I am grateful to Andrew Sharpe, Stuart Jamieson, Bob Davis, Han Jie Davis, Bob Urquhart, Paul McSweeney, Ben Bexley, Chris Adams, Amol Virkar, Judy Andrews, Tim Woolerson, Sarah Sleight, David Cohen, Craig Walker, Bill Swetman, Paul Barnett, Andrew Orford, Lee-Anne Springer, Melanie Ansell, Brooke Jenkinson, Mel Virgo, Vanessa Subramoney and Joyce Williams for their cooperation and authorisation to conduct the surveys in Albany, Geraldton, Karratha and Broome airports and libraries in Albany, Geraldton and Karratha, visitor centres in Albany and Broome, South Regional TAFE in Albany, Central Regional TAFE in Geraldton, Central North regional TAFE in Karratha and Broome and TRANSWA in Albany and Geraldton. Additionally, I would like to extend my great thanks to my colleagues Keone Kelobonye, Yunchen Liu, and Bin Jia who helped me to distribute and collect survey questionnaires at the regional airports.

My thanks also go to all the staff and postgraduate students in the School of Earth Sciences, Curtin University for providing help and encouragement in times of need. Thank you to Professor Bert Veenendaal, Associate Professor Jon Kirby, Dr. Ashraf Dewan, Lori Patterson and Caroline Rockliff, for their assistance during the life of my study. Also, I thank my friends Prof. Jie Sun, Dr. Honglei Xu, Keone Kelobonye, Dr. Qian Sun, Dr. Chunmei Chen, Dr. Ting Lin, Yu Zhou, Arfanara Nanjnin, and all fellow lab-mates of geo-spatial research lab 310A, for their friendship and providing a tolerant, friendly and very collaborative environment.

Lastly, but not the least, I would like to say a heartfelt thank you to my family, especially my parents. It is because of you that I have been able to reach my goal. Thanks for always being there.

## RELATED PUBLICATIONS

*Project report:*

Xia, J., **Zhou, H**., Hughes, B., Norman, R., Nikolova, G., Falkmer, T. (2019). *Review of Regulated Regular Public Transport Air Routes in Western Australia Regional Airports Intercept Survey*.

*Journal papers included in this thesis:*

**Zhou, H**., Xia, J., Luo, Q., Nikolova, G., Sun, J., Hughes, B., . . . Falkmer, T. (2018). Investigating the impact of catchment areas of airports on estimating air travel demand: A case study of regional Western Australia. Journal of Air Transport Management, 70, 91-103. doi:https://doi.org/10.1016/j.jairtraman.2018.05.001

**Zhou, H**., Norman, R., Kelobonye, K., Xia, J., Hughes, B., Nikolova, G., . . . Falkmer, T. (2019). Market Segmentation Approach to Investigate Existing and Potential Aviation Markets. Manuscript, [Revised version resubmitted to Journal of Transport Policy]

**Zhou, H**., Xia, J., Norman, R., Hughes, B., Nikolova, G., Kelobonye, K., . . . Falkmer, T. (2019). Do air passengers behave differently to other regional travellers?: A travel mode choice model investigation. Journal of Air Transport Management, 79, 101682. doi:https://doi.org/10.1016/j.jairtraman.2019.101682

**Zhou, H**., Norman, R., Xia, J., Hughes, B., Kelobonye, K., Nikolova, G., & Falkmer, T. (2020). Analysing travel mode and airline choice using latent class modelling: A case study in Western Australia. Transportation Research Part A: Policy and Practice, 137, 187-205. doi:https://doi.org/10.1016/j.tra.2020.04.020

*Journal papers not included in this thesis:*

**Zhou, H**., Sun, Q., Blane, A., Hughes, B., Falkmer, T., & Xia, J. C. (2019). Investigating on-road lane maintenance and speed regulation in post-stroke driving: a pilot case-control study. Manuscript, [Submitted to Journal of Transportation Safety & Security, revision requested]

Kelobonye, K., **Zhou, H**., McCarney, G., Xia, J., & Swapan, M. S. H. (2019). Measuring job, education and shopping accessibility with competition using the cumulative opportunities measure. Manuscript, [Revised version resubmitted to Journal of Transport Geography]

Xia, J., Cui, H., **Zhou, H**., Zougheibe, R., O'Connor, K., Chen, F., Hughes, B., & Yang, J. (2019). Competition or regulation: regional airfares in Western Australia. Manuscript, [Submitted to Journal of Transport Geography, revision requested]

Kelobonye, K., McCarney, G., Xia, J. C., Swapan, M. S. H., Mao, F., & **Zhou, H.** (2019). Relative accessibility analysis for key land uses: A spatial equity perspective. Journal of Transport Geography, 75, 82-93. doi:https://doi.org/10.1016/j.jtrangeo.2019.01.015

Kelobonye, K., Xia, J. C., Swapan, M. S. H., McCarney, G., & **Zhou, H.** (2019). Drivers of Change in Urban Growth Patterns: A Transport Perspective from Perth, Western Australia. Urban Science, 3(2), 40. doi:10.3390/urbansci3020040

# TABLE OF CONTENTS


**DECLARATION** .......... **i**
**ABSTRACT** .......... **ii**
**ACKNOWLEDGEMENTS** .......... **iv**
**RELATED PUBLICATIONS** .......... **vi**
**TABLE OF CONTENTS** .......... **viii**
**LIST OF FIGURES** .......... **xiv**
**LIST OF TABLES** .......... **xvi**
**LIST OF APPENDICES** .......... **xviii**
**LIST OF ACRONYMS** .......... **xix**
**CHAPTER 1 INTRODUCTION** .......... **21**
1. 1 Introduction .......... 21
1. 2 Research Objectives .......... 23
1. 3 Research Significance and Contribution .......... 24
1.3.1 Significance .......... 24
1.3.2 Contributions .......... 25
1. 4 Research Methodology .......... 26
1. 5 Thesis outline .......... 26
1. 6 Summary .......... 29
**CHAPTER 2 LITERATURE REVIEW** .......... **30**
2. 1 Aviation in Western Australia .......... 30
2.1.1 Aviation market .......... 30
2.1.2 Airline and travel mode competition .......... 32
2.1.3 Airline regulation and deregulation .......... 34
2. 2 Air travel demand forecasting .......... 37
2.2.1 What is air travel demand? .......... 37
2.2.2 Gravity model .......... 38
2.2.3 Gravity modelling studies and limitations .......... 39
2. 3 Air passenger market segmentation .......... 43
2.3.1 What is market segmentation and why is it important? .......... 43
2.3.2 Cluster analysis .......... 45
2.3.3 Aviation market segmentation studies and limitations .......... 46
2. 4 Travel mode choice .......... 48

2.4.1 What is travel mode choice? ..... 48

2.4.2 Travel mode choice models ..... 50

2.4.3 Travel mode choice studies and limitations ..... 56

2. 5 Summary ..... 57

**CHAPTER 3 RESEARCH FRAMEWORK ..... 58**

3. 1 Introduction ..... 58

3. 2 Study area ..... 58

3. 3 Research workflow ..... 61

3. 4 Research data collection ..... 62

3.4.1 Geo-spatial data collection ..... 62

3.4.2 Air travel and stated preference survey data collection ..... 63

3. 5 Air travel demand forecasting ..... 65

3. 6 Exploratory data analysis for exploring aviation market ..... 67

3.6.1 Visualisation and exploration of air passenger data ..... 68

3.6.2 Aviation market segmentation analysis ..... 69

3. 7 Travel mode and airline discrete choice modelling analysis ..... 71

3.7.1 Estimating travel mode and airline choice using Multinomial and Nested Logit modelling ..... 73

3.7.2 Analysing travel mode choice behaviour using latent class modelling ..... 75

3. 8 Major Software ..... 77

3.8.1 GIS software ..... 77

3.8.2 Programming languages and SPSS software ..... 78

3.8.3 Ngene and NLOGIT software ..... 78

**CHAPTER 4 FORECASTING THE AIR PASSENGER DEMANDS BETWEEN RPT AIRPORTS USING MODIFIED GRAVITY MODELS ..... 80**

4. 1 Introduction ..... 80

4. 2 Research Context ..... 80

4. 3 Methodology ..... 82

4.3.1 Gravity model and Poisson Pseudo-Maximum Likelihood Estimator ..... 83

4.3.2 Catchment area definition ..... 84

4.3.3 Data used in this study ..... 85

4. 4 Results ..... 85

4.4.1 Visualisation of total available seats in regional Western Australia ..... 85

4.4.2 Gravity models based on the Thiessen polygon catchment area ..... 87

4.4.3 Gravity models based on the 2.5 hour driving catchment areas ........................... 89
4. 5 Discussion ........................................................................................ 91
4.5.1 Key determinants ........................................................................... 92
4.5.2 Impact of size of catchment area on modelling ................................... 92
4.5.3 Load factors in Western Australia ..................................................... 95
4.5.4 Accuracy of online flight information ................................................ 96
4. 6 Summary .......................................................................................... 97
**CHAPTER 5 DISCRETE CHOICE EXPERIMENTAL DESIGN ................................ 98**
5. 1 Introduction ...................................................................................... 98
5. 2 Research Context ................................................................................ 98
5.2.1 Stated and revealed preference survey ............................................... 99
5.2.2 Previous work on experimental design .............................................. 100
5.2.3 Implausible and dominant alternatives and unrealistic choice tasks ......... 101
5.2.4 Gaps and aims ............................................................................... 102
5.2.5 How to form the experimental design ................................................ 103
5. 3 Methodology of Experimental Design ................................................... 104
5.3.1 Experimental design constructing processes ....................................... 104
5.3.2 Stage 1: Research problem and choice criteria identification ................. 104
5.3.3 Stage 2: Model specification and design properties .............................. 106
5.3.4 Stage 3: Generate experimental design .............................................. 115
5. 4 Implementing D-efficient experimental design for this research ................ 121
5.4.1 Choice criteria identification ............................................................ 121
5.4.2 Model specification and design consideration ..................................... 123
5.4.3 Generate experimental design using EMFA method ............................. 125
5.4.4 Questionnaire construction ............................................................... 134
5. 5 Discussion ........................................................................................ 135
5. 6 Summary .......................................................................................... 137
**CHAPTER 6 AIR PASSENGER SURVEY DATA VISUALISATION ....................... 138**
6. 1 Introduction ...................................................................................... 138
6. 2 Research Context ................................................................................ 138
6. 3 Data used in this chapter .................................................................... 139
6. 4 Characteristics of Western Australia regional air passengers ..................... 139
6.4.1 Regional Western Australian air passenger profiles .............................. 139
6.4.2 Trip origin and destination ............................................................... 141

6.4.3 Access mode, reasons to choose the air travel mode and trip purpose .............. 142
6.4.4 Travel cost and frequency .................................................................................. 146
6.4.5 Travel cost of one-way flight ............................................................................. 146
6.4.6 Reasons to choose flight, trip purpose and booking time in advance of departure ........................................................................................................................ 148
6.4.7 Comparison of air passenger characteristics across the four regional airports .. 150
6. 5 Key findings ............................................................................................................ 157
6. 6 Summary ................................................................................................................. 158
**CHAPTER 7 AVIATION MARKET SEGMENTATION ANALYSIS ........................ 159**
7. 1 Introduction ............................................................................................................ 159
7. 2 Research Context..................................................................................................... 159
7. 3 Methodology ........................................................................................................... 160
7.3.2 Mode-based clustering with EM estimator ....................................................... 161
7.3.3 Number of clusters and clustering software ...................................................... 165
7. 4 Results ..................................................................................................................... 166
7.4.1 Air passenger market segmentation ................................................................... 166
7.4.2 Non-air passenger market segmentation ........................................................... 168
7. 5 Discussion ............................................................................................................... 170
7.5.1 General characteristics of air passenger market segments ................................ 170
7.5.2 General characteristics of non-air passenger market segments.......................... 170
7.5.3 Existing and potential aviation markets ............................................................ 171
7.5.4 Results validation ............................................................................................... 173
7. 6 Summary ................................................................................................................. 173
**CHAPTER 8 ESTIMATING TRAVEL MODE AND AIRLINE CHOICE USING LOGIT MODELS ................................................................................................................. 174**
8. 1 Introduction ............................................................................................................ 174
8. 2 Research Context..................................................................................................... 174
8. 3 Methodology ........................................................................................................... 175
8.3.1 Stated preference survey design......................................................................... 175
8.3.2 Data used in this study ....................................................................................... 176
8.3.3 MNL model ........................................................................................................ 177
8.3.4 NL model............................................................................................................ 178
8.3.5 Direct and cross elasticities................................................................................ 181
8.3.6 Model fit statistics .............................................................................................. 182

8. 4 Results ..... 184
8.4.1 Modelling results of air passenger respondents ..... 184
8.4.2 Modelling results of non-air passenger respondents ..... 185
8.4.3 Elasticities ..... 187
8. 5 Discussion ..... 189
8.5.1 Factor analysis ..... 189
8.5.2 Elasticity analysis ..... 191
8.5.3 Willingness to pay ..... 191
8. 6 Summary ..... 192
**CHAPTER 9 ANALYSING TRAVEL MODE AND AIRLINE CHOICE USING LATENT CLASS MODELLING ..... 194**
9. 1 Introduction ..... 194
9. 2 Research Context ..... 194
9. 3 Methodology ..... 196
9.3.1 Data used in this study ..... 196
9.3.2 Latent class model ..... 196
9. 4 Results ..... 200
9.4.1 Latent class modelling results ..... 200
9.4.2 Characteristics of each segment ..... 202
9.4.3 Elasticities ..... 204
9. 5 Discussion ..... 206
9.5.1 Key factors ..... 206
9.5.2 Elasticity analysis ..... 207
9.5.3 Willingness to pay ..... 208
9.5.4 Characteristics of segments ..... 209
9.5.5 Number of segments ..... 210
9. 6 Summary ..... 211
**CHAPTER 10 EVALUATION AND CONCLUSIONS ..... 213**
10. 1 Introduction ..... 213
10. 2 Summary of Research Findings ..... 213
10.2.1 Domestic air travel demand forecasting ..... 215
10.2.2 Stated preference efficient experimental design ..... 215
10.2.3 Regional aviation market segmentation ..... 216
10.2.4 Travel mode and airline choice estimation ..... 217

10. 3 Limitations of the Research .......... 219
10.3.1 Limitations of data collection .......... 219
10.3.2 Limitations of modelling methods .......... 220
10. 4 Future Research Directions .......... 222
10. 5 Achievement of research objectives .......... 223
10. 6 Conclusions .......... 224
**REFERENCES .......... 226**

## LIST OF FIGURES

Figure 1-1 Thesis framework ... 29
Figure 2-1 Discrete choice question example of travel mode choice ... 50
Figure 3-1 Study area of Western Australia ... 60
Figure 3-2 Research workflow of the thesis ... 61
Figure 4-1 Spatial visualization of air passenger flows in Western Australia ... 87
Figure 4-2 Spatial distribution of towns and tourists in Western Australia ... 94
Figure 5-1 Experimental design matrices ... 103
Figure 5-2 Main stages for constructing constrained SP experimental design ... 104
Figure 5-3 Efficiency of design with respect to (a) sample size, and (b) different designs ... 114
Figure 5-4 Modified Federov Algorithm ... 116
Figure 5-5 Extension of the Modified Federov Algorithm for stage three ... 117
Figure 5-6 The t-value of the attributes related to each of the 10 simulation tests ... 134
Figure 5-7 Stated preference survey question of travel mode choice ... 135
Figure 6-1 Monthly income of air passengers in regional Western Australia ... 140
Figure 6-2 Education level of air passengers in regional Western Australia ... 141
Figure 6-3 Last stop before the airport ... 141
Figure 6-4 Popular destinations of air passengers in regional Western Australia ... 142
Figure 6-5 Travel companion of air passengers in regional Western Australia ... 143
Figure 6-6 Reasons for flying ... 144
Figure 6-7 Importance of key factors affecting travel mode choice ... 145
Figure 6-8 Trip purposes of air passengers ... 146
Figure 6-9 One-way travel cost ... 147
Figure 6-10 Booking time in advance to the flight ... 148
Figure 6-11 Chart matrix of Albany, Geraldton, Karratha and Broome air passengers reasons for flying ... 152
Figure 6-12 Chart matrix of trip purpose of Albany, Geraldton, Karratha and Broome air passengers ... 153
Figure 6-13 Chart matrix of one-way travel cost of Albany, Geraldton, Karratha and Broome air passengers ... 154
Figure 6-14 Chart matrix of trip frequency of Albany, Geraldton, Karratha and Broome air passengers ... 155

Figure 6-15 Chart matrix of booking time in advance of flight by Albany, Geraldton, Karratha and Broome air passengers ........ 157
Figure 7-1 Flowchart for EM algorithm ........ 162
Figure 7-2 Prominent characteristics of existing aviation and potential aviation market ...... 172
Figure 8-1 Structure of the MNL model ........ 181
Figure 8-2 Structure of the NL model ........ 181

# LIST OF TABLES

Table 2-1 Proposed and final Western Australia Government regulation policies ................. 35
Table 2-2 Properties of listed key studies using gravity models ............................................ 39
Table 2-3 Summary of previous mode choice studies ........................................................... 52
Table 3-1 Summary table of geo-spatial data collection ........................................................ 63
Table 3-2 Airport respondent data collection .......................................................................... 65
Table 3-3 Non-airport respondents data collection ................................................................ 65
Table 4-1 Descriptive statistics of the variables ...................................................................... 87
Table 4-2 PPML estimation results of gravity model based on the Thiessen polygon catchment area ................................................................................................................................ 88
Table 4-3 Descriptive statistics of the variables ...................................................................... 89
Table 4-4 PPML estimation results of Gravity model based on the 2.5 hour driving catchment area ................................................................................................................................ 90
Table 4-5 Correlation between available seats and independent variables ............................. 93
Table 4-6 Monthly load factor of Western Australia popular routes between June and September, 2016 ................................................................................................................................ 95
Table 4-7 Estimates of domestic aviation activities on Geraldton – Perth route ..................... 96
Table 4-8 Evaluation of online flight information – difference in seat and flight data ........... 96
Table 5-1 Full Factorial design and correlation matrix ........................................................ 108
Table 5-2 Orthogonal design ................................................................................................. 109
Table 5-3 Attribute table of ticket price ................................................................................. 118
Table 5-4 Attribute table of journey time ............................................................................... 118
Table 5-5 Similar choice tasks ............................................................................................... 119
Table 5-6 Plausibility, Dominance and Realism constraints specification ........................... 120
Table 5-7 Dimensions for generating the efficient design of this case study ........................ 122
Table 5-8 Parameter priors .................................................................................................... 125
Table 5-9 Pseudocode for generating attribute table of costs ............................................... 125
Table 5-10 Initial PDR constrains ......................................................................................... 126
Table 5-11 Pseudocode for generating candidate set ............................................................ 126
Table 5-12 Candidate set generated based on initial constraints .......................................... 128
Table 5-13 Efficient design A generated based on initial constraints ................................... 128
Table 5-14 Full PDR constraints ........................................................................................... 129
Table 5-15 Efficient design B generated based on full constraints ....................................... 130

Table 5-16 Theoretical minimum sample size ........ 131
Table 5-17 Pseudocode for generating random simulation results table ........ 132
Table 5-18 MNL modelling results based on the random simulation ........ 133
Table 5-19 Comparison of D-errors with non-zero parameter priors ........ 136
Table 6-1 Total RPT flights and total available RPT seats (weekly) ........ 139
Table 6-2 Access modes to the four airports ........ 143
Table 6-3 Air travel frequency to current destination ........ 147
Table 6-4 Main reason for selecting air travel mode based on trip purpose ........ 148
Table 6-5 One-way travel cost by booking time in advance to flights ........ 149
Table 6-6 Trip purpose by booking time in advance to flights ........ 150
Table 6-7 Final destinations for air travellers ranked by popularity ........ 151
Table 7-1 Information criteria for determining optimal number of air passenger segments . 167
Table 7-2 Air passenger market segments ........ 168
Table 7-3 Information criteria for determining optimal number of non-air passenger segments ........ 168
Table 7-4 Non-air passenger market segments ........ 169
Table 8-1 Airport respondent profiles ........ 176
Table 8-2 Non-airport respondent profiles ........ 177
Table 8-3 Modelling results for air passenger respondents ........ 185
Table 8-4 Modelling results for non-air passenger respondents ........ 186
Table 8-5 Direct elasticities from NL model for air passenger respondents ........ 187
Table 8-6 Direct elasticities from NL model for non-air passenger respondents ........ 188
Table 8-7 Cross elasticities for air passenger respondents ........ 188
Table 8-8 Cross elasticities for non-air passenger respondents ........ 189
Table 8-9 Information criteria of nested logit models ........ 190
Table 8-10 WTP of air and non-air passenger respondents ........ 192
Table 9-1 Information criteria for number of segments ........ 200
Table 9-2 Statistical results of multinomial and latent class models ........ 201
Table 9-3 Characteristics of each segment ........ 203
Table 9-4 Summary of direct elasticities ........ 205
Table 9-5 Cross elasticities of the latent segments ........ 206
Table 9-6 Willingness to pay for each of the segments ........ 209

# LIST OF APPENDICES

**APPENDIX A PERMISSION FROM CO-AUTHORS .................................................... 243**

**APPENDIX B INFORMATION SHEET .............................................................. 249**

**APPENDIX C ETHIC APPROVAL .................................................................... 250**

**APPENDIX D INTERCEPT SURVEY QUESTIONNAIRES ....................................... 252**

Appendix D1 Air Travel Information ...................................................................... 252

Appendix D2 Travel Mode and Airline Choice Block 1 ........................................... 254

Appendix D3 Travel Mode and Airline Choice Block 2 ........................................... 256

Appendix D4 Demographic Information ................................................................. 258

**APPENDIX E PEARSON CORRELATION MATRICES ........................................... 260**

Appendix E1 Pearson Correlation Between Independent Variables By Classifying Catchment Area Using Thiessen Polygons ............................................................................. 260

Appendix E2 Pearson Correlation Between Independent Variables By Classifying Catchment Area Using 2.5 Hour Drive Distance ............................................................................ 261

**APPENDIX F AIR PASSENGER MARKET SEGMENTATION ................................ 262**

Appendix F1 Information Criteria ........................................................................... 262

Appendix F2 Air Passenger Market Segments Block Two ......................................... 262

**APPENDIX G NON-AIR PASSENGER MARKET SEGMENTATION ....................... 263**

Appendix G1 Information Criteria ........................................................................... 263

Appendix G2 Non-air Passenger Market Segments Block Two .................................. 263

# LIST OF ACRONYMS

| | |
|---|---|
| ABS | Australian Bureau of Statistics |
| AIC | Akaike Information Criterion |
| APEC | Asia-Pacific Economic Cooperation |
| AVC | Asymptotic Variance–Covariance |
| BFGS | Broyden Fletcher Goldfarb Shanno |
| BIC | Bayesian Information Criterion |
| BITRE | Bureau of Infrastructure, Transport and Regional Economics |
| CUR | Clustering Using Representatives |
| DMP | Department of Mines and Petroleum |
| DoT | Department of Transport |
| EM | Expectation Maximisation |
| EMFA | Extention of Modified Federov Algorithm |
| FIFO | Fly-In Fly-Out (working arrangements) |
| GIS | Geographic Information Systems |
| IIA | Independence from Irrelevant Alternatives |
| IID | Independently and Identically Distributed |
| IV | Inclusive Value |
| LC | Latent Class |
| LGA | Local Government Area |
| MARS | Multivariate Adaptive Regression Spline |
| ME | Main Effect |
| MFA | Modified Federov Algorithm |
| ML | Mixed Logit |
| MNL | Multinomial Logit |
| MNP | Multinomial Probit |
| NL | Nested Logit |
| PAM | Partitioning Around Medoids |
| PDR | Plausibility, Dominance and Realism |
| PPML | Poisson Pseudo-Maximum Likelihood |
| RADS | Regional Airports Development Scheme |
| RP | Revealed Preference |
| RPT | Regular Public Transport |

| | |
|---|---|
| RSE | Residual Standard Error |
| RU2 | Random Utility 2 |
| SP | Stated Preference |
| TRA | Tourism Research Australia |
| VC | Variance-Covariance |
| WTP | Willingness To Pay |

# CHAPTER 1 INTRODUCTION

## 1. 1 Introduction

Aviation is a fast growing and dynamic industry that plays a vital role in contributing to travellers' wellbeing and facilitating economic development. The provision of ample opportunities for high-quality air movement services with affordable prices is, therefore, particularly important to the economy and society. From an Australian perspective, its geography means that aviation is effectively the only way to move people to many regional cities and towns domestically. According to government published aviation reports, Australian regular public transport (RPT) domestic air passenger demand has experienced a substantial increase in recent years, from 44.2 million in 2006 to 61.1 million in 2018 (Bureau of Infrastructure Transport and Regional Economics, 2007, 2018). The domestic aviation industry connects major cities with remote and regional areas, which often are geographically dispersed and isolated. However, aviation use in Western Australia has shown the opposite trend, with a 21% reduction in passenger volume across the top six regional air routes, from 2.33 million in 2012 to 1.84 million in 2018. The drivers of this decline remain uncertain; whether it is due to travellers switching to other modes of transport, travelling becoming less frequent (due to population change), the underlying economic demand or differing work practices, is not clear. One explanation is that the commodities boom in the 2000s fuelled a significant growth in aviation movements, with the subsequent end to the boom resulting in a parallel fall in aviation use. More than that, the mining construction downturn in Western Australia and the increasing number of jobless people in the state may also have reduced air passenger movements (Deloitte Access Economics, 2014a; Australian Bureau of Statistics, 2015b). Previous literature suggested that competition between air and road transportation, (car, coach), is increasingly intense due to increasing passenger numbers and the continuing development of the road transportation system (D'Alfonso et al., 2015; Jiang and Zhang, 2016b). In Western Australia, the distances between regional centres, (e.g., Albany, Esperance, Broome, Kununurra and Newman), and the metropolitan area (Perth) are relatively long compared with more populated areas of Australia or around the world. For example, the distance between Albany and Perth is 410 km, between Esperance and Perth is 715 km, and between Newman and Perth is 1195 km, while Kununurra is the most distant at over 3200 km by road. Western Australia is relatively unusual in that a significant majority of the population are located in a metropolitan hub, but with the remaining population very widely dispersed and needing appropriate transport options. Because of these distances, flying for regional trips is generally more time-efficient and more comfortable compared to road transport. However, it is generally more, and sometimes much more, expensive to fly. As previous research has indicated, the mode and airline choices of travellers may be subject to the influences of various factors. These include travel time, travel cost, accessibility, seat comfort and service frequency (e.g., Hess et al., 2007; Van Can, 2013; Chen and Chao, 2015; Ding et al., 2017), and socio-demographic factors such as gender and age (e.g., Wen and Lai, 2010; Ma et al., 2015; Román et al., 2017). However, little to no attention has been given to the investigation of which key parameters significantly affect travel mode and airline choice in regional Western Australia, and how these parameters affect that choice.

In addition, several other key aviation issues have been identified by the government transport agencies in Western Australia. High regional airfares compared to the rest of Australia is the most serious aviation issue and, even though the price-insensitive business-corporate sector has a relatively large share of the Western Australia market, high airfares still impede the regional public's experience of air travel service and curb economic growth (Department of Transport, 2015a). Disparity of air traffic demand appears on some regional air routes, with the air passenger flow concentrated on one direction only, which may affect the development of the aviation industry and hence reduce the air service quality. Additionally, some regional airports are unable to upgrade their infrastructure, such as runway construction and security screening, owing to a lack of finance. This restricts the introduction of larger aircraft as Commonwealth regulations require security screening of passengers and baggage for aircraft exceeding 20 tons.

In order to deal with these issues, the Western Australia State Government has tried to develop policies such as regulating/deregulating air routes, encouraging low-cost carriers, and developing planning schemes for upgrading regional airport infrastructure and stimulating economic growth (Department of Transport, 2014, 2015b). It is of crucial importance that these policies are targeted towards those airports and air routes where addressing these issues has the greatest potential benefits to the local community, and the state overall. Therefore, this thesis aims to investigate air travel demand and the factors that drive that demand, to identify and explore the regional and potential aviation markets and to estimate the travel mode and airline choices of regional travellers, especially how the key parameters, (travel time and service quality factors), affect their choices and behaviour. The expected outcomes, including air travel demand forecasts, regional air passenger market characteristics, as well as the competition between air and non-air travel modes and between airlines, should provide a more reliable and comprehensive insight for state government and airlines. With this guidance, they could more effectively and accurately implement the policies at those airports where there is a pressing demand, and hence better address aviation issues.

Previous literature has explored international aviation markets and investigated travel mode and airline choice between cities, states and or countries. Limited attention has been given to the regional area of Western Australia. Further, a limitation of the existing literature around aviation market segmentation and flight travel preferences is that much of it is based on respondents recruited at airports and/or train stations, which may be valuable for certain questions but is subject to selection bias and may not generalise to the rest of the population (e.g., Mason & Gray, 1995; Wen et al., 2008; Koo, Wu, & Dwyer, 2010; Van Can, 2013; Jung & Yoo, 2014). In this thesis, respondents were surveyed at both regional airports and other settings likely to involve those who do not frequently choose to fly, (e.g., shopping centres, libraries, on the street, technical education colleges and visitor centres). As a whole, this thesis will firstly estimate the air travel demand of airport-pairs in Western Australia, and subsequently develop an innovative market segmentation method for identifying existing and potential air passenger markets. Finally, it will investigate the competition between travel modes and airlines by estimating passenger travel mode choice and behaviour on competitive routes serviced by air transport in regional Western Australia.

## 1. 2 Research Objectives

The aim of the research is to investigate the regional aviation market and the competition in Western Australia, particularly the aggregate air travel demand, the identification of existing, (frequent air transport users), and potential, (non-frequent air transport users), aviation markets and the identification of key parameters, (e.g., travel cost, journey time and seat comfort), affecting travel mode and airline choice.

Therefore, this thesis provides useful information for government policymakers and airlines to access the parameters that can significantly affect air travel demand on the regional air routes, as well as the characteristics of target/core aviation markets. It will also broaden the understanding of regional travelling behaviours and competition in the aviation industry. With such guidance, policymakers and air carriers could more effectively implement policies and strategies to encourage public transport and airline usage and address the aviation issues. Thus, a relatively good air service could be ensured to cater for regional businesses and communities. In order to achieve these objectives, the following key tasks/objectives were developed for the project:

I. To forecast air passenger demands of RPT (regular public transport) airport-pairs using a modified gravity model (paper 1 - published);

II. To develop a novel framework for generating a more optimal stated preference (SP) experimental design;

III. To investigate regional air passengers' characteristics and the aviation market:
   a. Visualise and explore the regional air passengers' profiles and trip characteristics;
   b. Identify and explore regional aviation market using market segmentation techniques (paper/manuscript 2- resubmitted);

IV. To develop discrete choice models for identifying and investigating the key factors that affect regional travellers' mode and airline choices:
   a. Investigate and compare airport and non-airport passengers' travel mode and airline choices based on multinomial and nested logit models (paper 3 - published);
   b. Estimate traveller's travelling behaviours with preference heterogeneity accommodated using a latent class modelling approach (paper 4 - published).

## 1. 3 Research Significance and Contribution

### 1.3.1 Significance

Investigating regional air travel demand, the aviation market, travel mode and airline choice is important for air transport planning regulation/deregulation policymaking, which can help with facilitating growth in the aviation industry. This research develops a robust and novel methodology framework for achieving the above, which would result in a range of benefits as shown below.

- The modified gravity modelling analysis (as presented in Chapter 4), allows air travel demand on regional air routes to be forecast, with key drive factors identified, which would be useful for government and policy-makers in formulating and implementing aviation policies more effectively;
- The model-based market segmentation (as applied in Chapter 7) provides the ability to identify and investigate the core and potential aviation markets for more targeted advertising and marketing. Hence air transport usage could be promoted which could both reduce airfares and improve aviation services; and
- The transport related discrete choice analysis based on SP surveys (as presented in Chapters 8 and 9) provides insight into exploring and quantifying passengers' sensitivities to the key parameters. Thus, the government transport agencies and airlines could more effectively develop and deploy strategies, (such as airfare structures and service frequencies), to attract passengers and hence increase public air transport usage.

### 1.3.2 Contributions

In addition to investigating the regional aviation market in Western Australia, the major contributions of this research are:

- Extending spatial modelling analysis to provide more accurate prediction of bilateral air passenger demands (airport-pairs) by considering the impact of different sizes of airport catchment areas;
- Developing a novel methodology procedure for generating an efficient SP survey, whereby a high statistical efficiency and an appropriate plausibility and realism level of the experiment is maintained;
- Extending the mixture model-based market segmentation approach by introducing an SP experiment technique to identify both existing and potential aviation markets; and
- Developing a set of discrete choice models to estimate and compare air and non-air passenger travel preferences, thus providing a more reliable and comprehensive

understanding of passengers' travel mode choices and sensitivity to key factors such as travel cost, travel time and service quality.

## 1. 4 Research Methodology

This thesis proposes a rigorous and easily computed modified gravity model and a mixture of clustering travel mode and airline choice models. These models not only identify the determinants affecting air travel demand (such as geo-economic and service-related factors), but also explain regional passengers' travel mode and airline choice and reveal the competition in the regional aviation market. Primarily, modified gravity models with the Poisson pseudo maximum likelihood (PPML) estimation method are applied to forecast bilateral air travel demand of airport-pairs in regional Western Australia. Subsequently, the mixture model-based clustering approach with the Expectation Maximisation (EM) algorithm estimator is utilised to explore both existing and potential aviation markets in regional Western Australia. Next, discrete choice models including multinomial logit (MNL), nested logit (NL) and latent class (LC) models are developed for more comprehensively estimating individual travel mode and airline choices.

## 1. 5 Thesis outline

The framework of the thesis, constituting ten chapters, is shown in Figure 1-2, with a description for each chapter listed below:

**Chapter 1** introduces the key issues around aviation in regional Western Australia and describes how the thesis will approach some key questions in the area. In order to help government and regional airlines to better tackle the aviation issues, a set of objectives and the research significance are then proposed.

**Chapter 2** is the literature review which firstly introduces the background of aviation industry in regional Western Australia including the significance of the regional air service, market status and competition as well as a key aviation issue (high regional airfares). It then reviews the existing research on investigating aviation markets, the aggregate and disaggregate methods for travel demand forecasting, market segmentation, travel mode and airline choice modelling.

**Chapter 3** describes the study area and presents the research workflow for data collection and analysis. Data analysis comprises the methodology for RPT air passenger demand forecasting, regional aviation market exploration, as well as the passenger travel model and airline choice estimation.

**Chapter 4** introduces modified gravity models with PPML estimator for forecasting the bilateral air travel demand of airport-pairs in regional Western Australia. On the basis of gravity model outputs, the impact on air passenger flows by distance, airfare, catchment area, population, mining and tourism sectors are investigated. This chapter implements research task/objective I.

**Chapter 5** describes the entire procedure of experimental design for generating the SP survey that can be used to investigate individual choice behaviours, such as travel and consumption. In this thesis, it is used to explore regional aviation markets and estimate travel mode and airline choice. This chapter not only generates the efficient design for the subsequent SP survey, but also develops a novel approach by extending the Modified Federov Algorithm, which can effectively generate efficient discrete choice experimental design while maintaining an appropriate behavioural plausibility and realism. This chapter implements research task/objective II.

**Chapter 6** uses *python version 3.6* programming language (Van Rossum and Drake Jr, 1995) for pre-processing and analysing the air passenger survey dataset, which subsequently generates a visualization for preliminary understanding and comparing regional air passenger profiles, (e.g., socio-demographics and economics), and trip characteristics, (e.g., trip purpose and reasons to choose air travel). This chapter covers research task/objective III-a.

**Chapter 7** develops a mixture model-based market segmentation approach for exploring the regional aviation market in Western Australia. The model identifies and compares existing and potential aviation market segments based on air and non-air traveller characteristics and stated preference/probability for air and non-air travel modes. This chapter implements research task/objective III-b.

**Chapter 8** develops MNL and NL models to estimate air and non-air traveller travel mode and airline choice, respectively. The modelling results compare the air and non-air traveller choice behaviours by measuring their sensitivity to the key factors of travel cost, journey time, accessibility, service frequency and seat comfort. This chapter covers research task/objective IV-a.

**Chapter 9** builds an LC model to investigate travel mode choice behaviour within, and among, latent passenger segments using SP data. The LC model accommodates unobserved preference heterogeneity by assuming a discrete distribution of the unobserved preference, (represented by latent segments), that may provide more critical and reliable estimates of traveller mode choice behaviours. This chapter covers research task/objective IV -b.

**Chapter 10** is a conclusion chapter which comprises the summaries of the major findings and corresponding limitations with respect to the objectives, and recommendations for future studies.

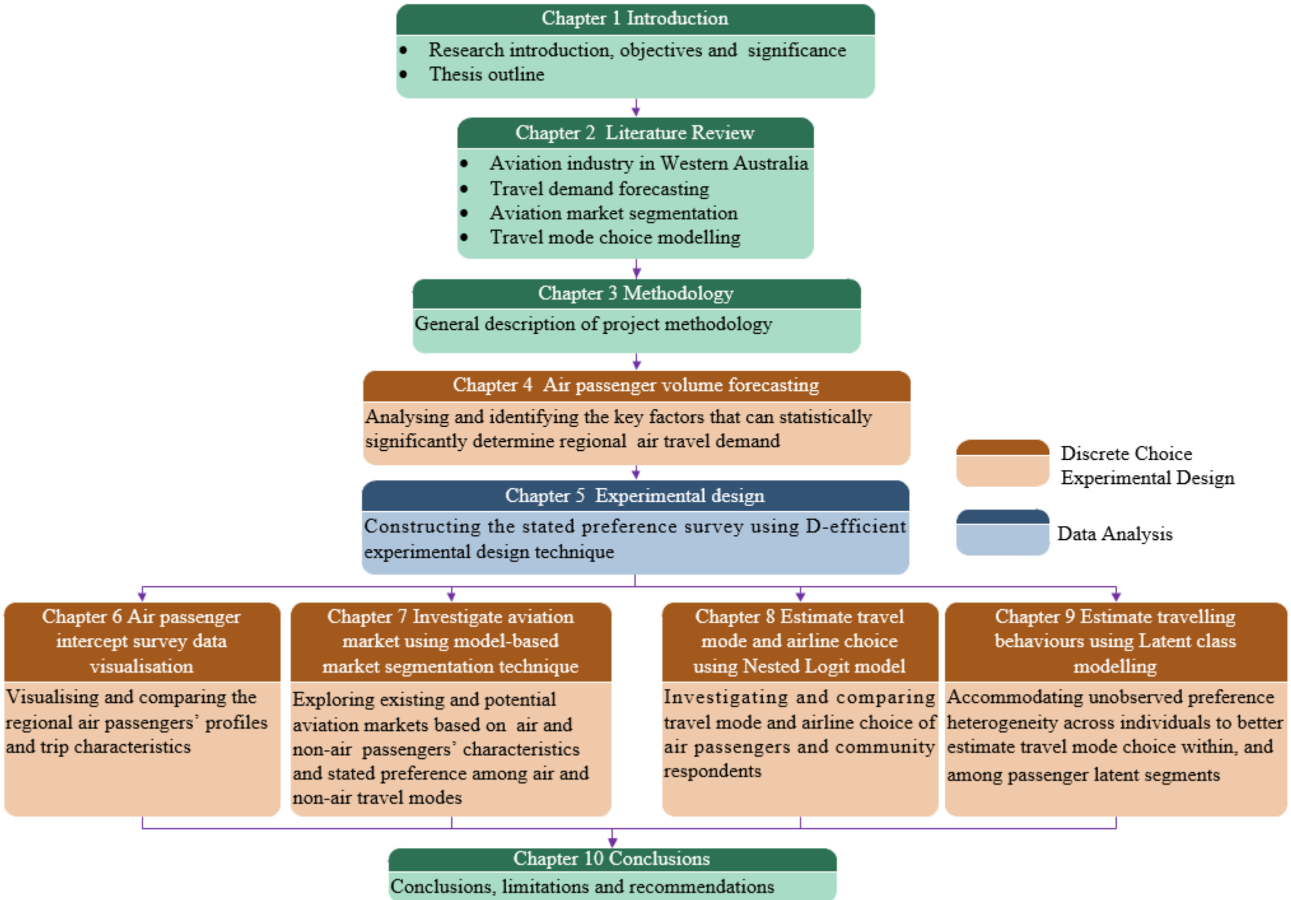



*Figure 1-1 Thesis framework*

## 1. 6 Summary

This chapter has introduced the key aviation policy issues in regional Western Australia. A set of objectives and the rationale for the research have been established for providing a more comprehensive investigation into the competitive regional aviation markets in Western Australia that therefore could assist the government and regional aviation industry to better deal with the aviation issues. The next chapter will review the literature relevant to the aviation market in Western Australia and the modelling methods of travel demand, market segmentation and travel mode, and airline choice.

# CHAPTER 2 LITERATURE REVIEW

This chapter presents a background of the regional aviation industry in Western Australia, as well as a review of previous research relevant to the exploration of regional aviation market demands, characteristics and travel mode and airline choices. It begins with a description of the regional air transportation market in Western Australia in terms of current status, competition, aviation issues and the State Government's policies and strategies. It subsequently discusses statistical modelling approaches commonly used in the literature to help governments and airlines to better counter these aviation issues. The specific statistical approaches considered are for travel demand forecasting (section 2.2), aviation market segmentation (section 2.3) and travel mode and airline choice estimation (section 2.4). While describing each modelling approach, corresponding definitions, significances, related key factors and identified research gaps to date are illustrated. These gaps will then be addressed in this thesis.

## 2. 1 Aviation in Western Australia

### 2.1.1 Aviation market

Air transportation services perform a pivotal role in connecting remote and rural regions in Western Australia with the hub city of Perth, especially for accelerating economic and social development of the state such as facilitating the resources flowing among local communities and catering for the corporate, leisure and mining sectors (Department of Transport, 2015c). In total, Western Australia covers more than 2.5 million square kilometers, (approximately one-third of the country), and with a population of just 2.5 million, is heavily reliant upon air transport services. The particularly high importance of aviation in Western Australia can be explained by the geospatial status of the enormous state - dispersed and isolated regional towns with very limited rail connections within the state. Therefore, the State Government has been in the past, and continues to be, concerned with providing safe, reliable, efficient and affordable intrastate RPT aviation services, as far as feasible. In line with this, some vital policies have been developed for targeting the objectives, such as the *Western Australian Transport Coordination Act 1966* and the *Transport Coordination Regulations 1985* regarding the licensing of aircraft and placing of conditions.

With respect to the regional aviation market, in 1997 the Government established the Regional Airports Development Scheme (RADS), which endeavours to ensure that that '*regional aviation infrastructure and airport services are developed and maintained to facilitate air access and enhance economic growth in Western Australia*' (Department of Transport, 2015c, p. 11). Particularly for the small airports in regional Western Australia, it is necessary to upgrade some assets and to replace others. For instance, Kununurra airport requires replacement of its time-expired runway in order to meet the general aviation transportation requirements. Without prompt upgrades, extra costs and restrictions may be imposed upon the social and economic development of the state. In the previous five years, regional airports have invested $145 million on airport infrastructure such as the upgrading of runways and service infrastructure (Department of Transport, 2015c). The small airports in regional Western Australia such as Albany, Busselton and Esperance also play an important role in transportation to Perth or other places in Australia. These airports are typically serviced by regulated routes, as the Government's transport agency considered these airports too small to be able to sustain direct competition (Department of Transport, 2015c). However, as the Department of Transport (DoT) realised, it is impossible to invest in all the regional airports in Western Australia due to insufficient funding. Therefore, it is important to use the limited funding for those airports where the aviation benefits to the local community can be maximised, and especially for those small airports with passenger flows that are inadequate to enable them to be financially self-sustaining. Thus, reliable air traffic demand forecasts become more and more crucial for guiding the planning of regional airport investment.

The vast majority, (more than 90%), of interstate and international travellers to Western Australia do so by air (Department of Transport, 2015c). The cost of intrastate travel in Western Australia is higher than in most other states in Australia. This may be due to the modest aviation traffic and the resulting low economies of scale, (e.g., the relatively high average security cost per passenger at the small regional RPT airports), the absence of low-cost carriers, a large proportion of price-insensitive customers and the high costs operating in remote areas. Specifically, in comparison with leisure aviation passengers, business travellers account for a large component of air passengers and are price-insensitive, as their tickets are paid for by their employers (Department of Transport, 2015c). However, the aviation issue of high airfares conflicts with the State Government's objectives and will not only limit the growth of regional aviation services but also affect community satisfaction with their air services.

Therefore, to more comprehensively understand air travel demand in Western Australia, its key determinants and the parameters related to the RPT airport service infrastructure, (e.g., accessibility), will be considered in this thesis because the inappropriate development of airport infrastructure or application of airline regulating policies may impose significant cost and other limitations that could impact negatively on economic growth.

### 2.1.2 Airline and travel mode competition

As reported in the Western Australian State Aviation Strategy (Department of Transport, 2015c), aviation is a fast growing and dynamic industry that plays a vital role in the economic and social development of Western Australia. However, as previously mentioned, between 2012 and 2018, the statistical reports show that Western Australia experienced a considerable decrease in domestic airline passengers movements (Bureau of Infrastructure Transport and Regional Economics, 2012, 2018), even though overall aviation traffic in Australia experienced considerable growth over the same period. Due to the significant decrease in air passenger movements, the airlines may introduce more competitive policies, such as reducing airfares and improving service quality, in order to attract more passengers (Gautam, 2002; Gaggero and Piga, 2010). From the State Government’s perspective, one critical objective is to foster competition between regional airlines, in order to reduce airfares and maintain sufficient air services for the local communities and continuously stimulate economic development for the resource sector. Corresponding actions such as deregulating air routes and airlines and introducing charter flights, (without detracting from the RPT air traffic), are being implemented. RPT air services are operated through a published flight schedule that caters for the needs of the public communities, whereas charter flights are ‘closed’ air services that precisely are not RPT services and normally devoted to satisfying the needs of the resource industries[1].

Although the two-airline policy was terminated in Australia in 1990 by the Airline Agreement Termination Act, the domestic/regional aviation industry still retains an effective duopoly with choice on many routes limited to Virgin and Qantas airlines (Douglas, 1993; Quiggin, 1997; Zhang et al., 2018). It indicates a relatively fierce competition between two airlines that operate on the same air routes. However, the competition between air and road transportation (e.g., highway) is growing in intensity due to the continuing development of road transportation systems, especially for short or mid-distance travel route (less than 1000 km) (D'Alfonso et al., 2015; Jiang and Zhang, 2016a). For example, the distance between Geraldton and Perth is around 400 km and the drive time, (assuming the average driving speed is 90 km/h), and flying time, (assuming the average flying speed is 400 km/h), are approximately four hours and one hour, respectively. However, if a passenger chooses to travel by air, the total travel time will include access time, (access to the departing airport), waiting time in the airport, flying time from origin to destination, (including landing time and take off time), and time from airport to final destination. Additionally, the modest cost of road travel compared to the relatively high regional airfares in Western Australia may be a critical factor affecting the mode choice of regional travellers. Therefore, an individual passenger will make a choice of the travel mode based on several parameters such as total travel time, cost, accessibility, service quality and convenience, as well as the trip purpose and their personal preferences around transport modes  (Wen and Lai, 2010; Jung and Yoo, 2014; Román et al., 2017; Wu and So, 2018).

The opportunities for airport substitution leads to airport and service competition, but the degree of substitution opportunities usually varies in different regions (Starkie, 2002). There is little competition between airports in regional Australia due to long distances between the airports

[1] The Western Australia Transport Co-ordination Regulations 1985 (TCR) define the Regular Passenger Transport (RPT) services as those that "operate according to a published schedule', while the charter service is as "air service that is not an RPT service". (Department of Transport, 2015b).

(Forsyth, 2002), which normally operate as natural monopolies. Therefore, this study will not consider competition between airports in regional Western Australia.

### 2.1.3 Airline regulation and deregulation

Airline regulation for economic purposes, (as opposed to safety, environment or other purposes), includes the control of air travel in order to serve the interest of the public (Gautam, 2002; Button et al., 2019). Most air routes in Australia are unregulated and not subject to economic regulation. Travel routes, airlines, service quality factors, ticket prices or other factors may be regulated in order to ensure reliability or quality of air services, fair prices or affordable airfares. Since the aviation industry is complex and contains both regulated and competitive mechanisms, different regulation policies could influence its competition level. For example, the antitrust policy which was released by a local transport government to regulate a monopoly may foster free competition between airlines and thus the airfare may be reduced, due to less concentration of airline markets and freer competition (Gautam, 2002; Günster and van Dijk, 2016).

The motivation of the Western Australia State Government in regulating regional aviation is to ensure a sufficient level of air service affordability, reliability, frequency and quality to the public. In some cases, significant air route regulation is required to protect those routes with insufficient passenger demand for more than one airline to operate commercially. This would help to ensure an appropriate and efficient air service for the regional communities and industries (Department of the Prime Minister and Cabinet, 2013; Department of Transport, 2015a). Nevertheless, route regulation that is poorly applied may also lead to improper practices or unintended consequences, such as higher than necessary airfares due to limited market competition, high dependence on the exclusive airline and inefficiency of innovation and creativity (Department of Transport, 2014).

In contrast, airline deregulation is the process of removing government-imposed restrictions on airlines that affect, in particular, the carriers permitted to serve specific routes (Sandell, 1978). The 1978 Airline Deregulation Act significantly boosted the aviation industry in America by fostering a free aviation market that reduced airfares and increased air service frequency and air passenger movements (Sandell, 1978; Sepp and Aiello, 2018). Notably, Schipper et al. (2003) found that the deregulation of airlines could result in the reduction in airfares and increased flight frequency due to airline competition for passengers. In line with this, Quiggin (1997) applied a price indices theory-based empirical analysis and found that airline deregulation could have decreased air travel costs in Australia. They used Australian air travel data including market share and price changes following deregulations, provided by Bureau of Transport and Communications Economics (1994). However, it is not clear that regulation necessarily results in better or worse outcomes, dependent on the circumstances, timing and the quality of regulation.

The regulatory approach in Western Australia grants the operating RPT service and monopoly rights on the regulated routes to a single operator, which is annotated on the specified airline's airplane licence (Department of Transport, 2015a). All other airlines cannot operate RPT services on these regulated routes. Additionally, the Western Australia State Government has proposed a light-handed regulatory approach to ease the air route regulation, where reasonably practicable, that is economically and administratively efficient in fostering competition, lowering airfares and creating more choices for the regional communities (Department of Transport, 2014). Deregulation of air service was introduced to Geraldton and limited competition to Exmouth in 2011, respectively. It has significantly increased passenger numbers and reduced the cost of airfares (Department of Transport, 2015c). The proposed and final regulation policies beyond February 2016 are summarised in Table 2-1 (Department of Transport, 2015a, p. 24).

*Table 2-1 Proposed and final Western Australia Government regulation policies*

| Air Route (Fly to/from Perth) | Proposed regulation | Final recommendation |
|---|---|---|
| **Albany** | Reduce regulation and open route to potential competition but ensure Government oversight of scheduling, pricing and community engagement if appropriate | **Amended** – Regulate for up to five years and award rights to a single operator subject to a route review of airline performance |
| **Derby** | Flexibly regulate by allowing the potential entry of other airlines should there be a significant change in the region's economic circumstances | **Amended** – Regulate the Perth Derby (Curtin Airport and/or Derby General Aviation Airport) air route and award rights to a single RPT operator for up to five years subject to a route review of airline performance |
| **Esperance** | Reduce regulation and open route to potential competition but ensure Government oversight of scheduling, pricing and community engagement if appropriate | **Amended** – Regulate for up to five years and award rights to a single operator subject to a route review of airline performance |

| | | |
|---|---|---|
| **Learmonth (Exmouth)** | Fully deregulate | **Amended** – Reduce regulation but maintain minimum Government oversight by allowing other airlines to enter the market and operate on the route at any time; and, for as long as the route is serviced by a sole operator, place conditions on the airline's licence that requires the reporting of key route statistics |
| **Leinster** | Discontinue Leinster as an RPT airport beyond 2016 or earlier | **No change** |
| **Monkey Mia-Carnarvon** | Regulate | **No change** – regulate for up to five years and award rights to a single operator subject to a route review of airline performance |
| **Northern Goldfields (Meekatharra, Mount Magnet, Laverton, Leonora and Wiluna)** | Regulate either by maintaining the existing route connections or establishing new route configurations, direct or triangulated | **No change** – Award rights to a single operator subject to a route review of airline performance |
| **Charter air services** | Regulate charter operations on unregulated or regulated RPT routes including the placement of a special charter condition | **No change** |

*Note: Copied from Department of Transport (2015b), page 24*

As previously mentioned, the decline of regional air passenger volumes and the high airfares in regional Western Australia are the key problems that seriously hinder the regional public air service quality, (e.g., unaffordable travel for communities) and the economic growth (e.g., mining industries and tourism) of the state (Department of Transport, 2015c). Directional fluctuations in air passenger demand of the intrastate RPT air routes can only ensure a desired patronage for the airlines in one direction, which increases the running cost in disguise for the airlines and airports and hence results in a further increase in the airfare. The low financial capacity of some regional RPT airports has restricted their infrastructure maintenance, development and expansion for accommodating larger airplanes and improving air services. In order to assist the government and airlines to tackle these issues, this thesis constructs a robust and entire methodology framework to investigate the regional aviation markets, as well as the external and internal competitions, while addressing relevant research gaps. Consequently, the expected outcomes could guide the government and airline policy-makers to better understand the regional aviation markets and hence to more effectively respond to, and provide for, the target market.

## 2. 2 Air travel demand forecasting

### 2.2.1 What is air travel demand?

Air travel demand is commonly measured as the number of potential air passengers on the unidirectional trips from origin to destination, (e.g., airport-pair and city-pair), within a fixed time period, which therefore is a particular application of origin-destination market demand. Air passenger demand forecasting, which endeavours to predict the bilateral number of passengers between cities or airports, has been accessed by air transport government agencies and the aviation industry to explore potential passenger travel behaviours and to conduct economic planning (Bain, 1976; Grosche et al., 2007; Becken and Carmignani, 2020). Airlines and local governments are becoming more and more interested in the travel demand modelling studies (Valdes, 2015). Due to the dispersed regional location status in Western Australia, regional air transportation plays a significant role in linking remote regions with the hub city (Perth) but currently has experienced a considerable decline in air movements with direct competition from other airlines and ground travel modes. Therefore, accurate air travel demand estimations are of critical importance for guiding policymaking by air transport agencies, (i.e. regulation/ deregulation policy), and future infrastructure planning of airports (Scarpel, 2013). Reliable bilateral passenger demand forecasts also provide a good evidence base that can contribute to an airline's overall success (Grosche et al., 2007; Chang, 2014). Airlines can more efficiently deploy operating strategies and make decisions in order to satisfy demand and attract air passengers, such as airline fleet planning, starting new routes and introducing low-cost carriers (Doganis, 2009; Srisaeng et al., 2015; Becken and Carmignani, 2020).

A variety of theoretical approaches has been applied to forecast air travel demand that can be broadly split into two major categories, focusing on macro- and micro-approaches (Rengaraju and Arasan, 1992). Macro-approach based demand analysis attempts to forecast system-wide air travel activities by considering air transport network and airline-specific variables rather than the characteristics of particular links of nodes in the entire air transport network (Ghobrial, 1992). Therefore, this approach may be limited when aiming to more reliably predict air passenger volumes of specific air routes or links in a focused study area. Conversely, micro-approach based demand analysis intends to forecast specific origin-destination passenger demand, such as predicting the air passenger volumes of airport-pairs or city-pairs (Rengaraju and Arasan, 1992). To forecast the air travel demand of an airport-pair, the micro-approach model normally would include the socio-economic characteristics of the cities the airports are located in and the corresponding airline service factors as the independent variables to estimate the demand, with an assumption that the air movements are dependent upon these characteristics (Verleger Jr, 1972; Kanafani, 1983).

### 2.2.2 Gravity model

The gravity model is an aggregate demand analysis model that is often employed in micro-approach analyses, with some adaptions, to forecast bilateral origin-destination activities, such as trade flows (Anwar and Nguyen, 2011; Kabir and Salim, 2016; Çekyay et al., 2020), migration flows (Poot et al., 2016), passenger flows (Grosche et al., 2007; Asmael and Waheed, 2018), cargo flows (Hwang and Shiao, 2011), tourist flows (Marrocu and Paci, 2013) and other spatial interaction-based activities (Drakos et al., 2014). Gravity models are effective for understanding spatial structure and interaction (Nijkamp, 1997; Kabir et al., 2017; Çekyay et al., 2020). They are one of the most successful types of empirical models in economics that have been extensively applied in quantitative traffic demand research over the past 30 years (Hazledine, 2009; Anderson, 2010). The theoretical foundations of gravity can be incorporated into a spatial phenomenon to accurately evaluate and interpret spatial relationships. Gravity models are built upon the notion of Newton's law of gravitation whereby the attractive force between two objects is directly proportional to the product of their masses but inversely proportional to the distance between the objects. The physical concept has been widely transferred and applied in many branches of science, including economic geography, regional and human sciences and transportation planning (Nijkamp, 1997; Mikkonen and Luoma, 1999). The traditional gravity model has been applied in estimating spatial interaction, such as forecasting passenger movements between two areas (Chang, 2012; Binova, 2015; Nicolas et al., 2018) per Equation 2-1 below,

$$F_{ij} = \mathrm{g}\frac{P_i A_j}{f(d_{ij})} \quad \text{2-1)}$$

where $F_{ij}$ is the trips from origin area i to destination area j, $P_i$ is the production factor of area i, $A_j$ is the attraction factor of area j, $d_{ij}$ is the distance between area i and j, and g is a constant.

### 2.2.3 Gravity modelling studies and limitations

Many previous studies have used gravity models to analyse and model air travel demand and its determinants (e.g., Buraga and Rusu, 2014; Bínová, 2015; Kabir et al., 2017; Boonekamp et al., 2018). Most of these studies applied linear and parametric regression methodologies to identify the determinants of air travel demand between regions or countries. However, non-linear and non-parametric regression, such as the multivariate adaptive regression spline (MARS) model, have also been used to estimate air travel trips (Friedman, 1991; Chang, 2014). MARS allows different functions over different intervals; each interval using a regression slope to represent the nonlinear correlation between the dependent and independent variables (Chang, 2014). For these linear and non-linear models, the observation units are normally airports, major regions, airport-pairs, suburb-pairs, city-pairs and country-pairs, while the explanatory variables are airfare, flight frequency, aircraft size, air travel time, population, average income, GDP, employment rate and some specific factors that may influence the costs and probability of trading relationships between regions, such as dummy variables of flight destination within or out of the county and the language difference in a destination country (Wei and Hansen, 2006; Endo, 2007; Grosche et al., 2007; Hazledine, 2009; Hsiao and Hansen, 2011; Buraga and Rusu, 2014; Chang, 2014). These studies used the city, regions or the country where airports are located as the catchment area instead of airport centred catchment areas. Some key studies related to air travel demand forecasting using gravity models are summarised in Table 2-2.

*Table 2-2 Properties of listed key studies using gravity models*

| Authors | Study area | Explanatory factors | Major Results | Gaps filled |
|---|---|---|---|---|
| (Long and Uris, 1970) | 23 major cities in 1960s in America except New York and Washington | Population, distance and intervening opportunities | Air passenger trips are directly proportional to population, while distance and intervening opportunities are negatively correlated with air passenger trips. | Developing Synthesis and Intervening Opportunities models based on gravity model in order to consider the influence of alternative destinations on air trips to the given destination city. |
| (Long, 1970) | 23 major cities in 1960s in America except New York and Washington | Population, distance, intervening opportunities and disaggregation of intervening opportunities | Population is positively correlated with air passenger trips between city-pairs, distance and intervening opportunities are negatively correlated with the air trips, while assuming uniform | Considering the spatial structure effects in the modified gravity model, as well as the diversity effects among population of different cities. |

| | | | | |
|---|---|---|---|---|
| | | | population effects. | |
| (O'Kelly et al., 1995) | 25 cities in the United States | Distance and nodal attractions | Optimized nodal attraction includes propulsion and attraction factors were estimated for the twenty-five American cities, which shows acceptable estimation results. | Applying linear programming technique to estimate the optimized nodal attractions variables, in order to improve the accuracy of air passenger trip data analysis. |
| (Jorge-Calderón, 1997) | International air routes between cities in Europe in 1989 | Distance, population, income, dummy variables of whether the destination is hub, whether a hub airport exists and whether air route cross the sea | Population, income, dummy variables of hub airport and sea-crossed air route are positively correlated with air trips, while distance is negatively correlated with the air passenger trips. | Considering the diversity of exogenous demand effects on estimating travel trips in Europe. |
| (Wei and Hansen, 2006) | Major hubs in Unite States | Flight frequency, aircraft size, airfare, distance, local passenger population, number of spokes, income and aircraft arrival capacity | Fight frequency, aircraft size, distance and aircraft arrival capacity are positively correlated with air passenger trips, while airfare, number of local passengers and income are negatively correlated with air passenger trips. | Modelling the air travel demand on a hub-and-spoke network perspective. |
| (Endo, 2007) | United States, Japan and other countries | GDP, distance, network size, language and open sky policy. | GDP and common language are proportional to the international air travel demand whereas distance and larger network size both have a significant negative impact on the demand. The influence of open sky policy is proportional to air travel demand but can reduce the export demand. | Addressing trade in air travel demand of United States and Japan aviation market, and examine the key drivers. |
| (Grosche et al., 2007) | 28 European countries | Population, catchment, buying power, GDP, distance and travel time | All the factors have positive influences on the air passenger trips, except GDP and travel time which are inversely proportional to air passenger trips. | Providing more accurate unconstrained demand based on the extensive research. |
| (Hazledine, 2009) | Major cities in Canada | Distance, population, income, dummy factors of language and destination difference, | Distance and dummy variable language difference are negatively correlated with air passenger trips, while population is positively correlated with the air passenger trips. | Considering border effects to air travel by adding more evidence such as input dummy variables to the model. |
| (Buraga and Rusu, 2014) | 385 airports in Europe | Business trip cost, GDP and number of airports serviced in European destinations | Business trip cost and GDP both statistically affect the air passenger trips. | Providing an elaborate advanced understanding of the airport’s interactions/air passenger trips by modelling spatial discontinuities effect. |
| (Chang, 2014) | Asia-Pacific Economic Cooperation | Population, distance, income, GDP, annual export/import value, | Income, annual import value and unemployment rate play positive roles in determining | Introducing non-parametric and nonlinear regression method instead of parametric |

| | | | | |
|---|---|---|---|---|
| | (APEC) | language difference, unemployment rate and customer price index | the bilateral air travel demands. Distance and unemployment rate are negatively correlated with bilateral air travel demands. | and linear regression methods to identify the factors which can significantly affect air travel demands. |
| (Boonekamp et al., 2018) | Air travel demand of intra-European flights (11,619 origin-destination pairs) | Service frequency, airfare, population, GDP, airport connectivity, distance, domestic traffic, presence of low-cost carrier, number of hotel nights, ethnicity, public service obligation and aviation-dependent employment | Service frequency, population, GDP, airport connectivity, presence of low-cost carrier, number of hotel nights, ethnicity, public service obligation and aviation-dependent employment are positively correlated with the air travel demand, while airfare has a negative impact on the air travel demand. | Included two innovative variables (ethnicity and aviation-dependent employment) into the air travel demand forecasting analysis. The results indicate these two factors both have a statistically positive effect on the air travel demand. |
| (Becken and Carmignani, 2020) | Global air travel demand forecasting from 2020 to 2070. | GDP, airfare and three levels (low, high and extreme) of climate change mitigation, | Travel cost (negative effect) and GDP (positive effect) are the two key factors affecting global air travel demand. High level of climate change mitigation could more properly balance the variation of both GDP and travel cost, whereby it can provide a relatively better outcome for the sustainable growth of the global aviation industry. | Extended the literature on air travel demand analysis, by introducing and investigating the omitted key factor/constraint of climate change mitigation. |

The results from most previous studies that applied the gravity model for air passenger trips forecasting show that common factors, such as average per capita income, employment rate, distance, population and travel time, were found to be correlated with air passenger demand (Long, 1970; Grosche et al., 2007; Dobruszkes et al., 2011; Chang, 2012; Binova, 2015; Boonekamp et al., 2018; Becken and Carmignani, 2020). Generally, these driving factors can be divided into two classes; geo-economic and service-related factors (Jorge-Calderón, 1997; Grosche et al., 2007; Binova, 2015). Geo-economic factors can be further classified into activity-related and geographic variables. For example, population and average per capita income of the airline serviced area are the two activity-related variables most widely used by many researchers (Long, 1970; Grosche et al., 2007; Chang, 2012; Binova, 2015). Some studies also considered other activity-related factors, such as the full time employment rate, employment composition and Gross Domestic Product (Jorge-Calderón, 1997; Grosche et al., 2007; Buraga and Rusu, 2014). The most commonly used geographic variable influencing the trip demand between two regions is distance (Long and Uris, 1970; Jorge-Calderón, 1997; Chang, 2014). Grosche et al. (2007) noted that increasing distance can not only decrease the competitiveness between air travel and other travel modes, but also reduce social interactions between the two regions. Apart from that, service-related variables refer to service and air transport system characteristics, such as airfare, frequency and aircraft size (Jorge-Calderón, 1997; Grosche et al., 2007; Boonekamp et al., 2018; Becken and Carmignani, 2020). Travel time is an important factor that can be related to airline service quality. It partly depends on the flight frequency, since more frequent flights increase the probability that air passengers can find a flight closer to their preferred departure time and reduce their waiting time. On the other hand, higher airfares can decrease the air passenger numbers, especially on those shorter length air routes, as more passengers would be likely to choose other travel modes, such as car, buses and trains instead (Grosche et al., 2007). The literature shows that some researchers ignored airfare as a factor when estimating air passenger trips because the airfare usually has a high correlation with geographic distance and thus multicollinearity will appear (Rengaraju and Arasan, 1992; Grosche et al., 2007).

The network structure of flights between airports is another factor that can have an influence on air travel demand. The structure of the network is itself an outcome of economic, political and geographical factors (Guimerà et al., 2005). As such, Guimerà et al. (2005) further found that air transport is a scale-free network in which *"the most connected cities are not necessarily the most central"* (p. 7794). In the post-deregulation industry, airlines choose hub locations that can give them a competitive edge (Wei, 2014). This can improve flight frequency and passengers' preference for the link as schedule delays are reduced and service quality improves (Givoni and Rietveld, 2009). At the same time, flight frequency needs adequate demand to sustain it, which implies a 'feedback effect' (Wei, 2014) that may only be possible in some hubs within the network.

However, studies on the spatial extent of the factors affecting air passenger demand in regional areas are limited. Furthermore, inappropriately defined catchment areas of airports may cause some issues in air travel modelling. Previous research mostly tended to use administrative boundaries, such as county, city and region as the catchment area of airports (Wei and Hansen, 2006; Hazledine, 2009; Buraga and Rusu, 2014; Chang, 2014). The limitation of this method might be arbitrary without considering the spatial distribution of airports. For example, the people who live in one city but are located closer to the airport in an adjacent city may not belong to the catchment area/city they live in. Therefore, the forecast results could have some errors if the catchment areas of the airports cannot be classified appropriately. In order to counter this problem, two kinds of catchment determination methods have been applied in this thesis, based on the location of airports. The first method is generating catchment areas based on Thiessen polygons, which ensure that the people who live in a particular airport's catchment area are closer to that airport than to all other airports. The second way to create catchment areas is based on 2.5 hours' driving distance. Geographic Information Systems (GIS) techniques were used to implement these two catchment area determination methods.

## 2. 3 Air passenger market segmentation

### 2.3.1 What is market segmentation and why is it important?

The concept of market segmentation was first proposed by Smith (1956), and has since been frequently used in market theory and practice. In historical literature, a number of definitions for market segmentation can be found, however, the most accurate definition to date is from Smith (1956, p. 6), who defined it in the following way: "*Market segmentation involves viewing a heterogeneous market as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desire of consumers for more precise satisfaction of their varying wants*". In general, it is the process of separating a market of consumers into a number of homogenous subsets or segments with similar characteristics and needs (Sarabia, 1996; Wedel and Kamakura, 2002; Kieu et al., 2018; Ahani et al., 2019) and a way to identify target markets by which more attractive and competitive marketing strategies can be developed (Cahill, 1997; Wen et al., 2008). Market segmentation is important to industries and policy-makers in assisting them to identify the core segments of consumers and to more accurately evaluate the importance of each core segment, and hence more appropriately meet customers' needs for services/products and resources (Dibb, 1998; Hollywood et al., 2007; Ekinci et al., 2018). Normally, as indicated by Freathy and O'Connell (2000), the general outcomes from market segmentation beneficial to industry and society are related to stabilized pricing, enterprise sustainability under competition and reducing competitors. Similarly, as Dibb (1998) emphasized, market segmentation is the keystone in modern business marketing and plays a central part of marketing strategy. It therefore can facilitate a narrowing of the gap in terms of unbalanced supply between customer demand and commercial resources.

Historically, there has been a great deal of segmentation practice applied in a wide range of fields including retailing markets (Quinn et al., 2007), electronic business (Wu and Chou, 2011; Kamthania et al., 2018), education (Angulo et al., 2010), tourism and movements (Xia et al., 2010; Sánchez-Fernández et al., 2019; Ahani et al., 2019), and transportation (Beirão and Cabral, 2008; Harrison et al., 2015; Ekinci et al., 2018). Generally, there are two major categories of segmentation. The first segmentation approach is to identify the segments on a macro level through generalized recognition based on the easily identifiable variables (Laughlin and Taylor, 1991; Rao and Wang, 1995; Freathy and O'Connell, 2012). Based on this, a large market can be disaggregated into a number of sub-segments (Hassan and Craft, 2005). Dissimilarly, the second segmentation approach is a process to aggregate the consumers into homogenous segments based on managerially relevant “micro-segments” like consumers’ choice-making style and other attitudinal responses (Rao and Wang, 1995; Barry and Weinstein, 2009; Kieu et al., 2018). These methods can be applied to identify the segments either before the research is undertaken, (using simple variables such as gender and age) – namely *a-priori* – or after the data and responses have been collected and analysed, (using the interrelated variables like psychographic attributes) – namely *post-hoc* (Green, 1977; Wind, 1978). Normally, demographic, behavioural, beneficial, psychographic and geographical variables are the five basic categories that can be used to identify the market segments (Reid and Bojanic, 2009).

Regional airlines are facing competition not only from other airlines but also from road transportation alternatives such as cars and coaches, or rail and ferries in other contexts. It is critical for the air carriers to develop more targeted and competitive marketing strategies to better satisfy the needs of passengers, and hence to increase patronage and potentially to reduce airfares. Thus, regional aviation market segmentation that can assist in identifying the core and target segments of customers with similar needs and characteristics could provide some valuable insights for helping the regional airlines customise their strategies.

### 2.3.2 Cluster analysis

Cluster analysis, (*post-hoc* method), is an efficient way to identify distinct market segments, where the cases/customers are homogenous and cohesive within each segment but heterogeneous across segments (Fraley and Raftery, 2002; Wedel and Kamakura, 2012; Jacques et al., 2013). It is a recognised market segmentation approach for analysing both categorical and numerical data (Norusis, 2010; Witten et al., 2016), that has been widely applied in a broad range of areas including data mining, signal processing, machine learning and some more practical applications, such as text summarisation and customer segmentation.

Hierarchical (Maimon and Rokach, 2005), partitional (Äyrämö and Kärkkäinen, 2006) and model-based (Melnykov and Maitra, 2010) clustering methods are the main streams of cluster analysis. The hierarchical clustering method breaks data into a nested sequence of clusters that are conceptually shaped in a tree structure, such as Clustering Using Representatives (CURE) (Guha et al., 1998) and CHAMELEON (George et al., 1999) algorithms. The main objective of this method is to generate a dendrogram that can clearly illustrate the hierarchical relationship between the clusters. Partitioning-based clustering methods simply classify a set of observations into a pre-defined number of clusters based on their similarity, each observation is uniquely assigned to only one cluster, such as the k-means (Lloyd, 1982) and partitioning around medoids (PAM) (Kaufman and Rdusseeun, 1987) methods. The algorithms of these two kinds of clustering methods are heuristic and relatively simple to use but, as there are no formal models underlying the algorithms, it is not possible to imply formal inference. However, model-based clustering is an appropriate alternative that endeavours to segment the data by optimising the fit between the given data and some mathematical model (Fraley and Raftery, 2002). Mixture modelling, also known as model-based clustering, has been extensively applied to identify the homogenous segments of given data sets (McLachlan and Chang, 2004; Lai et al., 2018).

Cluster analysis assumes that sample data originate from a distribution that is a mixture of finite components or sub-groups (clusters). Each cluster of data is therefore generated from a probability density function with a weight in the mixture. Typically, the probability density functions are usually assumed to have a multivariate normal/Gaussian distribution, as has been suggested by the majority of researchers, and can provide relatively good results in most instances (Witten et al., 2016; Tan, 2018). In the mixture model, each data observation has a weighted probability of belonging to each of the clusters. Expectation Maximization (EM) is an efficient clustering algorithm used for estimating the mixture model parameters, such as the mixture component weights and means, by optimising the corresponding local log-likelihood (Yeung et al., 2001; Witten et al., 2016). As indicated by many researchers (Neal and Hinton, 1998; Xu and Wunsch, 2005; Kishor and Venkateswarlu, 2016), EM can handle a variety of data and produce exceptionally good clustering results although it demands more time to execute. Therefore, this thesis will use EM as the mixture model clustering technique to identify segments for the regional aviation market in Western Australia.

### 2.3.3 Aviation market segmentation studies and limitations

Existing research has segmented the international aviation market through clustering analysis. Bruning et al. (1985) aimed to investigate which factors among demographics, socio-economic and trip characteristics can be used to divide national and regional air passengers into distinct markets. Mason and Gray (1995) generated a conceptual benefit segmentation approach to identify the market segments of international and domestic air passengers at Stansted airport in Britain. They found three distinct market segments and generalised some market strategy insights accordingly. Jacques et al. (2013) applied a multi-step segmentation framework to investigate the commuting trip market for ground public transit, private automobile and active travel mode, (walking and cycling), users surveyed at McGill campus in Canada and found four distinct market segments. They firstly identified mode-based clusters by a two-step cluster analysis that was then segmented by the partitional clustering method of k-means. Wen et al. (2008) used factor and k-means clustering analysis to explore the international aviation market for air passengers interviewed at Taoyuan International Airport of Taiwan and identified three representative market segments. Harrison et al. (2015) proposed a new partitioning-based market segmentation model by including the notion of passenger core values, such as time sensitivity and trip purpose, to identify groups of international air passengers at Brisbane International Airport. Kieu et al. (2018) modified classical Affinity Propagation algorithm for segmenting the large-scale transit passenger market with spatial (geodetic coordinates) and behavioural characteristics (such as Randomness of travel behaviour and the frequency of transport usage) in New South Wales (Australia), which has effectively identified a set of frequent and non-frequent passenger segments in both Northern Beaches and Inner South Sydney area. Lu et al. (2019) used multinomial function with the input variables of passengers' demographics (e.g., gender, age and travel experiences) or curiosity to the new service of air-bridge-air route to segment the passenger market from 11 pre-selected Chinese mainland cities that relatively far (more than 800 kilometers) from Hongkong, three distinct market segments were finally identified. Other studies have applied the mixture model-based segmentation approach to describe different types of tourists who visited Penguin Island, Victoria, Australia (Xia et al., 2010) and personal travel market (Hruschka et al., 2004). However, none of these researchers used a statistical model-based algorithm to apply the clustering analysis for a regional aviation market. Importantly, the literature review has identified a lack of aviation market research that considers non-airport respondents in market segmentation analysis. This is a significant omission as non-airport respondents are the people that airlines seek to reach by changing policies and services, since they may not frequently fly and hence represent latent air passenger demand. Therefore, the non-airport respondents have a significant value in understanding the potential aviation market.

Specifically, although some previous studies have explored the characteristics of international aviation markets through market segmentation approaches (Mason and Gray, 1995; Wen et al., 2008; Harrison et al., 2015), the area of market segmentation for exploring regional air travel markets remains under-researched. In contrast to international air travellers, regional air travellers may more easily use an alternative transport method in some contexts, such as a bus, train or car for their regional travel. Furthermore, as previously stated, most of the existing studies have focused on cluster analysis of the air passenger respondents interviewed in airports. This may be valuable for investigating the characteristics of the existing aviation market. However, there may be a proportion of air passenger respondents who use the airline occasionally for a particular reason e.g., emergencies, but normally use non-air transport more frequently. There may also be a significant component of non-air travellers with other characteristics who have a low or moderate preference for air transport for their regional trips, and there could be potential value in understanding their attitudes, both for the airlines and for policy-makers. As described in Chapter 7, a better characterisation of the entire market of air travel is required, covering the entire spectrum of likelihood to fly. Therefore, as described later in this thesis, a more holistic recruitment strategy than that of most previous studies was used, with surveys collected at both airports and in other places (e.g., visitor centres, libraries, parks, and shopping centres).

## 2. 4 Travel mode choice

### 2.4.1 What is travel mode choice?

Travel mode choice is central to many topics within the area of transportation planning and modelling, by which policy-makers can estimate the choices that travellers may make from among a set of available travel modes for a given trip or origin-destination-pair (Ben-Akiva and Bierlaire, 1999), as well as the number of travellers selecting the specified travel mode for the given trip (Leung and Lai, 2002). As indicated by Ben-Akiva and Bierlaire (1999), air travel choice and behaviour is a vital stage of travel demand analysis since individuals' mode choices can be aggregated to predict overall travel demand. However, the emphasis on travel mode choice has been on why an individual chooses a particular travel mode from among the transport modes (Pels et al., 2003; Van Can, 2013). A clearer understanding of this could assist airlines and government transport agencies to more precisely estimate the market share of a particular travel mode (Ben-Akiva and Bierlaire, 1999; Van Can, 2013; Inoue et al., 2015), as well as to investigate the internal competition within airlines and exogenous competition between air and non-air travel modes. Prior research suggests that travellers consider various factors when deciding on a travel mode for their trip, such as the travel characteristics (e.g., travel cost and time) and service quality factors (e.g., service frequency and seat comfort) (Pels et al., 2000; Theis et al., 2006; Jung and Yoo, 2014; Chen and Chao, 2015). Individual travellers may have different trade-offs across these factors based on their observed socio-demographics such as income and gender (Hess and Polak, 2006a; Balcombe et al., 2009), as well as unobserved characteristics. Consequently, for this competitive passenger market, it is crucial for the airlines to identify the choice criteria of individual travellers and the decision rule in making their travel mode choice (Pels et al., 2009; Chen and Chao, 2015). Understanding the reasons for choosing alternative modes of transport can therefore help policy-makers and airlines to formulate more appropriate market strategies in attracting passengers by meeting their mode choice criteria.

A large number of existing transportation studies in the literature concerned with travel mode and airline choice modelling attempted to more accurately estimate travellers' mode choice behaviours, as well as to identify the key factors that were statistically significant in affecting their decision making (Psaraki and Abacoumkin, 2002; Hess et al., 2007; Alhussein, 2011; Jung and Yoo, 2014; Wu et al., 2018). The nature of these travel mode and airline choice studies is a discrete choice modelling analysis that is based on the assumption of random utility theory that the individual traveller will always select the travel mode with the highest utility (Ben-Akiva et al., 1985; Anderson et al., 1992; Hensher et al., 2015a). As Ben-Akiva and Bierlaire (1999) proposed, the discrete choice modelling framework is based on four general assumptions, namely decision-maker, alternatives, attributes and decision rule. The decision maker is the individual who defines the entity of the decision making. Since the discrete choice model is regarded as a disaggregated model, the individual in this case can either be a single person or a group of people subject to particular applications. Alternatives are the finite available options that the individual can choose among and thus, based on the chosen and un-chosen alternatives, the individual's decision-making process can be investigated. Attributes are the numerical or nominal values that can measure the costs and benefits of an alternative to the individual that will be used to determine the alternative's utility. Decision rule is the procedure that the individual follows to balance the trade-off between attributes and hence to evaluate the provided alternatives and make the final decision. A concrete example of a discrete choice question for travel mode choice is shown in Figure 2-1, where an individual can choose the most appealing option from the four listed alternatives/options, described using five attributes and a set of different attribute-levels.

| Trip 1 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$150 | A$175 | A$350 | A$200 |
| Time to bus-station or airport-terminal | N/A | 15 mins | 30 mins | 15 mins |
| Journey or travel time | 12 hrs | 15 hrs | 2 hrs | 2 hrs |
| Service frequency (weekly) | Any time | 30 buses/coaches a week | 16 flights a week | 2 flights a week |
| Seat comfort level | High (leg room 90 cm) | Medium (leg room 80 cm) | Medium (leg room 80 cm) | Low (leg room 70 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

*Figure 2-1 Discrete choice question example of travel mode choice*

### 2.4.2 Travel mode choice models

There have been numerous studies investigating travel mode and airline choice using the discrete choice modelling approach (Hensher, 1994; Bolduc, 1999; Van Can, 2013; Wu et al., 2018; Wu et al., 2018; Peyhardi, 2020). To the author's knowledge based on the literature, probit, MNL, NL, cross-sectional NL, mixed logit (ML) and LC models have been used for modelling travel mode and airline choice. Probit models assume that the error terms, (or unobserved utilities), of utility functions follow a multinomial Gaussian distribution motivated by the Central Limit Theorem (Ben-Akiva and Bierlaire, 1999; Train, 2009). It has the ability to accommodate potential correlations among alternatives. In comparison to logit models, it may require more effort to compute the choice probabilities owing to the higher complexity of its formulation (Ben-Akiva and Bierlaire, 1999). However, logit models are more tractable and hence more applications and derivations have been developed, as the integral for choice probabilities of logit models has a closed form. It is the most common type of discrete choice model, which is derived from the assumption that the error terms of the utility functions are Gumbel distributed (Gumbel, 2012; Hensher et al., 2015a).

### 2.4.2.1 MNL and NL modelling

Logit models first arose in the field of binary choice modelling, where the MNL model is a generalisation of the binary model when it exceeds two alternatives (McFadden, 1973; Ben-Akiva et al., 1985; Ben-Akiva and Bierlaire, 1999). The MNL model is the simplest and most widely applied logit model in travel mode and airline choice studies with a closed-form (Horowitz, 1991; Train, 2009; Van Can, 2013; Jung and Yoo, 2014) that assumes the error terms of the alternatives' utilities are Independently and Identically Distributed (IID) and are drawn from a Type I Extreme Value distribution (Ben-Akiva and Bierlaire, 1999; Bolduc, 1999). However, the IID distribution assumption contributes to the property of Independence from Irrelevant Alternatives (IIA) that may be violated in practice in some cases, especially when the alternatives may have some similarities. To tackle this limitation, another logit model, namely the nested logit (NL) model, can be applied. This is an extension of the MNL model and was first introduced by Ben-Akiva (1973). It allows for the correlation of, or substitution amongst, similar alternatives by partially relaxing the IID assumption - the error terms of alternatives in the same nest can have some degree of correlation.

MNL and NL models are the most widely used methods for quantifying preferences around travel mode and airline choice (Hess, Adler, & Polak, 2007; Van Can, 2013; Jung & Yoo, 2014; Wu et al., 2018). Table 2-3 summarises twelve previous papers that used discrete logit and probit models to examine individuals' travel mode and airline choice behaviours regarding overseas, domestic, short-haul and urban travel in different study areas, including the key findings on which factors may affect their choice. Of these twelve papers, five applied the MNL model, seven the NL model and three the multinomial probit (MNP) model. The major factors these studies considered included travel time, travel cost, service frequency, seat comfort and accessibility. All of these studies found that travel cost and time have a statistically significant influence on passengers' travel mode choices; a higher travel cost or longer journey time significantly reducing the number of travellers that chose that particular transportation mode. Moreover, Hess and Polak (2006a), Jovicic and Hansen (2003), Jung and Yoo (2014), Van Can (2013) and Masoumi (2019) concluded that service frequency, extra cost for luggage, access time and comfort, (which could be affected by a range of parameters and factors), could affect passengers' travel mode choice. Additionally Bolduc (1999) identified that the socio-demographic variable of gender played a significant role in affecting mode choice, although Johansson et al. (2006) found that gender and age were not significant. Shen et al.

(2020) applied NL modelling approach and noted that income, occupation and education were all been found have a significantly positive impact on attracting young travellers to choose a certain travel mode. These papers used different types of survey data, focusing either on SP (e.g., Chang and Sun, 2012; Jung and Yoo, 2014; Inoue et al., 2015) or revealed preference (RP) data (Hess and Polak, 2006a; Wang et al., 2014), These data were collected in different countries all over the world, (e.g., Japan, China, South Korea, Denmark, Chile, Sweden and the United States). Thus, combining the key results, (as shown in Table 2-3 below), can provide a relatively general understanding about the influence of various travel attributes, but any generalisations drawn from the location specific results should be done with caution. However, few of these papers clearly indicated how they constructed the experimental designs in terms of orthogonal, efficient designs or even full factorial design, which is a tremendously important point that needs to be elaborated on before estimating the discrete choice models. This is because an efficient and realistic experimental design can help to increase the reliability of the final estimating results. Additionally, only a few of the papers that applied NL models, e.g., Wang et al. (2014) mentioned whether they had tried or tested other possible nesting structures in addition to the structure they used in the study. This thesis (Chapters 5 and 8), not only gives a relatively detailed illustration of how the efficient SP experimental design was constructed, but also the testing of other possible nesting structures.

*Table 2-3 Summary of previous mode choice studies*

| Authors | Study area | N[a] | O[b] | Model Used | Key findings – factors affecting travel mode and airline choice |
|---|---|---|---|---|---|
| (Inoue et al., 2015) | Domestic travel between a metropolitan area and four regional areas in Japan. | 1,500 | 3,000 | NL | Increased travel cost or time statistically significantly reduced the probability of business and non-business travellers choosing a travel mode. Increased service frequency statistically significantly increased the chance of mode choices of non-business travellers. |
| (Chang and Sun, 2012) | Oversea air travel between China and Taiwan. | 286 | 1,710 | MNL | Increased ticket fare, extra cost for luggage or additional transit airport statistically significantly reduced the probability of business and non-business passengers choosing the related travel mode, while the service frequency was found to be statistically insignificantly correlated with passengers' travel mode choices. |
| (Jung and Yoo, 2014) | Short-haul domestic travel in South Korea. | 3,834 | 3,834 | MNL&NL | Increased ticket fare, access time and journey time statistically significantly reduced the probability of business and non-business passengers selecting the related travel mode, but the service frequency was found to be statistically insignificant. |
| (Wang et al., 2014) | Domestic travel on three regional routes in China. | 2,821 | 2,039 | MNL&NL | Increased travel time and trip costs statistically significantly reduced the probability of business and non-business passengers using the |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | corresponding travel mode. |
| (Van Can, 2013) | Domestic travel to Nha Trang in Vietnam. | 402 | 1,206 | MNP | Increased travel cost or travel time statistically significantly reduced sales. Increased comfort level, safety level price against quality level or punctuality level statistically significantly increased the sales. |
| (Qiao et al., 2016) | Domestic travel between Chengdu and Longquan in China. | 955 | 2,145 | NL | Increased arrival time, off-vehicle time, in-vehicle time or travel cost statistically significantly reduced the probability of passengers choosing the relevant travel mode. |
| (Jovicic and Hansen, 2003) | Short-haul travel within the Copenhagen area in Denmark. | 1,460 | 19,989 | NL | Increased travel cost, access time, waiting time, in-vehicle time statistically significantly reduced number of commuters, leisure, education, and business travellers. |
| (Hess and Polak, 2006a) | Air travel from the San Francisco area. | 5,091 | 5,091 | MNL&NL | Increased flight frequency has a statistically significant positive effect in attracting people to use airlines, while longer in-vehicle time statistically significantly reduced the probability of passengers choosing the related travel mode. |
| (Bolduc, 1999) | Urban travel for working trips within Santiago of Chile | 1,299 | 6,731 | MNP | Increased travel cost, journey time, time to access to and waiting for the transport led to a statistically significant decrease in choosing the particular transport mode. Demographic variable of gender was been found have a significant impact on influence mode choice of working trips. |
| (Johansson et al., 2006) | Commuting travel within Stockholm and Uppsala, Sweden | 1,708 | 1,708 | MNP | Increased travel cost, journey time statistically significantly decreased the probability of using corresponding transport mode, while individuals' preferences for flexibility and comfort are also important and safety preferences are insignificant. |
| (Masoumi, 2019) | Urban travel mode choice in the Middle East and North Africa | 8,284 | 8,284 | MNL | Long walking distance, cultural problems, lack of biking infrastructures, personal preference towards car travel diminished the proportion of travellers from walking, cycling and using public travel modes. Improve service quality of comfort and convenience will attract passengers to use the certain travel mode (car or public transport). |
| (Shen et al., 2020) | Young people's travel mode choice in Nanjing, China | 349 | 2,792 | NL | Increased in-vehicle travel time, travel cost, packing cost and waiting time led to a statistically significant decrease in choosing the certain transport mode, while improving bus or tube convenience had significantly increased the proportion of young travellers to use particular transport mode. Demographic variables of income, occupation and education were been found have a significantly positive impact on attracting young travellers to choose a certain travel mode, especially for E-hailing and conventional car-based travel modes. |

*Table notes:*

*[a] N is the sample size, [b] O is the number of observations*

#### 2.4.2.2 Latent Class discrete choice modelling

In the last 20 years, some studies have used more advanced models instead of the MNL model for investigating travellers' mode and airline choice behaviours. These relaxed the IIA assumption and accommodated preference heterogeneity to some extent, such as the cross-sectional NL model (Jovicic and Hansen, 2003; Hess and Polak, 2006b; Inoue et al., 2015; Wu et al., 2018) and the ML model (Espino et al., 2008; Lee et al., 2016; Qiao et al., 2016; Mehdizadeh et al., 2018; Monchambert, 2020). The ML model is a common approach with relatively greater flexibility in accommodating unobserved heterogeneity of preference by assuming the model parameters lie within a specified continuum and random distribution (Revelt and Train, 1998; McFadden and Train, 2000; Hensher and Greene, 2003). One challenge of the ML model is that it requires the analyst to pre-specify the distribution of preferences across the population and the model is not efficient in illustrating the heterogeneity among sources (Lee et al., 2003; Monchambert, 2020), where the sources may be correlated with individuals' characteristics (Boxall and Adamowicz, 1999). However, an LC model is able to provide sufficient accommodation of preference taste heterogeneity by performing as a semi-parametric version of an MNL model (Greene and Hensher, 2003)[2]. It actually assumes a discrete distribution of parameters across individuals to account for the heterogeneity in a population and, hence, there is no need to know and specify a distribution of parameters, which is an advantage compared to an ML model. The LC approach has been widely applied in market segmentation research, as it assumes and identifies a discrete number of latent segments or classes, (e.g., passenger segments), where individuals' preferences are homogenous within the segment but heterogeneous across the segments (Boxall and Adamowicz, 2002; Shen, 2009). In the LC model, EM algorithms have been widely and popularly used to identify the latent segments. The membership probability of each of the classes is estimated by finite iterations of expectation and maximization routines with weighted log likelihood, until convergence occurs, as well as the model parameters corresponding to each LC. The membership of each class represents a latent component where each individual will have a probability of being in each LC, which thus can be considered as a form of soft clustering. However, the analyst needs to predefine the number of latent classes, which could be a critical issue in the application of the model (Román et al., 2017).

[2] Also named as LC-MNL model, but to simplify, this thesis will use LC model instead of LC-MNL Model.

Greene and Hensher (2003) compared the ML model with the LC model in a case study estimating road type choice for long-term travel, where they indicated that both models have their merits and shortcomings. The paper showed that the LC model is an efficient method when researchers are not sure about the distribution of taste heterogeneity across individuals, which also provides an intuitive interpretation for policymakers and investigators. On the other hand, Hess et al. (2009) applied MNL, ML and LC models to analyse the SP data of travel mode and departure time. They found that the LC model results were an improvement relative to the MNL model and the improvement in log likelihood was comparable to the gains through the ML model. They also proposed that the LC model significantly outperformed the ML model in terms of the interpretability of the results, as the correlation between taste heterogeneity and demographic indicators could be more easily interpreted in the LC model. Thus, the accommodation of unobserved heterogeneity can be further improved, as well as the choice estimation. In line with this, the previous studies of freight transport choice conducted by Massiani et al. (2007) and Greene and Hensher (2013), used the LC model to account for discrete segments of unobserved heterogeneity across respondents. They found that the LC model not only outperformed in terms of model fit but also generated more reliable and significant parameter estimates when compared with either the basic MNL model or the ML model. There is an increasing number of studies applying the LC model in stated and/or observed real preference experiments in a broad range of disciplines including transportation (Román et al., 2017), urban development (Jiao et al., 2015), accident analysis (Cerwick et al., 2014) and health (Greene et al., 2014), which confirms the efficiency of the LC model in accommodating unobserved heterogeneity of respondents' preferences, as well as checking for attribute non-attendance (Román et al., 2017).

The LC model framework offers an effective segmentation approach in capturing individuals' unobserved preference heterogeneity for travel mode and airline choice modelling. Bhat (1997) used an LC model with endogenous segmentation to estimate Canadian travellers' intercity mode choice preferences among car, train and airline. The data were collected in airports and road stations within the Toronto-Montreal corridor. Teichert et al. (2008) used an LC modelling approach to capture individuals' preference heterogeneity for investigating frequent air flyers' flight choice behaviours for European short-haul trips, and found the LC model performed significantly better than the priori segmentation-based MNL model. Shen (2009) applied an LC model to accommodate individuals' unobserved heterogeneity in preferences for their road transport mode choice including cars, buses and trains in metropolitan Japan, focusing specifically on the Osaka region. Wen and Lai (2010) applied the LC model to investigate air passengers' international airline choice, and found that the LC model outperformed the standard MNL model as it properly accommodated travellers' preference heterogeneity through the LC model-based market segmentation approach. Vij et al. (2013) used an LC framework to evaluate individuals' mode choices, (auto, public transit, bicycle and walk), for work and non-work trips in Karlsruhe, Germany, with the taste heterogeneity across individual's modality styles accounted for. Additionally, Molesworth and Koo (2016) introduced an LC modelling approach to examine individuals' interstate flight choices between piloted and remotely piloted aircraft in Australia, whereas Seelhorst and Liu (2015) employed the LC models to accommodate the unobserved preference heterogeneity for investigating the effects of Frequent Flyer Program membership on airline choice in America. However, as few of these studies used such methodology to investigate travel mode and airline choice in regional areas and none of them considered non-air passenger respondents (represented by non-airport respondents) in the analysis, generalisability to the broader population is likely to be difficult.  In this thesis, the literature is extended by applying the LC modelling method to explore and compare travel mode and airline choice preferences for regional travel within and among different market segments, using the survey data collected not only in the regional airports but also in other locations and residential areas where respondents may not use airlines frequently.

### 2.4.3 Travel mode choice studies and limitations

As mentioned previously, many studies have investigated passenger travel mode and airline choice between cities, states and or countries (e.g. Hess et al., 2007; Lapparent et al., 2009; Chang and Sun, 2012). However, limited research has applied discrete choice modelling to estimate individual mode choice behaviour within regional Western Australia. A further limitation of the existing literature around flight travel preferences is that much of it is based on respondents recruited at airports and/or train stations, which may be valuable for certain questions but is subject to selection bias and may not generalise to the rest of the population (e.g., Hess and Polak, 2006a; Hess et al., 2007; Koo et al., 2010; Van Can, 2013; Jung and Yoo, 2014).

To extend the travel mode and airline choice modelling literature and help to fill the gaps, this thesis collected SP survey data from respondents interviewed at both regional airports and at other settings likely to involve those who do not frequently choose to fly. In this thesis, applied MNL and NL models are used to estimate travel mode and airline choice for regional travel in Western Australia. In addition to this, to accommodate unobserved preference heterogeneity among individuals, an LC model was deliberately introduced as a flexible semi-parametric variant of the standard multinomial logit choice modelling approach. It was used to investigate regional travellers' travel mode and airline choice behaviours within each of the latent segments, as well as the general demographic, economic and trip characteristics of individuals in each segment. Therefore, individuals' travel mode and airline choice preferences can be further and more completely evaluated. The State of Western Australia is in the unusual position of having a significant majority of its population in a metropolitan hub, (Perth), with the remaining population very widely dispersed and, thus, needs appropriate transport options. This geographical pattern means that the findings may generalise, (with some caveats), to regions with similar characteristics elsewhere, whereas the methodology of statistical and spatial modelling analysis implemented in Western Australia could also be applied to other regions.

## 2. 5 Summary

This chapter has provided a general background of the aviation market in regional Western Australia including market status, air service significances and the existing key aviation problems. In order to assist the government transport agencies and regional airlines tackle these aviation issues and improve air passenger movements, it subsequently reviewed the literature on spatial and statistical analysis related to the investigation of aviation markets. Specifically, research on air travel demand, aviation market segmentation and travel mode choice modelling were discussed, including the definitions, significances, key methods and potential predictors, as well as the limitations.

To fill the identified research gaps, as well as to investigate the regional aviation markets and competitions, the next chapter will develop a theoretical methodology and framework relating to the statistical modelling of travel demand, market segmentation and travel mode and airline choice.

# CHAPTER 3 RESEARCH FRAMEWORK

## 3. 1 Introduction

The previous chapter identified a number of research gaps related to regional air travel demand modelling, market segmentation and travel mode and airline choice estimation. Further, it explained the value of establishing a robust methodology to more comprehensively understand the regional aviation market, both for the government and industry. This chapter provides an overview of the study area and the research framework developed to address these issues.

## 3. 2 Study area

The study area of this research is Perth city and other regional centres/towns in Western Australia with RPT airports. Perth is the metropolitan hub and the state capital city. In 2016/2017, the state had 22 regional airports providing RPT air services that were available to the public. The state of Western Australia occupies almost a third of Australia's landmass, which makes it the largest state in the country, and one of the largest, least populated and most isolated worldwide. According to the Australian Bureau of Statistics (2016), the population of Western Australia in 2016 was about 2.5 million. While the majority of the population is found in Perth, there are a number of towns widely dispersed across the state. By virtue of the state's large size, and the long distances between these towns, air travel plays an important part in connecting regional towns and the hub city. Following discussions with the Department of Transport Western Australia, four regional towns and their airports, (Albany, Geraldton, Broome and Karratha) were selected as the key study areas of this research, in order to explore the regional aviation market, as well as to estimate the travel mode and airline choice in regional Western Australia. The locations of these four regional airports, (and Perth), are shown on Figure 3-1. The reasons to choose these four towns are outlined below.

Albany is the oldest colonial settlement in Western Australia and is one of the most popular retirement and tourist areas for Western Australia. It is located on Western Australian southern coast, 414 km south of Perth, and has a population of 33,145 (2016 Census). Albany airport is 11 km from the Albany Central Business District. Rex Airlines operates daily flights between Perth and Albany (Regional Airport, 2018). Geraldton is located in the Midwest region of Western Australia, 424 km north of Perth, with a local population of 37,432 (2016 Census). It is a coastal city well-known for tourism, fishing, mining, wheat, sheep and minerals. Geraldton airport is 10 km from Geraldton town centre. Virgin and Qantas Aviation operate daily flights between Geraldton and Perth, and the airport also provides other services including General Aviation Charter flights, Royal Flying Doctor Services, RAAF deployments, and fly in fly out mining services to Regional Western Australia (Geraldton Airport).

Karratha is an important mining town, located in the Pilbara region, 1,535 km north of Perth with a local population of 15,828 (2016 Census). Virgin and Qantas operate daily flights between Karratha and Perth (Karatha Airport, 2018). Broome is an international tourism town, centre for regional services and natural resources, and a minor aviation hub in Western Australia located in the Kimberley region, 2,240 km north of Perth. It has a local population of 13,984 (2016 Census), which can grow to over 45,000 during the peak tourist season between June and August. Qantas, Airnorth, Virgin Australia and Slippers Aviation operate RPT daily air services from Broome to Perth, Kununurra/ Darwin, Fitzroy Crossing and Halls Creek (Broome International Airport, 2018).

There are two main reasons for selecting these four towns. Firstly, these towns are not only the major regional towns in Western Australia but also cover a variety of geographical locations in the state, involving different levels of isolation. Karratha and Broome lie far to the north of Perth while Albany and Geraldton are much closer, as shown in Figure 3-1. This range of locations allows us to explore whether and if so, how well, distance correlates with travelling behaviours and preferences. Secondly, they also epitomise (cover) the various key socio-economic characteristics of the state, in particular the mining industry, tourism and general business such as agriculture, animal husbandry and manufacturing.

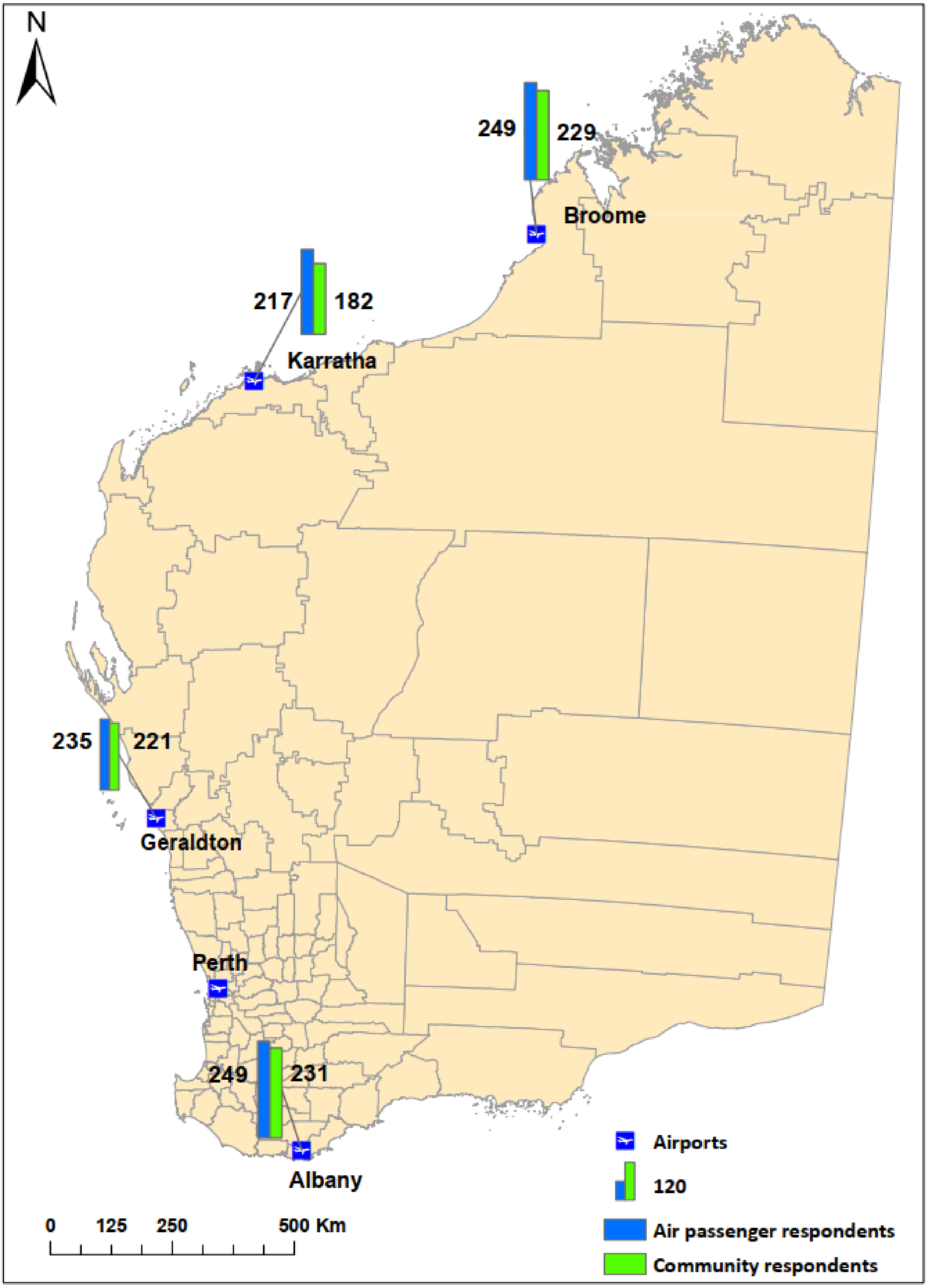



*Figure 3-1 Study area of Western Australia*

## 3. 3 Research workflow

The methodology of this research for predicting regional air travel demand, exploring the regional aviation market and estimating travel mode and airline choice and behaviours is comprised of four main steps: 1) Research data collection and manipulation, 2) Air travel demand forecasting, 3) Exploratory data analysis for exploring the regional aviation market and 4) Travel mode and airline discrete choice modelling.

The four main steps of the methodology are illustrated in sequence in Figure 3-2, with the corresponding descriptions given below.

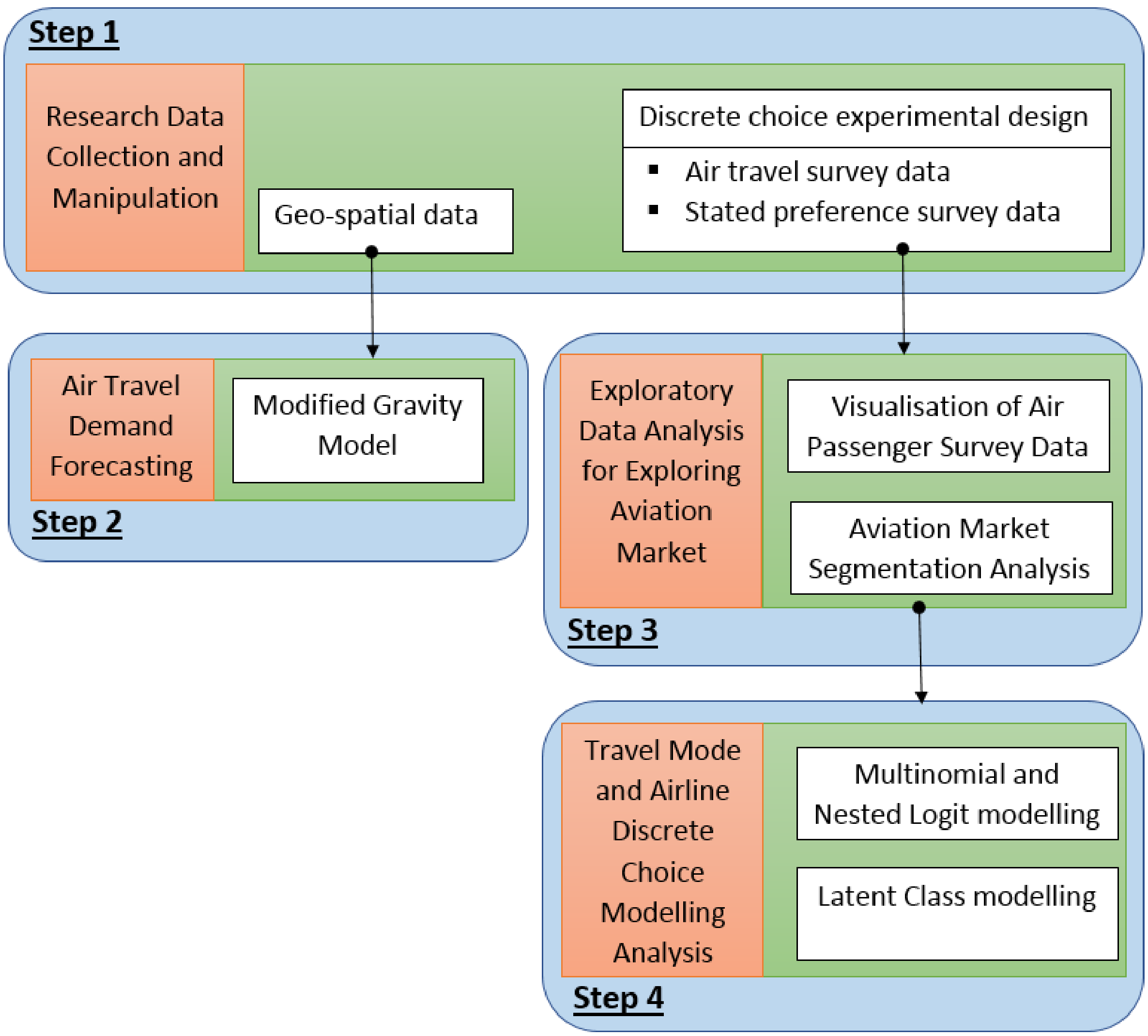



*Figure 3-2 Research workflow of the thesis*

## 3. 4 Research data collection

The data collected to conduct the research can be classified into two categories:

- geo-spatial data; and
- air travel survey and SP survey data.

The geo-spatial data were mainly used to visualise and forecast the regional air travel demand across the RPT air routes in regional Western Australia. The air travel and SP survey data were used to explore the regional aviation market and to identify and estimate the key drivers that significantly affect regional passengers' travel mode and airline choice.

### 3.4.1 Geo-spatial data collection

The geo-spatial data collected are categorised into two types; flight data and geo-economic data. Flight data include airfares and real time bilateral flight information, (flight number, travel time, aircraft type and passenger seat number), between RPT airport-pairs in regional Western Australia, which were collected from multiple sources. Airfare pricing is dynamic and complex and can be influenced by many factors such as the booking time, ticket surplus and fare classes such as the economy and business classes. Economy class airfare data are used in this research, since the number of business class seats are normally quite limited compared to that of economy class seats, especially for short-term regional air flights. However, economy class can still have different fare types, such as the restricted, (less flexibility to change, refund or cancel the booked flights), and fully flexible, (full flexibility to change, refund or cancel the booked flights). In order to minimise these impacts and for simplicity, only fully flexible economy class airfares were used and with prices obtained four months in advance of the assumed travel date, (i.e. the airfares of the flights by airlines on each RPT air route for travel date of 30th October 2017 were collected on 30th June 2017). The airfare data were collected from the official websites of the operating airlines, (Virgin Australia Airlines, Qantas Airlines, Regional Express Airlines, Skippers Aviation and Airnorth Airlines). The bilateral flight information was extracted from the *Flightradar24* and *Plane Finder* websites that provide global real time flight tracking services and information. Geo-economic data were obtained from the Australian Bureau of Statistics (ABS) online database (Australian Bureau of Statistics, 2016), the Department of Mines and Petroleum (DMP) online database (Department of Mines and Petroleum, 2017), Tourism Research Australia's (TRA) online database (Tourism Research Australia, 2017) and Main Roads Western Australia (provide by the government). These data include statistical boundaries of local government areas (LGAs) in Western Australia, geographical distances between the airport-pairs, population, average per capita income, number of operating mine sites, tourist population and the road network with speed limits. A detailed summary of the geo-spatial data collection sources and time durations is provided in Table 3-1.

*Table 3-1 Summary table of geo-spatial data collection*

| **Geo-spatial data** | | **Source** | **Time duration** |
|---|---|---|---|
| **Flight data** | Flight information | Extracted from: https://www.flightradar24.com & https://planefinder.net | 01/07/2016 – 30/09/2016 & 01/01/2017 – 30/06/2017 |
| | Full service airfare | Collected from: https://www.virginaustralia.com; https://www.qantas.com; www.rex.com.au; https://www.skippers.com.au & https://www.airnorth.com.au | 30/06/2017 |
| **Geo-economic data** | Statistical boundaries | Download from ABS (http://www.abs.gov.au) | 2016 |
| | Distance | Download from ABS (http://www.abs.gov.au) | 2016 |
| | Population | Download from ABS (http://www.abs.gov.au) | 2016 |
| | Average Income | Download from ABS (http://www.abs.gov.au) | 2016 |
| | Operating mine sites | Provided by DMP | - |
| | Tourist number | Provided by TRA | 2016 |
| | Road network | Provided by Main Roads Western Australia | - |

### 3.4.2 Air travel and stated preference survey data collection

Field surveys, including air travel questionnaires and an SP survey, were conducted to understand regional air passenger characteristics and to identify the key factors affecting travel mode and airline choice. Specifically, the air travel survey questionnaire was designed to obtain information on air passenger trips, including trip origin and destination, access mode to the airport, travel group, reasons to select air travel mode, trip purpose, travel cost and travel frequency (see Appendix D1). The SP survey was designed to examine the regional air and non-air traveller’s decisions about travel mode choice alternatives, including car, bus and regional airlines. The questionnaire was constructed using the modified D-efficient SP experimental design technique, which is fully described in Chapter 5. In the SP survey, each respondent was presented with six hypothetical travel mode choice questions, with each question containing four travel mode options/alternatives, (car, bus and two regional airlines), described by several attributes, (e.g., travel cost, journey time and seat comfort). The respondent was asked to select a preferred option for each choice question. Additionally, there were several general survey questions to obtain the respondent's socio-demographic information, (e.g., gender, age and education background). The questionnaires are shown in Appendix D with more details of these surveys presented in Chapter 5. In total, fifty pilot surveys were undertaken in Perth to refine and revise the survey questionnaires prior to conducting the surveys across the four sites.

The field survey data were divided into two parts, based on whether the respondent was interviewed at an airport or at non-airport locations. For the airport respondents, (i.e. those interviewed at an airport departure lounge), both air travel information and SP data were collected. For the non-airport respondents, (i.e. those intercepted at other locations in the town), SP survey data alone were collected. In Figure 3-1, the blue bars represent the number of airport/air passenger respondents surveyed in the related airports and the green bars show the number of non-airport/community respondents from the airports' towns.

**a) Airport respondent sample**

The surveys filled out by the air passenger respondents, (represented by airport respondents), were conducted in the departure lounges of the airports at the four selected towns (Albany, Geraldton, Broome and Karratha). A total of 950 airport respondents completed the surveys, as summarised in Table 3-2. At the start of the survey, the researchers asked the respondents what their current trip purposes (business/work or personal) were and then handed them an SP survey questionnaire relevant to their trip purpose. A total of 621 business airport respondents answered the air travel survey and a set of hypothetical SP travel mode choice questions assuming they were travelling for business, while the remaining 329 non-business airport respondents completed the air travel survey and the SP survey developed for non-business travellers (See Appendix D for samples of both types of questionnaire).

*Table 3-2 Airport respondent data collection*

| Airport respondents | Business group | Non-business group | Time period |
|---|---|---|---|
| Albany regional airport | 167 | 82 | 21/05/18 – 28/05/18 |
| Geraldton regional airport | 170 | 65 | 19/06/18 – 01/07/18 |
| Broome regional airport | 104 | 145 | 30/07/18 – 08/08/18 |
| Karratha regional airport | 180 | 37 | 10/08/18 – 18/08/18 |
| Total | 621 | 329 | |

#### b) Non-airport respondent sample

Non-air passenger respondent (represented by non-airport/community respondents) SP survey data were collected in Albany, Geraldton, Broome and Karratha at a number of locations, including public libraries, town streets, shopping centres, regional colleges and town parks. A total of 863 non-airport respondents completed the survey, distributed as shown in Table 3-3. Similarly, before starting the SP survey, respondents were asked their usual trip purposes and how they preferred to travel, and then were given the appropriate questionnaire.

*Table 3-3 Non-airport respondents data collection*

| Non-airport respondents | Business group | Non-business group | Time period |
|---|---|---|---|
| Albany town | 94 | 137 | 21/05/18 – 28/05/18 |
| Geraldton town | 92 | 137 | 19/06/18 – 01/07/18 |
| Broome town | 37 | 184 | 30/07/18 – 08/08/18 |
| Karratha town | 42 | 140 | 10/08/18 – 18/08/18 |
| Total | 265 | 598 | |

## 3. 5 Air travel demand forecasting

Air travel demand is an important indicator for airlines to improve service quality and reduce airfares. However, the research that provides accurate forecasts for regional passenger movements is limited (refer to section 2.2.3). Further, these studies did not incorporate the impacts of airport catchment areas but tended to use administrative boundaries without considering that the spatial distribution of the airports may reduce the accuracy of the demand forecast.

Therefore, modified gravity models were developed to forecast the bilateral air passenger seat numbers on RPT air routes in Western Australia. Two kinds of airport catchment determination methods were used in this study to incorporate catchment effects and therefore to improve reliability of the demand forecast. The four steps involved in the air travel demand modelling analysis are listed below, with the methodology implementation details demonstrated in sections 4.3 and 4.4. One noteworthy point is that, since the domestic air passenger historical data are sensitive data and not available publicly, total number of seats of the domestic flights carried (air passenger seat numbers), were used as a proxy to represent air travel demand.

**Step 1: Developing a modified gravity model**

As described in Chapter 2, gravity models are efficient tools that have been widely used to investigate spatial structure and interaction. Therefore, the first step in estimating air travel demand was to design and modify the gravity model, to meet the modelling requirements of the present research. Initial factors that may affect air travel demand in the gravity model were:

- Predefined catchment areas for the regional airports;
- Populations of the catchment areas;
- Average incomes of the population within the catchment areas
- Numbers of operating mine sites in the catchment areas;
- Tourist populations in the catchment areas during the data collection period;
- Fully flexible economic class airfare for each RPT air route; and
- Average distances between airport-pairs.

**Step 2: Defining catchment areas**

In the case of Western Australia, the catchment area of each regional airport incorporated into the modified gravity model was determined based on two different criteria. The first criterion was that all locations within a particular airport's catchment area should be closer to that airport than to any other airport. This criterion was met by developing a Thiessen polygon-based catchment area for each airport using ArcGIS 10.2.2 software. In contrast to the first criterion, the second criterion was that all locations inside the catchment area should be within 2.5 hours driving distance of the airport. The travel time-based threshold can be set based on a variety of factors (e.g., road network or spacing/proximity of adjacent airports) and subject to the study area and objectives (see section

4.3.2). The Network Analyst function of ArcGIS 10.2.2 software was used to generate the 2.5 hours driving distance catchment areas. Thus, two separate, and in some cases quite different, catchment areas were developed for each airport.

**Step 3: Manipulating data**

Once the airports' catchment areas were defined and generated, the next step was to prepare and transform the data for the gravity modelling. Geo-spatial data (as mentioned in section 3.4.1) were used for this study purpose. However, the geo-economic data, such as the population and income, were LGA boundary-based data. The created catchment areas normally contained one or more whole LGAs as well as parts of adjacent LGAs. Therefore, an area-weighted average method was applied to convert the LGA boundary-based data into the corresponding catchment areas using ArcGIS 10.2.2 software.

**Step 4: Air passenger seat number estimation**

As the data relating to each of the initial factors were manipulated, Pearson Correlation tests were applied to check the multicollinearity among these factors. Based on the multicollinearity results, the gravity model was modified, and four sub-models were derived, where each sub-model contained a set of different uncorrelated factors. Finally, these modified gravity models were estimated using the Poisson regression estimator provided in *R*, a programming language (R Core Team, 2017), and, thus, the air passenger seat numbers on RPT air routes in Western Australia were estimated.

## 3. 6 Exploratory data analysis for exploring aviation market

Exploratory analysis aims to understand the regional air passengers' characteristics such as socio-demographics and trip characteristics, and to identify the existing and potential regional aviation markets. In the next section, the exploratory data analysis conducted in this thesis is discussed.

### 3.6.1 Visualisation and exploration of air passenger data

Although a few studies have explored domestic and international air passenger characteristics in Australia, little to no attention has been given to the exploration of regional air passenger characteristics in Western Australia. This section provides a way to explore and compare regional air passenger profiles and trip related characteristics based on the survey data collected in the four selected regional centres (see as Chapter 6).

**Step 1: Collecting and collating the data**

As indicated in section 3.4.2, air passenger survey data were collected using a paper-based questionnaire (Appendix D), with air passenger respondents randomly intercepted in the four regional airport departure lounges (within the four selected regional towns). In total, 902 respondents completed the surveys. *Python* programming software (Van Rossum and Drake Jr, 1995) was employed to transform and format the data.

**Step 2: Visualise data for exploring and comparing air passengers' characteristics**

Once the air passenger survey (collected at the regional airports), database was created from the previous step, the next step was to explore and compare the air passenger characteristics related to:

- Air passenger profiles (e.g., age, gender, income and education background);
- Trip origin and destination;
- Access mode to airport;
- Travel group;
- Reason to choose air travel;
- Rank of the travel-related factors affecting people's travel mode choice;
- Trip purpose;
- Travel cost (one-way);
- Travel frequency; and
- Booking time preference in advance of the flight.

*Python* language was used to interrogate the air passenger database and create appropriate graphs for initially identifying the air passenger characteristics. Different categories of graphs including pie charts, word cloud image, bar and stacked bar charts were used to visualise the data, subject to the data types. In addition to this, cross tables were also created to examine whether the air passengers' personal characteristics (e.g., trip purpose) could affect their travel mode behaviours.

### 3.6.2 Aviation market segmentation analysis

As described in Chapter 2, several researchers have applied the market segmentation approach to explore international aviation markets. However, few such analyses have been conducted for regional aviation markets, which can be quite different. Furthermore, these studies mostly identified the market segments based on surveying airport respondents, which is valuable for describing the existing aviation market, but is incomplete when used to evaluate the potential aviation market. Extending the segmentation analysis to include non-air travellers, (community respondents), could contribute to understanding the potential aviation markets. Thus, based on the airport and non-airport respondents' SP data collected from the intercept surveys at the four selected regional towns, this section develops a more holistic methodology to identify and investigate the existing and potential aviation markets in regional Western Australia, using the mixture model-based market segmentation approach (Chapter 7). Three steps were involved, which are described briefly here, and more fully in section 7.3:

**Step 1: Determination of variables used for segmenting aviation markets**

The first step in applying a mixture modelling analysis to identify aviation market segments is to determine the factors that will be used for segmenting the markets. These factors cover socio-demographics, trip characteristics and stated probabilities/preferences and are:

- Gender (Male; female);
- Age (under 25; 25 to 44; 44 or more);
- Education background (Basic education; tertiary education);
- Income level (Low income; middle income; high income);
- Trip purpose (Business purpose; non-business/personal purpose) and
- Individual-specific stated probabilities for choosing air (airlines) and non-air travel (car and bus) modes.

All the factors are categorical variables with two or three attribute categories/levels, except the stated probabilities, which are numerical values specific to each of the respondents.

**Step 2: Generation of aviation market segments**

The mixture model-based market segmentation approach is one of the main streams in clustering analysis that handles both categorical and numerical data. It assumes that the sample data, (air and non-air traveller samples), originate from a distribution that is a mixture of finite clusters, (also called components or segments). Furthermore, there is an underlying probability density function for each cluster. The cluster probability densities contribute to the mixture, but with a corresponding weight. The weight is the cluster/segment size with the sum of the weights equal to 1. Expectation Maximization (EM) is an effective algorithm that has been extensively used for estimating the mixture model parameters by optimising the criterion of log-likelihood and, thus, the market segments can be identified. In this stage, the EM algorithm iteratively runs the expectation (E) and maximisation (M) steps to compute the posterior probability of a respondent belonging to each of the clusters, based on his/her attributes, (i.e., gender, trip purpose and stated preferences), that were then used to re-estimate and finalise the mixture model parameters. Thus, the distinct market segments of airport and non-airport respondent samples were both identified.

**Step 3: Identification of existing and potential aviation markets**

Once the market segments of the airport and non-airport passengers were respectively identified from the estimated mixture models, the next step was to classify these market segments for eliciting the existing and potential aviation markets based on the mean stated preference for/probability of selecting the air travel mode. The segments with a significantly high preference for air transport were assigned to the existing aviation markets, while the segments with a low to moderate preference for air travel mode were classified as the potential aviation market. Therefore, it would provide a more comprehensive insight for policymakers and airlines into both the existing and potential aviation markets, by examining the dominant characteristics of those market segments. Of paramount interest are the characteristics of the potential aviation market that represents a high potential value to the aviation industries. The airlines could focus on advertising targeting these travellers who might be more easily attracted to use air transport, which could further increase airline patronage and therefore improve air service quality and reduce airfares.

### 3. 7 Travel mode and airline discrete choice modelling analysis

In the field of transportation, discrete choice analysis has been extensively applied to the investigation of travel mode and airline choice behaviours. One major reason for employing discrete choice techniques is the assumption that regional travellers have the capacity and knowledge to identify and select the alternative/option with the highest utility, which is consistent with the discrete choice modelling assumption.

Discrete choice models are used to estimate the probability of a specific choice being made by the decision maker, (in this case the travellers), from a finite set of alternatives, with the underlying theory assuming that his/her choice is determined by the order of preference scale over the alternatives, which could be represented by a utility function (Anderson et al., 1992; Hensher et al., 2015a). Specifically, for travel mode choice, given possible $J$ alternatives to an individual $n$ and the utility provided by alternative $j$ ($j \in J$) to the individual $n$ is $U_{nj}$. then individual $n$ will choose the alternative $j$ that provides the largest utility, $U_{nj} > U_{ni}, j \neq i, i=1,2,\ldots,J$. However, the total utility $U_{nj}$ is known to the individual/choice-maker but not to the analyst. The analyst can only know the observed utility component ($V_{nj}$) that is determined by the alternative related attributes, (e.g., ticket fare or journey time of a particular travel mode in the mode choice analysis), namely the observed utility. Thus, the total utility $U_{nj}$ is comprised of two components:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \qquad 3\text{-}1)$$

where $V_{nj}$ is the observed utility; and $\varepsilon_{nj}$ is the unobserved utility (or error term) that accounts for the component that is not included in the $V_{nj}$ and is not known to the researcher.

The density of the unobserved utility is noted as $f(\varepsilon_n)$. The probability that an individual $n$ chooses alternative $j$ is (Train, 2009):

$$\begin{aligned}\pi_{nj} &= prob(U_{nj} > U_{ni}, \forall i \in i = 1,...,I,\ j \neq i) \\ &= prob(V_{nj} + \varepsilon_{nj} > V_{ni} + \varepsilon_{ni}, \forall i \in i = 1,...,I,\ j \neq i) \\ &= prob(\varepsilon_{ni} - \varepsilon_{nj} < V_{nj} - V_{ni} +, \forall i \in i = 1,...,I,\ j \neq i)\end{aligned} \tag{3-2}$$

This probability is a cumulative distribution function integrated based on the density $f(\varepsilon_n)$ that can be rewritten as (Train, 2009):

$$\begin{aligned}P_{nj} &= \int_{\varepsilon}^{+\infty} I(\varepsilon_{ni} - \varepsilon_{nj} < V_{nj} - V_{ni} +, \forall i \in i = 1,...,I,\ j \neq i) f(\varepsilon_n) d\varepsilon_n \\ &= \int_{\varepsilon}^{+\infty} I(\varepsilon_{ni} < \varepsilon_{nj} + V_{nj} - V_{ni} +, \forall i \in i = 1,...,I,\ j \neq i) f(\varepsilon_n) d\varepsilon_n\end{aligned} \tag{3-3}$$

Where the $I(\cdot)$ is the indicator function, that will be 1 if alternative $j$ has been chosen (term in parentheses is true), else will be 0.

Logit models, (i.e., MNL model), are the most common category of discrete choice model. They assume that the unobserved utility $\varepsilon_{nj}$ is drawn from Type I Extreme Value (EV1) distribution and the corresponding variances are equal for each of the alternatives. However, in order to be able to estimate the model parameters, all these variances are normalised to be constant and the scale parameters λ of the EV1 distributions are normally set to 1 (Hensher et al., 2015a).

As part of this thesis, discrete choice models for estimating individuals' travel mode and airline choice behaviours in regional Western Australia were developed. The purposes of the models are twofold; 1) to identify which key factors, (e.g., journey time and travel cost), can significantly affect travel mode and airline choice and 2) to quantify individuals' sensitivity to these factors by estimating the effects of changes in the parameters that influence the choice behaviours. MNL, NL and LC models were developed respectively to investigate the regional travellers' travel mode and airline choice from various aspects with different assumptions relevant to individuals' preferences. Thus, a more comprehensive and reliable understanding of and insight into regional travellers' travel behaviours, as well as the competition between different travel modes, can be investigated.

### 3.7.1 Estimating travel mode and airline choice using Multinomial and Nested Logit modelling

As mentioned in section 2.4.3, although a number of existing studies have explored travel mode choice between hub cities, states or countries, little to no attention has been given to the investigation of travel mode choice in regional Western Australia. Additionally, most of the mode and airline choice related studies were based on data collected at airports and/or train stations, which may not be fully representative of the entire population. Therefore, this section estimated travel mode and airline choice for regional travel in Western Australia based on the SP survey data, collected face-to-face at both airports and other locations, (e.g., shopping centres, libraries and visitor centres), and using MNL and NL models. Three steps were involved that are briefly described below, with full details and results presented in Chapter 8.

**Step 1: Determination of alternatives and attributes**

The first step to apply the discrete choice modelling approach for estimating travel mode and airline choice is to determine the alternatives (e.g., airline and bus) and the key factors (or attributes) that may affect travellers' mode choices. On the basis of the real transportation status in regional Western Australia, the travel mode alternatives were car, bus and two unnamed regional airlines. The key factors are given below, selected based on the previous travel mode literature, the pilot study and focus group discussion.

- Travel cost (Ticket fare or cost of driving);
- Access time (Access time to bus station or airport);
- Journey time (Travel time from origin to destination);
- Service frequency (Weekly service frequency of bus or airline); and
- Seat comfort level (Measured based on the leg room distance).

The SP survey choice question provided the travel mode alternatives, with each alternative described by a set of attribute/factor levels , and the respondent was required to select the alternative that appealed to him/her the most (see Chapter 5). These key factors were used to construct the observed utility for each of the travel mode alternatives, which could then be used to predict the probability that an alternative would be selected by a regional traveller.

**Step 2: Develop Multinomial and Nested logit models**

This section presents the MNL and NL models that were developed for the estimation of travel mode and airline choice using the observed utility function. MNL model is the simplest discrete choice logit model that restricts the covariance of the unobserved utilities $\varepsilon$ across the alternatives to zero. In other words, it assumes that the $\varepsilon$ across all of the alternatives are identically and independently distributed (IID) and that the alternatives are independent and irrelevant. Thus, the choice probabilities can be derived by the IID and EV1 (scale parameter $\lambda$ normalised to be 1) distribution assumption represented as Equation 3-4 (Hensher et al., 2015a).

$$\varepsilon \sim IID \,\&\, EV1\left[\begin{pmatrix}0.57721\\0.57721\\0.57721\\0.57721\end{pmatrix},\begin{pmatrix}\pi^2/6 & 0 & 0 & 0\\0 & \pi^2/6 & 0 & 0\\0 & 0 & \pi^2/6 & 0\\0 & 0 & 0 & \pi^2/6\end{pmatrix}\right] \quad \text{3-4)}$$

Therefore, the preference for one alternative will not be influenced by any other alternatives in the choice question. However, the strict IID assumption in the MNL model is likely to be unrealistic in some cases, especially when the alternatives may have some correlations. In this study, there is likely to be a correlation or substitution for the preferences between the two regional airline alternatives. Thus, NL models were developed to accommodate the possible correlations by allowing for a partial relaxation of the IID assumption which thus may provide more reliable estimations. The NL model has a hierarchical tree-like structure that links the alternatives within different nests. The alternatives under the same nest share a common non-zero covariance of the unobserved utilities, but zero covariance of the alternatives across different nests. For example, an NL model with four alternatives and two nests could be represented by the configuration of Equation 3-5, where the scale parameter is normalised to 1 for notational convenience (Hensher et al., 2015a).

$$\varepsilon \sim EV1\left[\begin{pmatrix}0.57721\\0.57721\\0.57721\\0.57721\end{pmatrix},\begin{pmatrix}\sigma^2 & \sigma_{nest\ a} & 0 & 0\\\sigma_{nest\ a} & \sigma^2 & 0 & 0\\0 & 0 & \sigma^2 & \sigma_{nest\ b}\\0 & 0 & \sigma_{nest\ b} & \sigma^2\end{pmatrix}\right] \quad \text{3-5)}$$

Thus, it can allow a degree of correlation or substitution between the two alternatives, (airlines in this thesis), within the same nest. One notable point here is that the thesis designed and tested different nesting structures for the travel mode alternatives, (car, bus and two regional airlines), in order to find the most appropriate nesting structure for estimating the travel mode and airline choice in regional Western Australia.

**Step 3: Estimation of travel mode and airline choice**

The SP data of the airport and non-airport passengers were respectively used to estimate the parameters of the MNL and NL models. The model fit statistics, (e.g., Akaike information criterion (AIC) (Akaike, 1998)), were then used to identify which model was more appropriate in estimating the travel mode and airline choice. Based on the parameter estimates of the optimal logit models, the key factors that statistically significantly affected regional travellers' travel mode and airline choice were identified, as well as the scales of those factors. Additionally, the SP data were classified into four subgroups; a) business air passenger group; b) non-business air passenger group; c) business non-air passenger group and d) non-business non-air passenger group. The logit models were estimated based on these four subsets of data. Therefore, the differences in travel mode choice behaviour between business and non-business passengers, and between air and non-air passengers, were also investigated, including a comparison of marginal elasticities and willingness to pay for variations in service quality.

### 3.7.2 Analysing travel mode choice behaviour using latent class modelling

Although MNL and NL models have been applied widely in the literature for analysing travel mode choice behaviours, it is usually assumed that preferences across respondents are homogenous, which is a well-acknowledged limitation. In the previous section, travel mode and airline choice among four pre-classified groups of travellers were estimated. Therefore, it assumes that travellers within the same group, (such as the business air traveller group), have homogenous preferences. Although trip purpose is a significant observed characteristic for identifying the preference heterogeneity that has been popularly used in the literature, other observed characteristics such as age, income and gender, or even the unobserved personal character, may also influence preferences. Generally, different choice modelling methods have different angles and/or purposes to explore the travel mode choice behaviours. MNL models are the fundamental models with several shortcomings due to the strict IID assumption, while NL models further improve the MNL by capturing the potential correlations between alternatives.

On the other hand, if there is preference heterogeneity across the travellers with the same trip purpose, the LC model can help to more effectively accommodate preference heterogeneity and relax some assumptions of the MNL model (Hensher et al., 2015a). It assumes that a traveller's choice behaviour is determined by choice related key factors, as well as the latent heterogeneity owing to individual-specific characteristics, (e.g., socio-demographics), which therefore may provide a more comprehensive angle into understanding regional traveller travel behaviour. In this section, two steps were involved to apply the LC modelling analysis, which are described briefly here and more fully in section 9.3:

**Step 1: Developing latent class model**

The key factors used to develop the utility function for the LC model were the same as the variables used in building the MNL model in previous section. In this study the LC model performed as a semi-parametric variant of the MNL model that postulates and identifies a predefined discrete number of latent segments/classes across the travellers, where the travellers' preferences are homogeneous only in the same segment. Based on an initial "guess" of the parameter values (normally the means) for each of the segments, a finite iteration of expectation and maximisation routines with weighted log likelihood was used to re-estimate and finalise the segment-specific parameter estimates. In estimating the LC model, one important point was to determine the most appropriate number of segments. The information criteria AIC and the Bayesian information criterion (BIC) were used to assist in identifying the optimal number of segments and to develop the LC model.

**Step 2: Estimating regional travel behaviour**

The LC model with the most appropriate number of latent segments was developed from the previous step. This step investigates the travel behaviour using the model outputs. The model outputs have three important components:

- A multinomial formatted class membership function that indicates the segment/class assignment memberships;
- Segment-specific choice model parameter vectors; and
- Respondent-specific posterior class probabilities.

The assignment memberships indicate the size of the corresponding segment - the proportion of respondents that the segment accounted for. The segment-specific parameters explain the choice probabilities and travel behaviours of the travellers within that segment, which are assumed to be generated based on the MNL model. For each segment, the corresponding parameter estimates can be further analysed, (i.e., willingness to pay and elasticity) to quantify traveller sensitivity to the key factors, (e.g., travel cost and seat comfort). However, in the LC model, each respondent has a posterior class probability of being in each segment; the sum of the set of posterior class probabilities therefore equals one. The respondent-specific posterior class probabilities can be used to evaluate the characteristics of each latent segment, such as the dominant demographics and trip purposes. Consequently, based on the LC modelling results, not only the travellers' travel mode and airline choice preferences within each of the distinct segments, but also the major characteristics of each segment, can be understood.

## 3. 8 Major Software

### 3.8.1 GIS software

The spatial data storage and processing were implemented using ArcGIS Version 10.2 (Esri., 2017). Geo-spatial data with attributes information, such as the LGA boundary data with attributes of population, average income and tourist number, were stored in Esri shapefiles and geodatabase with the Western Australian projected coordinate system - Geocentric Datum of Australia 1994 Map Grid of Australia Zone 50 (GDA 1994 MGA 50). The spatial data processing of the study area (Figure 3-1) and air travel flow visualisation (Figure 4-1), definition of the catchment areas, the road network analysis and the area-weighted spatial data conversion (section 4.3.2) were completed using the functions and tools provided by the ArcGIS software.

### 3.8.2 Programming languages and SPSS software

The *Python* programming language was used throughout the whole research in terms of collecting, manipulating, exploring and simulating the geo-spatial and survey data. The details of its use at various points of the thesis are given below:

- Extracting real time flights information, (i.e., flight number, aircraft type and seat numbers), and regional travel data online, (i.e., travel time and cost of different travel modes of regional trips in Western Australia);
- Manipulating the geo-spatial and survey data to create formal format data tables for exploring regional aviation markets and further statistical analysis (i.e., regression & discrete choice modelling analysis) (Chapter 4, Chapter 6, Chapter 7, Chapter 8 and Chapter 9); and
- Assisting the construction of the SP experimental design by creating appropriate candidate set tables and providing random simulation test for the generated discrete choice experiment (Chapter 5).

R language provided a Poisson regression estimator for the gravity models developed to forecast air travel demand (Chapter 4). Statistical package for the Social Science (SPSS) software was used for applying the Pearson Correlation test (Chapter 4).

### 3.8.3 Ngene and NLOGIT software

*Ngene* is the software that offers concise and flexible functionalities for generating experimental designs that are used in SP experiments for the purpose of estimating choice models, particularly of the logit type (ChoiceMetrics, 2018). In this research *Ngene 1.2.0* was used to identify the realistic SP questions for constructing the discrete travel mode and airline SP survey.

*NLOGIT* is an extension of the software package LIMDEP. It provides sufficient functionalities and programs for simulating and estimating statistical models, especially for the discrete choice models (e.g., logit models) with SP panel data. This research used *NLOGIT 5* to estimate the travel mode and airline choice related MNL, NL and LC models (Chapters 8 and 9).

### 3. 9 Summary

This chapter has presented the methodology used to investigate the characteristics and competition of regional aviation markets in Western Australia following the research methodology framework set out in section 3.3. Data collection including geo-spatial and a field survey (air travel and SP survey), was introduced. Then, based on the geo-spatial data, modified gravity models were developed to forecast the air passenger demand on RPT air routes in Western Australia, and to compare the effects of the drive factors on demand. Exploratory data analysis was then applied to explore and compare the regional aviation markets characteristics using the field survey data. At the end of the methodology framework, by analysing the SP data, discrete choice models, including multinomial, NL and LC models, were developed and estimated, in order to identify and investigate the key factors affects regional traveller travel mode and airline choice. The key steps of the methodology for each analysis were described individually. The chapter concluded with the software packages utilised for the statistical modelling analysis.

The next chapter applies the modified gravity models for estimating and comparing the regional air travel demand on RPT air routes in regional Western Australia. The concepts, methods and findings are discussed in detail.

# CHAPTER 4 FORECASTING THE AIR PASSENGER DEMANDS BETWEEN RPT AIRPORTS USING MODIFIED GRAVITY MODELS

## 4. 1 Introduction

The previous chapter described the methodological framework for achieving the objectives of the research, specifically spatial modelling of air passenger demand forecasting, mixture modelling for aviation market segmentation and discrete choice modelling for travel mode and airline choice. The first objective is to forecast air passenger demand and is the focus of this chapter. This chapter develops modified gravity models using PPML estimator to estimate the air travel demand between airport-pairs in regional Western Australia.

The present chapter is based on the first published work[3] resulting from this thesis published in the *Journal of Air Transport Management* (Zhou et al., 2018). In this chapter, section 4.2 provides a background to air passenger flows and forecasting, and the factors that may affect passenger flows. Section 4.3 focuses on the framework and methodology of the air passenger volume estimation. The interpretation and discussion of the results are illustrated in sections 4.4 and 4.5, respectively. Section 4.6 provides a summary of the findings from this chapter.

## 4. 2 Research Context

Air transportation connects remote and regional areas and provides key services and resources to local communities and tourism and mining industries. Demand is a key driving force for providing high quality and affordable air services. However, recent accurate passenger movement estimates are currently not available to policy makers due to a lack of relevant historical air travel data (Regional Aviation Association of Australia, 2013). This chapter therefore aims to estimate air passenger demands, (represented by total available air passenger seat numbers) with a method that can be applied to other routes based on available information.

---

[3] Zhou, H., Xia, J., Luo, Q., Nikolova, G., Sun, J., Hughes, B., . . . Falkmer, T. (2018). Investigating the impact of catchment areas of airports on estimating air travel demand: A case study of regional Western Australia. Journal of Air Transport Management, 70, 91-103. doi:https://doi.org/10.1016/j.jairtraman.2018.05.001

Many different methods and techniques have been developed to forecast air passenger demands but gravity models are the most commonly employed method (Grosche et al., 2007; Chang, 2012; Buraga and Rusu, 2014; Zhang and Zhang, 2016). In this study, a modified gravity model with PPML estimator based on online flight information is presented. It is used to forecast bilateral air passenger seat numbers on regional air routes in Western Australia, as well as to explore how determinant factors influence air passenger seat numbers. The contribution of this chapter to existing modelling and aviation transport literature is that it considers the impact of different sizes of the airport catchment areas on air passenger seat modelling. Two different methods to define airport catchment areas were used; 1) Thiessen polygons and 2) 2.5 hour driving distance. The size and shape of these two catchment areas are indeed different. Therefore, the factors within the catchment area vary, such as population, tourist numbers and the number of operating mine sites. This leads to differences in the air travel demand models. Furthermore, in the past few years, Western Australia has experienced a mining downturn from a previous mining 'boom' and the number of jobless people in the state has increased by one-third (Deloitte Access Economics, 2014b; Australian Bureau of Statistics, 2015a). At this critical moment, tourism has been considered as one of the major driving forces in boosting Western Australian economy (Tourism Western Australia, 2012; Hall, 2015). Consequently, this chapter tests the magnitude of the influences of the mining industry and the tourism sector on influencing air seat numbers.

The reason to choose RPT air routes is because the focus is on air aviation to serve the general community and business as well as the mining industry, which is also served through closed charter flights, provided on a contract basis and not available for general travel. The outcomes of this study should be useful for understanding the key parameters of aviation services and guiding policy development[4].

---

[4] In order for research findings to be as significant as possible, the research was conducted in close collaboration with the Department of Transport Western Australian Aviation Policy and Projects branch, which supports the objectives and approach.

## 4. 3 Methodology

The historical data concerning domestic air passenger numbers on RPT routes in regional Western Australia are not available due to commercial and confidentiality reasons. Instead, this chapter use the total seats available on the domestic flights, (air passenger seat numbers), as a proxy. The total available seats can be estimated using gravity models as one form of spatial interaction. In this chapter, a modified gravity model was developed to forecast bilateral total available seats, (i.e., between airport pairs). Based on the previous studies, geo-economic factors including catchment areas of airports, distance between airports, airfare, population, income, and the number of operational mine sites and tourists within the catchment area of airports were used in the model for forecasting the bilateral total available passenger seats. Western Australia has a strong mining industry, which employs a considerable amount of fly-in fly-out (FIFO) workers (Baker et al., 2015). In addition, tourism is growing in the state. The recent tourism strategy aims to increase the value of tourism in Western Australia to $12 billion by 2020 (Tourism Western Australia, 2012). Therefore, tourist air travel may become a growing area worth investigating. Consequently, this study developed separate models to investigate these two major effects, (mining and tourism), on air passenger movement in Western Australia.

In addition, studies on the spatial extent of the factors affecting air passenger volumes in regional areas are limited. Catchment areas should be selected to represent areas from which travellers access the air routes being modelled. Inappropriately defined catchment areas of airports may result in poor quality air travel modelling. Previous research tended to use administrative boundaries, such as county, city and region as the catchment area of airports (Wei and Hansen, 2006; Hazledine, 2009; Buraga and Rusu, 2014; Chang, 2014). The limitations of this method might be arbitrary if the spatial distribution of airports has not been carefully considered when defining the catchment areas. In other words, the administrative boundaries do not necessarily align well to air travel catchment areas. For example, some people who live in one city but are closer to the airport in an adjacent city may not belong to the catchment area/city in which they live. Therefore, the forecast results could have some errors if the catchment areas of the airports cannot be accurately defined. In order to counter this problem, two types of catchment determination methods were applied, based on the location of airports. The first method was generating catchment areas based on Thiessen polygons, which ensure that the people who live in an airport's catchment area are closer to that airport than to any other airport. The second way to create catchment areas was based on a given driving distance. This chapter has adopted a 2.5 hour driving distance threshold based on an assessment of the driving distances between regional airports in Western Australia, average distances from mine sites and townships to the closest airports in Western Australia, generally acceptable driving time to airports, and also at the suggestion of the government transport agencies (Department of Infrastructure and Regional Development, 2003; Williams, 2015; Grylls, 2016). GIS techniques were used to implement these two catchment area determination methods.

### 4.3.1 Gravity model and Poisson Pseudo-Maximum Likelihood Estimator

Gravity models have been used widely in understanding spatial structure and interaction (Nijkamp, 1997), such as migration movement (Christian and Braden, 1966; Karemera et al., 2000), tourist passenger trips (Congdon, 2000), air travel (Grosche et al., 2007; Zhang and Zhang, 2016) and road freight movement (Bergkvist and Westin, 1998). This study estimated total available seats for each RPT regional Western Australia airport-pair by seven possible factors affecting air travel based on a gravity model using a PPML estimator. The PPML estimation technique is consistent in the presence of heteroscedasticity and has a capability of dealing with zero values of the dependent variable. Further detail on PPML estimation methods can be found in Shepherd, (2012); Silva & Tenreyro, (2006); Yotov, Piermartini, Monteiro, & Larch, (2016). These factors were chosen based on the availability of relevant data, Western Australia economic and geographical characteristics and previous research. The modified gravity model is given by Equation 4-1:

$$F_{ij} = G\frac{C_i^{a_1} C_j^{a_2} \times P_i^{b_1} P_j^{b_2} \times I_i^{d_1} I_j^{d_2} \times M_i^{e_1} M_j^{e_2} \times T_j^{f} \times A_{ij}^{h}}{D_{ij}^{\theta}} \qquad 4\text{-}1)$$

A logarithmic transformation is applied to transform the gravity model into a multivariate linear equation, to simplify the parameter estimation, per Equation 4-2:

$$\begin{aligned} \ln F_{ij} &= \ln G + a_1 \ln C_i + a_2 \ln C_j + b_1 \ln P_i + b_2 \ln P_j + d_1 \ln I_i + d_2 \ln I_j + e_1 \ln M_i + e_2 \ln M_j \\ &+ f \ln T_j + h \ln A_{ij} - \theta \ln D_{ij} \end{aligned} \qquad 4\text{-}2)$$

Where $F_{ij}$ represents the total air travel passenger seat number from airport *i* to airport *j*.

- $C_i$ and $C_j$ are the catchment areas of airports *i* and *j*;
- $P_i$ and $P_j$ are the populations of catchment areas where the airports *i* and *j* are located;
- $I_i$ and $I_j$ are the average incomes of the catchment areas where airports *i* and *j* are located;
- $M_i$ and $M_j$ are the numbers of operating mine sites in the catchment areas where airports *i* and *j* are located;
- $T_j$ is the number of tourists in the catchment area where airport *j* is located;
- $A_{ij}$ is the highest fully flexible economic class airfare from original airport *i* to destination airport *j*;
- $D_{ij}$ is the average driving distance in kilometres between airports *i* and *j*;
- *G* is a constant parameter; and
- $a_1$, $a_2$, $b_1$, $b_2$, $d_1$, $d_2$, $e_1$, $e_2$, $f_1$, *h, and* $\theta$ are the coefficients which control the influences of variables $C_iC_j$, $P_iP_j$, $I_i I_j$, $M_iM_j$, $T_iT_j$, $A_{ij}$ *and* $D_{ij}$, respectively.

### 4.3.2 Catchment area definition

As mentioned previously, two methods were used to define the catchment area of each regional RPT airport. The first created Thiessen polygons[5] of each RPT airport in regional Western Australia, where any location inside the polygon was closer to that airport than any other airport, (as shown in Figure 4-1). The Thiessen polygons were generated via a two-step process: 1) using straight-line segments to connect airport point locations into a triangulated irregular network and 2) creating perpendicular bisectors to all these straight-line segments (Croley and Hartmann, 1985; Burrough et al., 2015). Thus, these perpendicular bisectors form the Thiessen polygons (catchment areas) for each RPT airport. ArcGIS 10.2.2 software was used to create the Thiessen polygons of the RPT airports.

[5] Thiessen polygons are created based on a given sample points, where each polygon contributes to an area surrounding one of the sample points, any inside location is closer to the particular sample point than any other sample points.

The second method is the catchment area defined by a 2.5 hour driving distance threshold from each airport. The driving distance was the road network distance determined using the service area function of the ArcGIS 10.2.2 software. The Western Australia road network, provided by Main Roads Western Australia, contained road information, such as georeferenced road networks, road name and speed limit. The 2.5 hour driving catchment areas of Western Australia RPT airports are shown as hollow dotted polygons in Figure 4-1, with the coloured polygons the LGAs of Western Australia. The source data of the geographic and economic factors were collected based on LGAs. Therefore, area-weighted average methods were then used (Cohen et al., 1988; Mueller et al., 2012) to convert the data in the LGA boundary into the corresponding catchment areas using ArcGIS software, with the assumption that the factors are equally distributed within each LGA. For instance, if an airport catchment area covered 100% of LGA *a*, 50% of LGA *b* and 40% of LGA *c*, then the population in the airport catchment area equals to the sum of population in LGA *a*, 50% population in LGA *b* and 40% population in LGA *c*.

### 4.3.3 Data used in this study

This study was conducted using the collected geo-spatial data including flight data and geo-economic data. Flight data are the real time bilateral flight information between regional RPT airports in Western Australia and the fully flexible economic class airfares. The geo-economic data included statistical boundaries of LGAs in Western Australia, geographical distances between the airport-pairs, population, average per capita income, number of operating mine sites, tourist population and the road network with road names and speed limits. Details of the data collection process are provided in section 3.4.1.

## 4. 4 Results

### 4.4.1 Visualisation of total available seats in regional Western Australia

Figure 4-1 illustrates the total available seats on regional Western Australia RPT air routes. Perth and Broom airports are the hubs of the Western Australia regional air travel network connecting a number of regional spoke airports. The total available seats distribution varied across different air routes in Western Australia. The map shows that the air routes connecting airports that service major mine sites have relatively higher number of total available seats, such as, the air routes between Perth and Karratha, Perth and Port Hedland and Perth and Kalgoorlie airports.

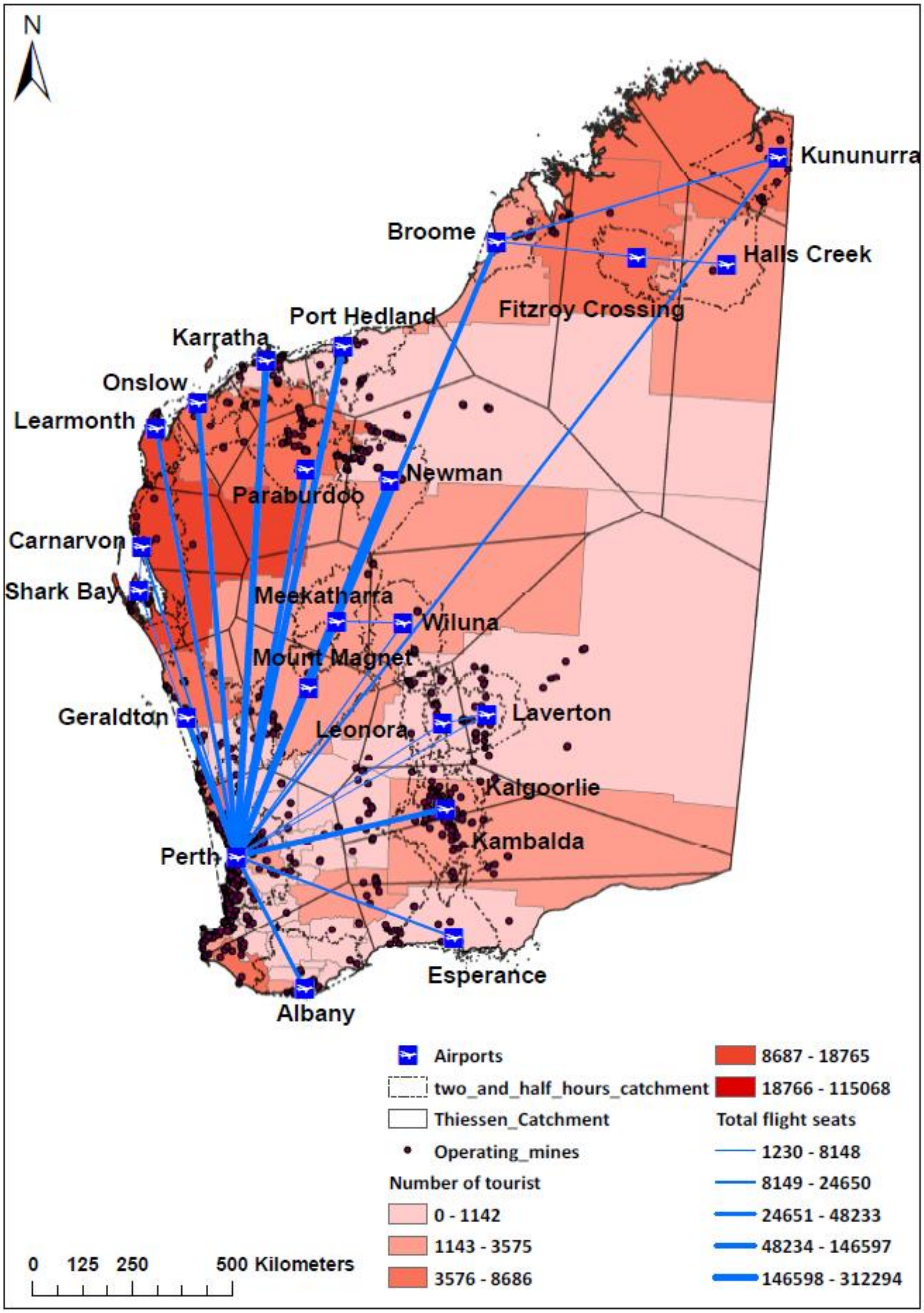


*Figure 4-1 Spatial visualization of air passenger flows in Western Australia*

### 4.4.2 Gravity models based on the Thiessen polygon catchment area

An exploration based on the visualisation presented in Figure 4-1 suggests that the number of operating mine sites and the number of tourists have a positive influence on air passenger seat numbers. In contrast, the null hypothesis $H_0$ of this study is that the independent factors are not correlated with the total available passenger seats on the domestic air routes in Western Australia, (significance level: 0.05). For the 22 airports analysed, 46 out of 462 airport-pairs were identified as having RPT services. Table 4-1 shows the descriptive statistics of the variables derived based on the Thiessen polygon catchment areas.

*Table 4-1 Descriptive statistics of the variables*

| Descriptive Statistics | | | | | |
|---|---|---|---|---|---|
| Variables | N | Minimum | Maximum | Mean | Std. Deviation |
| LN_Total_seats | 462 | 0.0000 | 12.6517 | 0.9822 | 3.0108 |
| LN_Mine_sites_origin | 462 | 1.3863 | 5.9081 | 3.3194 | 1.0744 |
| LN_Mine_sites_destination | 462 | 1.3863 | 5.9081 | 3.3194 | 1.0744 |
| LN_Travel_time (mins) | 462 | 0.0000 | 5.3706 | 0.4395 | 1.3338 |
| LN_Population_origin | 462 | 6.7142 | 14.4746 | 9.1014 | 1.6208 |
| LN_Population_destination | 462 | 6.7142 | 14.4746 | 9.1014 | 1.6208 |
| LN_Driving_distance | 462 | 4.8259 | 8.5113 | 7.0765 | 0.6879 |
| LN_Average_income_origin | 462 | 10.5428 | 11.4216 | 10.9940 | 0.2733 |
| LN_Average_income_destination | 462 | 10.5428 | 11.4216 | 10.9940 | 0.2733 |
| LN_Catchment_area_origin ($km^2$) | 462 | 10.0947 | 13.0259 | 11.3183 | 0.7359 |
| LN_Catchment _area_destination ($km^2$) | 462 | 10.0947 | 13.0259 | 11.3183 | 0.7359 |
| LN_Tourists_DES | 462 | 4.7536 | 12.6159 | 8.3524 | 1.4861 |
| LN_Airfare_full_service | 462 | 0.0000 | 6.5889 | 0.6047 | 1.8255 |

As mentioned in section 4.3.1, a logarithmic transformation was applied to transform the gravity model to a linear function. The Pearson Correlation test showed that population of origin airport's catchment area ($P_i$) was highly correlated with number of operating mine sites in the same area ($M_i$) (0.570, $p$-value < 0.05); population of destination airport's catchment area ($P_j$) was highly correlated with the number of operating mine sites in the destination airport's catchment area ($M_j$) (0.570 $p$-value < 0.05); and travel time was strongly correlated with airfare ($A_{ij}$) (0.996, $p$-value < 0.05). The Pearson Correlation test results also indicated that average income ($I_i$, $I_i$) and size of the catchment area ($C_i$, $C_i$) are both weakly or insignificantly correlated with total available seats, (correlation coefficients are all smaller than 0.100). Taking into consideration the multicollinearity and Pearson Correlation results, this chapter adjusted the gravity model (Equation 4-2) to derive four separate models with different combinations of factors affecting the available air passenger seat numbers (see Table 4-2). The coefficients of PPML regression results in the table indicate that variables such as population in the catchment area of origin and destination airports, number of operating mine sites in the catchment area of destination airports and airfare were significantly positively correlated with the total available air passenger seats in all the models, *p*-values < 0.05. In other words, the results illustrate that an increase in any of these variables is related to an increase in total available air seats on the related RPT air routes. Conversely, the results in all the four models show that distance has a statistically significant negative correlation with total available seats. In model T2 the coefficients of the number of operating mine sites in a destination airport's catchment area ($M_i$) is 0.124 (*p*-value = 0.00), which indicates that it has a significantly positive relationship with the total available seats. In addition, operating mine sites and tourist numbers in destination airport's catchment area were included in the model T4 and found that the coefficient of the operating mine sites was larger than that of tourist numbers. This might mean that the mine site parameter may contribute more to the estimation of the total available seats than the tourist parameter.

However, in Table 4-2 the AIC value of model T2 is 1612.51 and the Residual Standard Error (RSE) is 0.2650 on 457 degrees of freedom. RSE is a measure of how well a model fits the data, with a lower RSE indicating that the model fits the data more accurately. Although model T2 doesn't have the smallest RSE and AIC, all its four independent variables are statistically significant at a level of 0.001. Therefore, considering the information criteria results in conjunction with the general statistical significance of parameter estimates (e.g., Greene and Hensher 2013; Vij et al., 2013), model T2 is considered more appropriate for the prediction of total available seats.

*Table 4-2 PPML estimation results of gravity model based on the Thiessen polygon catchment area*

| **Model T1** | | | | | |
|---|---|---|---|---|---|
| y ~ ln_distance + ln_Pop_ORI + ln_Airfare_Highest + ln_Pop_DES | | | | | |
| Variables | Coefficients | Std-error | z-value | p-value | Residual Std-error |
| Intercept | -2.65602 | 0.67847 | -3.915 | 0.000*** | 0.2872 |
| LN_Distance | -0.41650 | 0.05350 | -7.785 | 0.000*** | Degrees of freedom |
| LN_population_origin | 0.04838 | 0.01934 | 2.502 | 0.012* | 457 |
| LN_Airfare_full_service | 1.08227 | 0.03872 | 27.950 | 0.000*** | AIC |
| LN_population_destination | 0.05122 | 0.05122 | 2.770 | 0.006** | 1686.84 |
| **Model T2** | | | | | |

| y ~ ln_distance ln_Pop_ORI + ln_Airfare_Highest + ln_Mine_sites_DES | | | | | |
|---|---|---|---|---|---|
| **Variables** | **Coefficients** | **Std-error** | **z-value** | **p-value** | **Residual Std-error** |
| **Intercept** | **-2.45715** | **0.54823** | **-4.482** | **0.000*** ** | **0.2650** |
| **LN_Distance** | **-0.46011** | **0.05262** | **-8.743** | **0.000*** ** | **Degrees of freedom** |
| **LN_population_origin** | **0.04779** | **0.01402** | **3.410** | **0.000*** ** | **457** |
| **LN_Airfare_full_service** | **1.10621** | **0.03720** | **29.734** | **0.000*** ** | **AIC** |
| **LN_mine_sites_destination** | **0.12389** | **0.02690** | **4.606** | **0.000*** ** | **1612.51** |
| **Model T3** | | | | | |
| y ~ ln_distance+ ln_Pop_ORI + ln_Airfare_Highest + ln_Tourists_DES | | | | | |
| Variables | Coefficients | Std-error | z-value | p-value | Residual Std-error |
| Intercept | -2.70013 | 0.73140 | -3.692 | 0.000*** | 0.2909 |
| LN_Distance | -0.41090 | 0.05106 | -8.047 | 0.000*** | Degrees of freedom |
| LN_population_origin | 0.04713 | 0.02049 | 2.300 | 0.021* | 457 |
| LN_Airfare_full_service | 1.07918 | 0.03807 | 28.347 | 0.000*** | AIC |
| LN_Tourists_destination | 0.06035 | 0.02692 | 2.242 | 0.025* | 1698.67 |
| **Model T4** | | | | | |
| y ~ ln_distance + ln_Pop_ORI + ln_Airfare_Highest + ln_Mine_sites_DES + ln_Tourists_DES | | | | | |
| Variables | Coefficients | Std-error | z-value | p-value | Residual Std-error |
| Intercept | -2.35739 | 0.61412 | -3.839 | 0.000*** | 0.2623 |
| LN_Distance | -0.45020 | 0.04500 | -10.005 | 0.000*** | Degrees of freedom |
| LN_population_origin | 0.04057 | 0.01824 | 2.225 | 0.026* | 456 |
| LN_Airfare_full_service | 1.11603 | 0.03448 | 32.367 | 0.000*** | AIC |
| LN_mine_sites_destination | 0.14495 | 0.04737 | 3.060 | 0.002** | 1604.03 |
| LN_Tourists_destination | -0.02353 | 0.03813 | -0.617 | 0.537 | |

*Significant at the 5% level
**Significant at the 1% level
***Significant at the 0.1% level

### 4.4.3 Gravity models based on the 2.5 hour driving catchment areas

Table 4-3 presents the descriptive statistics of the variables based on the 2.5 hour driving distance catchment areas.

*Table 4-3 Descriptive statistics of the variables*

| Descriptive Statistics | | | | | |
|---|---|---|---|---|---|
| Variables | N | Minimum | Maximum | Mean | Std. Deviation |
| LN_Total_seats | 462 | 0.0000 | 12.6517 | 0.9822 | 3.0108 |
| LN_Mine_sites_origin | 462 | 0.6931 | 5.6204 | 3.2622 | 1.1572 |
| LN_Mine_sites_destination | 462 | 0.6931 | 5.6204 | 3.2622 | 1.1572 |
| LN_Travel_time (mins) | 462 | 0.0000 | 5.3706 | 0.4395 | 1.3338 |
| LN_Population_origin | 462 | 5.4638 | 14.4173 | 8.4017 | 1.8812 |
| LN_Population_destination | 462 | 5.4638 | 14.4173 | 8.4017 | 1.8812 |
| LN_Driving_distance | 462 | 4.8259 | 8.5113 | 7.0765 | 0.6879 |
| LN_Average_income_origin | 462 | 10.4882 | 11.4510 | 10.9613 | 0.3163 |
| LN_Average_income_destination | 462 | 10.4882 | 11.4510 | 10.9613 | 0.3163 |
| LN_Catchment_area_origin ($km^2$) | 462 | 8.7416 | 10.9906 | 10.3176 | 0.5137 |
| LN_Catchment _area_destination ($km^2$) | 462 | 8.7416 | 10.9906 | 10.3176 | 0.5137 |
| LN_Tourists_DES | 462 | 4.2341 | 12.4945 | 7.6536 | 1.8239 |

| | | | | | |
|---|---|---|---|---|---|
| LN_Airfare_full_service | 462 | 0.0000 | 6.5889 | 0.6047 | 1.8255 |

A multicollinearity test using the Pearson correlation test was carried out before applying PPML estimation of the gravity models. The results of collinearity, (shown in the Appendix E), are similar to the results in section 4.4.2. Therefore, due to the collinearity test results, this chapter adjusted the gravity model (Equation 3) to produce four new models per the models in section 4.4.2.

Table 4-4 shows the regression results in terms of the four models. The results show that the second model has the smallest AIC value and RSE. All four independent variables in the model are statistically significantly correlated with total available seats. Thus, model C2 was considered to have the best fit among the four models in estimating total available air seats. The coefficients from model C2 indicate that all the independent variables, i.e., population of origin airport's catchment area $P_i$ (coefficient: 0.039, *p*-value=0.000), number of operating mine sites in destination airport's catchment area $M_j$ (coefficient: 0.118, *p*-value=0.000) and airfare $A_{ij}$ (coefficient: 1.121, *p*-value=0.000) are directly proportional to the number of total available air passenger seats. The results also show that distance $D_{ij}$ is significantly inversely proportional to total available seat number, which is consistent with the findings in section 4.4.2.

The direction of the coefficients of the parameters in sections 4.4.2 and 4.4.3 are similar. However, one noticeable difference between model T2 and C2 is that the magnitudes of coefficients of distance between airports, population of origin airport's catchment area and number of operating mine sites in destination airport's catchment area in model T2 are larger than those in model C2. Further, the coefficient of number of tourists in destination airport's catchment area in model T3 is larger than in model C3. This indicates that the catchment area of airports plays a role in affecting the modelling results and the relative impacts (dominance) of the factors.

*Table 4-4 PPML estimation results of Gravity model based on the 2.5 hour driving catchment area*

| **Model C1** | | | | | |
|---|---|---|---|---|---|
| y ~ ln_distance + ln_Pop_ORI + ln_Airfare_Highest + ln_Pop_DES | | | | | |
| Variables | Coefficients | Std-error | z-value | p-value | Residual Std-error |
| Intercept | -2.48352 | 0.59186 | -4.196 | 0.000*** | 0.2727 |
| LN_Distance | -0.44885 | 0.05020 | -8.942 | 0.000*** | Degrees of freedom |
| LN_population_origin | 0.04803 | 0.01338 | 3.590 | 0.000*** | 457 |
| LN_Airfare_full_service | 1.09688 | 0.03936 | 27.865 | 0.000*** | AIC |

| | | | | | |
|---|---|---|---|---|---|
| LN_population_destination | 0.05224 | 0.01210 | 4.318 | 0.000*** | 1638.97 |
| **Model C2** | | | | | |
| **y ~ ln_distance + ln_Pop_ORI + ln_Airfare_Highest + ln_Mine_sites_DES** | | | | | |
| **Variables** | **Coefficients** | **Std-error** | **z-value** | **p-value** | **Residual Std-error** |
| **Intercept** | **-2.41903** | **0.50684** | **-4.773** | **0.000*** ** | **0.2586** |
| **LN_Distance** | **-0.45563** | **0.05645** | **-8.072** | **0.000*** ** | **Degrees of freedom** |
| **LN_population_origin** | **0.03915** | **0.01102** | **3.552** | **0.000*** ** | **457** |
| **LN_Airfare_full_service** | **1.12127** | **0.03636** | **30.835** | **0.000*** ** | **AIC** |
| **LN_mine_sites_destination** | **0.11770** | **0.02380** | **4.946** | **0.000*** ** | **1589.92** |
| **Model C3** | | | | | |
| y ~ ln_distance + ln_Pop_ORI + ln_Airfare_Highest + ln_Tourists_DES | | | | | |
| Variables | Coefficients | Std-error | z-value | p-value | Residual Std-error |
| Intercept | -2.44584 | 0.61126 | -4.001 | 0.000*** | 0.2810 |
| LN_Distance | -0.44049 | 0.04833 | -9.114 | 0.000*** | Degrees of freedom |
| LN_population_origin | 0.04559 | 0.01408 | 3.238 | 0.001** | 457 |
| LN_Airfare_full_service | 1.08730 | 0.03877 | 28.048 | 0.000*** | AIC |
| LN_Tourists_destination | 0.05675 | 0.01662 | 3.414 | 0.000*** | 1666.68 |
| **Model C4** | | | | | |
| y ~ ln_distance + ln_Pop_ORI + ln_Airfare_Highest + ln_Mine_sites_DES + ln_Tourists_DES | | | | | |
| Variables | Coefficients | Std-error | z-value | p-value | Residual Std-error |
| Intercept | -2.42476 | 0.52862 | -4.587 | 0.000*** | 0.2593 |
| LN_Distance | -0.46186 | 0.04618 | -10.002 | 0.000*** | Degrees of freedom |
| LN_population_origin | 0.04168 | 0.01293 | 3.223 | 0.001** | 456 |
| LN_Airfare_full_service | 1.11746 | 0.03496 | 31.968 | 0.000*** | AIC |
| LN_mine_sites_destination | 0.10886 | 0.04072 | 2.674 | 0.008* | 1593.40 |
| LN_Tourists_destination | 0.00843 | 0.02550 | 0.330 | 0.741 | |

*Significant at the 5% level
**Significant at the 1% level
***Significant at the 0.1% level

## 4. 5 Discussion

This study developed four gravity models for forecasting the total available air passenger seat numbers based on Thiessen polygons and 2.5 hour driving catchment areas. Although eleven variables were initially considered in the study (Equation 4-2), the final models included just five variables (Tables 4-2 and 4-4). This was to ensure the validity of the models due to correlations between some variables (as shown in the Appendix E) and the research hypothesis proposed for this study.

### 4.5.1 Key determinants

The modelling results of the Thiessen polygon and 2.5 hour driving catchment areas both indicated that distance and population have a statistically significantly impact on affecting the total available seat numbers, which is consistent with the previous literature (e.g., Hazledine, 2009; Chang, 2014;Jorge-Calderón 1997). Additionally, the mining and tourism sectors were also found to have a significantly positive influence on the air travel demand, for the case of Western Australia.

Interestingly, the modelling results of the Thiessen polygon and 2.5 hour driving catchment areas both indicated that the full-service airfare had a statistically significant positive correlation with total available seat numbers. This finding is inconsistent with the outcomes of Wei and Hansen (2006), who found a significant negative impact of airfare on the aggregate air travel demand. Moreover, the coefficients of airfare in all of the tested models are larger than the coefficients of other independent variables. This means that the unit percent change in airfare can explain more of the total available seats than the unit change of the other variables.

### 4.5.2 Impact of size of catchment area on modelling

The major differences between the Thiessen polygons and the 2.5 hour driving distance catchment areas are the size and coverage of the areas. The Thiessen polygon catchment areas cover the whole of Western Australia, while the 2.5 hour driving distance catchment areas cover only 32 percent of the Western Australia region. Thus, the values of the independent variables derived from the two catchment areas are different. Therefore, a correlation analysis between the total available air passenger seat numbers and the other variables was conducted (Table 4-5) to understand the impact of the catchment area on these relationships. The correlation between total available air seats and number of tourists in the destination airport catchment areas, population in both destination and origin airport catchment areas, and number of operating mine sites in both destination and origin airport catchment areas in the Thiessen polygon catchment areas are slightly higher than those in the 2.5 hour driving catchment areas. Not surprisingly, increasing the size of the catchment areas is likely to result in the inclusion of more population, tourists, and mine sites, depending upon their spatial distribution. Nevertheless, the coefficients of population and number of operating mine sites in model T2, (Thiessen polygon catchment areas), are relatively larger than the coefficients in model C2, (2.5 hour driving catchment areas). On the other hand, in terms of predictability of these two types of models, the 2.5 hour driving catchment areas models were found to slightly outperform the Thiessen polygon catchment area models. As shown in Figure 4-2, this may be due to spatial concentrations of populations, tourists and mine sites within the 2.5 hour driving catchment area.

*Table 4-5 Correlation between available seats and independent variables*

| Correlation with total seats | | LN_Mine_site_origin | LN_Mine_site_destination | LN_Population_origion | LN_Population_destination | LN_Average_income_origin | LN_Average_income_destination | LN_Tourist destination |
|---|---|---|---|---|---|---|---|---|
| 2.5 hours driving catchment area | Pearson Correlation | 0.250* | 0.235** | 0.398** | 0.377** | -0.064 | -0.066 | 0.334** |
| | P-value | 0.000 | 0.000 | 0.000 | 0.000 | 0.170 | 0.154 | 0.000 |
| Thiessen polygon catchment area | Pearson Correlation | 0.302** | .0282** | 0.416** | 0.390** | -.090 | -0.092* | 0.360** |
| | P-value | 0.000 | 0.000 | 0.000 | 0.000 | 0.054 | 0.047 | 0.000 |

*Significant at the 5% level
**Significant at the 1% level

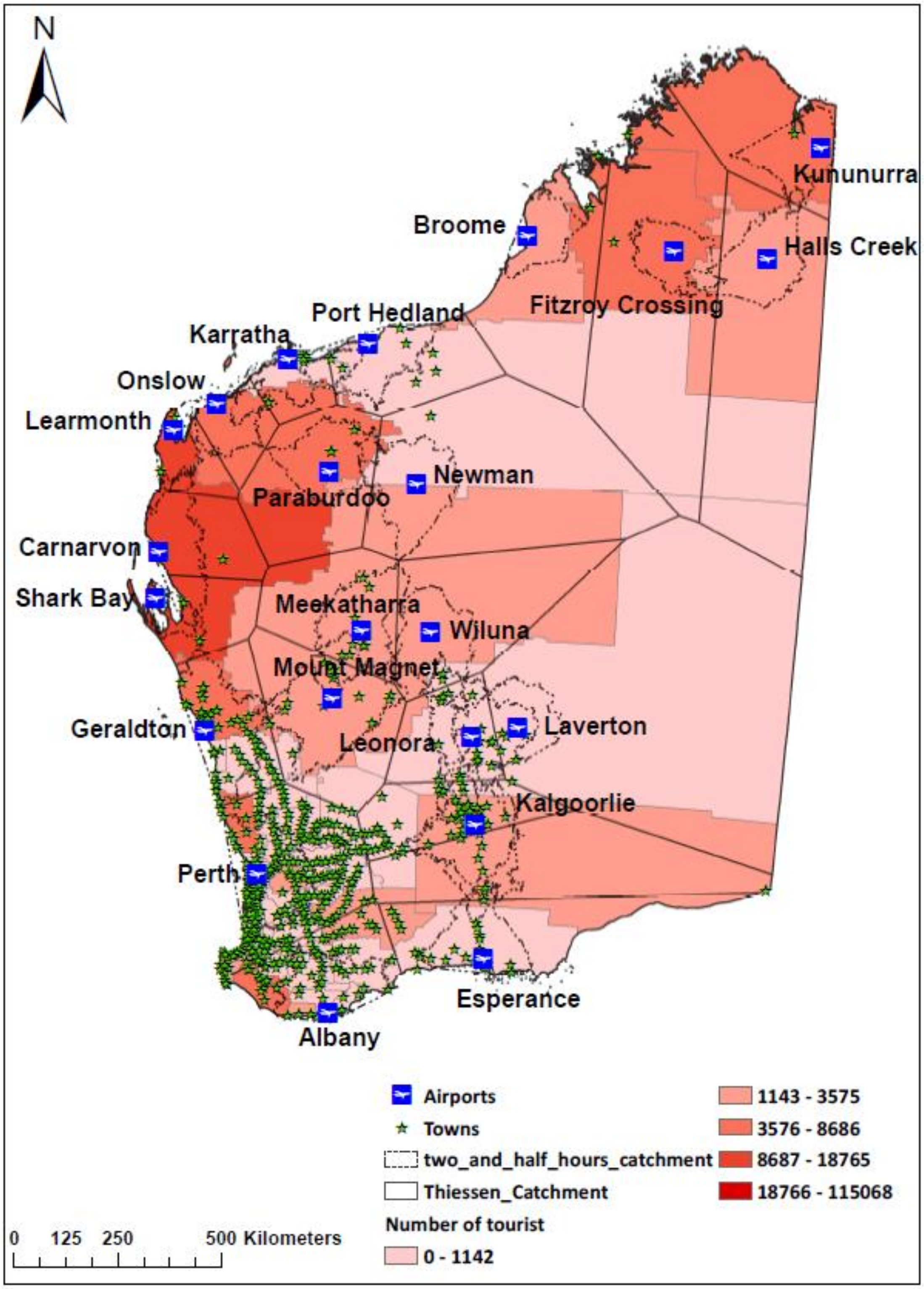



*Figure 4-2 Spatial distribution of towns and tourists in Western Australia*

### 4.5.3 Load factors in Western Australia

One data limitation of this study is the use of the available seats instead of total passengers carried, due to the load factors, (actual passenger numbers), for each route in Western Australia being unavailable. However, the Bureau of Infrastructure, Transport and Regional Economics (BITRE) has summarised monthly load factors for the 66 most popular domestic air routes in Australia (Department of Infrastructure and Regional Development, 2017). The load factors for the six most popular routes in Western Australia for the period of June to September 2016 are presented in Table 4-6, with the factors ranging from 47.0% to 82.9%. This level of variability may introduce some errors in forecasting the passenger movement based on the product of available seats and average load factor. Therefore, this study used available seat numbers as an indicator of air travel demand instead of the number of passengers carried.

*Table 4-6 Monthly load factor of Western Australia popular routes between June and September, 2016*

| | Load factor (percentage)-Monthly | | | |
|---|---|---|---|---|
| Air route (bidirectional) | Jun-16 | Jul-16 | Aug-16 | Sep-16 |
| Broome - Perth | 78.5% | 82.9% | 76.5% | 81.2% |
| Geraldton - Perth | 48.0% | 47.0% | 49.4% | 50.0% |
| Kalgoorlie- Perth | 59.4% | 57.8% | 61.3% | 62.3% |
| Karratha - Perth | 62.2% | 59.5% | 57.4% | 58.9% |
| Newman - Perth | 50.7% | 52.9% | 53.3% | 50.0% |
| Port Hedland - Perth | 56.3% | 59.0% | 57.4% | 56.1% |
| 66 most popular domestic air routes in Australia | 77.6% | 80.6% | 76.1% | 78.6% |

According to the House of Representatives Standing Committee on Transport and Regional Services (2003), in order for the flights to remain viable for the airlines, load factors need to be at least 60-70 percent. In 2016, only the Broome-Perth route was above this level, with three other routes at or just below the minimum level and two routes well below the minimum level. Consequently, it has been identified that this could contribute to the relatively higher cost of flights in Western Australia compared to other states (Department of Transport, 2015d). The airlines are able to maintain the services due to the dominance of the business-corporate sector in the Western Australia air travel market, a sector that is relatively price insensitive (Department of Transport, 2015d). This creates challenges to the price sensitive tourism and leisure markets. For example, Geraldton is located 430 km from Perth and is one of the major regional towns and service centres in Western Australia. Its main economic drivers include mining, agriculture, retail, manufacturing, tourism, building/construction and fisheries (Grylls, 2016). According to the survey of air route connectivity in mid and north west Western Australia, business trips contributed around 65% of travel along the Geraldton – Perth route with the remaining 35% being tourism and private travel (Grylls, 2016), even though this route has the lowest load factor compared to the other five popular routes (See Table 4-6). Table 4-7 shows the monthly flights, seats, passengers carried and load factors in February 2014-2017. In contrast, the Perth to Broome route, (the route between Western Australian hub airports), had an average load factor of around 80% between June and September, 2016 (Department of Infrastructure and Regional Development, 2017). This high level of demand may make passengers more vulnerable when there is a disruption in the airport such as cancellation of flights, which may lead to longer delays due to lower spare capacity (Rodrigue et al., 2013).

*Table 4-7 Estimates of domestic aviation activities on Geraldton – Perth route*

| | Feb 2014 | Feb 2015 | Feb 2016 | Feb 2017 |
|---|---|---|---|---|
| Flights | 284 | 263 | 225 | 174 |
| Seats | 17,298 | 15,228 | 18 696 | 17,400 |
| Passengers carried | 10,045 | 9,282 | 9,214 | 8,586 |
| Load factors | 0.58 | 0.61 | 0.49 | 0.49 |

### 4.5.4 Accuracy of online flight information

As mentioned in the description of the methodology, flight and seat data from the website www.Flightradar24.com were collected. This section evaluates the accuracy of the data collected based on estimates of monthly aviation activities, (available seats), on the most popular six routes in Western Australia as published by BITRE (Table 4-8) (Department of Infrastructure and Regional Development, 2017). The average seat number difference between the two data sets is 4.66% in July 2016 and 6.12% in August 2016. These differences may be due to the last-minute cancellations of flights and unavailable seating information for certain types of aircraft.

*Table 4-8 Evaluation of online flight information – difference in seat and flight data*

| | Seats | | | | | |
|---|---|---|---|---|---|---|
| | Reported available seats | | Collected available seats | | | |
| Air route (bidirectional) | Jul-16 | Aug-16 | Jul-16 | Aug-16 | % Difference Jul-16 | % Difference Aug-16 |
| Broome - Perth | 39,123 | 38,754 | 38,992 | 39,016 | -0.33% | 0.68% |
| Geraldton - Perth | 20,400 | 20,600 | 20,800 | 20,900 | 1.96% | 1.46% |
| Kalgoorlie- Perth | 33,772 | 35,310 | 32,789 | 34,781 | -2.91% | -1.50% |

| | | | | | | |
|---|---|---|---|---|---|---|
| Karratha - Perth | 68,443 | 72,449 | 76,205 | 79,554 | 11.34% | 9.81% |
| Newman - Perth | 48,393 | 53,217 | 52,271 | 59,554 | 8.01% | 11.91% |
| Port Hedland - Perth | 50,220 | 52,284 | 51,428 | 55,499 | 2.41% | 6.15% |
| Total | 260,351 | 272,614 | 272,485 | 289,304 | 4.66% | 6.12% |
| | | | Flights | | | |
| | Reported available flights | | Collected available flights | | | |
| Air route (bidirectional) | Jul-16 | Aug-16 | Jul-16 | Aug-16 | % Difference Jul-16 | % Difference Aug-16 |
| Broome - Perth | 309 | 311 | 311 | 317 | 0.65% | 1.93% |
| Geraldton - Perth | 204 | 206 | 208 | 209 | 1.96% | 1.46% |
| Kalgoorlie- Perth | 290 | 300 | 292 | 306 | 0.69% | 2.00% |
| Karratha - Perth | 566 | 605 | 661 | 687 | 16.78% | 13.55% |
| Newman - Perth | 350 | 374 | 379 | 422 | 8.29% | 12.83% |
| Port Hedland - Perth | 380 | 390 | 382 | 404 | 0.53% | 3.59% |
| Total | 2,099 | 2,186 | 2,245 | 2,356 | 6.96% | 7.78% |

## 4. 6 Summary

The key objective of this chapter was to estimate the total available seats using online flight information. It especially investigated the impact of the catchment area of airports on modelling the total available seats. The size of the catchment area can affect the magnitude of explanatory factors and therefore influence the modelling results. When deciding the catchment area for the study, it is important to take the spatial distribution of these explanatory factors into consideration, and the more appropriate determination of airport catchment areas the better modelling performance.

Based on the results of these models, the mining sector was found to have a greater influence on total available seats than the tourism sector. This was true for the gravity models based on both the Thiessen polygon catchment areas and the 2.5 hour driving catchment areas. The next chapter will elaborate on the novel procedure for generating the SP experiment, which will then be used to investigate the regional aviation market in Western Australia and to estimate travel mode and airline choice.

# CHAPTER 5 DISCRETE CHOICE EXPERIMENTAL DESIGN

## 5. 1 Introduction

The previous chapter identified the driving factors, (e.g., airfare and the mining sector), that can determine the air travel flows on RPT air routes. This chapter moves on to the second major component of this thesis, namely the discrete choice experiment, that is used to construct SP survey and explore the preferences for different aspects of air travel and can be used to estimate demand. Discrete choice experiments typically ask respondents to consider a series of hypothetical choices between options that are described by sets of dimensions, each of which takes one of a set of levels. One important problem in constructing SP experimental design is how to ensure the plausibility and realism of the generated choice questions. This chapter proposes a novel methodology framework by extending the Modified Federov Algorithm with the *Python* programming language to generate a D-efficient SP survey that maintains an appropriate behavioural plausibility and realism, as well as high statistical efficiency. The SP survey is then used for further analysis in this research, including investigation of the regional aviation market and estimation of travel mode and airline choice behaviour.

The present chapter is composed of one paper, which has not submitted to any journal yet. Section 5.2 introduces the research context describing the role of, and need for, experimental design in earlier research. Section 5.3 elaborates on the methodology procedure for constructing the discrete choice experiment. Section 5.4 implements the construction of the D-efficient experiment for this research, using the proposed methodology. Finally, section 5.5 provides discussion.

## 5. 2 Research Context

The objective of an SP analysis is to determine individuals' preferences/decisions between alternative choices, (such as different transport modes: car, bus, taxi, air), and quantify their sensitivities to the attributes of the alternatives. In order to derive a reliable survey, which can best reflect participants' choice-making, there is a need to design an SP experiment, a process that involves developing, testing and optimising a combination of attribute levels for different choice alternatives (Rose and Bliemer, 2009). SP experiments have been widely used in transportation research to provide statistically valid estimates of travel preferences and behaviours (Bliemer et al., 2009; Rose and Bliemer, 2009). These include travel mode and route choices. For instance, Jung and Yoo (2014) developed an SP experiment to examine individual travel mode choice behaviour related to short-haul domestic trips; Chang and Sun (2012) collected SP data to capture passengers' willingness to pay for air travel services; Chen and Chao (2015) applied SP data to forecast passengers' airline choice of international flights; Qiao et al. (2016) used SP data to estimate individuals' travel mode choices between public and private travel modes; Hess et al. (2007) used SP data to understand passengers' airport and airline choice for domestic travel in the United States; Lapparent et al. (2009) employed an SP experiment to examine individuals' travel mode choice behaviour for international travel within Europe. Additionally, Puckett and Hensher (2009) discussed heterogeneity processing of freight stakeholders' trip choice with an SP experimental design; Sener et al. (2009) used SP experiments to evaluate cyclists' route choice behaviour in Texas, United States.

Typically, in SP experiments, participants are presented with a series of hypothetical choice scenarios, (also named choice situations, choice tasks, sets and questions), where each scenario contains a finite set of alternatives described by several attributes, (or dimensions), and each attribute has a set of pre-defined possible attribute-levels (Bliemer et al., 2009). The respondents are required to choose one or more alternatives that appeal to them the most in each choice scenario. Generally, a respondent is required to choose between all the choice scenarios. All respondents' responses are collected and pooled to estimate individual sensitivities and preferences to attributes of alternative choices (travel mode choice).

### 5.2.1 Stated and revealed preference survey

Both SP and RP experiments can be used to understand individuals' travel behaviour. An RP experiment is based on real observations of the travel behaviour, while an SP experiment obtains individual choice preferences under hypothetical scenarios. Both techniques have strengths and weaknesses. For example, for air travellers, RP data can refer to respondents' real experiences, expanding on the information of the flights they actually used. It is obvious that RP data have some significant limitations as they cannot be used to capture the influence when a new alternative is provided, and they lack reliability due to the limited variations within the data (Louviere and Hensher, 1982; Louviere and Woodworth, 1983; Walker et al., 2018). In comparison, SP data can be linked to the travel mode and airline choice in some pre-designed hypothetical choice scenarios that may arise in the real world. SP data allow researchers to add some trade-off to the hypothetical nature of choice scenarios (Walker et al., 2018). However, the hypothetical nature may also lead to some limitations. For example, an experimental scenario may not fully reflect the actual behaviour in a real life situation. SP experiments are preferred in discrete choice modelling for estimating people's preferences to the existing alternatives, as well as non-existing alternatives (Hensher, 1982; Hensher and Louviere, 1983; Walker et al., 2018). In this research an appropriate SP experiment was generated for the further analysis.

### 5.2.2 Previous work on experimental design

Researchers mostly use computer-based search algorithms to implement experimental design procedures (Johnson et al., 2013; Burgess et al., 2015). For example, SAS software (SAS Institute, 2011) provides a number of experiment design macros for constructing experimental designs that allow user-specifications, such as blocking experimental design into subsets and imposing restrictions on alternatives to minimise the implausibility of choice tasks (Kuhfeld, 2010). Elsewhere, Sandor and Wedel proposed a method to generate efficient design through minimising D-error, (the lower the D-error, the higher the efficiency of the design: for details about D-error see section 5.3.3.6) for MNL model (Sandor & Wedel, 2001), as well as for a cross-sectional ML model (Sándor and Wedel, 2002; Sandor and Wedel, 2005). While constructing the experiment with the alternatives, attributes and levels identified and the effects coding constructed, Sándor and Wedel introduced a Bayesian distribution of prior parameters instead of an individual prior parameter estimate to calculate the D-error, which can capture the uncertainty of parameter estimates and thus reduce bias from misspecification of priors. The procedure for experimental design was written in the GAUSS programming language (Gauss, 2011), which can also be transferred to other programming software (Johnson et al., 2013).

Bliemer, Rose and other collaborators have extended the experimental design procedure for various different discrete choice models, such as the NL model (Bliemer et al., 2009) and the panel Mixed MNL model (Bliemer and Rose, 2010). Collectively, this means that the uncertainty due to different estimating models can be accounted for in the design process, which improves the performance and efficiency of the design (Bliemer et al., 2009). Particularly, Bliemer and Rose proposed a statistical measure derived from the asymptotic variance–covariance (AVC) matrix of discrete choice models that can be used to find and minimise the theoretically required sample size for the experiments (Bliemer and Rose, 2005; Rose and Bliemer, 2013). Bliemer and Rose (2006) and Rose et al. (2008) also extended the experimental design procedure by including covariates into the models, that allows the joint optimization of design efficiency in order to determine different segments of the sample, as well as optimising the design to account for heterogeneity among individual respondents. All these experimental design procedures and extensions can be applied in the *Ngene* software (Rose et al., 2009). The *Ngene* software offers concise and flexible functionalities, (e.g., constraints specification and effects coding), to generate experimental designs for the numerous general and advanced logit models.

### 5.2.3 Implausible and dominant alternatives and unrealistic choice tasks

In the SP experimental design, each alternative is formed from predetermined attribute-levels and it is frequently true that some combinations of attribute-levels are implausible, or one of the options in the choice task is dominant (Collins et al., 2014; Cherchi and Hensher, 2015). As an example of implausible alternatives, considering two alternatives for domestic travel mode choice; bus and airline. Implausibility could result if, in the choice task, the arrival time of one of the alternatives is earlier than the departure time. It also is possible to suggest mathematically or physically impossible levels for other factors, so respondents may struggle to evaluate such implausible alternatives that thus increases hypothetical bias and errors. The dominant alternative refers to one alternative that is at least as good in every attribute, and clearly better in one or more attribute(s). For instance, in an airline choice that has two alternative airlines, if all the attributes such as airfare, journey time, service frequency and seat comfort of one airline are better than those of the other, then the first airline is the dominant alternative. Consequently, the choice task with the dominant alternative would fail to capture respondents' trade-off preferences across attribute-levels, since all the respondents would choose the dominant alternative, regardless of their preferences.

Although there are some experimental design methods that allow the researcher to add some constraints to reject implausible or dominant alternatives from the design, the choice scenario can still be problematic when some normal alternatives are appended into one choice task. This is because the hypothetical choice task formed may be inconsistent with reality. For example, in the case of regional travel mode choice between bus and airline, if the departure time, arrival time and journey time of the bus is 6 am, 9 am and 3 hours respectively, while for the airline is 2 pm, 6 pm and 4 hours, the two alternatives are both reasonable in isolation. However, if these two alternatives were presented in one choice task, it would be inconsistent with reality, since the journey time of using a bus cannot reasonably be shorter than that of travelling by air. Additionally, take the example where, in the choice task, the journey time for the bus is 25 hours while the journey time for the airline is 1 hour. Although this sounds logical, as the airline is faster than the bus, it is still unrealistic in many cases that a bus would take 25 times longer than an airline for a regional or domestic trip.

### 5.2.4 Gaps and aims

"*In the past several years, an increasing number of analysts have questioned the plausibility and realism of choice tasks as commonly represented in stated choice experiments*" (Collins et al., 2014, p. 4). Existing algorithms are not that efficient in dealing with this challenging issue while constructing SP experimental designs (Johnson et al., 2013; Collins et al., 2014). Another notable problem is that the researchers may lack the ability and experience to generate all the required constraints for rejecting implausible and dominant alternatives, as well as unrealistic choice tasks, especially for complex problems. Therefore, the aim of this study is to develop a semi-systematic method for generating an efficient SP design based on Extending the Modified Federov Algorithm (EMFA) (Cook and Nachtrheim, 1980), that can help researchers more easily identify all required constraints and effectively remove not only the implausible and dominant alternatives but also the unrealistic choice tasks. Finally, this chapter set out to demonstrate and implement an efficient SP experimental design procedure as a specification for MNL step by step. The collected data are then used to estimate individual travel mode and airline choice, using more complex discrete choice models, in Chapters 8 and 9.

### 5.2.5 How to form the experimental design

Normally, the experimental design can be viewed as a matrix of attribute-level values, where the rows and columns form the options or choice questions in the SP surveys. There are two kinds of experimental design matrices that are popularly used by researchers. In the first category, each row of the experiment matrix represents one independent choice scenario, while each column states the specified attribute for each alternative (Bliemer and Rose, 2006; Rose and Bliemer, 2009). Therefore, in this kind of experimental design matrix, each set of columns describes an alternative option (Figure 5-1a). For the second layout of the experimental design matrix, each row stands for an alternative while the column in each row represents the specific attribute-level value of the relevant alternative (Huber and Zwerina, 1996; Carlsson and Martinsson, 2003) (Figure 5-1b). However, in such design matrices, choice tasks are shown by combining a set of rows instead. Generally, whichever experimental design matrix is chosen by the analyst, the final purpose of both is to allocate the attribute-level values to form the choice questions of the SP surveys (as shown in Figure 5-1c).

| Bus | | Airline | |
|---|---|---|---|
| Cost | Time | Cost | Time |
| 100 | 6 | 200 | 1 |
| 200 | 12 | 300 | 2 |
| ... | ... | ... | ... |
| ... | ... | ... | ... |

*a*

| Choice set 1 | Bus | Airline |
|---|---|---|
| Ticket fare (A$) | 100 | 200 |
| Journey time (hrs) | 6 | 1 |
| Choice set 2 | Bus | Airline |
| Ticket fare (A$) | 200 | 300 |
| Journey time (hrs) | 12 | 2 |

*c*

| | Cost | Time |
|---|---|---|
| Bus | 100 | 6 |
| Airline | 200 | 1 |
| Bus | 200 | 12 |
| Airline | 300 | 2 |
| Bus | ... | ... |
| Airline | ... | ... |

*b*

*Figure 5-1 Experimental design matrices*

*a. Experimental design matrix of columns-based alternative; b. Experimental design matrix of row-based alternative; c. Constructed choice tasks based on experimental design matrixes*

## 5. 3 Methodology of Experimental Design

### 5.3.1 Experimental design constructing processes

Figure 5-2 illustrates the three main stages used to generate constrained SP experimental designs for discrete choice models, (e.g. MNL and NL models), that was modified based on the Hensher et al. (2015b) and Rose and Bliemer (2009)'s design procedure by extending the Modified Federov Algorithm (MFA) in stage 3.

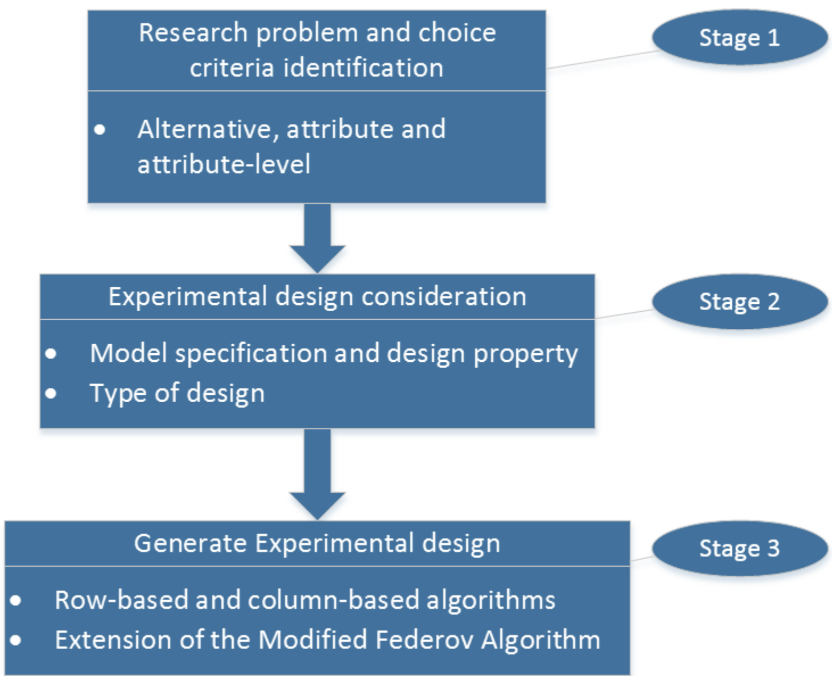



*Figure 5-2 Main stages for constructing constrained SP experimental design*

### 5.3.2 Stage 1: Research problem and choice criteria identification

The first stage requires the analyst to refine the choice related research problem and define the choice criteria including the identification of a list of relevant alternatives, a finite set of attributes and a finite number of attribute-levels. While defining the relevant alternatives, initially, qualitative studies, previous literature searches and focus group discussions with key stakeholders are important potential approaches. One notable point as mentioned by Hensher et al. (2015b) is that, if the number of identified alternatives is small, (not exceeding 10), there is no need to remove alternatives; otherwise the analyst may need to reject some unimportant alternatives or not label the alternatives (unlabelled experiment). For example, in a travel mode choice experiment, a labelled experiment would use cars, buses and trains as the alternatives, whereas an unlabelled experiment would use travel mode A, travel mode B and travel mode C as the alternatives. Once the choice alternatives that are to be considered by the decision makers are identified, the analyst is then required to determine the attributes and attribute-levels for these alternatives. The attributes are the factors that the decision makers consider when making their choice from among the listed alternatives. Previous literature may provide insights into the attributes criteria, especially the findings of related choice studies. For example, Van Can (2013) and Jung and Yoo (2014) found that attributes of travel cost, travel time and seat comfort have significant impact on passenger travel mode choice behaviours. Different alternatives may have common attribute and alternative specific attributes, the attribute-levels (ranges) of the same or similar attributes may differ from alternative to alternative. Marshall et al. (2010) identified that 70 percent of previous studies have between three and seven attributes, with the most common number being four. One noteworthy point is that the attributes can either have the same or a different number of attribute-levels. Hensher et al. (2015b) noted that the identification of more attribute-levels for the attribute allowed for the capture of more information in the utility space. The attribute-levels should be determined based on real-world values so that the choices appear reasonable to the decision makers. Otherwise, the decisions they make may not realistic reflect what they would do in the real world. For example, for a travel mode choice between bus and high speed train, the hypothetical trip times for high speed train should be less than, (and certainly no more than), the times for buses for the same route. Not having such a constraint is likely to make the respondent answer in a more casual, (less reliable), way. Additionally, the analyst needs to be cautious when setting the range of attribute-levels. This is because, in comparison with a narrow level range (e.g. 2 hours to 3 hours), setting a wider level range (e.g. 1 hours to 5 hours) can theoretically improve the parameter estimation from the discrete choice models, which suggests that wide range is relatively more appropriate than a narrow range (Rose and Bliemer, 2009). However, the analyst also needs to avoid using too wide a range of attribute-levels, since it may make it more likely to cause implausible and dominant alternatives in the choice questions.

### 5.3.3 Stage 2: Model specification and design properties

#### 5.3.3.1 Model specification

Stage two introduces the considerations of experiment design from which, in stage three, the analyst can generate the design. Model specification is a crucial property of the experiment design consideration. Once the alternatives, attributes and attribute-levels are all defined, the analyst needs to define the form of the utility function for the final discrete choice model, (e.g. MNL model), that will be used to analyse the collected survey data. In order to specify the utility function, three main steps are required. The first step is to determine whether the parameters are generic or alternative specific. Generic parameter refers to different attributes sharing one parameter while alternative specific parameter indicates that each alternative specific attribute is only associated with one parameter. The second step is to determine whether the utility function only accounts for main effects or, for example, includes interaction effects. As Hensher et al. (2015b, p. 210) proposed, "*main effect (ME) is defined as the direct independent effect of each attribute on the response variable, choice. The main effect, therefore, is the difference in the means of each level of an attribute and the overall or grand mean. An interaction effect is an effect on a response variable, choice, obtained by combining two or more attributes which would not have been observed had each of the attributes been estimated separately*". The third step is to check whether the utility function will include dummy effects. The dummy variable can add *L* minus 1 new variables to the function, where *L* is the number of attribute-levels of the dummy variable. For instance, if a dummy effect of seat comfort level has three levels, (low, middle and high), then two new variables would be added to the function. Specifically, the approach is to add middle and high seat comfort levels as the dummy variables into the utility function. If the alternative has a middle seat comfort, the coefficient of the middle seat comfort level will equal 1 and the coefficient of the high seat comfort level is 0, and vice versa for a high seat comfort. If the alternative come with a low seat comfort, both coefficients would be 0.

#### 5.3.3.2 Degrees of freedom

Degrees of freedom, required for the SP experimental design, is another vital point that needs to be considered. Simply put, "*a degree of freedom represents a single piece of information available* *to the analyst*" as explained by Hensher et al. (2015b, p. 212). For the experimental design, if there are $A$ alternatives and $S$ choice tasks, then the degrees of freedom would be $S \cdot (A-1)$. This number must be equal to or greater than the number of parameters in the utility function that will be estimated by the final discrete choice model, including the constants. For instance, for an experimental design with 4 alternatives and 12 parameters in the utility function, the degrees of freedom $S \cdot (4-1)$ should be equal to or greater than 12. Thus, the minimum value of experiment size $S$ is 4, (minimum number of choice tasks), in order to satisfy the degrees of freedom requirement.

#### 5.3.3.3 Attribute-level balance

Attribute-level balance is another property relevant to many experimental designs. It requires each alternative's attribute-levels to occur an equal number of times for each attribute (Rose and Bliemer, 2009; Hensher et al., 2015b). Restricting attribute-level balance can reduce the optimality of the experimental design but it can ensure that the parameters are estimated over the full range of the attribute-levels.

#### 5.3.3.4 Full factorial design and fractional factorial design

The analyst needs to consider which kind of experimental design will be generated. Full factorial design and fractional factorial design are two popular categories of design. Full factorial design is the design in which all possible choice tasks are enumerated and shown to the decision makers. However, the number of choice tasks can be too large to be generated, especially for a design with a relatively large number of alternatives, attributes and attribute-levels. Therefore, it is impossible to expect decision makers to complete all these choice tasks. For example, consider a full factorial design with $A$ alternatives, with each alternative having $B$ attributes and each attribute $L$ attribute-levels, (For simplicity, assuming each alternative has the same number of attributes and each attribute has same number of attribute-levels). If the full factorial design is a labelled design, the number of all possible choice tasks will equal $L^{AB}$. If the full factorial design is an unlabelled design, the experiment size will be $L^{B}$.

Fractional factorial design addresses this problem and is widely used by researchers. Fractional factorial design, as its name suggests, only picks a subset of choice tasks from the full factorial design and thus greatly reduces the number of choice tasks presented to decision makers. Fractional factorial design has many different types, the three main ones being random design, orthogonal design and efficient design. Random design entails randomly selecting a set of choice tasks from the full factorial design but this raises questions about the efficiency of the resultant designs and the capability of the design to allow exploration of specific effects.

#### 5.3.3.5 Orthogonal design benefits and problems

Orthogonal design is a widely known type of fractional factorial design that focuses on minimising the correlation between attribute-levels (Rose and Bliemer, 2009). In orthogonal designs, all the attributes are restricted to be orthogonal, thus all the attributes in the experimental design must be statistically independent. As a result, the impact on the observed choices caused by each attribute can be determined independently in the design. Generally, the orthogonal design is a subset of choice tasks from the full factorial design that requires a pairwise balance or proportional frequencies of attribute-levels, and there are no correlations between attributes in the design. As an example, assume there is a labelled full factorial design that has two alternatives (car and bus), and each alternative has two attributes with two levels for each of the attributes. Table 5-1 presents the main effects only full factorial design with a related correlation matrix. The correlation matrix indicates no correlation between attributes. Table 5-2 presents the orthogonal design generated based on the full experimental design, where the attribute-levels are balanced and the degrees of freedom for the model estimation and orthogonality are satisfied. It is clear to see that the number of choice tasks, (design size), has been halved and the attributes still remain uncorrelated.

*Table 5-1 Full Factorial design and correlation matrix*

| | **Full factorial design** | | | | | **Correlation matrix of the design** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **S (Choice task)** | **A1** | **A2** | **A3** | **A4** | | **A1** | **A2** | **A3** | **A4** |
| 1 | 1 | 1 | 1 | 1 | A1 | 1 | 0 | 0 | 0 |
| 2 | 1 | 1 | 1 | 2 | A2 | 0 | 1 | 0 | 0 |
| 3 | 1 | 1 | 2 | 1 | A3 | 0 | 0 | 1 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 1 | 1 | 2 | 2 | A4 | 0 | 0 | 0 | 1 |
| 5 | 1 | 2 | 1 | 1 | | | | | |
| 6 | 1 | 2 | 1 | 2 | | | | | |
| 7 | 1 | 2 | 2 | 1 | | | | | |
| 8 | 1 | 2 | 2 | 2 | | | | | |
| 9 | 2 | 1 | 1 | 1 | | | | | |
| 10 | 2 | 1 | 1 | 2 | | | | | |
| 11 | 2 | 1 | 2 | 1 | | | | | |
| 12 | 2 | 1 | 2 | 2 | | | | | |
| 13 | 2 | 2 | 1 | 1 | | | | | |
| 14 | 2 | 2 | 1 | 2 | | | | | |
| 15 | 2 | 2 | 2 | 1 | | | | | |
| 16 | 2 | 2 | 2 | 2 | | | | | |

*Table 5-2 Orthogonal design*

| | **Orthogonal design** | | | | | **Correlation matrix of the design** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **S (Choice task)** | **A1** | **A2** | **A3** | **A4** | | **A1** | **A2** | **A3** | **A4** |
| 1 | 1 | 1 | 1 | 1 | A1 | 1 | 0 | 0 | 0 |
| 2 | 1 | 2 | 1 | 1 | A2 | 0 | 1 | 0 | 0 |
| 3 | 1 | 1 | 2 | 1 | A3 | 0 | 0 | 1 | 0 |
| 4 | 1 | 2 | 2 | 1 | A4 | 0 | 0 | 0 | 1 |
| 5 | 2 | 1 | 1 | 2 | | | | | |
| 6 | 2 | 2 | 1 | 2 | | | | | |
| 7 | 2 | 1 | 2 | 2 | | | | | |
| 8 | 2 | 2 | 2 | 2 | | | | | |

The major reason for using an orthogonal design is that such a design can estimate the effects of attributes that affect a decision maker's choice independently, and the orthogonal design itself can be easily generated. Another reason is due to historical impetus, as the previous literature relating to experimental design was mainly associated with linear models that have orthogonality as a priority (Bliemer and Rose, 2011). This is because, for linear regression models, it not only eliminates the multicollinearity between the independent variables but also maximizes the t-values corresponding to parameter estimation. Therefore, if the design of a linear regression model is orthogonal, the off diagonals of the model's variance-covariance (VC) matrix will equal zero, which ensures that the parameter estimates are uncorrelated, whereby the possible standard errors related to the parameter estimation are minimised (maximize the t-values). However, unlike full factorial design, orthogonal design can only ensure orthogonality for certain effects, either main or interaction effects, depending on the selection of higher or lower order interaction terms (Rose and Bliemer, 2009). Therefore, one challenge of importance in constructing an orthogonal array is in picking the interaction effects that are considered to be negligible and hence do not need to be estimated separately.

Orthogonal design can generate key parameter estimates in a way that is not confounded, since the design has pre-restricted the key attributes to be statistically independent of each other. However, although orthogonality is essential for estimating independent impacts in linear models, the discrete models are actually non-linear models (Train, 2009). Furthermore, as mentioned by Lancsar and Louviere (2006), the orthogonality will only be maintained in the SP data for discrete choice analysis in some particular circumstances, even though the experiment design has already been forced to be orthogonal. This is because a number of choice studies have collected and included the data of non-design variables, (e.g. age and income), as the attributes in the estimating models. However, the covariate among these non-design variables is more likely to be non-orthogonal, as well as the covariate between these variables and the design attributes. For instance, if the analyst has introduced income and age as attributes to the estimating models, correlation is not only likely to appear between age and income but also between these non-design attributes and other attributes (Rose and Bliemer, 2009). In addition, it is easy to introduce implausible or dominant alternatives into the choice tasks in an orthogonal design (Hensher and Barnard, 1988). Therefore, it is necessary to remove these implausible or dominant alternatives that would break the design orthogonality (Hensher et al., 2015b). Due to these issues, an orthogonal design may not be appropriate for discrete choice models, especially when using non-linear models. In line with this, many researchers have argued over whether orthogonal design is appropriate for the discrete choice analysis using SP data (Rose and Bliemer, 2009, 2013; Iles and Rose, 2014). The key argument is that the desirable properties of logit models, (e.g. MNL and NL models), may be detracted by using orthogonal designs (Rose and Bliemer, 2009).

#### 5.3.3.6 Efficient design

Efficient design endeavours to minimise the elements of the AVC matrix of discrete choice models. As a result, it yields more reliable estimates of parameters for a fixed sample size. McFadden (1973) proposed that the AVC matrix for discrete choice models is different to that of linear models. For the discrete choice model, the Fisher Information matrix ($I_N$) is derived by taking the negative second derivatives of the log-likelihood function corresponding to the discrete choice model. Then the AVC matrix ($\Omega_N$) can be generated by taking the inverse of the $I_N$, where $N$ represents the number of respondents. An example of the determination of the AVC matrix corresponding to the MNL model with generic parameters is given below. The log-likelihood function of an MNL model can be calculated by the function in Equation 5-1:

$$L_N(\beta \mid X, y) = \sum_n^N \sum_s^S \sum_j^J y_{nsj} \log P_{nsj}(X \mid \beta) \qquad \text{5-1)}$$

where each respondent $n$ is required to answer $S$ choice tasks, each choice task $s$ has $J$ alternatives, $X$ is the attribute related to alternative $j$, in the choice task the value of $X$ is the attribute-level of itself. $\beta$ is the parameter corresponding to the attribute, $y_{jsn}$ is the SP survey observations and is 1 if the respondent chooses alternative $j$ in choice task $s$, otherwise is equal to 0. $P_{jsn}$ is the probability that respondent $n$ chooses alternative $j$ in choice task $s$, which can be calculated in the MNL model. One notable point here is that while estimating the likely AVC matrix, the set of parameter estimates $\beta$ are not known by the analyst. Therefore, a priori estimates of the parameters $\beta$ (also referred as parameter priors) are required to generate the AVC matrix (Bliemer et al., 2009; Bliemer and Rose, 2011; Hensher et al., 2015a). There are different ways to identify the parameter priors, for example from previous literature, focus groups or pilot studies (Bliemer and Collins, 2016).The parameter priors can also be set to zero but zero values may led to relatively large differences from the true values, therefore reducing the efficiency of the design (Bliemer et al., 2009). The first derivative of the log-likelihood function is given by Equation 5-2:

$$\frac{\partial L_N(\beta \mid X, y)}{\partial \beta_k} = \sum_{n=1}^N \sum_{s=1}^S \sum_{j=1}^J (y_{nsj} - P_{nsj}(X \mid \beta)) X_{knsj} \qquad \text{5-2)}$$

It’s second derivative yields the Fisher Information matrix per Equation 5-3,

$$I_N = \frac{\partial^2 L_N(\beta \mid X)}{\partial \beta_{k_1} \partial \beta_{k_2}} = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} X_{k_1 nsj} P_{nsj}(X \mid \beta)(X_{k_2 nsj} - \sum_{i=1}^{J} X_{k_2 nsi} P_{nsi}(X \mid \beta))$$

5-3)

where, in the second derivative of the log-likelihood function, the $y_{jsn}$ drops out. Thus, by considering alternatives, attribute-levels and parameters priors in the MNL model, the resultant choice probabilities $P_{jsn}$ can be calculated. Once the Fisher Information matrix has been generated, the AVC matrix ($\Omega_N$) of the MNL model with respect to $N$ respondents can be determined. Equation 5-4 below shows the model AVC matrix.

$$\Omega_N(\beta \mid X) = -I_N^{-1}(\beta \mid X) = -\frac{1}{N} I_1^{-1}(\beta \mid X) = \frac{1}{N} \Omega_1(\beta \mid X) \qquad 5\text{-}4)$$

where $\Omega_1$ is the AVC matrix based on a single respondent. The efficiency of the design can be then measured based on the AVC matrix. The most widely used way to measure the design efficiency reported in the literature is *D*-error, which can be computed from the determinant of the model's AVC matrix and using the number of parameters to scale this determinant value (Rose et al., 2008). The number of respondents was assumed to be 1 ($N$=1) while measuring the *D*-error statistic of the MNL model (Hensher et al., 2015b). Thus, it is given as in Equation 5-5 below,

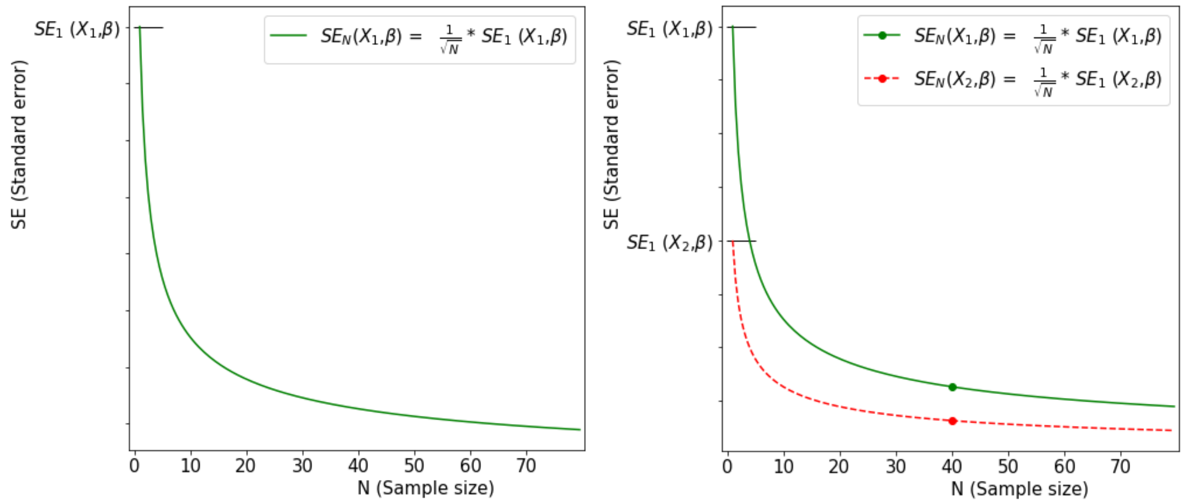
$$D - error = \det(\Omega_1)^{1/k} \qquad 5\text{-}5)$$

where $k$ is the number of parameters that can show the AVC matrix size. Hence, for the efficient design, the elements of AVC matrix can be minimised by minimising the corresponding *D*-error. This kind of design is also called *D*-efficient design. The lower the *D*-error, the higher the efficiency of the design. In other words, the expected asymptotic standard errors of an MNL model can be minimised if the *D*-error statistic for the design is minimised. Generally, the main merits of efficient design are twofold. Firstly, it reduces the asymptotic standard errors as the diagonal elements in AVC matrix are minimised, thereby increasing the *t*-values corresponding to the parameter estimates of the model, and thus the reliability of the modelling outputs are improved. Secondly, the confidence intervals of the parameter estimates become narrower due to the minimised *D*-error statistic. Due to these two benefits, it is possible to decrease the sample size but still maintain significance of the *t*-values. As shown in Equation 5-4, the AVC matrix of an MNL model is divided by the sample size $N$, (assuming each respondent completed all choice tasks). Therefore, the asymptotic standard errors can be derived, as the diagonal elements of the AVC matrix contribute to the variance of the parameter estimates. Equation 5-6, proposed by Rose and Bliemer (Rose and Bliemer, 2009, 2013), shows the standard error function of the discrete choice model.

$$se_N(\beta \mid X) = \frac{1}{\sqrt{N}} se_1(\beta \mid X) \tag*{5-6)}$$

Based on the standard error function, for a specified efficient design *X*, a given change of sample size will influence how much of the asymptotic standard error can be recognised. Additionally, based on the function, there are two effective ways to pre-minimise the expected asymptotic standard errors. Firstly, the graph in Figure 5-3a shows the function of asymptotic standard error against sample size, as a specified efficient design $X_1$ has been given. It is clear that increasing the sample size can reduce the asymptotic standard errors. It also shows that if the sample size is increased to a certain limit, the decrease in elasticity of asymptotic standard errors will be dramatically reduced. Secondly, the function graph of asymptotic standard error against sample size based on two different efficient designs (efficiency of $X_2$ is higher than $X_1$) as shown in Figure 5-3b, demonstrates that for a fixed sample size, (e.g., $N$=40), the standard error of an efficient design with a higher efficiency, (smaller *D*-error), is smaller than that of an efficient design with a lower efficiency. Apart from that, based on Equation 5-6, Rose and Bliemer (2013) indicated that the asymptotic *t*-values can be calculated based on the prior parameter estimates, as shown in Equation 5-7,

$$t_k = \frac{\beta_k}{\sqrt{se_1(\beta_k)^2 / n_k}} \tag*{5-7)}$$




*(a)* *(b)*

*Figure 5-3 Efficiency of design with respect to (a) sample size, and (b) different designs*

Therefore, assuming the prior parameters are correct, it is possible to determine the theoretical minimum sample size. For instance, the asymptotic *t*-values should be greater than 1.96, thus the 95% significance level of parameter estimates can be achieved. The equation determining minimum sample size can be derived based on rearranging the *t*-ratios as given in Equation 5-8 (Rose and Bliemer, 2013),

$$N_k \geq \left(\frac{1.96 \cdot se_1(\beta_k)}{\beta_k}\right)^2 \qquad 5\text{-}8)$$

It is noteworthy that Equation 5-8 calculates the minimum sample size ($N_k$) for corresponding parameter $\beta_k$ individually. Hence the analyst will need to calculate it for each parameter with the largest being the actual minimum sample size required. In addition, several other rules of thumb proposed by Orme (1998), Pearmain and Kroes (1990) and Lancsar and Louviere (2008) are of importance and need to be considered. These are introduced in section 5.4.3.5.

Many previous studies have found that efficient designs outperform traditional orthogonal designs in terms of providing better parameter estimates and higher levels of significance corresponding to the *t*-values. For example, Rose and Bliemer (2013) and Iles and Rose (2014) noted that for non-zero parameter priors, the parameter estimates based on D-efficient designs were more reliable than those based on random orthogonal designs, and the efficient designs could estimate the parameters at a statistically significant level with much smaller sample sizes. Ferrini and Scarpa (2007) found similar results for efficient designs. Bliemer and Rose (2011) also found that efficient designs could empirically produce better *t*-values compared to orthogonal designs. Therefore, this chapter generates a D-efficient experimental design for estimating individuals' travel mode choice for regional trips within Western Australia.

### 5.3.4 Stage 3: Generate experimental design

#### 5.3.4.1 Existing Row-based and Colum-based algorithms

Although attribute-level balance has been mentioned as a property of experimental design, it should be stressed that it is not mandatory for developing an efficient experimental design (Hensher et al., 2015b). Furthermore, a more efficient design may be found if there is no attribute-level balance. As Hensher et al. (2015b) indicated, the most basic and direct way to generate the most D-efficient design is to generate the full factorial design first, then extract each possible fractional factorial design and calculate the corresponding asymptotic *D*-error. The one with the lowest *D*-error is the most efficient design. However, this approach is too difficult to be implemented if the full factorial design is too large. For example, consider a labelled design with 4 alternatives, with each alternative having 4 attributes and each attribute 3 levels. The full factorial design would have $3^{4*4}$=43,046,721 choice tasks, making it practically impossible to determine all the possible fractional factorial designs. Instead of the basic method, two kinds of algorithms including row-based and column-based algorithms (Huber and Zwerina, 1996; Sandor and Wedel, 2001) have been applied widely by researchers to find the most efficient design. The row-based algorithm refers to finding the most efficient design from a pre-generated candidate choice task, (either a fractional factorial or full factorial design), based on pre-defined finite iterations (Hensher et al., 2015b). The column-based algorithm refers to "*creating a design by selecting attribute-levels over all choice situations for each attribute*" (Hensher et al., 2015b, p. 252). However, the row-based algorithm is more effective in finding a realistic experiment since the potentially unrealistic choice tasks can be priority filtered, whereas the column-based algorithm is easier to satisfy attribute-level balance.

The MFA is the most widely used row-based algorithm (Cook and Nachtrheim, 1980). The procedure was summarised by Hensher et al. (2015b) and, as shown in Figure 5-4, the first step in the procedure is to generate a candidate set, either a fractional factorial for complex choice problems or a full factorial for simple choice problems. After that, a design with balanced attribute-levels is developed by picking up choice situations from the candidate set. Next, the efficiency error, (*D*-error for this study), is calculated for the created design. The last step is to compare the efficiency error of the newly generated design with that of the current best design. If the *D*-error of the new design is found to be smaller, the current best design is replaced by this new one. The algorithm keeps iterating to find the most efficient design, (the smallest *D*-error), until all possible combinations of choice situations in the candidate set have been evaluated. However, it is not feasible to allow the algorithm to find and evaluate all the possible combinations of choice tasks. Therefore, the algorithm is usually set to terminate after a pre-specified finite number of iterations (Hensher et al., 2015b).

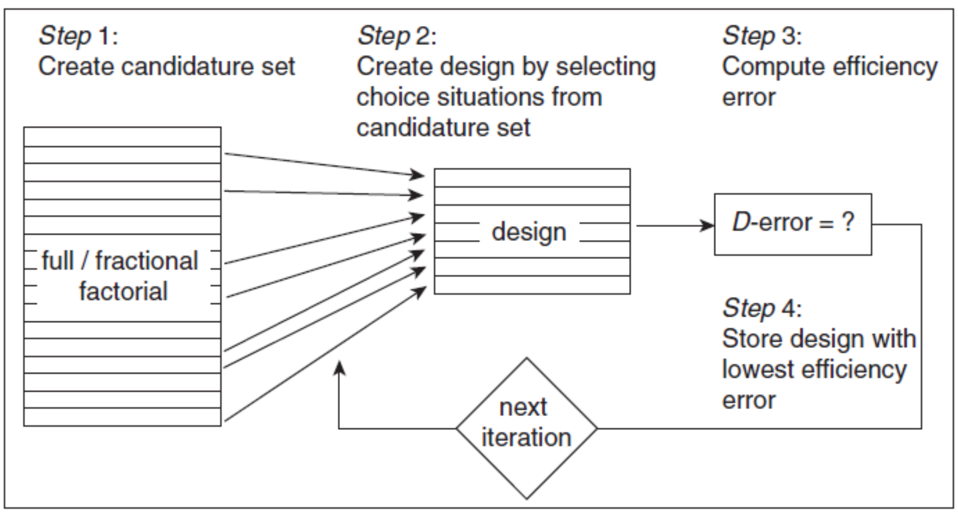



*Figure 5-4 Modified Federov Algorithm (source from Hensher et al. (2015b, p. 252).)*

### 5.3.4.2 Extension of the Modified Federov Algorithm

This chapter extends the MFA by adding steps to find and set up all required constraints for creating a relatively more realistic experimental design using a semi-automatic process. These seven steps, shown in Figure 5-5, address issues related to rejecting implausible attribute-levels for chosen alternatives, avoiding dominant alternatives and rejecting unrealistic choice tasks.

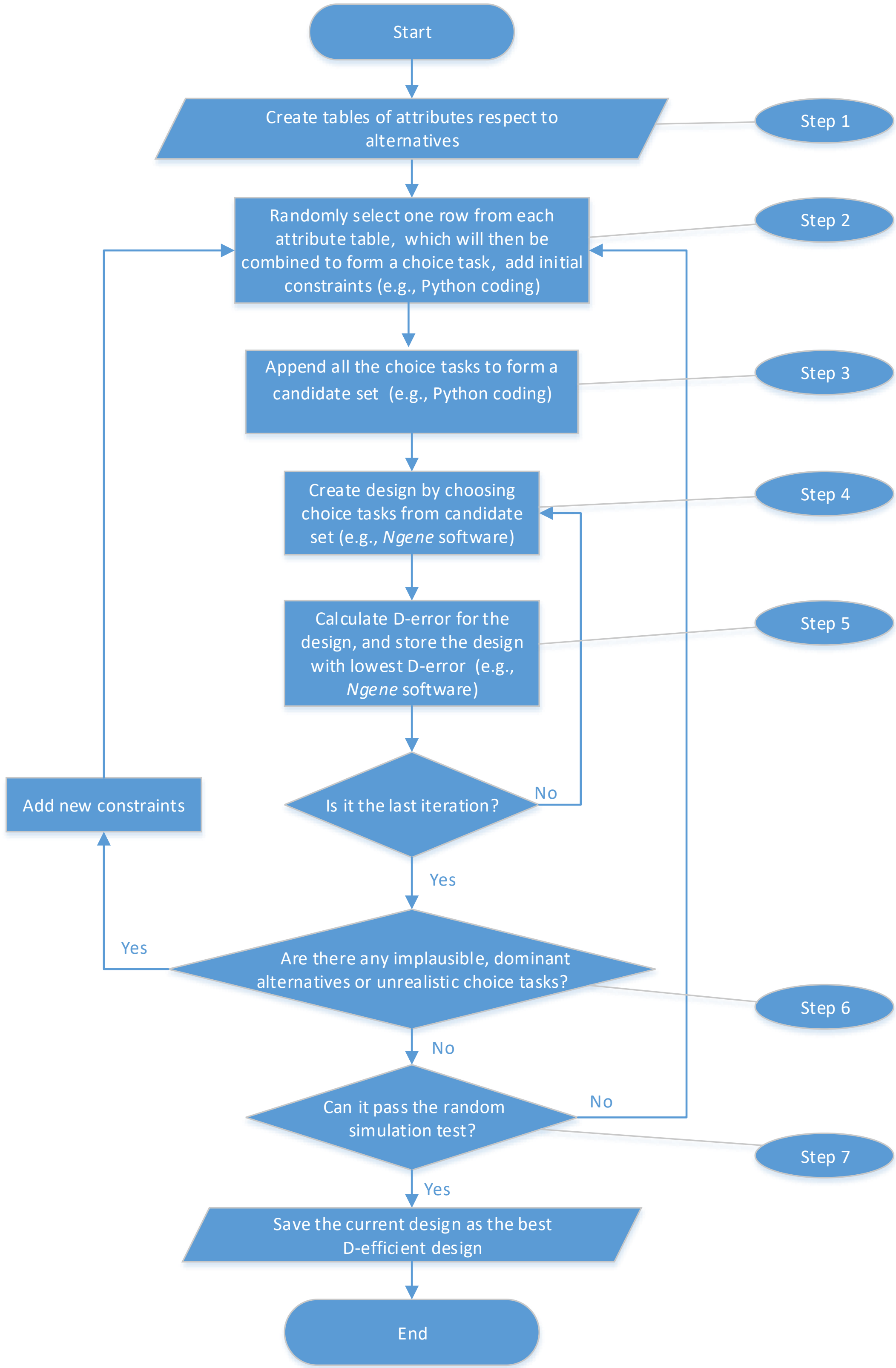



*Figure 5-5 Extension of the Modified Federov Algorithm for stage three*

The first step is to use attribute-levels to create an attribute table for each alternative, where each table contains all possible correlated attribute-level combinations. For example, in the regional travel mode alternatives of bus and airline, both alternatives may have only two attributes – ticket price and journey time. Attribute-levels of bus ticket price are assumed to be A$100 and A$300, and journey times are 4 hours and 14 hours. Attribute-levels of airline ticket price are A$200 and A$400, and journey times are 1 hour and 3 hours. Tables 5-3 and 5-4 illustrate all possible attribute-level combinations of ticket price and journey time respectively for travel mode alternatives bus and airline.

*Table 5-3 Attribute table of ticket price*

| | **Alternative: Bus** | **Alternative: Airline** |
|---|---|---|
| **Level combination of ticket price** | **Ticket price** | **Ticket price** |
| 1 | A$100 | A$200 |
| 2 | A$100 | A$400 |
| 3 | A$300 | A$200 |
| 4 | A$300 | A$400 |

*Table 5-4 Attribute table of journey time*

| | **Alternative: Bus** | **Alternative: Airline** |
|---|---|---|
| **Level combination of journey time** | **Journey time** | **Journey time** |
| 1 | 4 hours | 1 hour |
| 2 | 4 hours | 3 hours |
| 3 | 14 hours | 1 hour |
| 4 | 14 hours | 3 hours |

In the second step, *Python* coding was used to randomly select an attribute-level combination, (a row in the table), from each attribute table and combine them to form a complete choice task. However, it may not be possible to determine all the required constraints at the beginning, especially for complex problems. Thus, some initial constraints are added to try to prevent problematic choice tasks being constructed. One important reason for using steps one and two to generate a choice task is that some complex choice problems may have a huge number of choice tasks, (millions, billions or more). In such cases, randomly choosing choice tasks from the full factorial, instead of combining attribute-levels from attribute tables to form choice tasks, may lead to some selected choice tasks being very similar and thus losing the attribute-level variation. For instance, in Table 5-5, the choice tasks 1, 2, 3 and 4 are quite similar since all the attribute-levels are the same except attribute 3 of alternative B, which significantly reduces the attribute-level variation among choice tasks. Including such similar choice tasks to form candidate sets may lead to the frequency of some attribute-levels being significantly reduced or even not included. Therefore, the difficulty in finding attribute-level balanced design increases. Steps one and two can help to avoid this issue. This is because the number of level combinations of each attribute table is finite and not that large, hence, randomly choosing level combinations from these tables to integrate the choice task can assist in avoiding similar choice tasks.

*Table 5-5 Similar choice tasks*

| | Alternative A | | | Alternative B | | |
|---|---|---|---|---|---|---|
| **Choice task** | **Attribute 1** | **Attribute 2** | **Attribute 3** | **Attribute 1** | **Attribute 2** | **Attribute 3** |
| 1 | 1 | 2 | 4 | 3 | 1 | 2 |
| 2 | 1 | 2 | 4 | 3 | 1 | 3 |
| 3 | 1 | 2 | 4 | 3 | 1 | 4 |
| 4 | 1 | 2 | 4 | 3 | 1 | 5 |

"*The integration of constraints into stated choice experimental designs requires an unambiguous specification of what constraints are required*" (Collins et al., 2014, p. 7). The efficient design approach is likely to produce some dominant and some implausible alternatives, alternatives that are not appropriate to be included in the choice tasks, as they might reduce the reliability and significance of the parameter estimates (Rose and Bliemer, 2009; Hensher et al., 2015b). As indicated by Collins et al. (2014) and Cherchi and Hensher (2015), these alternatives should be removed from the experimental design by adding specified constraints, even though this may result in attribute-level imbalance. As mentioned earlier, the presence of unrealistic choice tasks is another issue that often arises in the experimental design field (Collins et al., 2014). The constraints for removing implausible alternatives, dominant alternatives and unrealistic choice tasks are classified as Plausibility, Dominance and Realism (PDR) constraints. The PDR constraints can be specified using *Python* conditionals. Three main types of PDR constraints and corresponding pseudocodes are summarised in Table 5-6 using this case study example, where A, B, C and D are logical expressions. One noteworthy point here is that the PDR constraints should be defined based on the attribute-levels and real-life situations.

*Table 5-6 Plausibility, Dominance and Realism constraints specification*

| **PDR constraints** | **General Pseudocode** | **Pseudocode for specifying the PDR constraints** |
|---|---|---|
| Cost by car smaller than cost by airline | *IF A is true THEN*<br>*SAVE*<br>*ENDIF* | *IF cost (driving) < cost (airfare) THEN*<br>*SAVE the choice task*<br>*ENDIF* |
| Cost per in-vehicle hour by car smaller than A$15/hr, larger than A$5/hr | *IF B is true THEN*<br>*SAVE*<br>*ELSE*<br>*REJECT*<br>*ENDIF* | *IF 5 <cost (driving) /journey time (car)< 15 THEN*<br>*SAVE the choice task*<br>*ELSE*<br>*REJECT the choice task*<br>*ENDIF* |
| No dominant alternative between airline 1 and airline 2, with respect to three attributes: travel cost, journey time and service frequency | *IF C is true THEN*<br>*REJECT*<br>*ELSE IF D is true*<br>*REJECT*<br>*ELSE*<br>*SAVE*<br>*ENDIF* | *IF cost (airline 1 ticket fare) < cost (airline 2 ticket fare) AND journey time (airline 1) < journey time (airline 2) AND frequency (airline 1) > frequency (airline 2) THEN*<br>*REJECT the choice task*<br>*ELSE IF cost (airline 2 ticket fare) < cost (airline 1 ticket fare) AND journey time (airline 2) < journey time (airline 1) AND frequency (airline 2) > frequency (airline 1)*<br>*REJECT the choice task*<br>*ELSE*<br>*SAVE the choice task*<br>*ENDIF* |

The third step appends the filtered choice tasks to form a candidate set, with each choice task in the candidate set satisfying the imposed PDR constraints. The size of the candidate set depends upon the number of choice tasks predefined by the researcher. The fourth step is then to create an experimental design by choosing choice tasks from the candidate set. The number of choice tasks refers to a pre-defined experiment size *S*, such as 12, which includes a combination of different attribute-levels across all the alternatives and satisfies the attribute-level balance as much as possible. The fifth step is to evaluate the experimental design from the fourth step using a *D*-error method. As it is an iterative process, a number of designs are created, with a *D*-error computed for each design. The best efficient design is defined as the one with the smallest *D*-error. The number of iterative processes for deriving the best design is predefined. In this research, it has been set to 20,000 iterations. Once the most efficient design has been generated, the sixth step checks whether there are any implausible or dominant alternatives within the choice tasks and if there are any unrealistic choice tasks. If the design does have any of these problematic choice tasks, new constraints that can reject these problematic choice tasks are generated and added to step two, and steps two to six are rerun. This iterative process ends when all choice tasks in the design are realistic, no dominant or implausible alternatives remain, and the D-error is optimised. The seventh step is to test if the identified design from step six can pass a random simulation test, in which random response data is generated and regressions explored to see if coefficients tend to be statistically significantly different from zero more frequently than would be expected by random chance. If this test fails, the procedure goes back to step two and the iterative process is restarted; otherwise, the resulting D-efficient experimental design can be used for the study.

In this chapter, step 1, 2 and 3 used the *Python* programming language skills and steps 4 and 5 were run using *Ngene* software. The random simulation test (step 7) was implemented using *Python* and *NLOGIT 5.0* (Greene, 2012). Step 6 is the only non-automatic step that requires a manual check for plausibility and realism of the generated choice tasks. The following section discusses the implementation of the efficient and realistic design method using a case study of travel mode and airline choice in regional Western Australia.

## 5. 4 Implementing D-efficient experimental design for this research

### 5.4.1 Choice criteria identification

For this case study, four main travel modes were considered as the labelled alternatives; car, bus and two regional airlines, (airline 1 and airline 2). The reason to include two airlines as alternatives is that most regional RPT airports in Western Australia only have at most two airlines operating on one regional air route. The two airline services are similar, thus both airlines were considered in order to make the SP experiment more reflective of actual choices faced in the real world and to make the questions represent the real situation more closely. The option of train was not considered in this study as it is not widely available in Western Australia (only Perth to Kalgoorlie has a passenger train service). Previous literature can provide insights into identifying attributes (Klojgaard et al., 2012), showing that travel time, travel cost, accessibility, service frequency and seat comfort are the major determinants of travellers' mode choice behaviour (Jovicic and Hansen, 2003; Hess and Polak, 2006a; Chang and Sun, 2012; Van Can, 2013; Jung and Yoo, 2014; Chen and Chao, 2015; Inoue et al., 2015; Qiao et al., 2016). Thus, these five key factors were set as the attributes for constructing the D-efficient design. The corresponding attribute levels were defined based on the regional air and non-air travel data of Western Australia by Rome2rio (extracted from https://www.rome2rio.com), which is a widely used worldwide online multimodal transport search engine. It provides data on travel time, travel cost and service frequency for a variety of domestic and international trips. Table 5-7 specifies the attribute-levels for these five attributes and four alternative travel modes in the Western Australia context. One notable point here is that the seat comfort can be measured depending on different aspects, (e.g. leg room, seat width and seat material); in this study leg room distance was used to measure seat comfort levels, as a proxy.

*Table 5-7 Dimensions for generating the efficient design of this case study*

| | **Car** | **Bus** | **Airline 1** | **Airline 2** |
|---|---|---|---|---|
| Travel cost (A$) | 25, 150, 275, 400 | 50, 175, 300, 425 | 200, 350, 500, 600 | 200, 350, 500, 600 |
| Access time (mins) | n/a | 15, 30, 45, 60 | 15, 30, 45, 60 | 15, 30, 45, 60 |
| Journey time (hours) | 3, 12, 21, 30 | 5, 15, 25, 35 | 1, 2, 3, 4 | 1, 2, 3, 4 |
| Frequency (weekly) | n/a | 2, 16, 30, 44 | 2, 16, 30, 44 | 2, 16, 30, 44 |
| Seat comfort level | Middle[b], High[c] * | Low[a], Middle[b], High[c] | Low[a], Middle[b], High[c] | Low[a], Middle[b], High[c] |

*Table notes:*
*[a] A low level of seat comfort, and leg room distance is 70 cm, [b] a middle level of seat comfort and leg room distance is 80 cm, and [c] a high level of seat comfort and leg room distance is 90 cm. * The seat comfort level of car was pre-restricted to start from the middle value, since the driving seat of a car can normally be adjusted flexibly.*

### 5.4.2 Model specification and design consideration

#### 5.4.2.1 Model specification

The D-efficient experimental design was based on a main-effect only MNL model with generic coefficients for the five key attributes/factors, as the LC model was performed as a semi-parametric variant of the MNL model in this thesis. Since the interest of this research is on how the different factors influence traveller mode choice rather than whether there are any minor differences within any specific/same factor, the same coefficients were used for the key factors across the travel modes rather than travel mode-specific coefficients. In line with this, generic coefficients were also used in a number of previous travel mode choice studies. For example, Jung and Yoo (2014) used a main-effect only and generic parameters utility function for estimating travel mode choice between air and non-air (high-speed train) travel modes. Inoue et al. (2015) applied a main–effect utility function with generic coefficients for travel cost, travel time and service frequency across the air and non-air (high speed train and bus) travel modes. Van Can (2013) set up generic parameters for all attributes such as travel cost, time and comfort across all three travel modes, (plane, coach and train). Equation 5-9 shows the observed utility function of the MNL model developed for the regional travel mode choice analysis.

$$\begin{aligned} V_j &= Constant_{mode} + \beta_1 TravelCost_j + \beta_2 AccessTime_j + \beta_3 JourneyTime_j \\ &+ \beta_4 Frequency_j + \beta_5 SeatComfortMid_j + \beta_6 SeatComfortHigh_j \end{aligned} \quad \text{5-9)}$$

where:

$V_j$ is the observed utility of alternative $j$,

$Constant_{mode}$ is the constant for car, bus and airline respectively, and accounts for the difference in the experience of different travel modes,

$TravelCost_j$ is the travel cost, (ticket fare or the cost of driving) to use alternative mode $j$ (A$),

$AccessTime_j$ is the access time to a bus station or an airport (mins),

$JourneyTime_j$ is the travel time from an origin to a destination (hours),

$Frequency_j$ is the number of operating buses or flights per week, and

*SeatComfortMid*$_j$ and *SeatComfortHigh*$_j$ are dummy variables representing middle and high seat comfort level, respectively.

#### 5.4.2.2 Design consideration

Since there are four alternatives defined for the experiment design, the degrees of freedom is $S{\cdot}(4-1)$ (where $S$ is the size of the experiment and refers to the number of choice tasks). In total 8 parameters (including 2 constants) were estimated in the utility function of the MNL model. Thus, the minimum experimental size $S$ required to meet the degrees of freedom requirement is 3 ($3S \geq 8$). Apart from that, Table 5-7 shows that the number of attribute-levels is 2, 3 or 4, giving a lowest common multiple of 12. Consequently, the experimental size was set to be 12, which not only meets the degrees of freedom requirement for the experiment design but also maintains the property of attribute-level balance. However, as it was considered too difficult for respondents to answer 12 questions with four alternatives, a blocking strategy was used as an efficient way to reduce the number of choice questions shown to each respondent. Blocking breaks down the design into several subsets, (also called blocks), with each subset requiring one respondent to complete it. Thus, the design was broken down into two blocks using the software package *Ngene* 1.2.0, with each block having 6 out of the 12 choice tasks.

#### 5.4.2.3 Priors specification

Parameter priors are required for the construction of an efficient experimental design, where relatively more accurate priors, (closer to true values of parameters), can help to improve the efficiency of the design (Bliemer et al., 2009; Bliemer and Rose, 2011; Hensher et al., 2015b). This thesis pre-defined the parameter priors based on previous mode choice related literature (Jovicic and Hansen, 2003; Hess and Polak, 2006a; Chang and Sun, 2012; Van Can, 2013; Jung and Yoo, 2014; Chen and Chao, 2015; Inoue et al., 2015; Bliemer and Collins, 2016; Qiao et al., 2016), as well as the pilot study and focus group discussions. Table 5-8 shows the pre-defined parameter priors information used for generating the efficient design.

Table 5-8 Parameter priors

| Generic parameters | Parameter priors |
|---|---|
| $\beta_1$ (Travel cost, Australian dollar) | -0.01 |
| $\beta_2$ (Access time, minutes) | -0.004 |
| $\beta_3$ (Journey time, hours) | -0.04 |
| $\beta_4$ (Service frequency, weekly) | 0.001 |
| $\beta_5$ (Seat comfort level-middle, dummy variable) | 0.2 |
| $\beta_6$ (Seat comfort level-high, dummy variable) | 0.3 |

### 5.4.3 Generate experimental design using EMFA method

#### 5.4.3.1 Creating attribute tables

Based on the attribute information reported in Table 5-7, five attribute tables corresponding to travel cost, access time, journey time, frequency and seat comfort in terms of car, bus, airline 1 and airline 2 were created using *Python* 3.6. Table 5-9 shows the pseudocode used to generate the attribute table of costs for the four alternatives.

Table 5-9 Pseudocode for generating attribute table of costs

**Generating attribute table of costs**

```
BEGIN
GET travel cost lists of car, bus, airline 1 and airline 2 AS car_cost, bus_cost, airline1_cost and airline2_cost
GET empty list of attribute cost AS attribute_table_of_cost
FOR each row IN car_cost list:
        FOR each row IN bus_cost list
                FOR each row IN airline1_cost
                        FOR each row IN airline2_cost
                                COMBINE row of car_cost list, row of bus_cost list, row of
                            airline1_cost list and row of airline2_cost list AS a new row
                            AND APPEND TO attribute_table_of_cost list
                      END FOR
              END FOR
       ENDFOR
ENDFOR
SAVE attribute_table_of_cost list AS an attribute table of cost in EXCEL file
END
```

#### 5.4.3.2 Generating candidate set

Once all the five attribute tables were created in step 1, the second step was to randomly select one row from each of the tables and combine them into one complete choice task. Some initial constraints, (as shown in Table 5-10), were added to restrict randomly selected attribute-level combinations in order to prevent implausible, dominant alternatives and unrealistic choice tasks. Each completed choice task was then appended to generate the candidate set, until the size of the candidate set reached 20,000 (20,000 choice tasks was predefined to form the candidate set). Table 5-11 is the *Python* pseudocode for constructing the candidate set; the details of constraints' pseudocode are not listed as the constraints are subject to different case studies.

*Table 5-10 Initial PDR constrains*

| Number | PDR constraints |
|---|---|
| 1 | Cost by car and bus lower than by airlines |
| 2 | Journey time by airlines shorter than by car or bus |
| 3 | Journey time by car shorter than by bus |
| 4 | Journey time by car or bus no more than 10 times that by airlines |
| 5 | No dominant alternative between the two regional airlines |

*Table 5-11 Pseudocode for generating candidate set*

***Generating candidate set***

```
BEGIN
READ the five attribute tables
GET empty list of candidate_set
SET number of choice tasks N=0
WHILE (number of choice tasks N< 20,000)
        DO RANDOMLY SELECT one row FROM each of the five tables
        COMBINE the rows to FORM a complete choice task row
        IF the complete choice task MEET initial PDR constraints
                APPEND the choice task row TO  candidate_set list
              N=N+1
      ELSEIF the complete choice task CANNOT MEET initial PDR constraints
             REJECT the complete choice task
      ENDIF
ENDWHILE;
```

*SAVE candidate_set list AS a table of candidate set in EXCEL file*
END

#### 5.4.3.3 Constructing and finding the most efficient design

The candidate set was generated in step 3, as shown in Table 5-12. Hence, in step 4 *Ngene 1.2.0* was used to create the efficient design by selecting the choice tasks from the candidate set, with respect to the MNL model utility function, (as given in Equation 5-9), developed for the travel mode choice study. The *D*-error relating to the created efficient design was then calculated in step 5 and the design with the lowest *D*-error stored in the *Ngene* software. In step 6, the number of iterations was set to 20,000 for finding the design with the smallest *D*-error. Table 5-13 shows the design (*D*-efficient design A) with smallest *D*-error (*D*-error=0.00504) found by the 20,000 iterations from the first candidate set.

*Table 5-12 Candidate set generated based on initial constraints*

| | Attributes of car | | | Attributes of bus | | | | | Attributes of airline 1 | | | | | Attributes of airline 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Choice task** | **Travel cost** | **Journey time** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Frequency** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Frequency** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Frequency** | **Seat comfort** |
| 1 | 275 | 21 | 1 | 425 | 60 | 25 | 2 | 1 | 650 | 60 | 3 | 2 | 2 | 350 | 45 | 3 | 44 | 0 |
| 2 | 150 | 21 | 2 | 425 | 15 | 25 | 30 | 2 | 650 | 45 | 4 | 16 | 2 | 650 | 60 | 4 | 30 | 0 |
| 3 | 275 | 21 | 1 | 175 | 30 | 35 | 2 | 2 | 650 | 60 | 4 | 16 | 2 | 350 | 45 | 4 | 30 | 0 |
| 4 | 150 | 12 | 1 | 175 | 60 | 15 | 44 | 0 | 350 | 60 | 2 | 30 | 0 | 650 | 15 | 2 | 44 | 0 |
| 5 | 25 | 3 | 1 | 175 | 30 | 5 | 30 | 1 | 200 | 45 | 1 | 2 | 0 | 200 | 60 | 1 | 44 | 0 |
| 6 | 150 | 21 | 2 | 175 | 15 | 25 | 44 | 0 | 350 | 60 | 4 | 2 | 2 | 350 | 30 | 3 | 30 | 1 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| 20000 | 25 | 30 | 2 | 175 | 30 | 35 | 2 | 1 | 500 | 30 | 4 | 44 | 0 | 200 | 30 | 4 | 2 | 0 |

*Table 5-13 Efficient design A generated based on initial constraints*

| | Attributes of car | | | Attributes of bus | | | | | Attributes of airline 1 | | | | | Attributes of airline 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Choice task** | **Travel cost** | **Journey time** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Frequency** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Frequency** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Frequency** | **Seat comfort** | **Block** |
| 1 | 150 | 3 | 1 | 175 | 15 | 5 | 30 | 2 | 200 | 60 | 1 | 44 | 0 | 200 | 45 | 1 | 2 | 1 | 2 |
| 2 | 275 | 12 | 1 | 425 | 15 | 15 | 16 | 2 | 650 | 30 | 2 | 44 | 2 | 500 | 15 | 2 | 30 | 0 | 1 |
| 3 | 275 | 3 | 1 | 175 | 60 | 5 | 44 | 2 | 350 | 15 | 1 | 16 | 2 | 500 | 45 | 1 | 44 | 2 | 2 |
| 4 | 400 | 12 | 1 | 300 | 30 | 35 | 30 | 0 | 500 | 45 | 2 | 44 | 1 | 650 | 30 | 2 | 16 | 2 | 1 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| 12 | 25 | 21 | 2 | 50 | 15 | 25 | 2 | 0 | 200 | 45 | 4 | 44 | 1 | 650 | 15 | 3 | 2 | 0 | 2 |

**D-error:** 0.00504

**Sample size estimate:** 122.6

### 5.4.3.4 Realism inspection

After generating the first efficient design (design A), implausibility, dominance of alternatives and realism of the twelve choice tasks were checked. Several problems were found. For example, the alternative car in choice task 3 was found to be implausible as the journey time was 3 hours and the driving cost (fuel cost) was $275. Furthermore, choice task 4 was found to be unrealistic as it had the journey time by bus (35 hours) nearly three times longer than that by car (12 hours). Based on these outcomes, more constraints were added, such as, '*cost difference between car and bus not more than $100'* and the design process re-run from steps 2 to 6. Better choice tasks were obtained, however, problems were still detected. This process was repeated until no implausible or dominant alternatives and no unrealistic choice tasks remained, with a *D*-error of 0.00576 (design B). Table 5-15 shows the optimal efficient design B and Table 5-14 illustrates the final constraints imposed, including the additional constraints generated from the multiple loops of steps 2 to 7.

*Table 5-14 Full PDR constraints*

| **Number** | **PDR constraints** |
|---|---|
| 1 | Cost by car and bus lower than by airlines |
| 2 | Journey time by airlines shorter than by car or bus |
| 3 | Journey time by car shorter than by bus |
| 4 | Journey time by car or bus no more than 10 times that by airlines |
| 5 | No dominant alternative between two regional airlines |
| 6 | Cost difference between car and bus not more than $100 |
| 7 | Cost difference between car and airlines, and between bus and airlines less than $600 |
| 8 | Journey time difference between car and bus not more than 10 hours |
| 9 | Journey time difference between two airlines not more than 2 hours |
| 19 | Ratio of cost by car to its journey time less than 15, but more than 5 |
| 11 | Ratio of cost by bus to its journey time less than 15, but more than 5 |
| 12 | Ratio of cost by airline to its journey time less than 500, but more than 50 |

*Table 5-15 Efficient design B generated based on full constraints*

| | **Attributes of car** | | | **Attributes of bus** | | | | | **Attributes of airline 1** | | | | | **Attributes of airline 2** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Choice task** | **Travel cost** | **Journey time** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Freque ncy** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Freque ncy** | **Seat comfort** | **Travel cost** | **Access time** | **Journey time** | **Freque ncy** | **Seat comfort** | **Block** |
| 1 | 400 | 30 | 2 | 425 | 30 | 35 | 16 | 2 | 500 | 60 | 4 | 30 | 0 | 650 | 60 | 4 | 44 | 1 | 1 |
| 2 | 400 | 30 | 1 | 425 | 15 | 35 | 2 | 1 | 650 | 45 | 4 | 44 | 0 | 500 | 15 | 4 | 16 | 2 | 2 |
| 3 | 275 | 30 | 1 | 300 | 60 | 35 | 30 | 0 | 650 | 60 | 4 | 30 | 2 | 650 | 45 | 4 | 44 | 0 | 2 |
| 4 | 275 | 30 | 2 | 300 | 30 | 35 | 16 | 2 | 350 | 15 | 4 | 2 | 0 | 350 | 60 | 4 | 44 | 1 | 1 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| 12 | 150 | 12 | 2 | 175 | 15 | 15 | 30 | 1 | 350 | 30 | 2 | 16 | 1 | 200 | 15 | 2 | 2 | 0 | 1 |

**D-error:** 0.00576

**Sample size estimate:** 123.8

#### 5.4.3.5 Sample size

For design B, the *Ngene 1.2.0* software was used to calculate the minimal theoretical sample size required to satisfy a two-tailed significance level of 0.05, (*t*-ratios is 1.96), corresponding to each pre-defined parameter prior, as shown in Table 5-16 (largest minimum sample size estimate=123.8). Thus, assuming each respondent was to complete all 12 choice tasks, 124 respondents would be the smallest sample size required for the design. However, the thesis pre-defined breaking down the experiment design into two blocks. Therefore, the theoretical minimum sample size of respondents required for the SP experimental design is 248, (124 x 2).

*Table 5-16 Theoretical minimum sample size*

| | Travel cost | Journey time | Seat comfort middle (dummy) | Seat comfort high (dummy) | Access time | Service frequency |
|---|---|---|---|---|---|---|
| Prior value | -0.01 | -0.04 | 0.3 | 0.2 | -0.004 | 0.01 |
| Minimal sample size $N$ | 2.2 | 13.5 | 37.4 | 83.8 | 123.8 | 24.1 |

Another widely used ‘rule of thumb’ for the minimum sample size of an SP choice experiment, proposed by Orme (1998), suggests that the sample size requirement for main effects estimation can be derived from equation 5-10:

$$N \geq \frac{500 \cdot l}{J \cdot S} \qquad \text{5-10)}$$

where $l$ is the highest number of levels of any attribute, $J$ is the number of alternatives and $S$ is the number of choice questions. Therefore, the minimum sample size from this rule of thumb is 42 ((500 x 4) / (4 x 12)), which again should be doubled due to the use of two blocks, giving a minimum sample size of 84. The rule of thumb proposed by Lancsar and Louviere (2008) suggests a minimum of 20 respondents for one choice task, whereas Pearmain and Kroes (1990) noted that 100 is the borderline sample size for choice modelling analysis. The survey of this thesis, (used in Chapters 8 and 9), exceeded the requirement of the theoretical minimum sample size (248) and these rules of thumb.

#### 5.4.3.6 Random simulation

The random simulation test aims to identify whether there are any biases within the experimental design. It assumes that each simulated respondent has an equal chance/probability of choosing any of the travel mode alternatives, i.e., random selection. Based on these random simulated data, the test is whether any of the estimated model parameters are insignificant. The outputs of the discrete choice modelling prove that none of the estimated parameters are significant. *Python 3.6* was used to simulate the answers of 400 respondents to the choice tasks, with every simulated respondent set to answer one block of the SP choice tasks (randomly making the choice). Table 5-17 shows the *Python* pseudocode for the random simulation.

*Table 5-17 Pseudocode for generating random simulation results table*

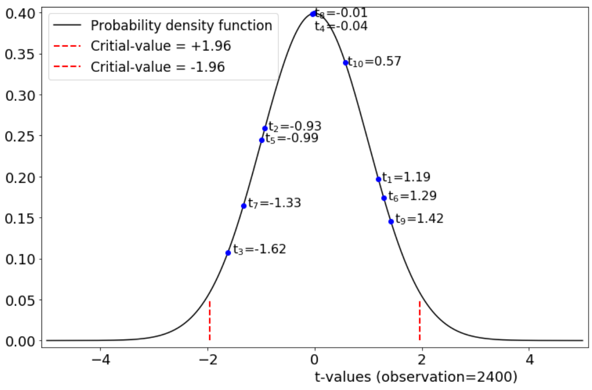
| ***Generating random simulation results table*** |
|---|
| *BEGIN* |
| *GET empty list of SP_survey_simulated_results* |
| *SET number of simulated respondent N=0* |
| *WHILE (number of simulated respondent N < 400)* |
|     *READ experimental design of choice tasks/questions* |
|     *FOR every choice task IN experimental design* |
|         *DO RANDOMLY selecting an alternative AMONG the four alternatives in the choice task* |
|         *APPEND randomly selection results TO the list of SP_survey_simulated_results* |
|         *N=N+1* |
| *ENDWHILE* |
| *SAVE SP_survey_simulated_results list AS a table in EXCEL file* |
| *END* |

After the simulated SP survey results table was generated, *NLOGIT* software was used to analyse the data based on the MNL model, (see utility function of Equation 5-9). The MNL modelling results of the random simulation are illustrated in Table 5-18, where the estimated parameters and the corresponding *t*-values (in parentheses) are reported. All the parameter estimates are statistically insignificant, since no *t*-values are significant at the 95% confidence level, or even at the 90% confidence level. In other words, the results indicate that all the attributes in terms of travel cost, access time, journey time, service frequency and seat comfort (dummy variable) were found to have no influence on people’s travel mode choice-making.

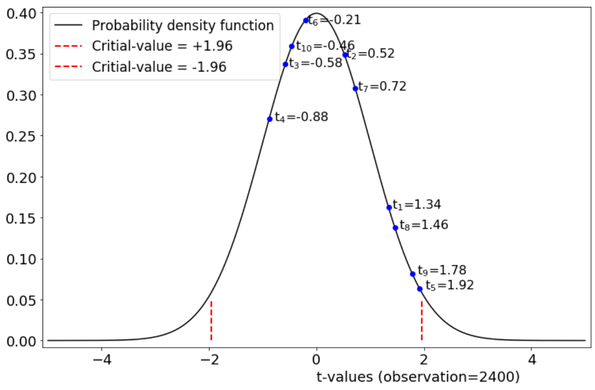
*Table 5-18 MNL modelling results based on the random simulation*

| Parameter | Multinomial logit model |
|---|---|
| | Simulation group |
| | Coefficient |
| Observation | 4800 |
| Constant (bus) | -0.0671 |
| | (-0.96) |
| Constant (airline 1 and airline 2) | -0.0467 |
| | (1.58) |
| **Variables** | |
| Travel cost (A$) | 0.0003 |
| | (1.19) |
| Access time (min) | 0.0010 |
| | (0.88) |
| Journey time (hour) | 0.0044 |
| | (1.34) |
| Service frequency (weekly) | 0.0017 |
| | (1.50) |
| Seat comfort_Middle (true=1, otherwise=0) | 0.0352 |
| | (0.82) |
| Seat comfort_High (true=1, otherwise=0) | 0.0326 |
| | (0.77) |
| **Model fit statistics** | |
| LL($\beta$): Log likelihood function | -6650.6 |
| AIC | 13317.2 |

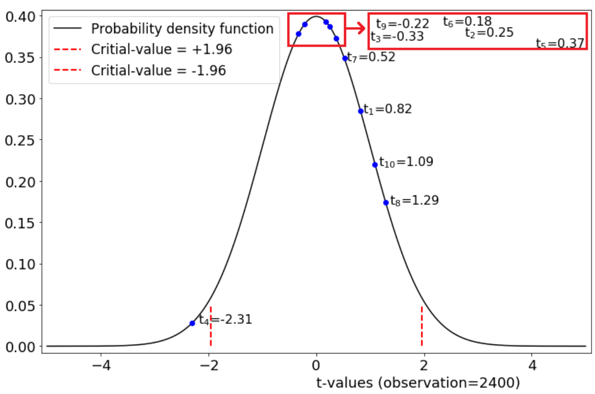
The random simulation test was run ten times, with the parameter estimates of the attributes recorded each run. As shown in Figure 5-6 (a to f), all the attribute parameter estimates computed in *NLOGIT* with respect to each of the 10 randomly simulation tests are statistically insignificant, (i.e., *t*-values are insignificant), except for the parameter estimate of mid-level seat comfort in the fourth simulation test, (Wald *t*-value $t_4$ in Figure 5-6c = -2.31 < -1.96). Therefore, the best *D*-efficient design B passed the random simulation test, and hence was used in this thesis.



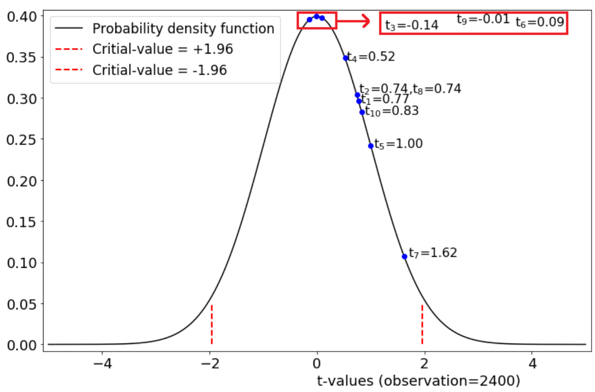

(a)



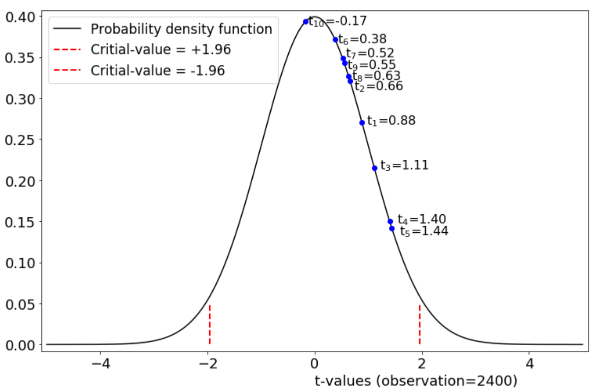

(b)



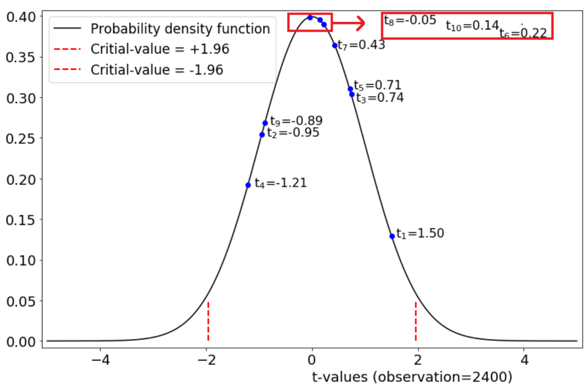

(c)




(d)




(d)




(e)

*Figure 5-6 The t-value of the attributes related to each of the 10 simulation tests*
*(a) t-value of cost, (b) t-value of travel time related to each of the 10, (c) t-value of seat comfort-mid, (b) t-value of seat comfort-high, (d) t-value of seat access time, (e) t-value of frequency*

### 5.4.4 Questionnaire construction

Based on the generated most appropriate *D*-efficient design, (design B), the questionnaires for blocks 1 and 2 of the design were then constructed, (see Appendices D2 & D3). Figure 5-7 presents an example of the SP choice questions. The questionnaires were printed in paper form and distributed to the respondents, (see section 3.4.2). Some commonly observable socio-demographic questions were also generated and are shown in Appendix D4. Additionally, as the survey data of both airport and non-airport respondents were collected, small changes were made. For the airport respondents, extra survey questions were asked concerning their current air trip information, such as the trip origin and the reason to choose air travel mode, as shown in Appendix D1.

| Trip 1 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$150 | A$175 | A$350 | A$200 |
| Time to bus-station or airport-terminal | N/A | 15 mins | 30 mins | 15 mins |
| Journey or travel time | 12 hrs | 15 hrs | 2 hrs | 2 hrs |
| Service frequency (weekly) | Any time | 30 buses/coaches a week | 16 flights a week | 2 flights a week |
| Seat comfort level | High (leg room 90 cm) | Medium (leg room 80 cm) | Medium (leg room 80 cm) | Low (leg room 70 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

*Figure 5-7 Stated preference survey question of travel mode choice*

## 5. 5 Discussion

The main contribution of this section is to develop an extended MFA to generate, in a semi-automatic way, realistic choice questions, without implausible and dominant alternatives. *Python* codes were developed to automate some steps of the EMFA that enabled the iterative process to effectively find and specify PDR constraints to eliminate implausible and dominant alternatives and unrealistic choice tasks. It was not possible to fully automate the whole process because it was still necessary to manually evaluate the choice questions output by the algorithm. However, the general constraint rules of mode choice, (such as, car, bus and airline) could be used by other researchers as initial constraints. Therefore, it might be useful to develop an open constraints database for certain choice studies, such as mode choice, to share with everyone, that would make experimental design much easier in the future. The procedures for specifying constraints and applying the simulation test were written in *Python* language. However, other software such as SAS and GAUSS could adjust the codes based on the proposed pseudocode for the same purpose.

This chapter found the phenomenon of the trade-off between statistical efficiency and plausibility and realism of the experimental design, as confirmed by Cherchi and Hensher (2015). Table 5-19 shows that efficient design C without considering any constraints has the lowest *D*-error, compared to efficient design A, (satisfies initial constraints), and design B, (satisfies full constraints). It appears that adding the constraints to the efficient design reduced its efficiency to some degree, (larger *D*-error), but made the choice questions more usable. In the future, it might be interesting to investigate how different types of constraints, the order of setting up constraints and how constraints set up for addressing adaptive behaviour of decision-makers (Keane, 1997) impact on the level of statistical efficiency achieved. The adaptive behaviour of decision makers means that the expectations of decision makers are formed adaptively, and the mode choice in a particular instance may be affected by the experience of the last, or a previous, choice made (Keane, 1997).

*Table 5-19 Comparison of D-errors with non-zero parameter priors*

| **Efficient design** | **Constraints** | ***D*-error** |
|---|---|---|
| Design A | Initial constraints | 0.00504 |
| Design B | Full constraints | 0.00577 |
| Design C | No constraints | 0.00391 |
| | | |
| Design A2 | Initial constraints | 0.00528 |
| Design B2 | Full constraints | 0.00601 |
| Design C2 | No constraints | 0.00386 |
| | | |
| Design A3 | Initial constraints | 0.00491 |
| Design B3 | Full constraints | 0.00545 |
| Design C3 | No constraints | 0.00394 |
| | | |
| Design A4 | Initial constraints | 0.00536 |
| Design B4 | Full constraints | 0.00608 |
| Design C4 | No constraints | 0.00390 |
| | | |
| Design A5 | Initial constraints | 0.00516 |
| Design B5 | Full constraints | 0.00560 |
| Design C5 | No constraints | 0.00366 |

**I ran the algorithm five times to generate optimised efficient designs and their D-errors, with different candidate sets each time.*

The EMFA performs well in generating efficient SP experimental designs, which also satisfy various constraints. However, the statistical property of attribute level balance may not be ensured, which is similar to the findings based on a row-based algorithm, as proposed by Collins et al. (2014). The *Python* coding techniques developed here are a relatively effective way to impose constraints for rejecting implausible and dominant alternatives and unrealistic choices tasks while forming candidate sets, especially for complex problems that may need more numerous and complex constraints. The EMFA method can also help researchers save time in finding and specifying the constraints.

### 5. 6 Summary

The focus of this chapter was on developing a semi-automatic approach to assist researchers to effectively find and specify all the required constraints while generating an efficient experimental design. Thus, an optimal efficient design with relatively high realism can be constructed. This new approach can make the process of constraint setting up and integration into the experimental design easier, especially for those choice problems with a huge size of full factorial design. A row-based algorithm EMFA was developed for this study that was then applied to the case study of constructing an efficient SP design for investigating passenger travel mode choice in regional Western Australia. The final generated efficient design was found to not only maintain a relatively high design efficiency, (small D-error), but also avoided implausible choice tasks. The field survey data collected based on this design is used for market segmentation, (in Chapter 7), and travel mode and airline choice analysis (in Chapters 8 and 9) in the present research. The next chapter provides a set of visualisations, (e.g., pie charts, bar charts and cross tables), based on the collected air travel survey data to initially explore air passenger key characteristics, which can provide some preliminary insights for governments and airlines to understand the aviation market.

# CHAPTER 6 AIR PASSENGER SURVEY DATA VISUALISATION

## 6. 1 Introduction

This chapter summarises air passenger profiles, trip origins and destinations, travel modes to access airports, trip purposes, travel groups, reasons for choosing to travel by air, air ticket booking timings, travel costs and frequency and factors affecting travel mode choice, based on visualisations of the air passenger survey data collected at four regional airports in Western Australia - Albany, Geraldton, Karratha and Broome.

Section 6.2 provides the research context and introduces the motivation for exploring the regional air passenger characteristics. Section 6.3 outlines the data used for this chapter. Sections 6.4 and 6.5 explore the regional air passenger characteristics and compare the air passenger dominant characteristics across the four selected regional towns. Key findings are presented in section 6.6.

## 6. 2 Research Context

As discussed in Chapter 2, based on a review of current literature, a research gap was identified, namely the lack of research into the characteristics of the aviation market in Western Australia. The filling of this gap contributes to improving the understanding of air passenger travel behaviour. In this chapter, on a macro perspective, section 6.4 explores the demographics, socio-economic and trip characteristics across all the regional air passenger respondents interviewed at Albany, Geraldton, Karratha and Broome RPT airports. On a micro perspective, section 6.5 compares the differences in demographics and some trip characteristics between the samples collected at each of the airports. The aviation market segmentation and travel mode choice estimation analysis are conducted separately in Chapters 7, 8 and 9.

Table 6-1 summarises total RPT flights and total available RPT seats (weekly) for the four airports. These four towns differ in a number of ways including in their proximity to Perth, population, history and key industries. By comparing the results of these four towns, a better picture of air travel behaviour in regional Western Australia can be derived.

*Table 6-1 Total RPT flights and total available RPT seats (weekly)*

| Airports | Total RPT flights (weekly) | Total available RPT seats (weekly) | Number of airlines |
|---|---|---|---|
| Albany | 23 | 782 | 1 |
| Geraldton | 23 | 2300 | 2 |
| Broome | 47 | 5271 | 2 |
| Karratha | 58 | 5800 | 2 |

## 6. 3 Data used in this chapter

The data used in this chapter is from the air travel information intercept survey collected at the four selected regional town airport departure lounges, from May to August 2018. In total, 950 air passenger surveys were collected. The surveys and data collection details were described in section 3.4.2.

## 6. 4 Characteristics of Western Australia regional air passengers

### 6.4.1 Regional Western Australian air passenger profiles

In this section, the regional airport respondents' demographic information, which includes age, gender, monthly income and education, is discussed. The information is also presented graphically in Figures 6-1 and 6-2.

#### 6.4.1.1 Age, gender, monthly income and education of the respondents

- **Age**: The majority of respondents (83.3%) were between 25 and 64 years old. Specifically, the age ranges between 25 and 34 years (21.9%), 35 and 44 years (22.3%) and 45 and 54 years (20.6%) were the three most popular age groups who travelled by air, followed by those aged between 55 and 64 (18.4%). Additionally, 7.2% of the respondents were aged from 18 to 24. The proportion of the total sample who were aged 65 and over, or under 18 was small (less than 2.5% in total).
- **Gender**: Except for 1.8% who skipped the gender question, 41.0% were female, and 57.2% were male.
- **Monthly income before tax**: The chart (Figure 6-1) below shows the distribution of the regional air travellers' income. In the sample, 29.7% of respondents had a high monthly income, ($8,700 or more), before tax. Following that, 14.3% had a monthly income between $6,500 and $8,699, 11.3% between $5,500 and $6,499 and 16.1% between

$3,500 and $5,499. Only 2.2% had zero income with 6.1% earning between $1 and $1,749 per month.

- **Education level**: As shown in Figure 6-2, the largest proportion of the respondents had either postgraduate (25.4%) or undergraduate (21.8%) qualifications. The remaining respondents had a college certificate/diploma (15.9%), vocational/technical certificate (13.1%) or at least senior high school diploma (17.8%). Just 3.4% of respondents had only primary or some secondary education.

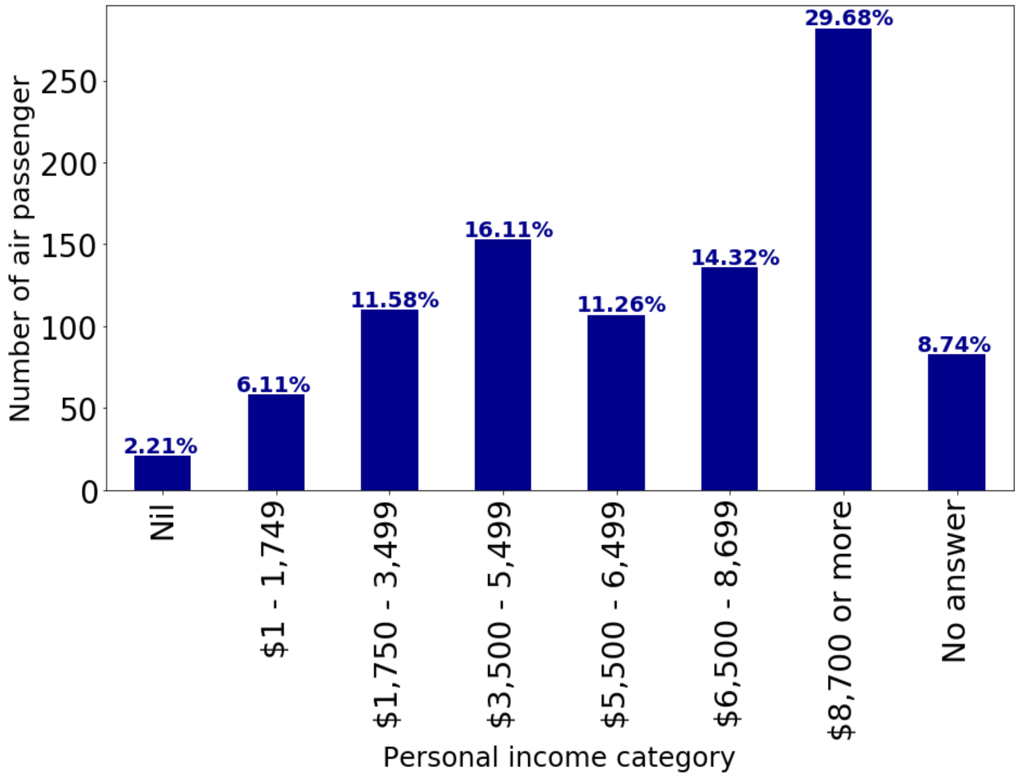



*Figure 6-1 Monthly income of air passengers in regional Western Australia*

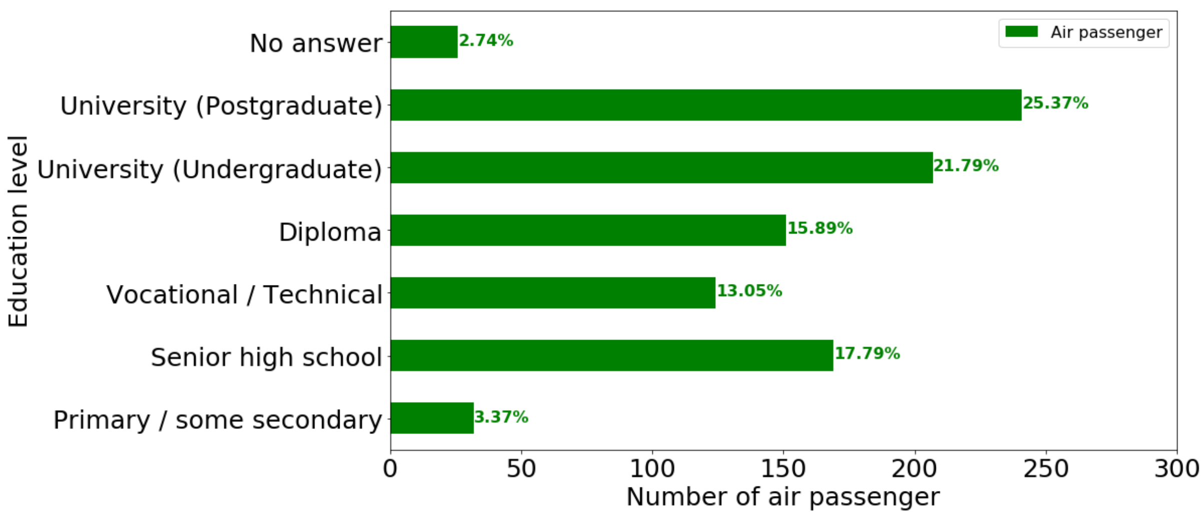


*Figure 6-2 Education level of air passengers in regional Western Australia*

### 6.4.2 Trip origin and destination

This section explores the trip origin and destination characteristics of the air passenger respondents. Note that trip origin is defined as the last stop before heading to the airport.

#### 6.4.2.1 Trip origin analysis

The pie chart below (Figure 6-3) demonstrates that most air passengers (34.4%) started their trip from their own home. Nearly 30% of respondents started their trip from a place of business or workplace. Slightly more than one fifth (21.6%) of the air passengers began their journey from private accommodation, (hotel, motel, inn, bed & breakfast, Airbnb, backpackers), and 7.7% from some else's home. The proportions of respondents who started their trip from a school, college or university, a restaurant or a tourist attraction are all less than 1%. 4.5% of the respondents specified other locations including campsites, boat at sea or offshore platforms.

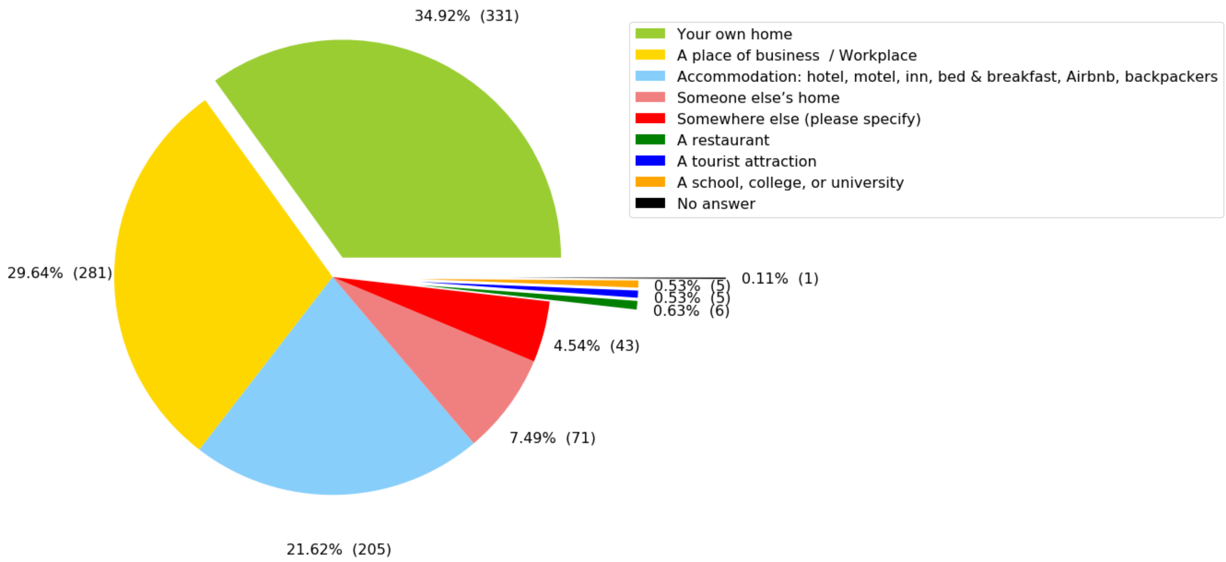



*Figure 6-3 Last stop before the airport*

#### 6.4.2.2 Trip destination analysis

The word cloud image below (see Figure 6-4), illustrates the regional air passenger final destinations, with the size of the placename reflecting its popularity. The most popular air passenger destination was Perth, as would be expected with it having the main hub airport in Western Australia and a population much larger than the rest of the state put together. Some other frequently mentioned regions within Western Australia were Broome, Karratha, Geraldton, Kalgoorlie and Bunbury. Major cities in Australia including Melbourne, Sydney, Brisbane and Adelaide were also popular destinations for air passengers, especially, Melbourne and Sydney. Furthermore, the word cloud image also illustrates that a small number of air passengers had final destinations overseas, including Bali, Africa, United Kingdom, USA and Japan.

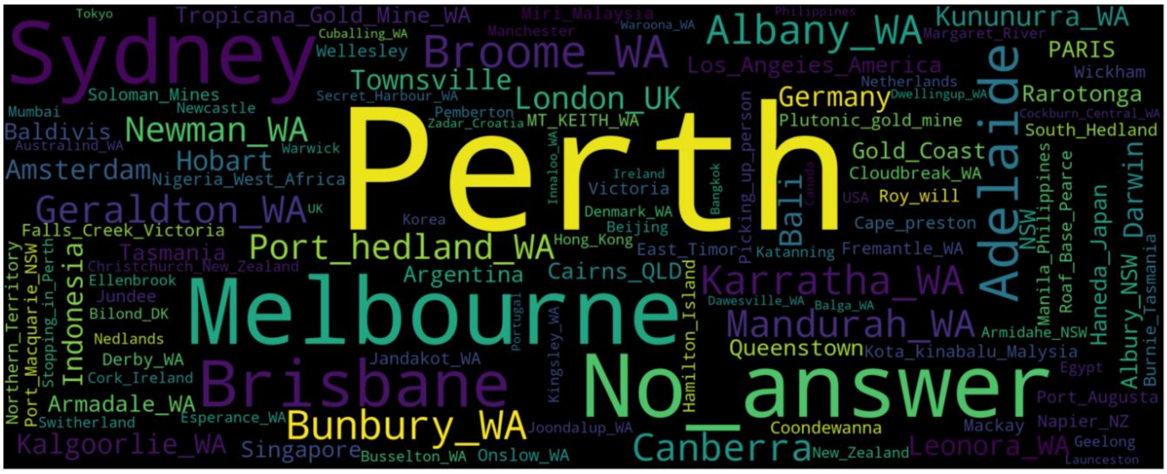



*Figure 6-4 Popular destinations of air passengers in regional Western Australia*

### 6.4.3 Access mode, reasons to choose the air travel mode and trip purpose

This section describes passenger aspects related to the air travel mode, including:

- ➢ Access mode to the airport
- ➢ Travel group
- ➢ Reasons for choosing air travel rather than road travel mode
- ➢ Rank of the travel-related factors affecting people’s travel mode choice
- ➢ Trip purpose

#### 6.4.3.1 Access mode to the airport

As shown in Table 6-2 below, around 30% of the respondents used their private car to get to the airport, followed by rental car (21.3%), company car (13.4%), taxi (11.3%) and ridesharing such as Uber (1.2%). 14.7% of the respondents were given a lift by friend/s, colleague/s, relative/s, or someone else to the airport. Only 2.8% of air passengers used public transport to access to the airports. Interestingly, 5.3% of the air passengers specified some other access mode to go to the airport, including by foot, by helicopter or shuttle bus. Overall, private car, rental car or sharing with others were the three most popular access modes for people to get to the airport.

*Table 6-2 Access modes to the four airports*

| Access mode to get to the airports | Percentage |
|---|---|
| Taxi | 11.30% |
| Ridesharing, such as Uber | 1.17% |
| Company car | 13.43% |
| Private car | 29.74% |
| Rental car | 21.32% |
| Public Transport (e.g. Bus) | 2.77% |
| Motorcycling | 0.00% |
| Cycling | 0.00% |
| Lift by friend/s, colleague/s, relative/s, or someone else | 14.71% |
| Other (please specify) | 5.33% |
| No answer | 0.21% |

### 6.4.3.2 Travel group

The pie chart in Figure 6-5 below shows with whom the air passenger respondents travelled. About 60.4% of the respondents travelled alone, with another one fifth travelling with their business associate/s or colleague/s, and 8.9% with a partner or spouse. Some 5.6% travelled with family and 2.3% with a friend/s.

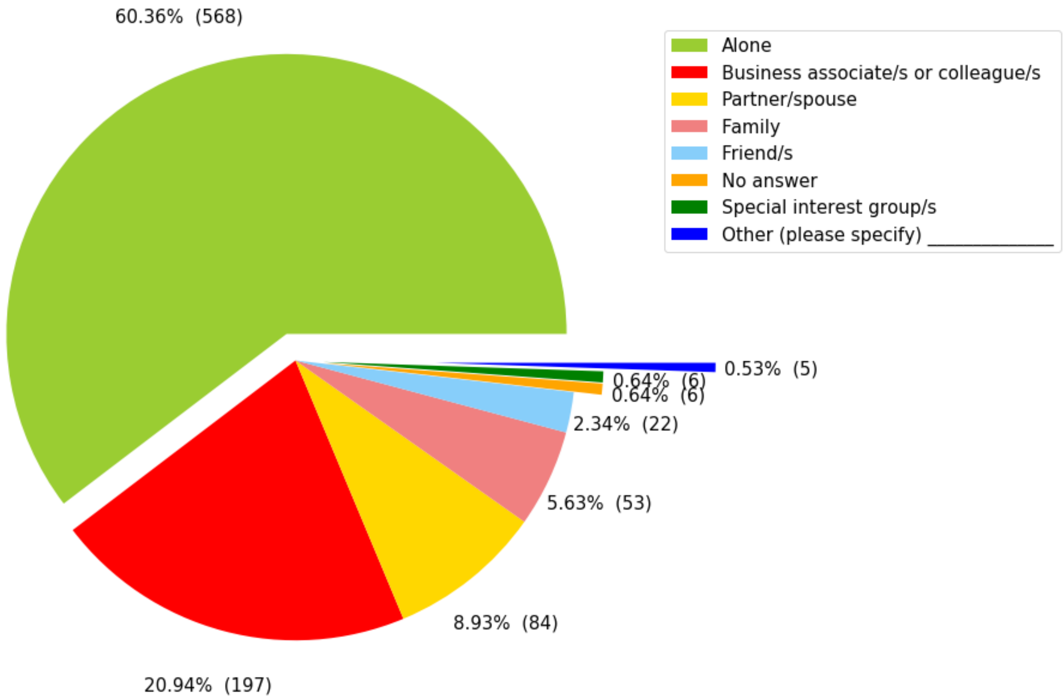



*Figure 6-5 Travel companion of air passengers in regional Western Australia*

#### 6.4.3.3 Reasons to choose air travel rather than road travel mode

The graph below (see Figure 6-6) shows the reasons why air passengers chose to travel by air rather than by car or coach. Most passengers (54.9%) chose to travel by air because of the distance involved to reach their destination being too long to drive. Following that, about 28.7% of respondents expressed that it was because of the convenience and being more time efficient. After that, 7.5% of the air passenger respondents travelled by air for a reason not listed on the questionnaire, the main ones being due to work requirements, going overseas, not owning a car or per a tour operator’s arrangements. Some 2.3% air passengers thought the airfare was cheap and affordable and thus chose to fly, with 2.3% of respondents deciding to fly because flying is more comfortable. Some 1.7% chose to fly because they were on an emergency trip and 1.3% considered flying to be safer. Overall, the majority of the respondents preferred to fly because their trip distance was too long to drive or for better time efficiency.

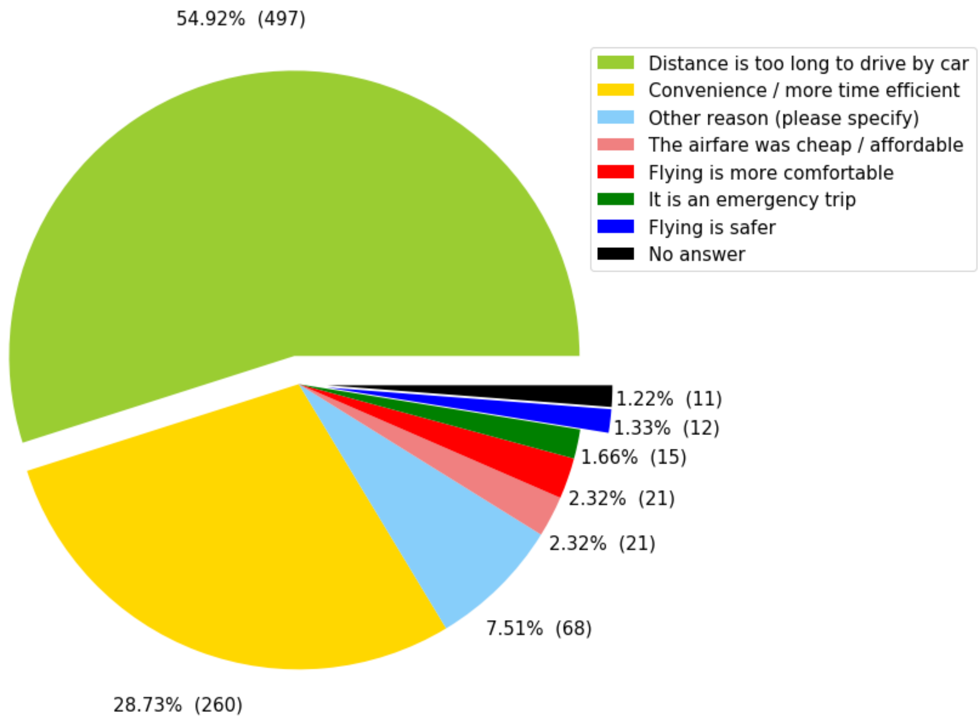



*Figure 6-6 Reasons for flying*

#### 6.4.3.4 Rank of the travel-related factors affecting people’s travel mode choice

The stacked bar chart (see Figure 6-7) describes how important the factors of ticket fare or driving cost, access time to bus-station or airport-terminal, journey time, service frequency and seat comfort are in affecting air passenger travel mode choice. The graph indicates that most air passenger respondents considered journey time to be the most critical factor in making their travel mode choice. Ticket fare or driving cost was also an important factor influencing their travel mode choice. Conversely, most of the air passenger respondents stated that access time to the airport was the least important factor.

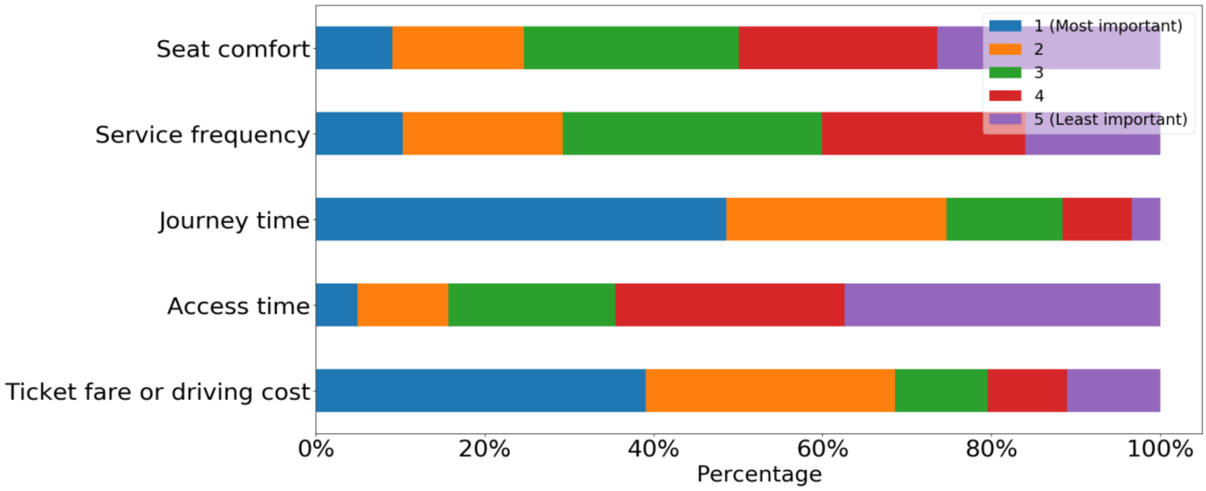



*Figure 6-7 Importance of key factors affecting travel mode choice*

### 6.4.3.5 Trip purpose

The pie chart of Figure 6-8 below shows the trip purposes of respondents. Most air passengers (63.9%) travelled for business, made up of other business (34%), fly in fly out (19.6%) and government work (10.3%). About 16.3% were on holiday, 9.2% visiting friends or relatives, 4.6% for medical or health reasons and 1.8% for education. Some 3.4% air passengers had another trip purpose, including returning home, exploration or attending important events, (e.g., concert, funeral or competition). In summary, the chart indicates that the vast majority of the surveyed regional air passengers were travelling for business, following a long way behind by a holiday/leisure then visiting friends or relatives.

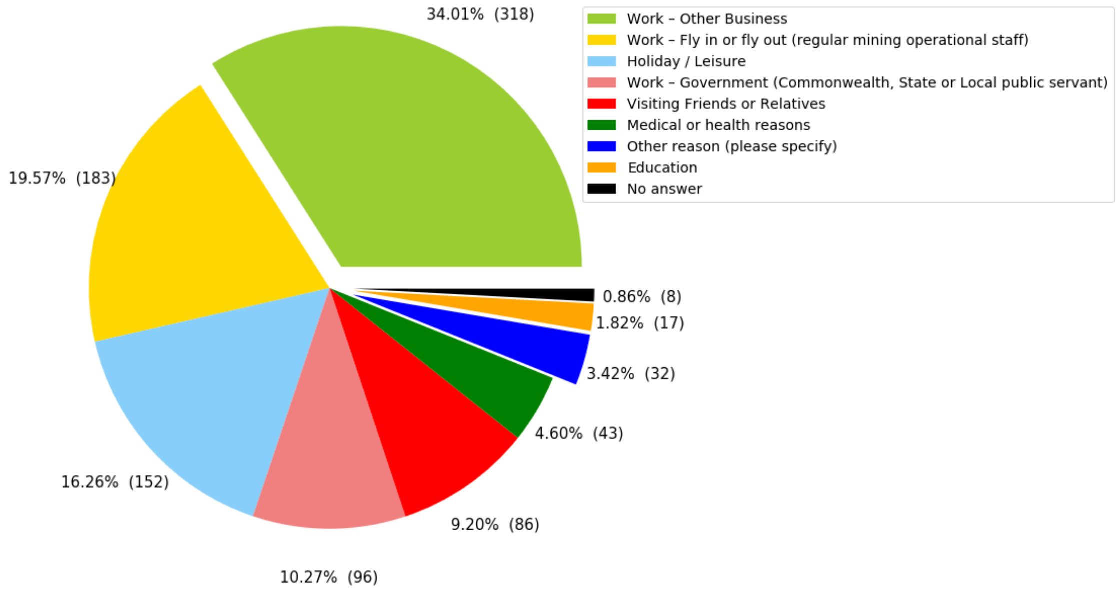



*Figure 6-8 Trip purposes of air passengers*

### 6.4.4 Travel cost and frequency

In this section, the regional air passenger respondents' travel cost and frequency are discussed. This includes:

- Travel cost one-way flight
- Air travel frequency during past year
- Booking time in advance of the flight

### 6.4.5 Travel cost of one-way flight

The pie chart (Figure 6-9) illustrates the airfare paid by the respondents for a one-way flight. It shows that slightly more than half (53.9%) of all respondents did not pay for the fare as it was paid by their employers. Some 10.1% air passengers paid between $200 and $299 for their one-way flight and 9.3% paid less than $199. A considerable proportion (19%) had to spend over $300 to buy their one-way air ticket; 3.1% paid $300 – $399, 5.1% paid $400 – $499 and 10.8% paid over $500.

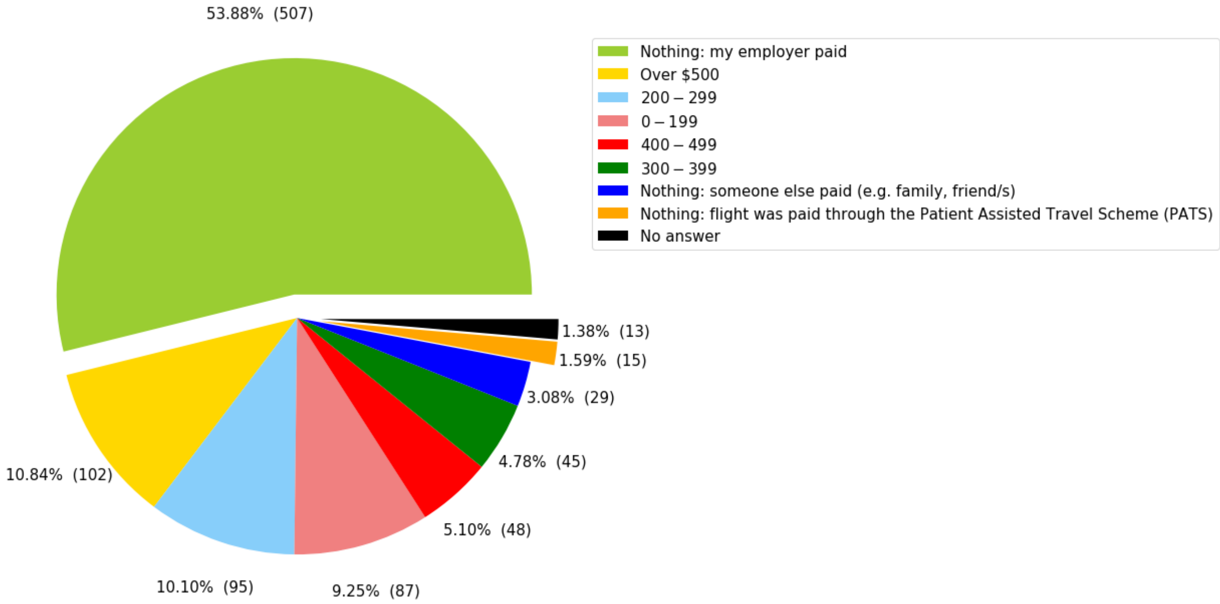



*Figure 6-9 One-way travel cost*

### 6.4.5.1 Air travel frequency during past year

Table 6-3 shows slightly more than one-third of the respondents only travelled once, (i.e., on the current trip), during the past 12 months to their current flight destination. Also, slightly more than one quarter (27.2%) of respondents had travelled six or more times during the past 12 months. Some 17.1% of respondents travelled twice and 18.3% three to five times in the past year.

*Table 6-3 Air travel frequency to current destination*

| Travel frequency in last year | Percentage |
|---|---|
| Once (current trip) | 36.56% |
| Twice | 17.07% |
| Three to five times | 18.34% |
| Six or more times | 27.19% |
| No answer | 0.84% |

### 6.4.5.2 Booking time in advance of the flight

The pie chart in Figure 6-10 shows the respondents' preferences for booking flights. Nearly one third of the respondents (30.3%) booked air tickets between one week and one month in advance of the flight. However, a slightly higher proportion (32.8%) preferred to book the ticket nearer to the departure date: 7.9% booking less than 24 hours before the flight, 7.2% less than two days before and 17.7% between two days and one week before. The pie chart also indicates that 14.3% preferred to book their ticket one to three months in advance of the departure date.

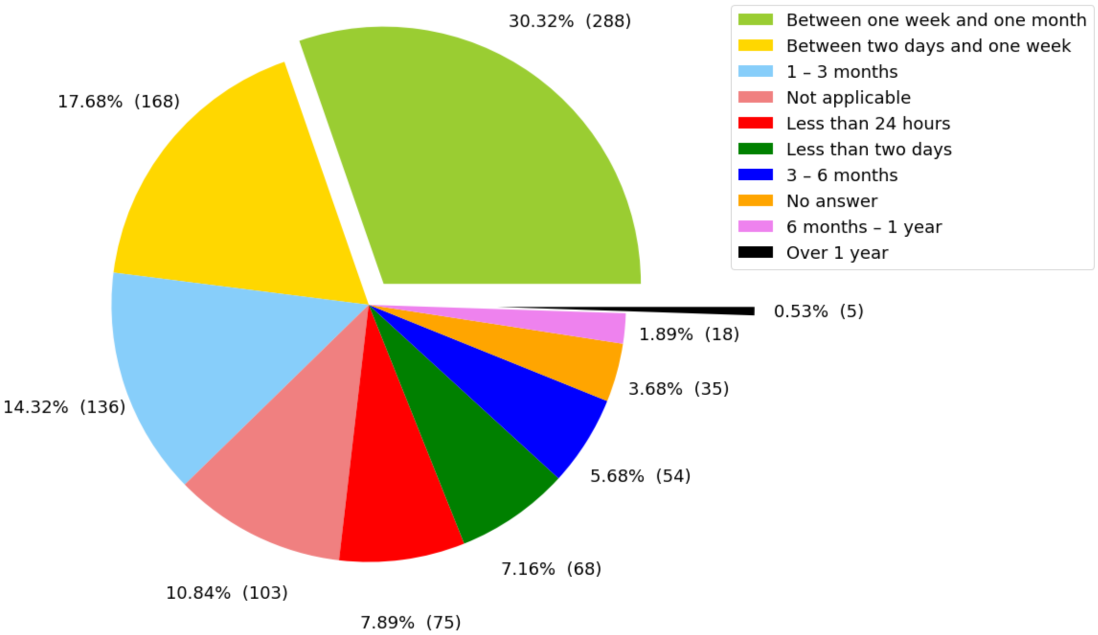



*Figure 6-10 Booking time in advance to the flight*

### 6.4.6 Reasons to choose flight, trip purpose and booking time in advance of departure

Table 6-4 shows the main reasons for regional air passengers choosing air transport rather than car or bus, broken down by trip purpose. Five categories of travellers considered the long distance and/or time efficiency are the most important reasons. They were travelling for holiday, education, mining business, other business or a medical purpose. Fly in fly out (FIFO) workers and those travelling for other business and the government mentioned that their companies arranged the air travel for them.

*Table 6-4 Main reason for selecting air travel mode based on trip purpose*

| Trip purpose / Reason to choose air travel | Holiday / Leisure | Visiting Friends or Relatives | Education | Work – Government | Work – Fly in or fly out | Work – Other Business | Medical or health reasons | Other reason | No answer | Sum |
|---|---|---|---|---|---|---|---|---|---|---|
| Distance is too long to drive by car | 114 | 34 | 13 | 31 | 96 | 169 | 16 | 16 | 1 | 490 |
| The airfare was cheap/ affordable | 4 | 3 | 0 | 1 | 5 | 2 | 3 | 2 | 0 | 20 |
| Convenience/more time efficient | 24 | 35 | 3 | 48 | 33 | 97 | 7 | 12 | 0 | 259 |
| It is an emergency trip | 1 | 4 | 1 | 1 | 0 | 5 | 3 | 0 | 0 | 15 |
| Flying is safer | 0 | 0 | 0 | 2 | 4 | 3 | 2 | 0 | 0 | 11 |
| Flying is more comfortable | 2 | 5 | 0 | 1 | 8 | 4 | 1 | 0 | 0 | 21 |
| Other reason | 5 | 4 | 0 | 7 | 24 | 20 | 7 | 1 | 0 | 68 |
| No answer | 0 | 1 | 0 | 1 | 4 | 3 | 2 | 0 | 7 | 18 |
| Total | 150 | 86 | 17 | 92 | 174 | 303 | 41 | 31 | 8 | 902 |

Table 6-5 below illustrates that the one-way flight airfares of the respondents varied depending upon the date they reserved/bought their ticket. With the exception of those respondents whose tickets were paid for by their employers, the majority of the air passengers who preferred booking the tickets less than 24 hours in advance of the flight paid less than $199. Those air passengers who booked their tickets less than two days, between two days and one week or between one week and one month, paid either less than $199 dollars or $200 – $299 dollars. Interestingly, the results also show that the air passengers who preferred to book their ticket relatively far in advance of the departure date, (one to six months), usually paid a relatively higher amount of money (over $300). These findings indicate that many regional air passengers booked their flights relatively close to the departure date, and the later, (closer to departure date), they booked their trip the lower price they usually paid. However, this may depend on the airlines fare pricing for individual routes, and the availability (scarcity) of seats.

*Table 6-5 One-way travel cost by booking time in advance to flights*

| Booking time in advance to flight / One-way airfare | Less than 24 hours | Less than two days | Between two days and one week | Between one week and one month | 1 – 3 months | 3 – 6 months | 6 months – 1 year | Over 1 year | Not applicable | No answer | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Nothing: my employer paid | 29 | 33 | 122 | 180 | 62 | 12 | 4 | 1 | 53 | 5 | 601 |
| Nothing: someone else paid (e.g. family, friend/s) | 1 | 3 | 2 | 8 | 4 | 1 | 1 | 0 | 7 | 2 | 29 |
| Nothing: flight was paid through the PATS* | 0 | 2 | 4 | 7 | 0 | 2 | 0 | 0 | 0 | 0 | 15 |
| $0 - $199 | 34 | 7 | 6 | 17 | 10 | 7 | 0 | 0 | 6 | 0 | 87 |
| $200 - $299 | 6 | 9 | 13 | 31 | 16 | 7 | 1 | 1 | 5 | 4 | 93 |
| $300 - $399 | 1 | 4 | 4 | 11 | 9 | 9 | 1 | 0 | 2 | 4 | 45 |
| $400 - $499 | 1 | 4 | 5 | 13 | 5 | 4 | 5 | 1 | 5 | 4 | 47 |
| Over $500 | 3 | 4 | 11 | 16 | 28 | 10 | 5 | 2 | 20 | 3 | 102 |
| No answer | 0 | 2 | 1 | 2 | 2 | 2 | 1 | 0 | 5 | 3 | 18 |
| Total | 75 | 68 | 168 | 285 | 136 | 54 | 18 | 5 | 103 | 25 | 937 |

** Patient Assisted Travel Scheme*

Table 6-6 presents the respondent booking times in advance of the flight with respect to their trip purpose. The respondents who travelled for government work, FIFO and other business work were more likely to book their air ticket less than one month in advance. For example, more than half of the FIFO respondents stated that they preferred to book their ticket between one week and one month in advance. However, one third of respondents who travelled for the purpose of leisure chose to book their ticket around one to six months in advance of the flight, while, for respondents who were visiting friends or relatives, between two days and 3 months was the most popular booking time.

*Table 6-6 Trip purpose by booking time in advance to flights*

| Booking time in advance to flight / Trip purpose | Less than 24 hours | Less than two days | Between two days and one week | Between one week and one month | 1 – 3 months | 3 – 6 months | 6 months – 1 year | Over 1 year | Not applicable | No answer | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Holiday / Leisure | 6 | 3 | 7 | 24 | 25 | 25 | 10 | 3 | 37 | 12 | 152 |
| Visiting Friends or Relatives | 6 | 8 | 10 | 29 | 20 | 7 | 2 | 0 | 3 | 1 | 86 |
| Education | 3 | 1 | 2 | 3 | 5 | 2 | 0 | 0 | 0 | 1 | 17 |
| Work–Government (Commonwealth, State or Local public servant) | 0 | 1 | 20 | 49 | 15 | 3 | 0 | 1 | 6 | 1 | 95 |
| Work – Fly in or fly out (regular mining operational staff) | 20 | 18 | 29 | 50 | 25 | 5 | 1 | 1 | 31 | 1 | 181 |
| Work – Other Business | 19 | 27 | 88 | 112 | 35 | 5 | 2 | 1 | 20 | 5 | 314 |
| Medical or health reasons | 11 | 4 | 5 | 13 | 2 | 4 | 1 | 0 | 2 | 0 | 42 |
| Other reason | 7 | 4 | 6 | 7 | 3 | 1 | 0 | 0 | 2 | 2 | 32 |
| No answer | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 2 | 2 | 8 |
| Total | 72 | 67 | 167 | 288 | 131 | 52 | 17 | 5 | 103 | 25 | 937 |

### 6.4.7 Comparison of air passenger characteristics across the four regional airports

#### 6.4.7.1 Demographic information

For Karratha airport, 79.9% of air passenger respondents were between the ages of 25 and 45 years, a much higher percentage than was observed at Broome (42.2%), Geraldton (41.6%), and Albany (35.3%). In contrast, Albany airport had 39.4% of air passenger respondents aged over 55, compared to 30.5% for Broome, 23% for Geraldton and 11.2% for Karratha. Respondents from Karratha and Geraldton airports were more likely to have higher monthly incomes than those from Albany and Broome airports. However, their overall education levels were slightly lower than those from Albany and Broome.

#### 6.4.7.2 Trip origin and destination

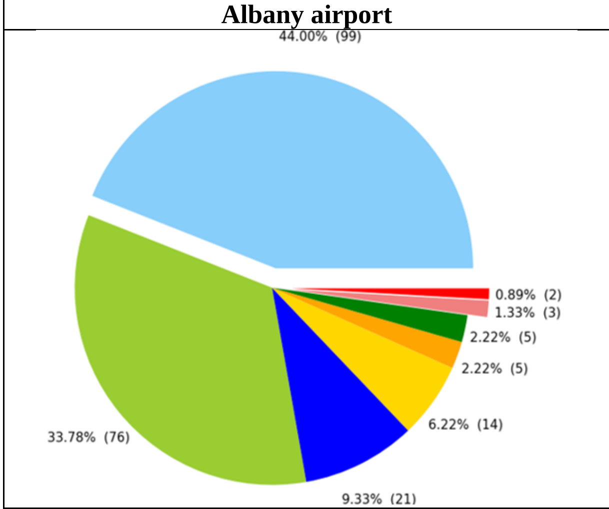
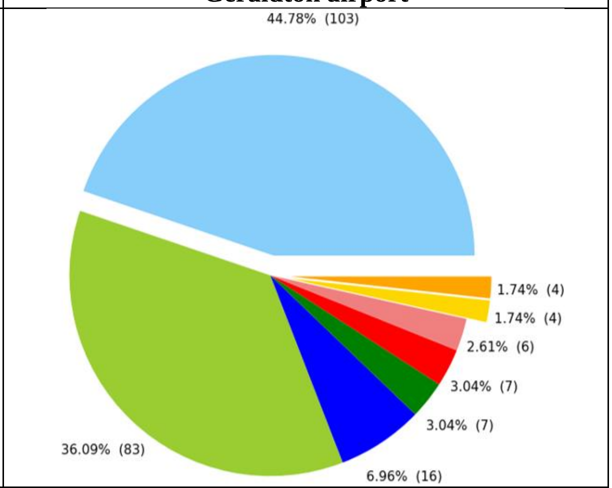
The most popular origin for respondents from Albany and Geraldton was their own home. However, it was accommodation for Broome and place of the business/workplace for Karratha, potentially highlighting the difference in markets. Table 6-7 lists the top final destinations for the four airports. Except for Albany, the other three have similar top destinations. Perth, Melbourne and Sydney are the top three. Broome, Brisbane and Adelaide are also popular. For Albany, the top three were Perth, Sydney and Brisbane.

*Table 6-7 Final destinations for air travellers ranked by popularity*

| Origin Airport | First | Second | Third | Fourth | Fifth |
|---|---|---|---|---|---|
| Albany airport | Perth (153/229) | Sydney (12/229) | Brisbane (7/229) | Broome (5/229) | Newman (5/229) |
| Geraldton airport | Perth (152/231) | Melbourne (12/231) | Sydney (10/231) | Broome (4/231) | Karratha (4/231) |
| Broome airport | Perth (124/231) | Melbourne (27/231) | Sydney (13/231) | Brisbane (8/231) | Adelaide (6/231) |
| Karratha airport | Perth (168/215) | Melbourne (6/215) | Sydney (6/215) | Brisbane (3/215) | New Zealand (3/215) |

### 6.4.7.3 Reasons for choosing air travel

The reasons for travelling by air are presented in Figure 6-11 for the four airports. Karratha and Broome are a long way by road from Perth, (1523km and 2239km, respectively). Therefore, ‘Distance is too long to drive by car’ was the dominant reason for choosing the air travel mode. For Albany and Geraldton, both ‘Distance is too long to drive by car’ and ‘convenience or more time efficient’ were popular reasons for choosing air travel.

| Albany airport | Geraldton airport |
|---|---|
|  |  |
| Karratha airport | Broome airport |

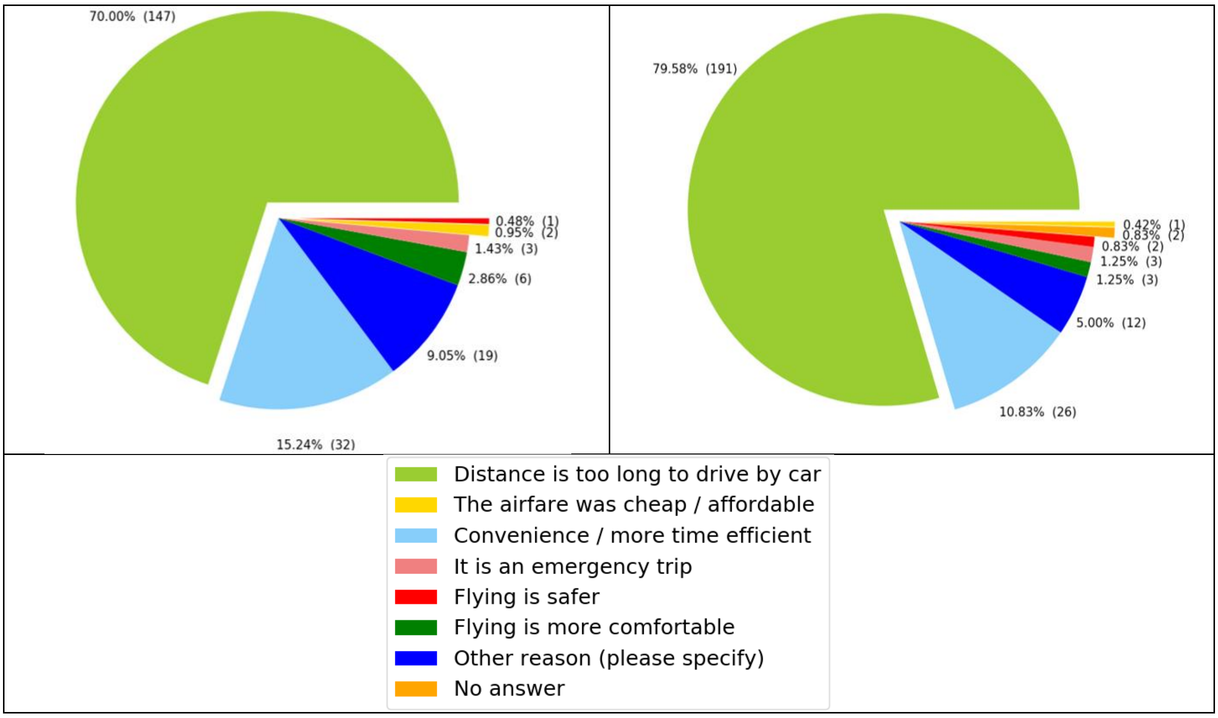
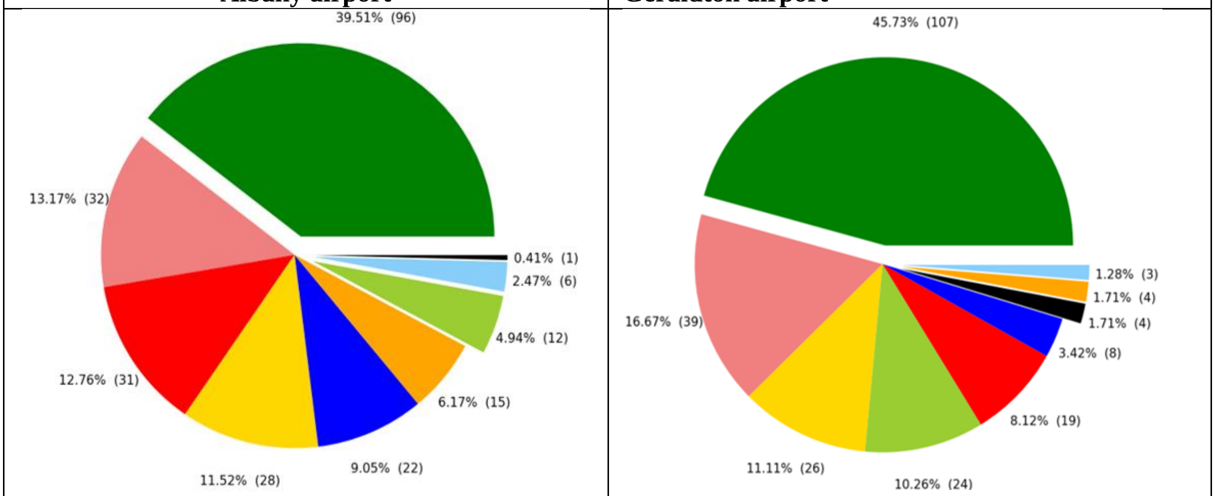



*Figure 6-11 Chart matrix of Albany, Geraldton, Karratha and Broome air passengers reasons for flying*

### 6.4.7.4 Trip purposes

The trip purposes of the air travellers at the four airports are presented in Figure 6-12. As can be seen, Broome is a tourist town with 42.3% of trips made for the purpose of 'Holiday/Leisure', the most popular trip purpose. As a mining town, respondents from Karratha chose 'Fly in or fly out for work' as the most popular travel purpose. For Albany and Geraldton, the top travel purpose was 'work-other business', with 'work for government' and 'visiting friends or relatives' also popular.

| **Albany airport** | **Geraldton airport** |
|---|---|




| **Karratha airport** | **Broome airport** |
|---|---|

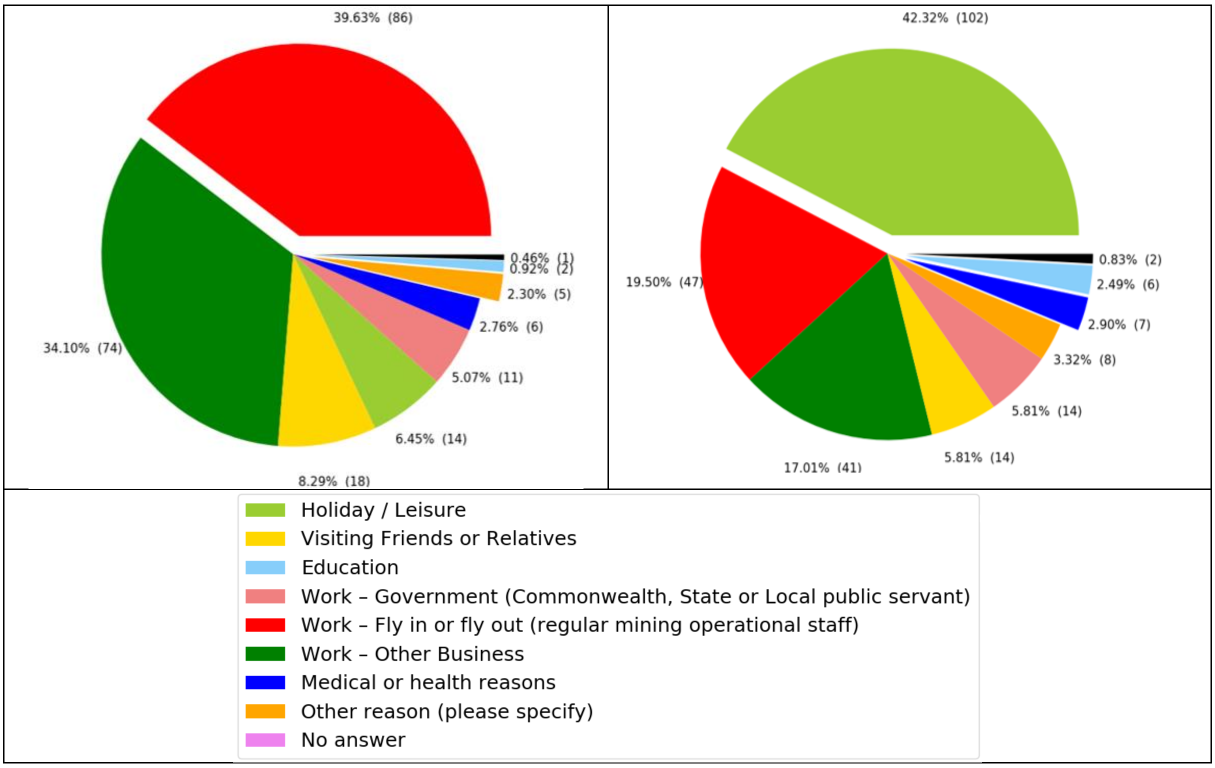



*Figure 6-12 Chart matrix of trip purpose of Albany, Geraldton, Karratha and Broome air passengers*

### 6.4.7.5 Travel cost and frequency

The one-way travel costs of the air travellers at the four airports are presented in Figure 6-13. ‘Paid by the employer’ is the most popular category of the travel payment for all respondents, especially for Karratha respondents, where over 75% of trips were paid by the employer. For Broome, it was only around 37%, reflecting the higher tourist and lower commercial components (see Figure 6-12). For both Karratha and Broome, if respondents paid the fare, the top cost category was over $500 for a one-way fare. Geraldton and Albany had a similar pattern. However, the top paid one-way fare category was $200-$299 for Geraldton and $0-$199 for Albany, which means that respondents from Geraldton airport paid a higher fare than those from Albany airport, although they are located a similar distance (around 450km) from Perth.

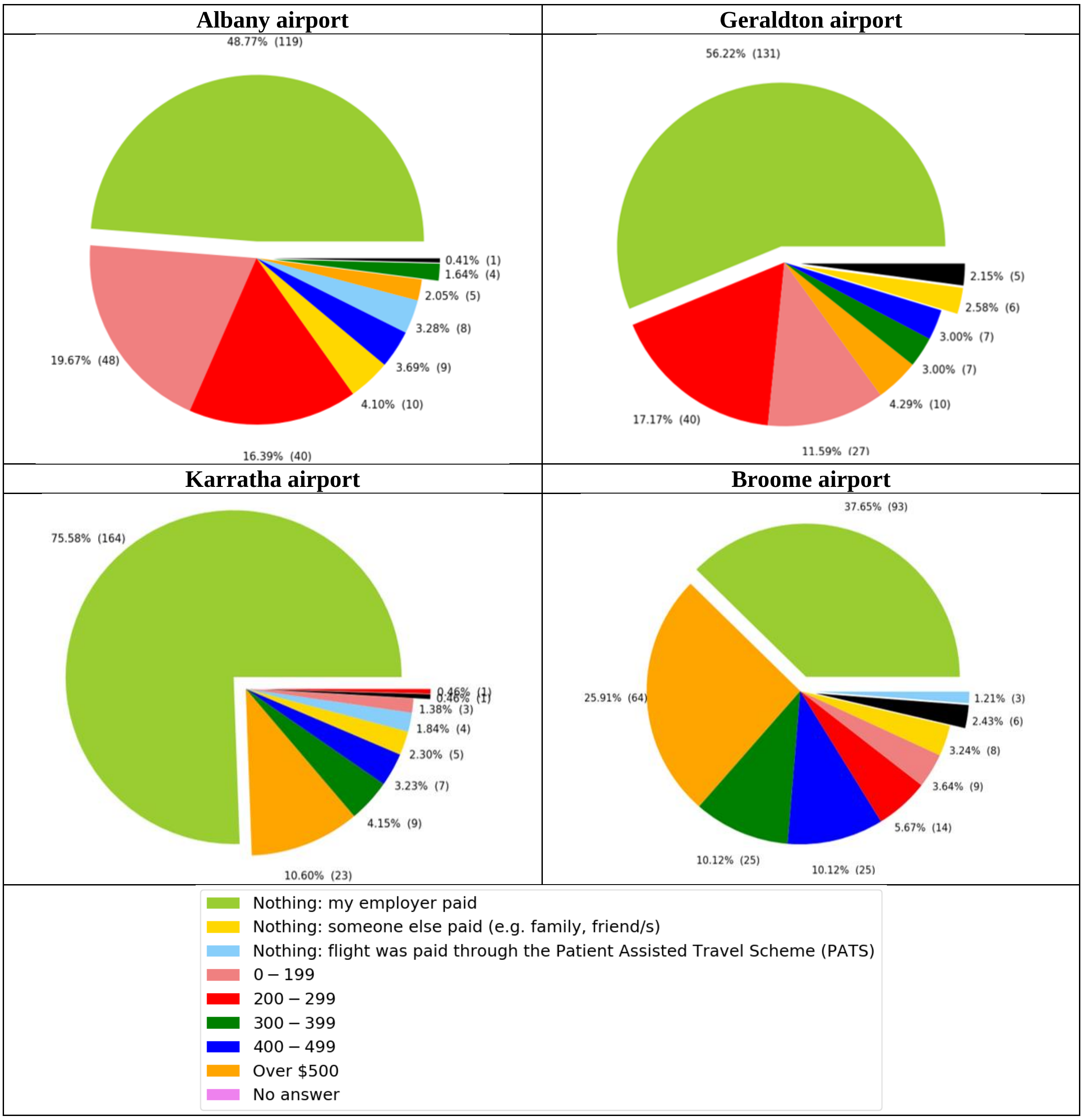



*Figure 6-13 Chart matrix of one-way travel cost of Albany, Geraldton, Karratha and Broome air passengers*

Figure 6-14 presents the trip frequencies of the air travellers at the four airports. At Karratha airport, six or more times in the past year was the most popular annual travel frequency for the respondents, while for other three airports, it was the first visit within the past year. This result indicates that the aviation market in a mining town (e.g., Karratha) is likely to comprise of more frequent air passengers, and especially if nearby mines are operating on a fly in fly out basis.

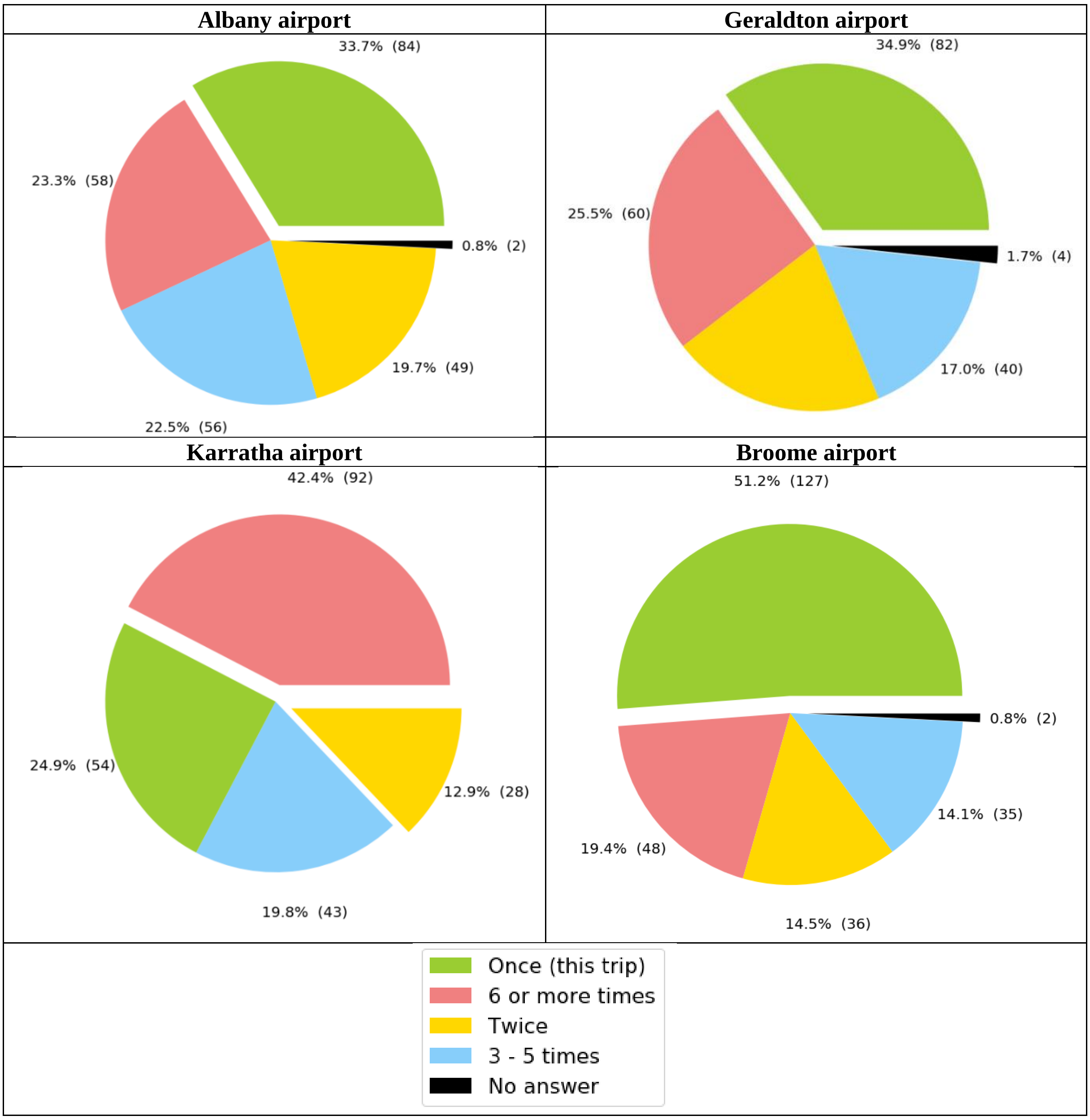
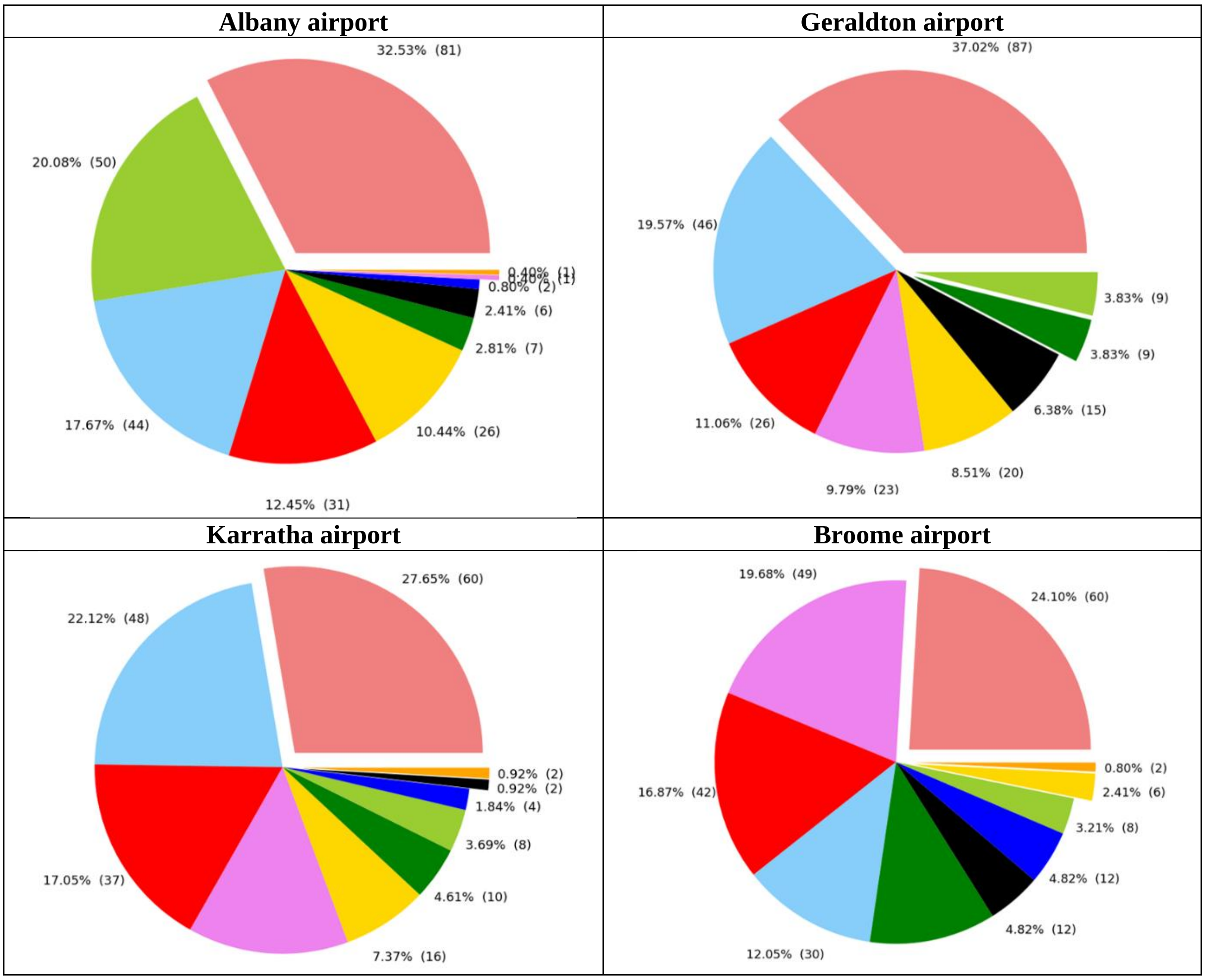



*Figure 6-14 Chart matrix of trip frequency of Albany, Geraldton, Karratha and Broome air passengers*

### 6.4.7.6 Booking time in advance of the flight and load factor

As Rex Airlines reached an agreement with the Shire of Albany for a special community fare of $129 one-way when booking a ticket within 24 hours, it has made the ticket price cheaper and more affordable. Respondents from Albany on average paid less than those from the other three airports and the average load factor of flights was relatively higher. According to the airport survey, the average load factor from Albany to Perth between 21 and 24 May 2018 was 73.53%. The aviation statistics from a government report for the same month in 2018, indicate load factors of 52.0% for Geraldton to Perth, 65.6% for Karratha to Perth and 82.2% for Broome to Perth (Department of Infrastructure Regional Development and Cities, 2018).

Figure 6-15 compares the booking time in advance of the flight for the four airports. 'Between one week and one month' was the most popular choice for all four airports. For Albany airport, the booking time was relatively shorter than others. 'Less than 24 hours' and 'Between two days and one week' were the second and third most popular choices, while for Geraldton and Karratha airport, they were 'between two days and one week' and '1-3 months'. Interestingly, for Broome airport, a large portion of the respondents (19.7%) picked the 'not applicable' choice. The third most popular choice for them was '1-3 months'. This means that respondents from Broome airports preferred to book their trips further in advance of their flight in order to secure their seats.



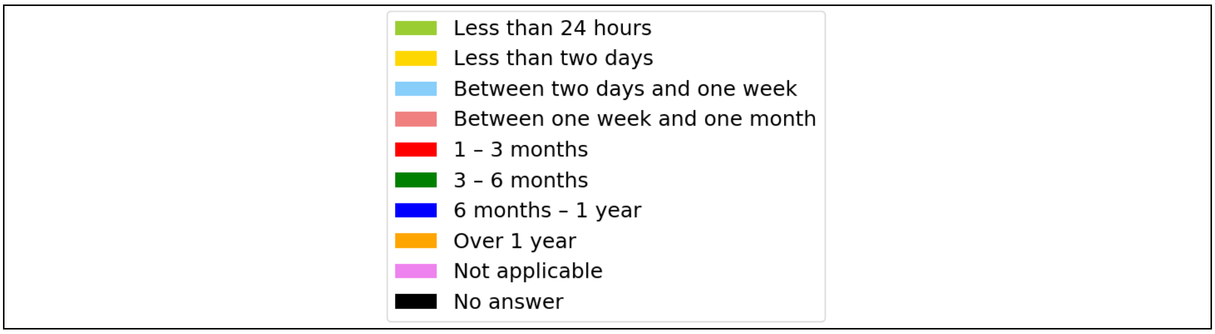



*Figure 6-15 Chart matrix of booking time in advance of flight by Albany, Geraldton, Karratha and Broome air passengers*

## 6. 5 Key findings

Previous sections (6.3 & 6.4) have provided visualisations and summary tables exploring the travel behaviour of the air passengers in regional Western Australia. The results reveal the overall profiles and trip characteristics of the regional air passengers. They also identify the differences in air travel behaviours across the Albany, Geraldton, Karratha and Broome regions, especially the demographics of the air passenger survey respondents, their reasons for choosing air transport, travel purposes, travel frequency and booking time in advance of the flight. Some key survey findings are reveals here:

1) In Karratha, 79.9% of air passenger survey respondents were between the ages of 25 and 45; it was 42.2% for Broome, 41.6% for Geraldton and 35.3% for Albany. However, Albany 39.5% of respondents were aged over 55, while it was 30.5% for Broome, 23.0% for Geraldton and 11.2% for Karratha.
2) Although the respondents from Geraldton and Karratha normally have a relatively lower education background in comparison with those from Broome and Albany, but their average income was slightly higher than those from Albany and Broome.
3) Place of the business/workplace was the most popular trip origin for the respondents from Karratha, while it was accommodation for Broome. By contrast, the most frequent origin place for the Albany and Geraldton respondents was the respondents their own home.
4) The air travel survey data also indicated that, for the respondents from Albany, more than 58% of people used a private car, while for people from Geraldton and Broome, the most popular choice was a rental car. For Karratha, it was a company car.
5) Travelling alone was the most popular travel group for all airports, for Broome, travel with partner or spouse was the second most popular one, but for other airports, it was the category of travelling with business associates or colleagues.

6) Karratha and Broome are further away from Perth (1523km and 2239km) than Albany and Geraldton are from Perth. As such, 'distance is too long to drive by car' was the dominant reason for choosing the air travel mode for Karratha and Broome survey respondents. In Albany and Geraldton, both 'distance is too long to drive by car' and 'convenience or more time efficient' are popular reasons for the respondent to choose air travel.
7) Respondents from Albany on average paid less for airfares than those from other regions. The average load factor of Albany flights was also relatively higher.
8) Except for Karratha, the first visit was the most popular annual travel frequency for the respondents from other airports. For Karratha it was six or more times.

Consequently, this chapter provides a preliminary understanding of the aviation market which could be used to inform transport industry management and influence government policies and strategies.

## 6. 6 Summary

In summary, this chapter has endeavoured to understand the socio-demographics and air travel information of regional air travellers. The same survey was conducted at four airports: Albany, Geraldton, Karratha and Broome. Albany and Geraldton are both important regional towns, located a similar distance (around 400km) from Perth, Albany to the south and Geraldton to the north. However, respondents from Geraldton paid higher fares to go to Perth than those from Albany due to different aviation fare structures, policies and regulation. Karratha and Broome are both further from Perth. Karratha is primarily a mining town and Broome is mainly a tourist town. The present chapter provides a preliminary understanding of the aviation which could provide input into further research and may facilitate aviation policy development in regional Western Australia.

The characteristics of regional air passengers were explored in this chapter. The next chapter provides a further investigation of the regional aviation market in Western Australia that extends the understanding of regional air passenger characteristics and the aviation market. A mixture model-based market segmentation approach is developed for identifying and investigating the existing and potential aviation markets in Western Australia, based on both airport and non-airport respondent survey data.

# CHAPTER 7 AVIATION MARKET SEGMENTATION ANALYSIS

## 7. 1 Introduction

The previous chapter summarised the characteristics of regional air passengers by creating a set of visualisations based on the air travel information survey data. In order to get a further understanding of the regional aviation market, this chapter provides a novel approach to identify and investigate existing and potential regional aviation markets, using a mixture model-based market segmentation approach.

The present chapter is based primarily on a paper[6] that was submitted to the Journal of Transport Policy and is currently under review. Section 7.2 provides a brief background to regional aviation market segmentation analysis. Section 7.3 describes the methods used to identify the aviation market segments. The segmentation results are interpreted in section 7.4, with a broader discussion in section 7.5. Finally, section 7.6 summarises the findings.

## 7. 2 Research Context

Customer preferences or needs for the same or similar service factors are heterogeneous (Kotler, 2009). It is therefore not possible for the airlines to satisfy all passenger preferences while maintaining commercially sound economic positions (Shaw, 2016). Market segmentation is a concept first proposed by Smith (1956) that has since been frequently used in market theory and practice. It breaks down the market into a finite number of homogenous subsets or segments with similar characteristics and preferences, that can then be used as a guide to specifically target sub-markets with tailored marketing strategies (Cahill, 1997; Wen et al., 2008). Segmentation of the regional air travel market may therefore be useful to research into the aviation market and allow the whole air travel market to be more completely characterised and better understood.

[6] Zhou, H., Norman, R., Kelobonye, K., Xia, J., Hughes, B., , Nikolova, G., . . . Falkmer, T. (2019). Market Segmentation Approach to Investigate Existing and Potential Aviation Markets. Manuscript, [Submitted to Journal of Transport Policy, under review]

This chapter conducts a market segmentation analysis of this important regional aviation market, with a view to identifying segments and thereby allowing improved alignment between the demands of consumers and the goals of industry and government in Western Australia. In particular, the mixture model-based market segmentation approach with EM algorithm estimator is applied to identify the market segments of regional airport and non-airport passenger respondent samples, (two samples), based on their socio-demographic information, trip purpose and stated preferences for air and non-air travel modes, (car and bus). This is followed by a discussion of the prominent characteristics of the identified market segments for both samples, respectively. Travellers' stated preferences for air transport are used to classify the segments into existing, (high preference for air transport), and potential, (low or moderate preference for air transport), aviation markets. This chapter subsequently compares the characteristics across the segments with a similar preference towards air travel, in order to better understand the existing and potential aviation markets. The findings shed light on the relative competitiveness of a regional airline relative to both other airline and road transportation competitors. They could assist airlines in developing customised, (more targeted and, hence, more efficient), strategies to satisfy the needs of passengers and, thereby, increase patronage, reduce airfares and improve the airline's viability and sustainability under competition.

## 7. 3 Methodology

### 7.3.1 Data used in this study

The regional aviation market segmentation analysis is conducted using the airport and non-airport respondents' SP survey data collected at the four selected regional towns in Western Australia. Details of the experimental design procedure for generating the SP survey questionnaire were described previously in Chapter 5. As the SP survey contains 12 choice questions and it was felt that 12 tasks was too many for some respondents, a blocking strategy was used to break the design into two blocks to prevent survey fatigue, (explained in section 5.4.2.2). Each block includes six out of the 12 questions.

In collecting the airport respondent data, the passengers in the airport departure lounges of the four selected towns were randomly approached. A total of 950 airport respondents completed the SP survey questionnaire, with 474 respondents answering the survey with block 1 mode choice questions and the remaining 476 answering the survey with block 2 questions. The non-airport respondent survey data were collected in community areas, (such as parks, public libraries, colleges and shopping centres) in the four selected towns. A total of 863 non-airport respondents answered the SP survey, with 441 respondents answering the survey with block 1 mode choice questions and the remaining 422 answering the survey with block 2 questions. For details of the survey data collection methods, sources and time duration please refer to section 3.4.2.

### 7.3.2 Mode-based clustering with EM estimator

The mixture model-based clustering approach assumes that there is an underlying probability density function for each of the clusters. The mixture model is shown in Equation 7-1 below.

$$P(x_n) = \sum_{c=1}^{C} P(c)P(x_n \mid c) \qquad 7\text{-}1)$$

where $P(x_n)$ is the total or unconditional prior probability of observed vector data point $x_n$ (i.e., customer $n$ with multiple attributes) over all clusters, $P(c)$ is prior class membership probability (also named mixture weight) of cluster (or segment) $c$, thus the sum will be 1. ($\sum_{c=1}^{C} P(c) = 1$). $P(x_n \mid c)$ is the conditional prior probability of taking $x_n$ from cluster $c$, which is explained by a set of vector parameters such as the vector mean and covariance matrix for the cluster that underlays a multivariate Gaussian distribution (explained in the following sections).

The EM algorithm is an iterative and maximum log-likelihood method for estimating these mixture model parameters, including $P(c)$ and $P(x_n \mid c)$, that therefore identify the clusters. After adopting random or initial guess values of the mixture model parameters, the expectation (E) step and maximisation (M) step are carried out iteratively to calculate the posterior cluster probabilities for each data point. These are then used to re-evaluate the model parameters, until the desired convergence of maximised likelihood occurs or, as Witten et al. (2016, p. 288) recommended, until the increase in log-likelihood becomes negligible. The final local optimised model parameters and the corresponding clusters can then be identified (Xia et al., 2010; Kishor and Venkateswarlu, 2016; Witten et al., 2016). The flowchart (Figure 7-1) below illustrates the six main stages of the EM algorithm, in sequence, used for estimating the mixture model parameters and identifying the clusters (i.e., market segments).

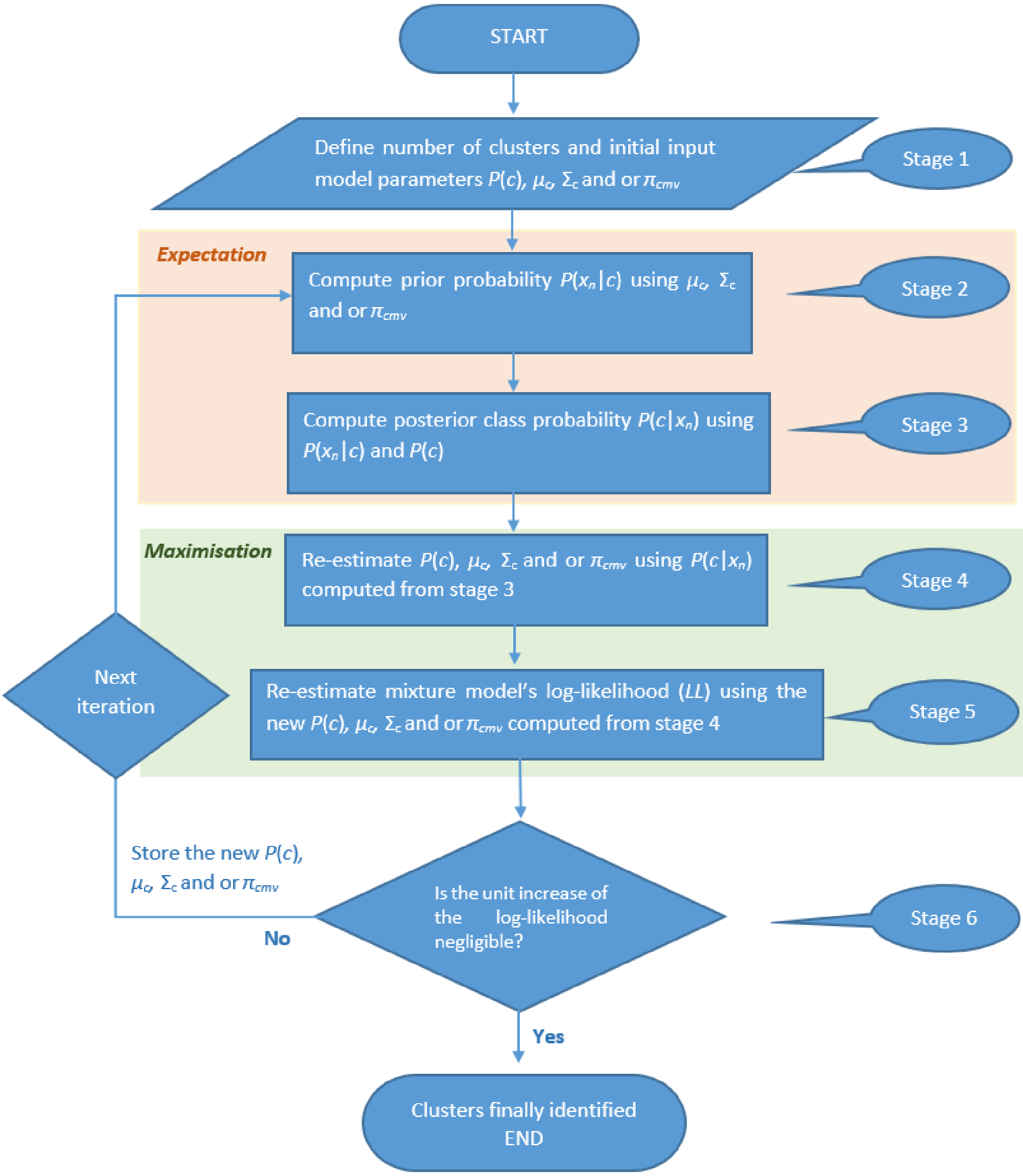



*Figure 7-1 Flowchart for EM algorithm*

#### 7.3.2.1 Stage 1

**Stage 1** is to predefine the number of clusters ($C$) that EM has to identify and make an initial guess of the model parameters for each cluster. These parameters are vectors and are different with respect to the numeral and nominal attributes of the observed vector data.

For a dataset of $N$, with multidimensional ($d$-dimension) numeral attributes, (e.g., quantified preference values for air and non-air transport), it is assumed that the numeral attributes under each cluster $c$ $(c \in C)$ follow a multivariate Gaussian distribution. Thus, the initial guesses for these parameter vectors are given as below:

- $P(c)$ is prior class membership probability of cluster $c$
- $\mu_c$ is the mean column vector corresponding to the multiple attributes conditional on cluster $c$
- $\Sigma_c$ is $d * d$ covariance matrix of cluster $c$

However, for a dataset of $N$, with multidimensional ($M$-dimension) nominal attributes, where the attributes are mutually independent, the assumption of a multivariate Gaussian distribution of the clusters is not valid. In such a case, the distributions of the clusters are assumed to be multi-way contingency tables of frequency counts of the attributes (Xia et al., 2010; Agresti and Kateri, 2011), and "*a nominal attribute with v possible values is characterized by v numbers representing the probability of each one*" (Witten et al., 2016, p. 289). That is, assuming the dataset has $C$ clusters and $M$ nominal attributes, (e.g., gender and education background), where each nominal attribute $m$ $(m \in M)$ (e.g., gender) contains $V_m$ possible category values, (e.g., male or female), the initial guesses of the parameter vectors are per below:

- $P(c)$ is prior class membership probability of cluster $c$ as previously mentioned
- $\pi_{cmv}$ is the cluster conditional probability of the observed data point $x_n$ from cluster $c$ with $v^{th}$ category value of nominal attribute $m$, for each $m$, $\sum_{v=category1}^{V_m} \pi_{cmv} = 1$

#### 7.3.2.2 Stages 2 and 3

**Stage 2** is the first part of the E step and uses the parameter vectors of each cluster to compute the conditional prior probability $P(x_n \mid c)$ of finding vector data point $x_n$ from cluster $c$.

Equation 7-2a shows the way to calculate the $P(x_n \mid c)$ for observed data point $x_n$ with multidimensional numeral attributes, where the $|\Sigma_c|$ is the determinant of covariance matrix $\Sigma_c$; index $T$ is a transpose of the column vector; and $\Sigma_c^{-1}$ is the inverse matrix of $\Sigma_c$.

$$P(x_n \mid c) = \frac{1}{\sqrt{(2\pi)^d \mid \Sigma_c \mid}} e^{-\frac{1}{2}(x_n - \mu_c)^T \Sigma_c^{-1} (x_n - \mu_c)} \qquad \text{7-2a)}$$

Equation 7-2b was used to calculate the $P(x_n | c)$ for observed data point $x_n$ with multidimensional nominal attributes, where $y_{nmv}$ equals 1 if the data point $x_n$ takes category value $v$ for $m^{th}$ nominal attribute, otherwise, equals 0.

$$P(x_n | c) = \prod_{m=1}^{M} \prod_{v=category1}^{V_m} (\pi_{cmv})^{y_{cmv}} \qquad 7\text{-}2b)$$

**Stage 3** is the last part of the E step and uses the old $P(c)$ and the $P(x_n | c)$ from the previous stage to compute $P(c | x_n)$ through Bayes' theorem as shown in Equation 7-3. Here $P(c | x_n)$ is the posterior class probability that the observed vector data point $x_n$ comes from cluster $c$.

$$P(c | x_n) = \frac{P(c)P(x_n | c)}{\sum_{c=1}^{C} P(c)P(x_n | c)} \qquad 7\text{-}3)$$

### 7.3.2.3 Stages 4 and 5

**Stage 4** is the first part of the M step and uses the $P(c | x_n)$ computed from stage 3 to re-estimate prior class membership probability $P(c)$ and the vectors $\mu_c$, $\Sigma_c$ and or $\pi_{cmv}$ for each cluster. The formulas are given in Equations 7-4 to 7-7, where $N$ is the number of data points/customers in the dataset.

$$P(c) = \frac{\sum_{n=1}^{N} P(c | x_n)}{N} \qquad 7\text{-}4)$$

$$\mu_c = \frac{\sum_{n=1}^{N} P(c | x_n) x_n}{\sum_{n=1}^{N} P(c | x_n)} \qquad 7\text{-}5)$$

$$\Sigma_c = \frac{\sum_{n=1}^{N} P(c | x_n)(x_n - \mu_c)(x_n - \mu_c)^T}{\sum_{n=1}^{N} P(c | x_n)} \qquad 7\text{-}6)$$

$$\pi_{cmv} = \frac{\sum_{n=1}^{N} P(c | x_n) y_{nmv}}{\sum_{n=1}^{N} P(c | x_n)} \qquad (v \in V_m, m \in M) \qquad 7\text{-}7)$$

**Stage 5** is the last part of the M step that subsequently re-estimates the mixture model's log-likelihood ($LL$). In this step, the new $\mu_c$, $\Sigma_c$ and/or $\pi_{cmv}$ re-estimated from stage 4 is used to re-estimate the conditional prior probability $P(x_n | c)$ for each cluster using Equations 7-2a and 7-2b. The mixture model's log-likelihood ($LL$) can then be re-estimated/re-maximised with the new $P(c)$ computed from stage 4 and a new $P(x_n | c)$ computed at the current stage, as given by Equation 7-8.

$$LL = \log P(x_1, x_2, ..., x_n) = \sum_{n=1}^{N} \log \sum_{c=1}^{C} P(c) P(x_n \mid c) \quad 7\text{-}8)$$

7.3.2.4 Stage 6

**Stage 6** is to test whether the increase in the re-maximised log-likelihood compared to the previous one is negligible - change in the improvement of the log-likelihood less than 1% or improvement less than 1.0E-6. If the test fails, the new $P(c)$, $\mu_c$, $\Sigma_c$ and/or $\pi_{cmv}$ is stored and the procedure returns to step 2. This iterative process continues until the convergence requirement is met, in which case the EM algorithm finishes and the clusters are identified, with the optimized mixture model parameters $P(c)$, $\mu_c$, $\Sigma_c$ and/or $\pi_{cmv}$ of each cluster determined.

7.3.3 Number of clusters and clustering software

One potential issue when applying an EM algorithm to estimate the mixture model parameters is that the converged log-likelihood obtained is a local maximum that may or may not also be the global maximum (Witten et al., 2016). A solution is to rerun the algorithm a number of times, with a different set of initial inputs, and choose the largest figure of the local maxima, that is more likely to be the global maximum. Defining the number of clusters is another issue that must be considered when using the EM algorithm. Log-likelihood will increase with the number of clusters but too many clusters may cause over-fitting of the model (Heckman and Singer, 1984). The AIC is one of the most widely used indexes for determining the optimal number of classes. It is calculated using $-2LL + 2K$, where $LL$ is the log-likelihood and $K$ is the number of free parameters in the mixture model. AIC gives a penalty for increasing the number of clusters that can therefore help to find the model that balances model fit and parsimony.

In this study, *WEKA 3.9* software was used to perform the EM clustering analysis. It is an open source tool written in Java language that provides a range of machine learning and clustering algorithms to users (Sharma et al., 2012). *WEKA* offers a Naïve Bayes related method that can effectively cluster vector data that contains both numeric and nominal attributes, with an assumption of independence between any attributes (Witten et al., 2016). It commonly assumes a diagonal covariance matrix for every mixture component that actually simplifies the calculations. In terms of the vector data that were mixed with numeric and nominal attributes, the joint distribution of each component is represented by the product of a normal distribution of each numeric attribute and frequency counts-based discrete distribution of each nominal attribute. However, $C$ was set to a different value to get different classification results based on the different cluster numbers using the EM algorithm. The random initialisation for computing the EM algorithm was set to a relatively larger number, i.e. 1,000. Eventually, the classification with the optimal cluster number was found by selecting the lowest AIC (Akaike, 1998; Xia et al., 2010).

Many existing studies defined transport market segments by socio-demographics, service-quality attributes and trip characteristics (e.g., Mason and Gray, 1995; Wen et al., 2008; Harrison et al., 2015). The selection of segmentation variables mainly depends on the purpose of the analysis (Xia et al., 2010). This chapter aims to identify the characteristics of the segments of regional travellers with different degrees of preference for air travel. Thus, the model-based market segmentation approach was used to uncover the segments based on socio-demographics, trip purpose and stated preference between air and non-air travel modes (measured as mean probabilities). The mode choice probability was calculated based on the six stated hypothetical choice questions of the stated preference survey. However, as described previously, each respondent was randomly allocated to one of the two blocks, each consisting of six choice questions. Thus, due to the choice questions in both blocks being different, this chapter investigates and compares the characteristics of identified market segments for airport respondents (474) and non-airport/community respondents (476) who answered the block 1 mode choice questions. The segmentation results for the remaining airport respondents (441) and community respondents (422), who answered the block 2 choice questions are reported in Appendices F & G as a reference for validation.

## 7. 4 Results

### 7.4.1 Air passenger market segmentation

This section segments the air passenger data (represented by the airport respondent sample), using the EM algorithm to estimate the mixture model parameters. Table 7-1 reports the converged log-likelihood and AIC results for the mixture models with between 1 and 5 segments. Of these, the 3-segment mixture model was found to have the lowest AIC value (-4,571.33) and was used as the optimal model for the air passenger market segmentation. One cautionary point here is that the log-likelihood results varied significantly with the different number of segments. One explanation is that the EM algorithm uses a Gaussian distribution probability density function to calculate the conditional prior probability for numerical attributes, whereas the probability density values may have a relatively large variation range and can be larger than 1.

*Table 7-1 Information criteria for determining optimal number of air passenger segments*

| No. of segments | N[a] | Log likelihood (LL) | K[b] | AIC [c] | Segment size |
|---|---|---|---|---|---|
| 1 | 474 | -978.40 | 13 | 1,982.80 | 100% |
| 2 | 474 | 1,145.95 | 27 | -2,237.90 | 21%,79% |
| **3** | **474** | **2,326.67** | **41** | **-4,571.33** | **54%,18%,28%** |
| 4 | 474 | 329.28 | 55 | -548.57 | 28%,22%,27%,23% |
| 5 | 474 | 387.13 | 69 | -636.27 | 20%,17%,26%,27%,10% |

*Table notes:*
*[a] N is the sample size, [b] K is the number of free parameters, [c] AIC = -2LL+2K.*

Table 7-2 reports the estimation results of the 3-segment mixture model for the air passenger respondent data. The segment size/mixture weight statistics indicate that segments A1, A2 and A3 contain 54%, 18% and 28% of the airport respondents from the sample, respectively. In terms of categorical variables, (e.g., gender), the $\pi_{cmv}$ values, (given as percentages in the table), show the probability that an air passenger belongs to a certain category/attribute, (e.g., male or female), of the variable for each of the segments. The category with the largest probability compared to the remaining corresponding categories is in bold font, and the cell is shaded if the category dominates.

Segment A1 accounted for 54% of the air passenger sample size; the travellers in this segment strongly preferred air travel for their regional trips. This group mainly comprised of air passengers who were male (61%), aged between 25 and 44 (70%), high income (66%), tertiary educated (78%) and travelling for business purposes (78%). Segment A2 was the smallest air market segment and contained 18% of the total sample population, with these passengers stating an almost equivalent preference for using air or non-air travel modes. Segment A2 was mainly comprised of passengers who were male (61%), middle or older age (55%), middle or high income (78%) and tertiary educated (62%). Similar to segment A1, segment A3 was also a market segment that highly preferred air travel, and made up 28% of the air passenger sample population. The majority of passengers in this segment were middle or older age (93%), high income (66%) and tertiary educated (69%). The larger ratio of standard deviation to the mean probability indicates that the corresponding mode choice probability density has more variation relative to the mean.

*Table 7-2 Air passenger market segments*

| Characteristics | Airport passenger | Segment A1 | Segment A2 | Segment A3 |
|---|---|---|---|---|
| Segment size | Proportion of sample | **54%** | 18% | 28% |
| Car probability | Mean | 0.048 | 0.364 | 0.044 |
| | Std. dev. | (0.076) | (0.277) | (0.068) |
| Bus probability | Mean | 0.000 | 0.111 | 0.003 |
| | Std. dev. | (0.000) | (0.169) | (0.007) |
| Airline probability | Mean | **0.952** | **0.525** | **0.954** |
| | Std. dev. | (0.076) | (0.254) | (0.070) |
| Trip purpose | Business | **78%** | 42% | **57%** |
| | Non-business | 22% | **58%** | 43% |
| Gender | Female | 39% | 39% | 47% |
| | Male | **61%** | **61%** | **53%** |
| Age | Under 25 | 7% | 13% | 2% |
| | 25 to 44 | **70%** | 33% | 5% |
| | 45 or more | 23% | **55%** | **93%** |
| Education background | Basic education | 22% | 38% | 31% |
| | Tertiary education | **78%** | **62%** | **69%** |
| Income | Low income[a] | 2% | 22% | 11% |
| | Middle income[b] | 32% | 32% | 24% |
| | High income[c] | **66%** | **46%** | **66%** |

*Table notes:*
*[a] Low income is defined as a monthly income between \$0 and \$1749 Australia dollars.*
*[b] Middle income is defined as a monthly income between \$1749 and \$5499 Australia dollars.*
*[c] High income is defined as a monthly income of \$5500 or more Australia dollars.*

### 7.4.2 Non-air passenger market segmentation

For the market segmentation of the non-air passenger data, (represented by the non-airport/community respondent sample), mixture models with the number of segments varying from one to five were also estimated. Table 7-3 summarises the converged log-likelihood and AIC values for the five mixture models. The mixture model with three segments again has the smallest AIC value, which indicates that it is the best model fit for the non-air passenger market segmentation.

*Table 7-3 Information criteria for determining optimal number of non-air passenger segments*

| No. of segments | N | Log likelihood (LL) | K | AIC | Segment size |
|---|---|---|---|---|---|
| 1 | 441 | -1,832.13 | 13 | 3,690.27 | 100% |
| 2 | 441 | -860.32 | 27 | 1,774.64 | 59%,41% |
| **3** | **441** | **-657.75** | **41** | **1,397.49** | **55%,22%,23%** |
| 4 | 441 | -709.39 | 55 | 1,528.78 | 19%,39,17%,25% |
| 5 | 441 | -971.43 | 69 | 2,080.86 | 16%,42%,16%,12%,14% |

Table 7-4 shows the three distinct segments identified based on the non-air passenger respondents' data. Segment N1 makes up 55% of the non-air passenger sample. Travellers in this segment were more likely to choose air travel for their regional trips than to drive a car or take a bus, with 61% of the members usually travelling for a non-business purpose. Additionally, the travellers in this segment were mainly middle or high income (73%) with a tertiary education level (72%). Segment N2 accounted for 22% of the total non-air passenger sample population. The majority of travellers in this segment were aged between 25 and 44 (50%), middle-income (57%) with a tertiary education level (65%). Eighty percent of the regional travellers were non-business travellers. Travellers in this segment stated a high preference for driving a car for their regional trips. Segment N3 contained 18% of the total sample population, where the travellers had a relatively equal probability of using air and non-air travel modes. Four fifths of the travellers in this segment were non-business travellers. The majority of travellers in this segment were under 25 years old (55%), low income (55%) and had a basic education (66%).

*Table 7-4 Non-air passenger market segments*

| Characteristics | Non-airport passenger | Segment N1 | Segment N2 | Segment N3 |
|---|---|---|---|---|
| Segment size | Proportion of sample | **55%** | 22% | 23% |
| Car probability | Mean | 0.122 | **0.735** | 0.256 |
| | Std. dev. | 0.125 | 0.268 | 0.204 |
| Bus probability | Mean | 0.001 | 0.061 | 0.234 |
| | Std. dev. | 0.007 | 0.090 | 0.217 |
| Airline probability | Mean | **0.878** | 0.204 | **0.510** |
| | Std. dev. | 0.126 | 0.239 | 0.240 |
| Trip purpose | Business | 39% | 19% | 20% |
| | Non-business | **61%** | **81%** | **80%** |
| Gender | Female | **54%** | 44% | **58%** |
| | Male | 46% | **56%** | 42% |
| Age | Under 25 | 15% | 11% | **55%** |
| | 25 to 44 | **47%** | **50%** | 24% |
| | 45 or more | 38% | 39% | 21% |
| Education background | Basic education | 28% | 35% | **66%** |
| | Tertiary education | **72%** | **65%** | 34% |
| Income | Low income | 27% | 24% | **55%** |
| | Middle income | **46%** | **57%** | 31% |
| | High income | 27% | 19% | 14% |

## 7. 5 Discussion

### 7.5.1 General characteristics of air passenger market segments

Based on the airport respondent dataset, three distinct market segments were identified by the EM algorithm-based mixture model. This is similar to the findings by Mason and Gray (1995), who found three representative segments for short haul aviation markets in the European Union. Segment A1, with the largest share (54%), can be considered as the predominant target for the regional aviation market and, therefore, airlines could customise more efficient programmes based on the characteristics of this market segment. For instance, as passengers in this target market were mainly business travellers aged between 25 and 44, airlines could adjust their air services to be more popular with the young to middle-aged business traveller. However, the remaining two segments comprise 46% of the total sample and should not be ignored. The results show that the travellers in segments A1 and A3 stated an extremely high preference for choosing air travel for their regional travel. Nevertheless, in contrast with the target segment A1, there was no dominance either in trip purpose or gender in segment A3, but the travellers in this segment were almost all middle-aged or older travellers. The travellers in segment A2 stated a relatively equal preference for choosing either air or non-air travel mode. In comparison with segments A1 and A3, the only significant difference in characteristics is that a larger proportion of low and middle-income persons were found in segment A2. Generally, relatively high income and education level are the common characteristics of the three segments, which is consistent with the finding of Wen et al. (2008) about segmentation of the international air market. However, Wen et al. (2008) discovered that youth and non-business travellers were the dominant common characteristics across the existing international airline market segments, that is contrary to the findings from this chapter on the regional air passenger market segments. The difference indicates that international travellers may have a significant variation of age group and trip purpose compared to domestic travellers.

### 7.5.2 General characteristics of non-air passenger market segments

Similar to the air passenger respondents, the non-air passenger (non-airport) respondents showed three distinct segments in the clustering analysis. As reported in Table 7-4, segment N1 was the largest target aviation market (55%). Interestingly, these travellers stated a high preference for choosing an airline (87.8%) as their regional transportation mode. In contrast, travellers in segment N2 (22% of the sample) were more likely to use non-air transport (car: 73.5%, bus: 6.1%) for their regional travel. The travellers in segment N3 were found to have a relatively equal preference for air or non-air travel modes. Segment N2 and N3 both stated a relatively high preference for road transport. There was also a high similarity of characteristics between the two segments, such as the low and middle-income levels and the large proportion of non-business travellers. On the other hand, segment N1 has a higher average air travel preference. One possible reason for this is that, compared to the other two segments, this segment has a larger percentage of travellers who are tertiary educated, have high income and would travel for business purposes. This finding reveals the importance of the non-airport/community travellers, as a considerable proportion of them may moderately or highly prefer to travel by air.

### 7.5.3 Existing and potential aviation markets

Figure 7-2 presents the main (dominant) characteristics of the existing aviation, (more likely to use air transport), and potential aviation markets, (less likely to use air transport), based on the segmentation results of the air and non-air passenger data in Tables 7-2 and 7-4 respectively. The existing aviation market contains three market segments (A1, A3 and N1), where the travellers from each of the segments stated a very high probability of choosing air travel for their regional trips. Segment N1 was the largest segment of the non-air passenger respondents, which reflects that, in addition to the segments of air passenger respondents, the non-airport respondent sample also contained a relatively large proportion of respondents who were likely to travel by regional air transport. The common characteristics among the three segments were the relatively high incomes and education levels.

The potential, or latent, aviation market includes the remaining three segments (A2, N2 and N3), where the respondents moderately or rarely preferred air travel, although segments A2 and N3 were distinctly different in most of their prominent characteristics. For example, segment A2 accounted for more middle-aged or older males who commonly had a tertiary education and a middle to high income, while segment N3 mainly comprised of young non-business travellers who had a basic education and a low to middle income. Both segments had a relatively equal probability of using air or a non-air travel mode, which indicates a high potential value for the aviation industry. The airlines could focus on investigating strategies and advertising that targeted these travellers who might be more easily attracted to use air transport. In contrast, segment N2 showed only a low probability of using an airline, but a high probability of using a car. These travellers were more frequent non-business travellers who were relatively young, low income and with a basic education level. Notably, although segment A2 was identified from air passenger respondents, it actually belongs to the potential aviation market since they have a moderate preference of air transport. As a whole, the findings of the existing and potential market could inform the airlines, so that they could customise more efficient strategies based on the characteristics of these segments, in order to not only compete for the existing aviation market but also to attract the potential aviation market as new air travellers.

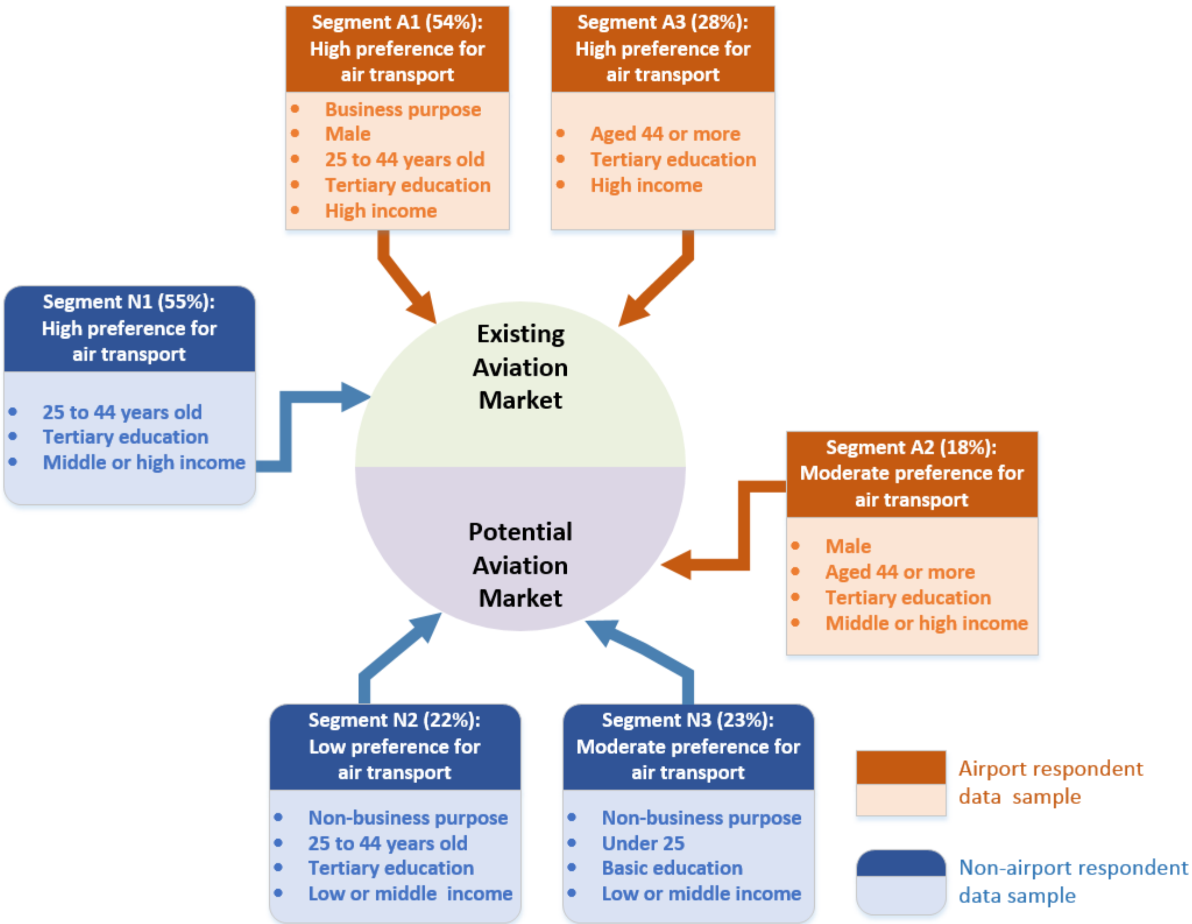



*Figure 7-2 Prominent characteristics of existing aviation and potential aviation market*

*Figure notes:*
*Low preference is defined as the mean probability of choosing a given travel mode between 0.000 and 0.300.*
*Moderate preference is defined as the mean probability of choosing a given travel mode between 0.301 and 0.600.*
*High preference is defined as the mean probability of choosing a given travel mode between 0.601 and 1.000.*

### 7.5.4 Results validation

The market segmentation of the airport and non-airport (community) respondents who completed the survey with block 2 questions were also assessed, with the results reported in the tables of Appendix F and G. The AIC indices also suggest that three distinct segments can be found in both the airport and non-airport respondent groups. The segmentation results in terms of segment size, characteristics and stated preferences for block 2 questions are similar and consistent with the market segmentation findings of respondents who did the survey with block 1 questions. Therefore, this result further confirms and validates the reliability of the aviation market segmentation findings in this chapter.

## 7. 6 Summary

This chapter investigated the regional aviation market using the EM algorithm-based mixture model as a market segmentation approach. This thesis is particularly interested in identifying the characteristics of existing and potential aviation market segments of airport and non-airport passengers. The SP survey was applied to estimate regional travellers' average selection probability for the air and road travel modes. The proposed mixture model uncovered three distinct market segments for both the airport and non-airport passenger groups of respondents. The prominent characteristics of travel mode preference, demographics, socio-economics and trip purpose between the market segments were subsequently compared. The segments with high air travel preference were assigned to the existing aviation market, while the rest of the segments were classified as potential aviation markets as they stated a moderate to high mean probability of choosing non-air transport (car or bus). While the market segmentation results were derived using a case study in regional Western Australia, the findings could potentially also inform airlines and government transport agencies in a similar regional context on the development of strategies and policies.

This and previous chapters have applied an exploratory analysis of the intercept survey data that has provided sufficient insights for the local government and airlines to better understand regional air and non-air travellers' characteristics, especially the identification of the existing and potential aviation markets in Western Australia. In order to achieve a more comprehensive and direct understanding of the competition within the regional aviation market, the travel mode and airline choice of regional travellers is of prime concern. Thus, the next chapter develops MNL and NL models to estimate the travel mode and airline choice.

# CHAPTER 8 ESTIMATING TRAVEL MODE AND AIRLINE CHOICE USING LOGIT MODELS

## 8. 1 Introduction

The previous two chapters have explored the characteristics of the regional aviation markets in Western Australia, based on the field survey. In this chapter, the focus moves to analyse the SP survey data in order to estimate the regional airport and non-airport passengers' travel mode and airline choice behaviours and preferences.

This chapter is based on the second published work[7] resulting from this thesis, published in the *Journal of Air Transport Management* (Zhou et al., 2019). Section 8.2 briefly introduces the research context including the motivation and purpose, and section 8.3 expounds the methodology for travel mode and airline choice estimation. The interpretation and discussion of results are presented in sections 8.4 and 8.5, respectively, the chapter concluding with a summary (section 8.6).

## 8. 2 Research Context

Although air transport is more time-efficient relative to road transport, it generally costs more, particularly the direct and marginal costs. More than that, the decision around travel mode choice depends upon a range of parameters such as travel time, travel cost, accessibility, seat comfort and service frequency (Pels et al., 2000; Hess et al., 2007; Chang and Sun, 2012; Van Can, 2013). The choice process may also differ based on the purpose of the trip (Van Can, 2013; Jung and Yoo, 2014). For instance, it may be that business travellers are less flexible and hence consider time as the primary concern. Conversely, for the non-business or leisure travellers, they may value travel cost as the priority in making their travel mode choice. Therefore, for this competitive passenger market, it is becoming particularly important for government policy makers and airlines to understand passenger sensitivity or preference to these key factors.

The aim of this chapter is to investigate the travel mode choice and behaviour of travellers on competitive routes served by air transport in regional Western Australia, using the SP data

[7] Zhou, H., Xia, J., Norman, R., Hughes, B., Nikolova, G., Kelobonye, K., . . . Falkmer, T. (2019). Do air passengers behave differently to other regional travellers?: A travel mode choice model investigation. Journal of Air Transport Management, 79, 101682. doi:https://doi.org/10.1016/j.jairtraman.2019.101682

collected at regional airports and other locations. In order to achieve this aim, a three-stage approach was developed. Firstly, discrete choice models including MNL and NL models were used to estimate and compare business and non-business traveller mode choice behaviour among car, bus and two unnamed regional airlines, for trips within Western Australia. Secondly, the model fit statistics from the MNL and NL models were compared, in order to find out which one provided a better overall model fit and thus would be more appropriate for estimating mode choice. Finally, the differences in travel mode choice behaviour between air (airport respondents) and non-air passengers (non-airport/community respondents) were then investigated by estimating their willingness to pay for different transport modes and the characteristics of each. Although this chapter modelled the travellers' modal choices in the Western Australian context, the findings of the study may be applicable to other geographically and socioeconomically similar regions, to provide insight for government policy makers and airlines seeking to influence travel behaviour.

## 8. 3 Methodology

### 8.3.1 Stated preference survey design

As discussed in section 5.4.1, travel cost, journey time, access time, service frequency and seat comfort were considered as the attributes when estimating individual regional travel mode choice in Western Australia. The four mode alternatives are car, bus and unnamed regional airlines 1 and 2. As shown in Table 5-7, the attribute-levels were defined based on air travel and non-air travel data, (i.e., travel cost and travel time), collected in regional Western Australia, and were intended to be plausible values. One important point is that seat comfort is multifaceted; this study used leg room distance as a measurement of seat comfort level. Also, the seat comfort level of a car was assumed to be either medium or high, as the driving seat normally can be adjusted to a relatively greater degree than a typical bus or airline seat.

SP surveys can be used for detecting and estimating the subjective preference of individuals, and thereby understanding people's choice behaviour, in a range of research areas, including transportation and health (Hess et al., 2007; Johnson et al., 2013; Shang and Zhang, 2013). As detailed in the SP experimental design chapter (Chapter 5), a D-efficient design method with EMFA was used to generate the SP survey questionnaire. This efficient design method can not only maximise t-ratios, which improves the statistical validity of the estimated parameters, but also minimise the sample size while still maintaining a statistically significant t-ratio. Indeed, some papers have already shown that efficient designs can produce more significant t-ratio values and generate more reliable model estimates (Kessels et al., 2006; Ferrini and Scarpa, 2007; Rose and Bliemer, 2009). As stated in section 5.4.3.3, *Ngene 1.2.0* was used to generate the SP survey with respect to MNL model, (i.e. it was optimized for the MNL model), whereas the variables/attributes in the choice scenarios were independent of each other. The efficient SP design optimized for an MNL model has also been found to perform well in analysing survey data using an NL model (Sándor and Wedel, 2002; Bliemer et al., 2007; Rose and Bliemer, 2009).

### 8.3.2 Data used in this study

The data used in this chapter are, as previously, the airport and non-airport respondent SP data collected in the four selected regional towns (Albany, Geraldton, Broome and Karratha). The air respondent SP surveys were collected in the airport departure lounges, with a total of 950 airport passengers completing the survey questionnaire, as summarised in Table 3-2 of section 3.4.2. Before the survey, the researchers asked each of the respondents what their current trip purpose was and then handed them a survey questionnaire tailored to that purpose. As previously indicated, a total of 621 business airport respondents answered a set of hypothetical SP travel mode choice questions assuming they were on a business trip, while the remaining 329 non-business airport respondents completed the non-business hypothetical SP choice questions. Table 8-1 presents the profiles of the airport respondents.

*Table 8-1 Airport respondent profiles*

| Variable | Frequency | Proportion |
|---|---|---|
| **Age** | | |
| 16-17 | 5 | 0.5% |
| 18-34 | 267 | 28.1% |
| 35-54 | 408 | 42.9% |
| 55 or older | 253 | 26.6% |
| Not Stated | 17 | 1.8% |
| **Gender** | | |
| Male | 543 | 57.2% |
| Female | 389 | 40.9% |
| Not Stated | 18 | 1.9% |
| **Monthly income** | | |
| Less than $3,499 | 189 | 19.9% |
| $3,500-$6,499 | 260 | 27.4% |

| | | |
|---|---|---|
| $6,500-$8,699 | 136 | 14.3% |
| $8,700 or more | 282 | 29.7% |
| Not Stated | 83 | 8.7% |

The non-airport respondent SP survey data were collected in Albany, Geraldton, Broome and Karratha, specifically from community locations including public libraries, town streets, shopping centres, regional colleges and town parks. A total of 863 non-airport respondents completed the survey, with the sample distribution of the SP survey shown in Table 3-3 of section 3.4.2. As previously stated, a total of 265 non-airport respondents answered the business purpose hypothetical SP choice questions and 598 answered the non-business purpose hypothetical SP choice questions. Each respondent was asked to answer six choice scenarios. Table 8-2 presents the profiles of the non-airport respondents.

*Table 8-2 Non-airport respondent profiles*

| Variable | Frequency | Proportion |
|---|---|---|
| **Age** | | |
| 16-17 | 87 | 10.1% |
| 18-34 | 317 | 36.7% |
| 35-54 | 234 | 27.1% |
| 55 or older | 203 | 23.5% |
| Not Stated | 22 | 2.5% |
| **Gender** | | |
| Male | 393 | 45.5% |
| Female | 447 | 51.8% |
| Not Stated | 23 | 2.7% |
| **Monthly income** | | |
| Less than $3,499 | 419 | 48.6% |
| $3,500-$6,499 | 219 | 25.4% |
| $6,500-$8,699 | 50 | 5.8% |
| $8,700 or more | 88 | 10.2% |
| Not Stated | 87 | 10.1% |

### 8.3.3 MNL model

As introduced in the reseach framework chapter (Chapter 3), logit models, (including MNL and NL), are a commonly used approach in discrete choice modelling. They assume that there is an underlying preference scale over the set of alternatives and that the individual will select the alternative with the highest utility (Anderson et al., 1992; Hensher et al., 2015a). The MNL model is the simplest and most widely used discrete choice model for understanding people’s choice behaviour (Chang and Sun, 2012; Van Can, 2013; Jung and Yoo, 2014). McFadden (1973) and Ben-Akiva et al. (1985) introduced the MNL model as a generalised binary logit model that can help to understand the respondent's preference among finite alternatives.

In the MNL model, the utility function of an individual *n* choosing alternative *j* among all the alternatives is given as follows:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad (j \in J_n) \qquad 8\text{-}1)$$

where:

$V_{nj}$ is the observed utility of alternative *j*,

$\varepsilon_{nj}$ is the unobserved utility (or error term) of alternative *j*, $J_n$ is the set of all the alternatives that an individual *n* can choose.

The observed utility $V_{nj}$ of alternative *j* is expressed as follows:

$$V_{nj} = \alpha_j + \sum_{k=1}^{K_j} \beta_k x_{njk} \qquad 8\text{-}2)$$

where

$\alpha_j$ is a constant that contributes to alternative *j*'s observed utility,

$x_{njk}$ is the $k^{th}$ attribute or explanatory variable that can influence the utility of alternative *j*,

$\beta_k$ is the parameter of the explanatory variable,

$K_j$ refers to the number of the explanatory variables related to alternative *j*.

Therefore, in the MNL model, the probability of an individual *n* choosing alternative *j* among all the alternatives $J_n$ is shown as follows:

$$P_{nj} = \frac{\exp(V_{nj})}{\sum_{j \in J_n} \exp(V_{nj})} \qquad 8\text{-}3)$$

### 8.3.4 NL model

The IID assumption used in the MNL model is likely to be unrealistic in a number of settings. For example, it is reasonable to assume that in the current experiment the preferences for the two airline options are correlated. The NL model allows for this type of correlation by permitting a partial relaxation of the IID and IIA assumption (Garrow, 2010; Hensher et al., 2015a). The alternatives within the same nest of the NL model share a common error term and the covariance between the alternatives in the same nest is not equal to zero, but the alternatives across different nests have independent error terms, (covariance between the alternatives from different nests is equal to zero). Thus, as a result of the relaxed IID assumption in the NL model, the alternatives in the same nest can have some degree of correlation or substitution, but are independent and irrelevant to each other across different nests.

In the NL model, the utility function of an individual *n* choosing alternative *j* under nest *b* is given by equation 8-4 (Hensher et al., 2015a):

$$U_{nj} = \mu_{j|b} V_{nj|b} + \varepsilon_{nj} \qquad 8\text{-}4)$$

where:

the $\mu_{j|\mathrm{b}}$ is a scale parameter estimated by the data, the value is inversely proportional to the error term variance,

$V_{nj|b}$ is the same as an observed utility function in an MNL model,

$\mu_{j|b} \cdot V_{nj|b}$ is an actual observed utility of alternative *j* under nest *b* in an NL model.

The utility function of nest *b* is calculated from the observed utility of all alternatives within the nest using Equation 8-5 (Hensher et al., 2015a):

$$V_b = \lambda_b \left(\frac{1}{\mu_{j|b}} \ln\left(\sum_{j \in b} \exp(\mu_{j|b} V_{nj|b})\right)\right) \qquad 8\text{-}5)$$

where the $\lambda_b$ is the scale parameter related to nest *b*.

Hence, the NL model is over-parameterised, which requires the normalisation of one or more parameters for the model identification (Ben-Akiva et al., 1985; Hensher et al., 2015a). Thus, as indicated by Hensher et al. (2015a), the $\lambda_b$ is normalised to 1[8], which leads to the NL model being normalised to a random utility 2 (RU2) NL model. Therefore, ratio $\frac{\lambda_b}{\mu_{j|\mathrm{b}}}$ will be normalised to $\frac{1}{\mu_{j|\mathrm{b}}}$, is give the name inclusive value (IV) parameter or logsum parameter of nest *b*, and will be estimated in the NL model (Garrow, 2010; Hensher et al., 2015a). The IV parameter or logsum parameter is also relevant to the correlation among alternatives under the

[8] Normalising $\mu_{j|\mathrm{b}}$ to 1 will produce a random utility 1 (RU1) model.

specific nest[9] and lies within the range of (0, 1). A value close to 1 indicates not only less correlation, or lower degree of substitution, between the alternatives within same nest, but also less difference in variance between adjoining levels in NL model. A value close to 0 implies the opposite. If the nest has only one alternative, the IV parameter will be fixed to 1 (Garrow, 2010; Hensher et al., 2015a). If the estimated IV parameters are statistically significant and also within a range of 0 to 1, it suggests that the developed NL model is an improvement over the MNL model (Hensher et al., 2015a). Consequently, the probability of an individual selecting alternative *j* under nest *b* in the NL model is calculated by Equation 8-7:

$$P_{nj} = P_{nj|b} \cdot P_{nb} = \frac{\exp(\mu_{j|b} V_{nj|b})}{\sum_{j \in b} \exp(\mu_{j|b} V_{nj|b})} \cdot \frac{\exp(\frac{1}{\mu_{j|b}} \ln(\sum_{j \in b} \exp(\mu_{j|b} V_{nj|b})))}{\sum_{b \in B} \exp(\frac{1}{\mu_{j|b}} \ln(\sum_{j \in b} \exp(\mu_{j|b} V_{nj|b})))} \qquad \text{8-7)}$$

where:

$P_{nj|b}$ is the conditional probability that individual *n* chooses alternative *j* in nest *b*,

$P_{nb}$ is the probability that individual *n* chooses nest *b*.

If all the nests in the model have only one alternative, (i.e., all the IV parameters are fixed to 1), the NL model reverts to an MNL model and the probability function degenerates to exactly the same as in the MNL model.

The observed utility function of the discrete choice model for this study was developed and described in section 5.4.2, but is repeated here for the convenience of the reader.

$$V_j = Constant_{mode} + \beta_1 TravelCost_j + \beta_2 AccessTime_j + \beta_3 JourneyTime_j$$
$$+\beta_4 Frequency_j + \beta_5 SeatComfortMid_j + \beta_6 SeatComfortHigh_j$$

where:

$TravelCost_j$ is the travel cost (ticket fare or the cost of driving) of alternative mode *j* (A$),

$AccessTime_j$ is the access time to a bus station or an airport (mins),

$JourneyTime_j$ is the travel time from origin to destination (hours),

$Frequency_j$ is the number of operating buses or flights per week,

[9] The correlation between the utility functions of any pair of alternatives under nest *b* in an NL model is represented by $1 - (\frac{\lambda_b}{\mu_{j|b}})^2$.

*SeatComfortMid$_j$* and *SeatComfortHigh$_j$* are dummy variables representing middle and high seat comfort level, respectively.

Figures 8-1 and 8-2 show the structures of the MNL and NL models developed in this chapter, respectively. For the two-level NL model, bus was put into the public non-air travel mode nest, airlines 1 and 2 into the public air travel mode nest, and car into the private non-air travel mode nest. Note that both the public and private non-air modes are degenerate nests as they contain only one alternative. Thus, the only effective difference between the two models is that the public air mode has two alternatives.

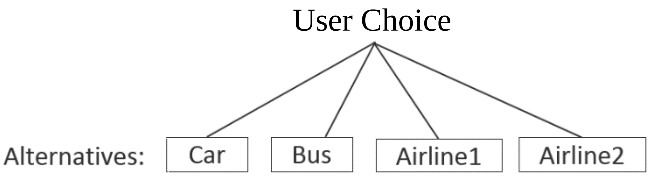



*Figure 8-1 Structure of the MNL model*

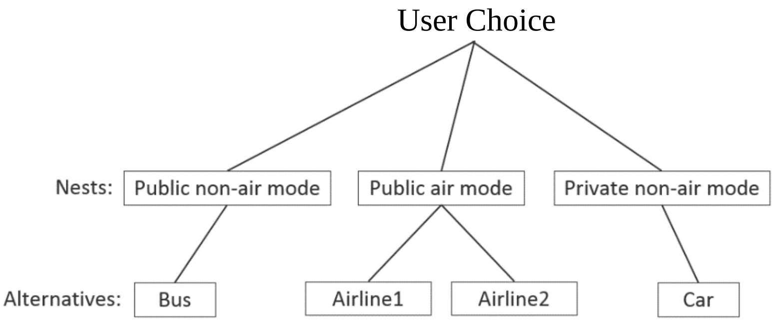



*Figure 8-2 Structure of the NL model*

### 8.3.5 Direct and cross elasticities

In order to further understand the competition between car, bus and air travel in regional Western Australia, this chapter calculated the direct and cross elasticities for travel cost, journey time and service frequency based on the NL model parameter estimates, using the SP data from both the airport and non-air respondents. A direct elasticity represents the percentage change in a dependent variable, (such as the probability of picking an alternative), caused by a one percent change in the explanatory variable (attribute) of interest. The direct elasticity function of an NL model is given by Equation 8-8 (Garrow, 2010):

$$E^{P_i}_{X_{ik}} = [(1-P_i)+(\mu_{i|b}-1)(1-P_{i|b})]\beta_k x_{ik} \qquad 8\text{-}8)$$

where:

$X_{ik}$ represents the explanatory variable $k$ of alternative $i$,

$P_i$ is the probability that the individual chooses alternative $i$,

$P_{i|b}$ is the conditional probability that the individual chooses alternative $i$ in nest $b$.

For the MNL model, $\mu_{i|b}$ is fixed to 1, giving the second term in the square brackets a value of zero, i.e., effectively disappears.

Additionally, a cross elasticity indicates the percentage change in the probability of a particular alternative being chosen due to a marginal change in a specified explanatory variable of another alternative. Thus, it is a vital indicator that represents competition between alternatives. The function of cross elasticity is shown in Equations 8-9a and 8-9b (Garrow, 2010):

$$E_{X_{ik}}^{P_j} = -P_i \beta_k x_{ik} \qquad \text{8-9a)}$$

$$E_{X_{ik}}^{P_j} = -[P_i + (\mu_{i|b} - 1)P_{i|b}]\beta_k x_{ik} \qquad \text{8-9b)}$$

Equation 8-9a calculates the cross elasticity of alternative $j$ concerning the per cent change of explanatory variable $k$ of alternative $i$, under the condition of $i$ and $j$ in different nests. Equation 8-9b calculates the cross elasticity, where $i$ and $j$ are in the same nest. For the MNL model, $\mu_{i|b}$ will be fixed to 1 instead.

### 8.3.6 Model fit statistics

#### 8.3.6.1 Log-likelihood function

The log-likelihood function is an important indicator of how well the model estimations match the observed choice probabilities determined from the survey data, (i.e. the model fit). The log-likelihood function of the estimated model is given in Equation 8-10 (Hensher et al., 2015a):

$$LL(\beta \mid x, Y) = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{j=1}^{J} Y_{nsj} \ln P_{nsj}(x \mid \beta) \qquad \text{8-10)}$$

where:

$\beta$ is the parameter of attribute $x$ estimated by the model,

$N$ is the number of respondents,

$S$ is the set of choice scenarios,

$Y_{nsj}$ is 1 if alternative $j$ is chosen, 0 otherwise,

$P_{nsj}$ is the estimated probability of alternative $j$ being chosen by individual $n$ from all the alternatives in choice task $s$.

The model uses optimization algorithm methods, such as Newton Raphson or Broyden Fletcher Goldfarb Shanno (BFGS), to find the best values for the $\beta$ parameters, i.e. the values that maximise the Log-likelihood function. Log-likelihood values cannot be used alone to evaluate the fitness of a model as they are a function of the sample size. Therefore, this chapter also uses Chi-square tests and McFadden pseudo rho squared for comparing model fit statistics.

#### 8.3.6.2 Chi-square test

The Chi-square test is a likelihood-ratio test that compares the estimated model's log- likelihood with its related base model's log-likelihood at convergence. If the calculated likelihood ratio is larger than the Chi-square table value, it suggests a rejection of the null hypothesis ($H_0$: restricted base model is statistically equivalent to the unrestricted estimating model). In other words, it indicates whether the estimated model statistically performs better than the base model. However, for discrete choice analysis, there are two kinds of base models that can be used for the test. The first comparison base model ignores all the information from the survey data and, therefore, is also called the null model. In the null model, all the alternatives have the same market share. The second comparison base model refers to the constant only model. Thus, the alternatives have the actual market share based on the real market survey data (e.g., RP survey data). In this chapter, it is considered to be more appropriate to use the null model than the base model for the Chi-square test, because the collected survey data are SP data, not the real market data, and therefore cannot be used to represent real market shares.

#### 8.3.6.3 McFadden pseudo rho squared

McFadden's pseudo rho squared, proposed by McFadden (1973), is calculated from the log-likelihood function of a base model and an estimated model. A null model is also used instead of a constant only model in calculating pseudo rho squared. The formula is given by Equation 8-11:

$$\rho^2 = 1 - \frac{LL(M_{\text{estimated model}})}{LL(M_{\text{null model}})} \qquad \text{8-11)}$$

However, for most sample data, the maximised log-likelihood is not realistically bounded at zero, because the data may have omitted attributes and idiosyncratic errors may arise in the data set that may contribute to the unobserved utility or the error term. As a result, the pseudo rho squared is not likely to be bounded at one (Hensher et al., 2015a). Hence, a small pseudo rho squared may only mean a lot of noise in the data set but, for the same data set, the larger the pseudo value, the better the model fit.

## 8. 4 Results

### 8.4.1 Modelling results of air passenger respondents

*NLOGIT 5* software was used to estimate the model parameters for the air passenger respondents, (represented by the airport respondent SP data). The parameters and the corresponding *t*-values (in parentheses) for both the MNL and NL models are shown in Table 8-3. For the models based on the business group, the log-likelihood function and rho-square ($\rho^2$) are both larger for the NL model than for the MNL model, indicating that the NL model has the better fit. In line with this, the IV value of the public air nest for the business group, (0.279, *t*-value: 3.77), is significant and within the 0 to 1 range at the 99% confidence level, which indicates that the NL model is an improvement on the MNL model. Similarly, for the model based on non-business trips, the log-likelihood function and rho-square of the NL model are also larger than those of the MNL model, and the IV value of the public air nest (0.844, t-value: 6.52) is significant and within the 0 to 1 range at the 99% confidence level. Thus, this also indicates that the NL model performed better than the MNL model.

The likelihood-ratio test results show that the chi-squared values of the business trip and non-business trip NL models, (5380.95 and 2218.09 respectively), are both larger than the critical value (27.88) at a significance level of 0.001. The two NL models both reject the hypothesis ($H_0$: estimated model and the base model are equivalent), which indicates that the two NL models better fit the SP data than the base models.

For both the NL models, travel cost, journey time, service frequency and seat comfort (dummy variables) were found to be statistically significant (see Table 8-3). The signs of the estimated parameters show that increasing the travel cost or journey time of a particular mode can reduce the probability of that mode being chosen, by both business and non-business air travellers. Conversely, increasing the service frequency or improving the seat comfort level of the mode could attract more travellers to that mode. However, for the NL model based on non-business trips, access time was found to be insignificant.

*Table 8-3 Modelling results for air passenger respondents*

| Parameter | Multinomial logit model | | Nested logit model | |
|---|---|---|---|---|
| | Business Group | Non-business Group | Business Group | Non-business Group |
| Observation | 3529 | 1823 | 3529 | 1823 |
| Constant (bus) | -1.3562** (-9.12) | -1.5770** (-8.81) | -1.3088** (-8.72) | -1.5569** (-8.68) |
| Constant (airline 1 and airline 2) | 1.1335** (9.59) | .92508** (6.41) | 1.13862** (13.07) | 0.9395** (6.73) |
| **Variables** | | | | |
| Travel cost (A$) | -0.0049** (-13.71) | -0.0077** (-16.49) | -0.0014** (-4.31) | -0.0069** (-8.06) |
| Access time (min) | -0.0010 (-0.53) | 0.0025 (1.09) | -0.0027** (-2.90) | 0.0015 (-0.75) |
| Journey time (hour) | -0.0810** (-12.63) | -0.0652** (-9.40) | -0.0818** (-14.33) | -0.0642** (-9.78) |
| Service frequency (weekly) | 0.0131** (8.18) | 0.0089** (4.07) | 0.0057** (3.66) | 0.0086** (3.80) |
| Seat comfort_Middle (true=1, otherwise=0) | 0.5580** (9.87) | 0.2118** (2.74) | 0.1709** (3.74) | 0.2077** (2.88) |
| Seat comfort_High (true=1, otherwise=0) | 0.8005** (13.78) | 0.4428** (5.67) | 0.2829** (3.66) | 0.3974** (4.66) |
| **Model fit statistics** | | | | |
| LL($\beta$): Log likelihood function | -3348.18 | -1897.34 | -3340.83 | -1896.71 |
| LL(base): Log likelihood of base model | -4636.63 | -2498.54 | -6031.30 | -3005.75 |
| Chi-squared | 2576.90** | 1202.40** | 5380.95** | 2218.09** |
| IV parameter of Public air nest | - | - | 0.2790** (3.77) | 0.8436** (6.52) |
| McFadden rho-squared | 0.2779 | 0.2406 | 0.4461 | 0.3690 |

*Significant at the 5% level.
**Significant at the 0.1% level.

### 8.4.2 Modelling results of non-air passenger respondents

Table 8-4 provides the MNL and NL model estimates for the non-air passenger respondents (non-airport respondent SP data). For the business trip group, the log-likelihood function (-1762.49) and $\rho^2$ (0.288) of the NL model are larger than the log-likelihood function (-1763.43) and $\rho^2$ (0.180) of the MNL model. Also, the IV value of the public air nest (0.759, t-value: 4.87) in the NL model is significant and between zero and one at the 99% confidence level. This indicates that the NL model performs better than the MNL model. For the non-business trip group, the results also indicate that the NL model performs better than the MNL model. As a result, this section focuses on interpreting the NL model parameter estimations.

From both NL models, the majority of the parameters estimated by the business group and non-business group NL models are significant at the 0.01 significance level, except access time which is not statistically significant. For both the business and non-business groups, the modelling results show that travel cost, journey time, service frequency and seat comfort are statistically significant. The sign of parameters in this model are the same as the ones in the model in section 8.4.1, only the magnitudes of these parameters are different. Therefore, the next few paragraphs investigate further these magnitude differences by comparing the estimated elasticities.

*Table 8-4 Modelling results for non-air passenger respondents*

| **Parameter** | **Multinomial logit model** | | **Nested logit model** | |
|---|---|---|---|---|
| | Business Group | Non-business Group | Business Group | Non-business Group |
| Observation | 1552 | 3511 | 1552 | 3511 |
| Constant (bus) | -1.4947** | -1.3459** | -1.4914** | -1.3478** |
| | (-8.73) | (-13.14) | (-9.01) | (-15.52) |
| Constant (airline 1 and airline 2) | 0.2320 | 0.0416 | 0.2775* | 0.0371 |
| | (1.58) | (0.43) | (1.98) | (0.43) |
| **Variables** | | | | |
| Travel cost (A$) | -0.0059** | -0.0080** | -0.0050** | -0.0049** |
| | (-12.18) | (-23.39) | (-6.16) | (-10.14) |
| Access time (min) | 0.0013 | 0.0007 | 0.0006 | -0.0008 |
| | (0.54) | (0.47) | (0.31) | (-0.76) |
| Journey time (hour) | -0.0667** | -0.0631** | -0.0656** | -0.0602** |
| | (-9.72) | (-14.64) | (-9.76) | (-14.21) |
| Service frequency (weekly) | 0.0057** | 0.0040** | 0.0055** | 0.0040** |
| | (2.46) | (2.52) | (2.57) | (3.85) |
| Seat comfort_Middle (true=1, otherwise=0) | 0.1075<br>(1.32) | -0.0090<br>(-0.16) | 0.1123<br>(1.62) | 0.0687*<br>(2.05) |
| Seat comfort_High (true=1, otherwise=0) | 0.2344**<br>(2.91) | 0.3231**<br>(5.84) | 0.1983 **<br>(2.56) | 0.2141**<br>(5.23) |
| **Model fit statistics** | | | | |
| LL($\beta$): Log likelihood function | -1763.43 | -4040.17 | -1762.49 | -4005.21 |
| LL(base): Log likelihood of base model | -2151.14 | -4795.02 | -2475.83 | -5298.97 |
| Chi-squared | 775.42 | 1509.70 | 1426.68** | 2587.53** |
| IV parameter of Public air nest | - | - | 0.7594** | 0.4348** |
| | | | (4.87) | (8.55) |
| McFadden rho-squared | 0.1802 | 0.1574 | 0.2881 | 0.2442 |

*Significant at the 5% level.
**Significant at the 0.1% level.

### 8.4.3 Elasticities

Table 8-5 presents the direct elasticities of the NL model built using the air passenger respondent data. The larger absolute values of the elasticities, (i.e., those greater than 1), imply that a one percent change in the explanatory variable of a given mode alternative may result in a greater than one percent change in its choice probability. For travel cost, the absolute values of the elasticities for the non-business group are higher than those for the business group. This means that the respondents who took non-business related trips were more price elastic than the ones who took business related trips. For example, an increase in travel cost by car of one percent is estimated to result in a reduction of 0.284 percent in the number of business travellers that choose car for transportation, with the corresponding figure for the non-business group being 1.260 percent. For the business group, respondents were more sensitive to journey time than travel cost and frequency, while for the non-business group, respondents were more sensitive to travel cost than journey time and service frequency. For both groups, in terms of travel cost, respondents were more sensitive in choosing air travel than the other travel modes. The travel mode of cars had the lowest direct elasticity. However, in terms of journey time, for both groups, respondents are more sensitive to the bus mode.

In contrast, when the same model was run for the non-air passengers (non-airport respondents), the elasticities for the air travel mode were greater (in absolute values) than the ones in the model for air passengers (airport respondents), as show in Table 8-6. This may mean that the non-air/community respondents were more sensitive to travel cost in terms of air travel mode choice than the air passenger respondents. For travel cost for the business group, the absolute values of the elasticities of community respondents were higher than those of air passenger respondents. This may mean that the community respondents were more sensitive to travel cost in terms of both the bus and car travel modes than the air passenger respondents. For the journey time of bus and car travel mode, community respondents for both the business and non-business groups were less sensitive than air passenger respondents.

*Table 8-5 Direct elasticities from NL model for air passenger respondents*

| | **Business group** | | | **Non-Business group** | | |
|---|---|---|---|---|---|---|
| **Travel Mode** | Travel cost | Journey time | Service frequency | Travel cost | Journey time | Service frequency |
| Car | -0.284 | -1.425 | - | -1.260 | -1.034 | - |
| Bus | -0.322 | -1.811 | 0.109 | -1.525 | -1.407 | 0.162 |
| Airline* | -1.012 | -0.425 | 0.254 | -1.842 | -0.167 | 0.149 |

*Table notes:*

* *For the airline direct elasticities , elasticity figures across the two airlines have been averaged for more intuitive interpretation*

*Table 8-6 Direct elasticities from NL model for non-air passenger respondents*

| | Business group | | | Non-Business group | | |
|---|---|---|---|---|---|---|
| **Travel Mode** | Travel cost | Journey time | Service frequency | Travel cost | Journey time | Service frequency |
| Car | -0.851 | -0.990 | - | -0.743 | -0.801 | - |
| Bus | -1.089 | -1.422 | 0.102 | -1.056 | -1.279 | 0.073 |
| Airline | -1.512 | -0.143 | 0.109 | -2.594 | -0.225 | 0.139 |

Table 8-7 presents the cross elasticities with respect to the air passenger respondents. It indicates that a one percent increase in travel cost by car could increase the probability of business passengers choosing bus or airline by 0.021 percent. However, a one percent increase in airfare (averaged by the two airlines) could encourage 0.238 percent of business air passengers and 0.991 percent of non-business air passengers to change to car or bus. Additionally, an increase of one percent in car journey time could encourage 0.167 percent of non-business car travellers to change to bus or airline.

*Table 8-7 Cross elasticities for air passenger respondents*

| | | Business group | | | Non-Business group | | |
|---|---|---|---|---|---|---|---|
| **Travel Mode** | | Travel cost | Journey time | Service frequency | Travel cost | Journey time | Service frequency |
| Car | Bus | 0.021 | 0.104 | - | 0.202 | 0.167 | - |
| | Airline 1 | 0.021 | 0.104 | - | 0.202 | 0.167 | - |
| | Airline 2 | 0.021 | 0.104 | - | 0.202 | 0.167 | - |
| Bus | Car | 0.005 | 0.024 | -0.002 | 0.037 | 0.035 | -0.006 |
| | Airline 1 | 0.005 | 0.024 | -0.002 | 0.037 | 0.035 | -0.006 |
| | Airline 2 | 0.005 | 0.024 | -0.002 | 0.037 | 0.035 | -0.006 |
| Airline 1 | Car | 0.218 | 0.095 | -0.066 | 0.960 | 0.074 | -0.095 |
| | Bus | 0.218 | 0.095 | -0.066 | 0.960 | 0.074 | -0.095 |
| | Airline2 | 0.846 | 0.367 | -0.259 | 1.181 | 0.090 | -0.117 |
| Airline 2 | Car | 0.258 | 0.115 | -0.067 | 1.021 | 0.079 | -0.080 |
| | Bus | 0.258 | 0.115 | -0.067 | 1.021 | 0.079 | -0.080 |
| | Airline1 | 0.995 | 0.441 | -0.263 | 1.255 | 0.096 | -0.098 |

Table 8-8 presents the cross elasticities for the non-airport/community respondents. A one percent increase in airfare (averaged over the two airlines) could increase the number of business passengers using car or bus by 0.661 percent. A one percent increase in car journey time could encourage 0.237 percent of business car travellers to select bus or airline. For the non-business group, an increase of one percent in airfare (averaged over the two airlines) could increase car or bus use by 0.528 percent, while adding one percent to car travel time could result in a 0.329 percent increase in the number of passengers selecting bus or airline.

*Table 8-8 Cross elasticities for non-air passenger respondents*

| Travel Mode | | Business group: Travel cost | Business group: Journey time | Business group: Service frequency | Non-Business group: Travel cost | Non-Business group: Journey time | Non-Business group: Service frequency |
|---|---|---|---|---|---|---|---|
| Car | Bus | 0.203 | 0.237 | - | 0.306 | 0.329 | - |
| | Airline 1 | 0.203 | 0.237 | - | 0.306 | 0.329 | - |
| | Airline 2 | 0.203 | 0.237 | - | 0.306 | 0.329 | - |
| Bus | Car | 0.039 | 0.051 | -0.005 | 0.062 | 0.077 | -0.006 |
| | Airline 1 | 0.039 | 0.051 | -0.005 | 0.062 | 0.077 | -0.006 |
| | Airline 2 | 0.039 | 0.051 | -0.005 | 0.062 | 0.077 | -0.006 |
| Airline 1 | Car | 0.648 | 0.070 | -0.055 | 0.513 | 0.054 | -0.033 |
| | Bus | 0.648 | 0.070 | -0.055 | 0.513 | 0.054 | -0.033 |
| | Airline2 | 0.727 | 0.099 | -0.078 | 1.607 | 0.164 | -0.104 |
| Airline 2 | Car | 0.673 | 0.073 | -0.046 | 0.543 | 0.057 | -0.026 |
| | Bus | 0.673 | 0.073 | -0.046 | 0.543 | 0.057 | -0.026 |
| | Airline1 | 0.961 | 0.103 | -0.066 | 1.688 | 0.172 | -0.084 |

For the NL models, two other possible nesting structures were also tested, namely:

A. private travel mode (car) and public travel mode (bus, airlines), and

B. air travel mode (airlines) and non-air travel mode (car, bus).

The performance of both nesting structures was relatively poor and the IV parameters were statistically insignificant and/or larger than 1, (they should be less than 1). The alternative nesting structures were therefore rejected.

## 8. 5 Discussion

### 8.5.1 Factor analysis

There are areas of agreement and areas of disagreement within the previous literature covering the factors that affect people's travel mode choice behaviour. This may be due to differences in the settings for the studies, including study area/location and travel category, (e.g., overseas travel, domestic travel or urban travel)). For example, Chang and Sun (2012) found that travel cost is a significant factor that affects both business and non-business traveller mode choice, and was more important than service frequency. Hess and Polak (2006a) found similar evidence on cost but they also found that service frequency has a significant influence on attracting travellers to use air travel. In this thesis, the NL modelling results indicate that increasing travel cost or journey time statistically significantly reduced sales for both business and non-business travellers. This result is consistent with the majority of the previous literature (Jovicic and Hansen, 2003; Hess et al., 2007; Wang et al., 2014; Qiao et al., 2016; Yang et al., 2018).

Additionally, the conclusion that improved seat comfort level would increase the number of passengers is consistent with the findings by Van Can (2013). In line with Hess and Polak (2006a), the results also show that service frequency has a statistically significant positive impact in attracting more passengers.

Interestingly, airport and or bus station access time was found to be a statistically insignificant factor for all the four traveller groups, with the exception of the air passenger business group, which may be because airports or bus stations are usually close to travellers' origins in regional Western Australia and they do not consider the access time to airports or bus stations important in making their mode choice for regional trips. Hence, this research recomputed the NL models for the 4 groups with access time removed. Table 8-9 below presents the AIC indices of the NL models with and without access time. On one hand, the NL model of the air passenger business group with access time is found to have an obviously lower AIC values compare to that without access time, the difference being around 11.5. The models of the remaining three groups with access time have slightly higher AIC values but, the differences are quite small, which suggests that the differences in model performance with or without access time are negligible. On the other hand, the t-value of access time in the NL model of the air passenger group is significant, although it is insignificant in the models for the remaining three groups. This finding indicates that different traveller groups may have different preferences relative to access time. Thus, it appears that researchers should be cautious of this attribute, as it can be either significant or insignificant.

*Table 8-9 Information criteria of nested logit models*

| NL models | Log likelihood function (LL($\beta$)) | AIC | Significance of access time |
|---|---|---|---|
| Air passenger business group (includes access time) | -3340.83 | 6699.66 | 0.0027** (t-value=-2.90) |
| Air passenger business group (excludes access time) | -3346.02 | 6708.04 | - |
| Air passenger non-business group (includes access time) | -1896.71 | 3811.42 | 0.0015 (t-value=-0.75) |
| Air passenger non-business group (excludes access time) | -1896.94 | 3809.87 | - |
| Non-air passenger business group (includes access time) | -1762.49 | 3542.97 | 0.0006 (t-value=0.31) |
| Non-air passenger business group (excludes access time) | -1762.53 | 3541.07 | - |
| Non-air passenger non-business group (includes access time) | -4005.21 | 8028.42 | -0.0008 (t-value=-0.76) |
| Non-air passenger non-business group (excludes access time) | -4005.51 | 8027.02 | - |

*Significant at the 5% level
**Significant at the 0.1% level

### 8.5.2 Elasticity analysis

For the NL model using the air passenger respondent data, the business group respondents were found to be more sensitive to journey time than travel cost and frequency. While for the non-business group, respondents were more sensitive to travel cost than journey time and service frequency, which is consistent with the findings in the literature (De Vany, 1974; Jung and Yoo, 2014; Inoue et al., 2015). Additionally, for both the business and non-business groups, respondents were more sensitive to travel cost when considering air travel compared to the other travel modes. This means that reducing airfares could potentially attract a higher percent of air passengers than decreasing travel cost by car or bus at the same rate. The results also shown that the car travel mode had the lowest direct elasticity, meaning that these travellers would be the least likely to change mode. In regional Western Australia, cars are still the one of the major transport modes and are heavily relied upon by the majority of people.

This study also compared the mode choice elasticity between air passenger respondents and non-air/community passenger respondents. Community respondents were found to be more sensitive to travel cost in terms of air travel mode choice than the air passenger respondents. This means that, in order to attract potential air travel mode users (community respondents), one effective way may be to reduce airfares. For example, for the non-business group, the direct elasticity indicates that a one percent reduction in airfare could lead to a 2.6 percent increase in air travel for the community respondent group, and about 1.8 percent growth for the air passenger respondent group.  When travelling on business, community respondents were more sensitive than the air passenger respondents to the travel cost of bus and car. With the vast distances between towns in regional Western Australia, most business trips were by air in order to travel efficiently, especially in terms of time. Interestingly, potential users had lower elasticities in choosing cars and buses than air passenger respondents in terms of journey time and service frequency. This could mean that influencing the travel mode preferences of road users by changes in journey time and or service frequency may be relatively more difficult.

### 8.5.3 Willingness to pay

Willingness to pay (WTP), also called ‘implicit prices’ (Hensher et al., 2015a), is a frequently reported output from discrete choice modelling, that is calculated as the ratio between marginal utility change for a particular attribute and the marginal utility change in the cost attribute. In this chapter, the marginal rates of substitution between travel cost and journey time (hours), service frequency (weekly) and seat comfort (dummy variable) were generated based on the NL modelling results for both air and non-air passenger respondents. As shown in Table 8-10, for the air passenger respondents, business travellers were willing to pay $56.80 to reduce the journey time by one hour, which was six times more than non-business travellers ($9.30). The business travellers were prepared to pay $3.90 for the weekly service frequency to increase by one, while the non-business travellers were only willing to pay $1.20. They would also be prepared to pay $118.70 to increase the seat comfort level from low to middle and $196.40 to increase it from low to high. However, non-business travellers were only prepared to spend $30.10 and $57.60, respectively. For the non-air passenger respondents, the result suggests that business travellers were willing to pay slightly more than non-business travellers. For example, the business non-air passengers were willing to pay $13.20 to reduce journey time by one hour, and the non-business passengers were willing to pay $12.20. Overall, the findings suggest that business travellers are willing to pay more money to reduce the time and improve the service qualities compared to the non-business travellers, which is consistent with the previous literature (Jovicic and Hansen, 2003; Jung and Yoo, 2014).

However, business air passengers were also willing to pay much more money than business non-air passengers to reduce time or improve the service quality attributes (e.g., service frequency and seat comfort). The results also found that the willingness to pay in terms of most factors for air passenger respondents were lower than the ones for non-air passenger respondents, except journey time.

*Table 8-10 WTP of air and non-air passenger respondents*

| | **Air passenger respondents** | | **Non-air passenger respondents** | |
|---|---|---|---|---|
| | Business group | Non-business group | Business group | Non-business group |
| Journey time ($/per hour) | $56.80 | $9.30 | $13.20 | $12.20 |
| Service frequency ($/per flight) | $3.90 | $1.20 | $1.10 | $0.80 |
| Seat comfort_Mid ($/increase to middle level) | $118.70 | $30.10 | $22.60 | $13.90 |
| Seat comfort_High ($/increase to high level) | $196.40 | $57.60 | $39.90 | $43.40 |

## 8. 6 Summary

The primary aim of this chapter was to investigate the travel mode choice behaviour of regional travellers for car, bus and regional airlines. Factors influencing this choice were assumed to include travel cost, travel time and service quality. One interest of this thesis was to find out whether or not there were any mode choice differences between air passenger respondents and non-air passenger respondents. The investigated mode choice behaviours of business and non-business travellers, air and non-air passenger respondents has shed some light that may be useful to the local government and regional airlines in Western Australia and may also be generalised to other regional cases. The examined mode choice results of elasticities and willingness to pay indicate that, the regional airlines can try to reduce airfare and/or offer extra incentives for leisure travellers, which could be a relatively more effective strategy to attract travellers, especially for non-business travellers. The air transport department can suggest and or regulate the regional airlines to reasonably increase the flight speed and seat comfort, which therefore can produce a relatively better outcome for the aviation industry with a considerable increase in the air passenger demand, especially for attracting the business travellers. In this context, on one hand, regional airlines can increase the fly speed as more as they can, since the willingness to pay results suggest that the airport passengers would like to pay AU$56.8 premium to reduce 1 hour travel time for business travel related ticket, and AU$9.3 for the non-business travel related ticket. On the other hand, the regional airlines could improve the seat comfort level for the business travel related ticket, as the airport business travellers are prepared to pay AU$196.4 premium to improve the seat comfort level from low to high. Additionally, the local government could consider upgrading the road construction to the airports for those mining or industry towns, which could attract some proportion of business travellers to use air transport. With such guidance, the government and airlines could understand the competition between not only regional airlines but also air and non-air travel modes and, thus, more effective policies and strategies could be developed to improve the regional travel services and airline patronage.

The next chapter extends the investigation of travelling behaviours of regional travellers, using an LC modelling approach that is valuable in accommodating preference heterogeneity across regional travellers. The methodology framework and findings will be illustrated.

# CHAPTER 9 ANALYSING TRAVEL MODE AND AIRLINE CHOICE USING LATENT CLASS MODELLING

## 9. 1 Introduction

The last chapter applied MNL and NL models to explore regional air and non-air traveller travel mode and airline choice. This chapter continues the travel mode and airline choice modelling analysis by accommodating travellers' preference heterogeneity in mode choice, using an LC modelling with market segmentation technique. It investigates the choice behaviour within, and among, traveller market segments using SP data. The market segments are the latent classes identified through the LC modelling procedure. Therefore, in conjunction with the modelling analysis of the previous chapter, a more comprehensive and reliable understanding of the travel behaviour of regional travellers can be achieved.

This chapter is based on the paper[10] published in Journal of *Transportation Research Part A: Policy and Practice* (Zhou et al., 2020). The chapter is structured as follows: section 9.2 provides the research context that mainly relates to the motivation for applying LC modelling; section 9.3 fully describes the LC framework and methodology and how unobserved preference heterogeneity is accommodated by using a market segmentation technique; section 9.4 subsequently applies the LC modelling approach to the case study in regional Western Australia and reports the estimation results of the travel mode and airline choice; section 9.5 discusses the LC modelling results that could assist in providing a more comprehensive understanding of the behaviour of regional travellers.

## 9. 2 Research Context

As found in the previous chapter, a traveller's decision about travel mode and/or airline is influenced by several key factors including cost, travel time, service frequency, accessibility and seat comfort. However, travellers with different demographic and trip related characteristics may have heterogeneous preferences or sensitivities that could also affect their decision-making process. For example, individuals with high incomes may be relatively less

[10] Zhou, H., Norman, R., Xia, J., Hughes, B., Kelobonye, K., Nikolova, G., & Falkmer, T. (2020). Analysing travel mode and airline choice using latent class modelling: A case study in Western Australia. Transportation Research Part A: Policy and Practice, 137, 187-205. doi:https://doi.org/10.1016/j.tra.2020.04.020

sensitive to travel cost. Equally, age may impact on a consumer's sensitivity to seat comfort. Therefore, for this competitive passenger market involving significant distances and cost, accommodating passengers' preference heterogeneity by identifying different traveller segments is particularly important when estimating travel mode and airline choice. This may help the government policy makers and airlines to tailor more targeted strategies (e.g., offering extra incentives to the group of travellers who were more price sensitive) to attract the passengers, based on the characteristics and mode choice behaviours on each of the market segments. For example, with the latent group/segment of the travellers identified and their characteristics and travel mode choice estimated, the regional airlines can try to offer extra incentives to the group of travellers who were relatively price-sensitive. In contrast, if the distinct group of price-insensitive travellers were also discovered, the regional aviation industry could improve other factors which most concerned them, such as reduce travel time and or improve service quality factors. As mentioned in the last chapter, the NL modelling results suggested that the regional airline could travel as quickly as possible, since the airport business passengers would like to pay AU$56.8 premium as the exchange to reduce 1 hour travel time for business travel related ticket, and willing to pay AU$196.4 to improve the seat comfort level from low to high.

As mentioned in the previous chapter, even though the MNL model has been widely used in previous studies its limitations are well-acknowledged; it assumes homogeneity of preferences across respondents, as well as independence of irrelevant alternatives. Therefore, the parameter estimates may lose some precision if preference heterogeneity does exist among the individuals. From this perspective, LC modelling has been found to be an appropriate approach to capture the potential preference heterogeneity, especially when there is uncertainty about the distribution of the preference heterogeneity across individuals and an intuitive interpretation for the policymakers and investigators is sought (Greene and Hensher, 2003).

Traditionally, market segmentation is conducted to identify the distinct segments or classes within a customer market, where customers with similar demographics or characteristics are assigned to the same segment. In this chapter, LC modelling mainly uses the stated choice observations of travellers to identify these segments, which is different to the traditional market segmentation approach applied in Chapter 7. Travellers' preferences are assumed to be homogenous in the same segment, but heterogeneous across different segments.

This chapter aims to estimate and compare traveller mode and airline choice behaviour for regional travel within, and among, the different latent segments. An LC model framework is used to accommodate travellers' unobserved preference heterogeneity and to identify membership of each latent segment or class. To achieve this goal, it firstly compares the model fit statistics between the MNL and LC models to check and confirm the appropriateness of LC modelling in estimating the travel mode choice of travellers within regional Western Australia. Secondly, it compares the differences in mode choice preference among the different traveller segments with respect to the transport service attributes and by estimating segment-specified direct elasticities and willingness to pay for different travel modes. Finally, it compares the respondent profiles for each of the latent traveller segments, in order to better understand possible relations between individual characteristics, (demographics, economics and trip characteristics), and their preference heterogeneity. The chapter captures the unobserved preference heterogeneity across regional travellers through the identification of latent traveller segments. Therefore, it can provide insights for policymakers and air carriers seeking to establish more nuanced policies and strategies to encourage public transport and airline use based on the characteristics and mode choice behaviours of each of the market segments.

## 9. 3 Methodology

### 9.3.1 Data used in this study

This chapter uses the same SP data as the previous chapter (Chapter 8), including the air and non-air passenger respondents. However, the LC model can identify the latent segments of respondents, i.e., where respondents within the same segment have a homogenous preference but with heterogeneity across the segments. Thus, there is no need to separately analyse the air and non-air passenger respondent data and the sample size is significantly larger than the theoretical minimum requirement. For details of the survey data collection methods, sources and time duration please refer to section 3.4.2.

### 9.3.2 Latent class model

The LC model for accommodating discrete preference heterogeneity was initially proposed by Lazarsfeld and Henry (1968) and was developed further for application to discrete choice analysis by Kamakura and Russell (1989) and then by Greene and Hensher (2003). As indicated by Greene and Hensher (2003), the model postulates that an individual's choice behaviour is determined by not only choice related attributes but also latent heterogeneity due to variations in individual-specified characteristics, (e.g., demographics), that are unobserved by the investigator. Specifically, the model allocates the respondents into a discrete number (*Q)* of latent segments/classes, while a parameter vector estimate for each segment is estimated. Thus, the unobserved preference heterogeneity can be captured and accommodated. The posterior probability that each respondent belongs to any of the segments is determined by the respondent's choice observations and characteristics.

In this paper, following the formulation of Greene and Hensher (2003), the initial assumption is that the central behaviour model has a basic MNL specification. Hence the conditional choice probability that individual $i$ chooses alternative $j$ among the set of all available alternatives $J$ in choice task $t$ within class $c$ is given as Equation 9-1:

$$P_{ijt|c} = P(y_{it} = j \mid c) = \frac{\exp(\beta_c x_{ijt})}{\sum_{j=1}^{J} \exp(\beta_c x_{ijt})} \qquad c = 1, 2, ..., C \tag*{9-1)}$$

where $x_{ijt}$ is the attribute vector of alternative $j$ in choice task $t$ ($t \in$ the set of choice tasks $T_i$), $\beta_c$ is the parameter vector of attribute vector $x_{ijt}$ conditional on latent class/segment $c$. However, the difference between any pair of classes is a kind of unobserved preference heterogeneity, and the $\beta_c$ can be written as Equation 9-2:

$$\beta_c = \beta + \delta_c \tag*{9-2)}$$

where $\delta_c$ is the unobserved heterogeneity following a discrete distribution. Furthermore, panel data including a set of $T_i$ choice tasks were assumed to be observed by (completed by) the same individual, and the $T_i$ choice tasks are independent (Greene and Hensher, 2003). Thus, an individual $i$'s contribution to the likelihood with respect to class $c$, is represented by the conditional joint probability of the individual $i$ who makes choices for the $T_i$ choice tasks in sequence $y_i = [y_{i1}, y_{i2}, ..., y_{iT_i}]$ conditional on class $c$ (given as Equation 9-3 below).

$$P_{i|c} = \prod_{t=1}^{T_i} P_{ijt|c} \tag*{9-3)}$$

The class assignment membership is measured by the prior membership probability of individual $i$ belonging to class $c$, as calculated by Equation 9-4 (Greene and Hensher, 2003):

$$H_{ic} = \frac{\exp(\theta_c z_i)}{\sum_{c=1}^{C} \exp(\theta_c z_i)} \tag*{9-4)}$$

where $z_i$ is a vector variable of the observed characteristics, such as demographics, that are regarded as the observed heterogeneity by the analyst and thus are built into the class membership probability function. For example, the analyst may consider individual's age distribution, (e.g., youth, middle age and old age), as the observed heterogeneity when investigating people's preference of entertainment mode. $\theta_c$ is an unknown vector of a parameter that is estimated by the model, while the $C^{th}$ parameter vector is normalised to 0 in order to allow model identification. If it is impossible to find the significant vector variable $z_i$, the only remaining element in $\theta_c z_i$ would be the constant term. In such cases, the prior membership probability of each class over individuals would be a constant. Therefore, the total or unconditional prior probability ($P_i$) that individual $i$ who makes choices for the $T_i$ choice tasks in sequence $y_i = [y_{i1}, y_{i2}, ..., y_{iT_i}]$ over all classes can be calculated per Equation 9-5.

$$P_i = \sum_{c=1}^{C} P_{i|c} H_{ic} \qquad 9\text{-}5)$$

The model parameters are estimated by a finite iteration of the expectation and maximization routines, (until convergence occurs), for weighted log likelihood using the unconditional prior probability $P_i$. Thus, the class membership prior probability $\widehat{H_{\mathbf{i}c}}$ and the conditional choice probability $\widehat{P_{\mathbf{i}|c}}$ are estimated. The individual $i$'s estimate of the latent class membership posterior probability can be then computed using Bayes theorem, Equation 9-6:

$$\widehat{H_{c|\mathbf{i}}} = \frac{\widehat{P_{\mathbf{i}|c}}\widehat{H_{\mathbf{i}c}}}{\sum_{c=1}^{C}\widehat{P_{\mathbf{i}|c}}\widehat{H_{\mathbf{i}c}}} \qquad 9\text{-}6)$$

The notation $\widehat{H_{c|\mathbf{i}}}$ represents the individual-specific posterior class probability, conditional on the individual's sequence of choices. $\widehat{H_{c|\mathbf{i}}}$ can be used to evaluate the characteristics of each latent segment, such as the distribution of demographics or trip purposes.

Thus, the formula for calculating the probability distribution of nominal demographic, economic or trip characteristic data, (e.g., gender distribution and income level distribution), related to latent segment $c$, was developed (Equation 9-7).

$$\pi_{cmv} = \frac{\sum_{i=1}^{N}\widehat{H_{c|\mathbf{i}}}y_{imv}}{\sum_{i=1}^{N}\widehat{H_{c|\mathbf{i}}}} \qquad (v \in V_m,\ m \in M,\ c \in C) \qquad 9\text{-}7)$$

where there are $C$ latent segments and $M$ nominal demographic attributes referring to the $N$ individuals; each attribute $m$ (e.g., gender) contains finite possible nominal values $v$ (e.g., female and male); $\pi_{cmv}$ is the probability that the nominal value $v$ (e.g., v=female) from latent segment $c$; $y_{imv}$ equals 1 if individual $i$ takes nominal value $v$ (e.g., v=female) for the $m^{th}$ attribute, or 0 otherwise. Therefore, from the individual-specific posterior class probability $\widehat{H_{\mathbf{c|i}}}$, it is possible to obtain the probability distribution of the demographic and/or trip characteristic attributes within each of the latent segments.

AIC and BIC are the most popular indices used to assist researchers determine the optimal number of segments/classes. The computational formulas of AIC and BIC have been widely used in LC modelling and are shown in Equation 9-8 and 9-9 (Louviere et al., 2000; Ruto et al., 2008; Shen, 2009; Kim et al., 2017),

$$AIC = -2LL + CK_{\beta} + (C-1) \tag{9-8)}$$

$$BIC = -2LL + (CK_{\beta} + (C-1)K) \cdot \ln(N) \tag{9-9)}$$

where *LL* is the log-likelihood value function calculated at convergence for the parameter estimates; *C* is the number of latent classes; $K_{\beta}$ is the number of elements in the utility function of the class-specified model; *K* is the number of estimated parameters in class classification model; and *N* is the sample size referring to the number of respondents for the panel data, which is consistent with the following statement by Louviere et al. (2000, p. 287): ”*In order to be conservative, we used N=209 (i.e., the number of respondents) rather than 209 x 16 (the number of choice observations) as the total number of observations in the analysis*”. The point of this thesis is that, in order to accommodate unobserved preference heterogeneity across individuals, the LC model for panel data incorporates an individual’s joint probability $P_{i|c}$ to likelihood rather than a set of probabilities $P_{ijt|c}$ being input one by one. Thus, using the number of respondents may be more reasonable. The value of *C* that minimizes the measured indices above is preferred. Meanwhile, as described by Walker and Li (2007), the BIC is frequently used to determine the best number of latent segments, as it includes a harsher penalty for the number of parameters than the AIC. Additionally, some researchers suggest calculating AIC and BIC using *-2LL + KF*,; where *F* is the penalty constant, with *F* = 2 for obtaining AIC and *F = ln(N)* for obtaining BIC (Scarpa and Thiene, 2005; Wen and Lai, 2010). The only difference here is the weights of penalty for the number of latent classes. The AIC (Equation 9-8) and BIC (Equation 9-9) functions recommended by Louviere et al. (2000, p. 287) were used in this chapter for the LC modelling, as they impose relatively harsher penalties for the number of latent classes. Estimators for the LC models are provided by *NLOGIT 5.0* (Greene, 2012).

## 9. 4 Results

### 9.4.1 Latent class modelling results

In this study, the LC models relating to different numbers of segments based on the specified utility function of a standard MNL model were estimated. Only segment-specific constants were considered in the latent class membership/classification functions[11]. The AIC, BIC and other measures for the models with between one and four segments are summarised in Table 9-1. Of these, the 4-segment LC model has the lowest AIC value. However, the BIC suggests three segments is a better approach. The modelling results show that both the 3-segment and 4-segment LC models contain a residual segment, (the final segment in each case consisting of approximately 7% of the sample). The parameter estimates in the two residual segments are almost all statistically insignificant. Apart from these residual segments, there are more insignificant parameter estimates in the remaining segments of the 4-segment LC model compared to the 3-segment LC model. In addition to this, there are some counterintuitive results in the 4-segment LC model. Therefore, after considering both information criteria in conjunction with the parameter estimates, the 3-segment LC model was selected for the analysis in this study.

*Table 9-1 Information criteria for number of segments*

| **Criteria** | **Number of segments** | | | |
|---|---|---|---|---|
| | **1 (MNL)** | **2** | **3** | **4** |
| Number of parameters | 8 | 17 | 26 | 35 |
| Log likelihood | -11,465.76 | -10,311.56 | -10,026.14 | -9,851.21 |
| AIC | 22,947.52 | 20,689.13 | 20,204.28 | 19,976.42 |
| BIC | 22,991.11 | 20,868.94 | 20,618.39 | 20,722.92 |
| Segment size | - | Segment 1: 68.6%<br>Segment 2: 31.4% | Segment 1: 33.1%<br>Segment 2: 60.1%<br>Segment 3: 6.8% | Segment 1: 32.2%<br>Segment 2: 22.2%<br>Segment 3: 38.6%<br>Segment 4: 6.9% |

Table 9-2 presents the MNL model and the three-segment LC model in terms of model statistics, (e.g., segment size and McFadden Pseudo R-squared), and the segment-specified parameter estimates, with corresponding t-values shown in parentheses. Residual segment 3 only accounts

[11] Gender, age, income, trip purpose were also included as variables in the latent class membership/classification functions for testing. The results showed that the parameter estimates were statistically insignificant in half or more of the identified latent segments, which indicates that these observed characteristics may only explain individuals' preference heterogeneity to a limited degree. Thus, only segment-specific constants were included in the membership functions.

for 6.8% of the respondents, with all parameters statistically insignificant at the 95% confidence level. This may indicate that these respondents did not understand the task. The remaining segments, 1 and 2, contain 33.1% and 60.1% of the sample population, respectively. For the travellers in segment 1, the signs of the parameter estimates and the corresponding t-values (in parentheses) suggest that an increase in the travel cost and/or journey time of a mode would cause a statistically significant reduction in use of that mode, while increasing its service frequency or improving the seat comfort would attract more travellers to that mode. However, the results also indicate that access time has no statistically significant impact on the mode choice of individuals in segment 1. In the larger segment 2, consisting of 60.1% of the respondents, the parameter estimates, including access time, are all statistically significant, indicating that access time was, however, important in determining choice for the travellers in segment 2. Generally, the signs of the coefficients in segment 2 are the same as those in segment 1, but not the values. This indicates that the travellers in the two segments may have different scales of sensitivity to the mode choice factors. Thus, elasticity and willingness to pay were computed and are discussed in the following sections.

*Table 9-2 Statistical results of multinomial and latent class models*

| **Parameter** | **MNL model** | **Three-segment LC model** | | |
|---|---|---|---|---|
| | | **Segment 1** | **Segment 2** | **Segment 3** |
| Constant (bus) | -1.4389**<br>(-21.29) | -0.6903**<br>(-6.91) | -3.1090**<br>(-20.80) | -7.6784<br>(-1.91) |
| Constant (airline 1 and airline 2) | 0.5403**<br>(9.25) | 0.3812**<br>(-3.31) | 0.2383*<br>(2.16) | 4.0136<br>(1.09) |
| **Variables** | | | | |
| Travel cost (AU$) | -0.0063**<br>(-32.98) | -0.0009**<br>(-22.22) | -0.0069**<br>(-20.80) | -0.0556<br>(-1.89) |
| Access time (min) | 0.0013<br>(1.40) | -.0005<br>(-0.30) | -0.0066**<br>(-3.76) | -0.0139<br>(-0.95) |
| Journey time (hour) | -0.0616**<br>(-22.77) | -0.0537**<br>(-10.20) | -0.3472**<br>(-12.59) | 0.4020<br>(1.44) |
| Service frequency (weekly) | 0.0075**<br>(8.32) | 0.0035*<br>(2.18) | 0.0181**<br>(11.51) | 0.0499<br>(1.45) |
| Seat comfort_Middle (true=1, otherwise=0) | 0.2423**<br>(7.56) | 0.1084<br>(1.83) | 0.4540**<br>(9.02) | -2.5470<br>(-1.01) |
| Seat comfort_High (true=1, otherwise=0) | 0.4845**<br>(15.19) | 0.2985**<br>(5.01) | 0.7615**<br>(14.54) | -0.0923<br>(-0.13) |
| **Model statistics** | | | | |
| Segment size/membership | - | 33.1% | 60.1% | 6.8% |
| Observation | 10,380 | 10,380 | | |
| Number of respondents | 1,718 | 1,718 | | |
| LL($\beta$): Log likelihood at convergence | -11,465.76 | -10,026.14 | | |
| McFadden Pseudo R-squared | 0.1976 | 0.2983 | | |

*Significant at the 5% level.
**Significant at the 1% level.

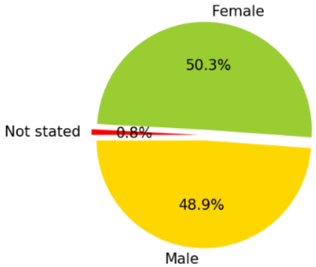
### 9.4.2 Characteristics of each segment

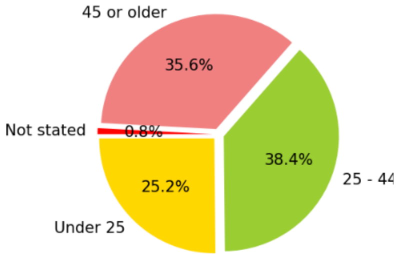
Based on the estimated individual-specific posterior segment membership probabilities from the three-segment LC model, the distributions of the demographic, economic and trip related characteristics of each segment were calculated using Equation 9-7 ,(developed for this study), with the exception of minor residual segment 3. In order to better visualise and compare the differences in the characteristics between segments 1 and 2, a series of pie charts were generated for illustration (Table 9-3). Segments 1 and 2 both contain seven pie charts showing the characteristics of gender, age group, education level, employment status, monthly income, trip purpose and preferences of airline booking time in advance. The results offer evidence that elder, higher income, business travellers were more likely to belong to segment 2 while, younger, lower income, non-business travellers were more likely to belong to segment 1. Segment 2 is comprised of more tertiary educated and employed, with basic educated travellers more likely to belong to segment 1. Travellers making short-term bookings are more likely to belong to segment 2, whereas segment 1 contains more travellers who made advance bookings (one month or more).

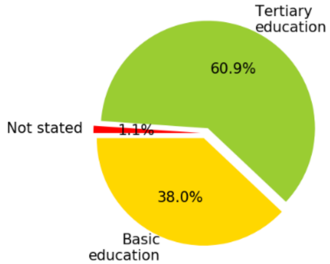
*Table 9-3 Characteristics of each segment*

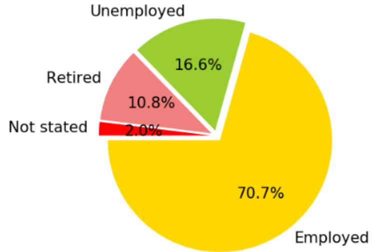
## Gender | Age group | Education level | Employment status

**Segment 1**

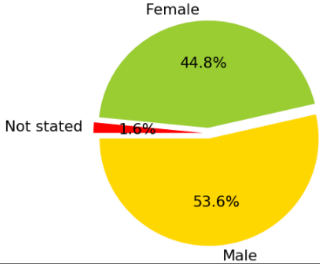


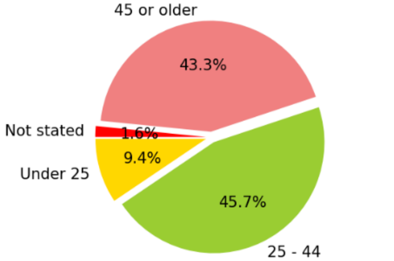



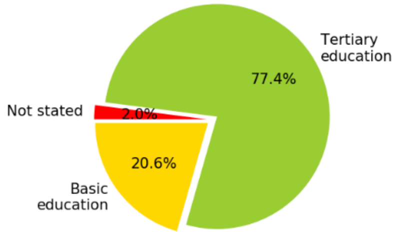



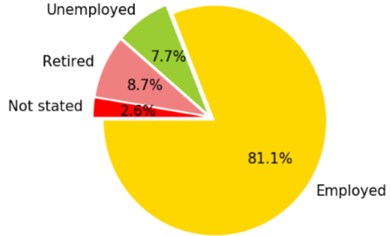




**Segment 2**



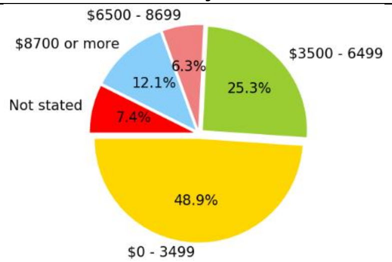



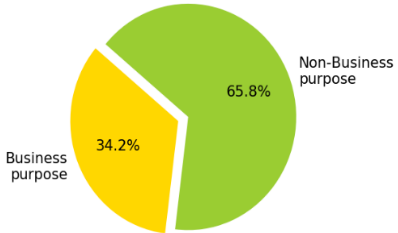



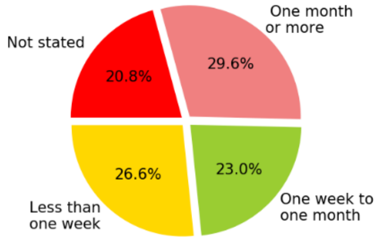




## Monthly income | Trip purpose | Booking time in advance

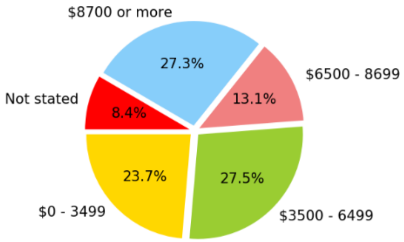
**Segment 1**

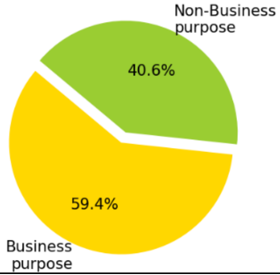


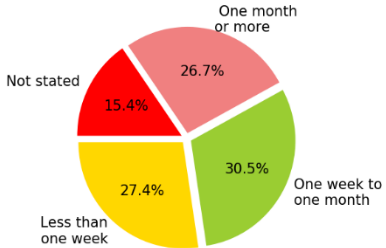




**Segment 2**



### 9.4.3 Elasticities

In order to better understand the differences in preference between the two latent segments, the direct elasticities with respect to the MNL model and each segment of the LC model were calculated and are presented in Table 9-4, with the exception of the insignificant parameters and residual segment 3. The direct elasticities reflect the impact of a unit change in a certain alternative's service attribute on the percentage change in the choice probability for a certain alternative/travel mode, (e.g., car, bus or airline). Notably, the direct elasticities of travel cost for all the travel modes with respect to segments 1 and 2 of the LC model, and the MNL model are smaller than negative 1, which indicates that given a 1% increase in travel cost of any of these travel modes would reduce its choice probability by more than 1%. Additionally, the absolute values of elasticity for bus and airlines in segment 1 are larger than those in segment 2, especially for the airlines. Hence, for segment 1, a 1% reduction in airfare is likely to increase sales by around 2.5%. This finding indicates that travellers from segment 1 are more sensitive to travel cost, (e.g., bus fare and airfare), compared to travellers in segment 2, especially to the airfare. Access time was found to be only significant in segment 2 but the small elasticities indicate that individuals from segment 2 are relatively less sensitive to access time that the other variables. Additionally, the direct elasticities of journey time show that travellers in both segments are much more sensitive to journey time by car and bus than to flight times. This finding is more obvious in segment 2 where a 1% increase in travel time by a car or bus could reduce their choice probabilities by up to 6.4% and 7.8%, respectively, while a 1% increase in travel time by air could reduce air ticket sales by only around 0.5%. In terms of bus and airlines, the direct elasticity results indicate that the two segments are both inelastic to the explanatory variable of service frequency, but the travellers in segment 2 were found to be more sensitive to the service frequency than the travellers in segment 1.

*Table 9-4 Summary of direct elasticities*

| Alternatives | MNL | Three-segment LC model | |
|---|---|---|---|
| | | **Segment 1** | **Segment 2** |
| **Travel cost** | | | |
| Car | -1.112 | -1.351 | -1.434 |
| Bus | -1.391 | -1.832 | -1.554 |
| Airline* | -1.507 | -2.469 | -1.379 |
| **Access time** | | | |
| Bus | - | - | -0.249 |
| Airline | - | - | -0.139 |
| **Journey time** | | | |
| Car | -0.953 | -0.715 | -6.402 |
| Bus | -1.341 | -1.086 | -7.813 |
| Airline | -0.106 | -0.106 | -0.505 |
| **Service frequency** | | | |
| Bus | 0.140 | 0.060 | 0.352 |
| Airline | 0.119 | 0.065 | 0.239 |

*Table notes:*

* *For the airline direct elasticities , elasticity figures across the two airlines were averaged for more intuitive interpretation*

In addition to the direct elasticities, cross elasticities of the latent segments were also computed. As previously mentioned, these reflect the change in choice probability of a particular alternative resulting from a unit change of another alternative's attribute that, thus, can help to explore the competition among the alternatives. As shown in Table 9-5, the cross elasticity results of segments 1 and 2 indicate that an increase in airfare would be more likely to reduce the competitiveness of the airlines, as it would significantly increase the market share of the ground travel modes. For the car alternative, an increase in travel time and cost would increase the competitiveness of bus and air, but to a lesser degree. The low cross elasticities of bus for all the service factors indicate that a small change of service quality of a bus would not considerably influence its market share or competitiveness. One noteworthy point is that, for the core market segment 2, improving the air service quality (e.g., service frequency and journey time) of an airline would increase its competitiveness relatively more than for segment 1.

*Table 9-5 Cross elasticities of the latent segments*

| Travel Mode | | Segment 1 Travel cost | Segment 1 Access time | Segment 1 Journey time | Segment 1 Service frequency | Segment 2 Travel cost | Segment 2 Access time | Segment 2 Journey time | Segment 2 Service frequency |
|---|---|---|---|---|---|---|---|---|---|
| **Car** | Bus | 0.551 | - | 0.293 | - | 0.022 | - | 0.110 | - |
| | Airline1 | 0.551 | - | 0.293 | - | 0.022 | - | 0.110 | - |
| | Airline2 | 0.551 | - | 0.293 | - | 0.022 | - | 0.110 | - |
| **Bus** | Car | 0.199 | - | 0.123 | -0.008 | 0.001 | 0.001 | 0.003 | -0.001 |
| | Airline1 | 0.199 | - | 0.123 | -0.008 | 0.001 | 0.001 | 0.003 | -0.001 |
| | Airline2 | 0.199 | - | 0.123 | -0.008 | 0.001 | 0.001 | 0.003 | -0.001 |
| **Airline1** | Car | 0.859 | - | 0.045 | -0.027 | 1.120 | 0.131 | 0.448 | -0.230 |
| | Bus | 0.859 | - | 0.045 | -0.027 | 1.120 | 0.131 | 0.448 | -0.230 |
| | Airline2 | 0.859 | - | 0.045 | -0.027 | 1.120 | 0.131 | 0.448 | -0.230 |
| **Airline2** | Car | 0.911 | - | 0.047 | -0.021 | 1.255 | 0.122 | 0.511 | -0.209 |
| | Bus | 0.911 | - | 0.047 | -0.021 | 1.255 | 0.122 | 0.511 | -0.209 |
| | Airline1 | 0.911 | - | 0.047 | -0.021 | 1.255 | 0.122 | 0.511 | -0.209 |

## 9. 5 Discussion

### 9.5.1 Key factors

For both segments 1 and 2, the key factors that influence travel mode and airline choice are travel cost, journey time, service frequency and seat comfort. These findings are consistent with the findings of the previous chapter and most previous studies. For instance, Jovicic and Hansen (2003), Hess et al. (2007), Teichert et al. (2008), Wen and Lai (2010), Seelhorst and Liu (2015), Molesworth and Koo (2016) and Lee et al. (2016) also found that an increase in travel time and cost would significantly reduce the demand for that travel mode. Van Can (2013) concluded that improving the seat comfort level would attract more passengers, Wen and Lai (2010) suggested that promoting the comfort in terms of seat space could significantly increase sales, while Teichert et al. (2008) and Van Can (2013) also produced similar findings. Further, Teichert et al. (2008), Shen (2009) and Wen and Lai (2010) noted that service frequency generally has a statistically significant impact on attracting passengers. Interestingly, access time is found to be only statistically significant in latent segment 2, while the study conducted by Jovicic and Hansen (2003) showed it to be statistically significant in all predefined segments. One explanation is that people from different study areas and traveller groups may have a distinct difference in the preference of access time. This finding suggests that researchers, airlines and government should be cautious about making firm conclusions regarding access time, as its significance on individuals' travel mode behaviours may be specific to traveller segments and study areas.

### 9.5.2 Elasticity analysis

This study calculated the direct elasticity statistics based on the parameter estimates of segments 1 and 2 independently (e.g., Wen and Lai, 2010; Román et al., 2017). Therefore, it can investigate not only the elasticity differences among travel modes within the segment but also the differences in elasticity of a certain travel mode between segments. One point to note is that, because the size of segment 2, (60.1% of sample population), is larger than that of segment 1, (33.1% of sample population), the impact on total travel use will be, ceteris paribus, larger if it is valued by segment 2. Regarding travel cost, the travellers from both segments 1 and 2 were found to be relatively sensitive, particularly segment 1, which is consistent with the outcomes by Teichert et al. (2008) and Wen and Lai (2010), who also found one small sized segment representing the most cost-sensitive respondents. This finding suggests that reducing airfares could be an efficient way to attract more air travellers, thus increasing the competitiveness of air travel generally, that may have commercial implications (Zhou et al., 2019). Conversely, travellers from both segments, (especially segment 2), were found to be more sensitive to travel time when travelling by car or bus, than by air. The reason for the higher sensitivity may be because the travel times by road for regional trips are normally far longer than by air, in which case, although the percentage increase of travel time by car, bus or airlines may be the same, the absolute increase in travel time by car or bus would be longer. Therefore, reasonably increasing the bus travelling speed or improving connectivity and infrastructure of the road network may be an effective way to encourage more people to choose public ground travel for regional trips.

Regarding the direct elasticities between the two segments, the choice probability changes reveal that segment 1 is more sensitive than segment 2 to the travel costs of all the travel modes. Similar results have been found in existing studies (Wen and Lai, 2010; Jung and Yoo, 2014; Román et al., 2017). The characteristics of the segments can help explain this finding, as segment 1 has a smaller proportion of middle or high-income respondents and business travellers compared to segment 2. However, the less price elastic segment 2 was also found to be more sensitive to other attributes, such as access time, journey time and service frequency. This finding indicates that price insensitive travellers would be more sensitive to other service-quality related factors while making a mode choice for regional travel, which is consistent with the finding by Wang et al. (2014) and Zhou et al. (2019), but contrary to the finding by Wen and Lai (2010). One explanation for the finding is that the price insensitive segment 2 also has a larger proportion of high-income individuals and these travellers may care more about service quality.

### 9.5.3 Willingness to pay

As shown in Table 9-6, segments 1 and 2 yielded distinctly different values of willingness to pay for improving the service factors. Travellers in segment 2 were willing to pay $50.70 and $2.60 to reduce journey times by one hour and increase weekly service frequency by one, respectively. However, the passengers in segment 1 were only willing to pay a relatively modest amount, $6.00 and $0.40 respectively. This finding is consistent with previous studies. For example, Seelhorst and Liu (2015) also discovered that the travellers in price-sensitive segments were only willing to pay a small amount to reduce the journey time, Teichert et al. (2008) and Wen and Lai (2010) found that the respondents from the price-sensitive segment were willing to pay less to improve service frequency. In price-insensitive segment 2, the mean willingness to pay for promoting seat comfort level from low to middle or high were $66.30 and $111.20 respectively, contrasting with $12.10 and $33.30 for travellers in price-sensitive segment 1. However, Wen and Lai (2010) found that, for one specific international air route, the respondents in the price-sensitive segment were willing to pay more for increasing air seat comfort level, which is contrary to the finding of this chapter. The suggestion they proposed is that different international air routes would yield significant variations in terms of willingness to pay for improving service quality factors. As a whole, travellers with high incomes and a business trip purpose are less price sensitive and willing to pay more to improve the service attributes. This finding is consistent with previous studies. For example, Wen and Lai (2010) found that passengers with high incomes are generally willing to pay more for service attribute improvements, while Pels et al. (2003), Jung and Yoo (2014) and Zhou et al. (2019) discovered that business travellers are willing to pay more for improvements in service factors, such as service frequency and accessibility.

Conversely, the results also indicate that travellers in segment 2 are willing to pay more ($58.20) to reduce access time by one hour compared to journey time ($50.70). This finding suggests that those travellers who significantly consider access time as a factor in making travel mode and airline choice are relatively more likely to prefer a reduction in access time over the same reduction in flight time. In spite of this, access time was found to be statistically insignificant in influencing passengers' travel mode and airline choice in segment 1. This information is potentially useful for governments, who could encourage more passengers to use bus and airlines by improving the accessibility of bus stations and airports.

*Table 9-6 Willingness to pay for each of the segments*

| | Three-segment LC model | | |
|---|---|---|---|
| | Segment 1 | Segment 2 | Segment 3 |
| Journey time ($/per hour) | $6.00 | $50.70 | - |
| Access time ($/per hour) | - | $58.20 | - |
| Service frequency ($/increase per flight by a week) | $0.40 | $2.60 | - |
| Seat comfort_Mid ($/increase to middle level) | $12.10[a] | $66.30 | - |
| Seat comfort_High ($/increase to high level) | $33.30 | $111.20 | - |

*Table notes:*

*[a] This parameter estimate is only significant at the 90% confidence level, t-value=1.83, p-value=0.067*

### 9.5.4 Characteristics of segments

This study used an LC model to accommodate preference heterogeneity across individuals, where the observed characteristics were not considered in the latent segment classification function, which is similar to existing research studies (Greene and Hensher, 2003; Wen et al., 2012; Kim et al., 2017; Román et al., 2017). One reason may be that the demographic and economic variables may have some correlations between each other. For example, people's age is likely to be correlated with their income level, and such a correlation may affect the segment classification results and therefore reduce the reliability of the model's estimation results. Another reason is that, although the distinct differences in the observed characteristics, such as demographics, economics and trip purpose, can influence individual preferences for travel mode and airline choice to some extent, they may not be enough to set these characteristic attributes as the explanatory variables for the latent segment identification. Therefore, in terms of including these attributes in the segment classification function, the corresponding parameter estimates were found to be statistically insignificant in at least half of the identified latent segments. However, from a counterview, it is because the LC model can account for the unobserved preference heterogeneity across the respondents and classify them into different latent segments. Thus, the general characteristic differences between the segments can be examined. Consequently, airlines and policymakers can deploy different strategies or policies to attract the passenger groups with different characteristics.

In this study, the characteristics of both segments were calculated based on aggregating each respondent's characteristics with the weights of his/her individual-specific posterior class probabilities, (as shown in Equation 9-7), rather than simply assigning respondents to one of the segments by the largest posterior class probability. In comparison with segment 1, the elasticity and the mean willingness to pay figures have shown segment 2, with a larger proportion of business trip purpose and high-income travellers, to be more sensitive to the service attributes, but less sensitive to the price. Moreover, in comparison with segment 1, segment 2 has a lower proportion of basic education and unemployed persons. One interesting finding is that the price-insensitive segment 2 mainly consist of middle-aged or older people (up to 90%) and more males, while the price-sensitive segment 1 contains more females and more than one quarter are young, (less than 25 years old). These findings suggest that reduced airfares would be more popular with the young, female and leisure travellers, but improved service quality would impact most on demand from older passengers and those travelling for business. Similar results were obtained by Wen and Lai (2010) with one exception; they found that reducing airfares may be more conducive to increasing male patronage.

### 9.5.5 Number of segments

Information criteria such as AIC and BIC are typically used by researchers to find the appropriate number of latent segments. As the log-likelihood increases with the number of segments, caution is required to ensure that the model doesn't become imprecise and over fitted, i.e., if too many segments are used (Heckman and Singer, 1984; Greene and Hensher, 2013). At this point, researchers should consider the general statistical significance of the parameter estimates and the rationality of the modelling results in conjunction with the information criteria results to determine the appropriate number of segments. There are many previous studies that consider these points, in addition to the information criteria, to find the appropriate number. For example, Greene and Hensher (2013) selected a two-segment LC model as it delivered a better performance of the statistical significance of the parameter estimates. Vij et al. (2013) considered the rationality of the modelling results, such as the signs and magnitudes of the parameter estimates, and finally chose a three-segment LC model with the second smallest AIC value but the smallest BIC value. Wen and Lai (2010) used a two-segment LC model for estimating individuals' airline choice because the parameter estimates were more stable, even though the corresponding AIC and BIC values were not the smallest. In line with these papers, the present study found that the three-segment LC model was the best fitting model even though the AIC index was the second smallest, (BIC value was the smallest). This is because the three-segment LC model not only provides a better performance of statistical significance of the parameter estimates but also filters out the residual segment 3, (which only accounts for a minor proportion of sample population with all parameter estimates statistically insignificant). The individuals in the residual segment may have struggled to understand the experiment despite significant prior information being provided, or completed the questionnaire without appropriate consideration. Interestingly, this residual segment also arose in the study conducted by Román et al. (2017), who found two small residual segments in their five-segment LC model. Ultimately, for researchers, it is hard to minimise this residual population, but this finding of a residual segment providing data likely to be low-quality, further demonstrates the need for data collection methods to be well considered before completion of cognitively challenging survey tasks such as this.

## 9. 6 Summary

Earlier studies have investigated individuals' travel mode and airline choice behaviour but few included a segmentation approach for the choice modelling. In this chapter, an LC model was used to accommodate unobserved preference heterogeneity across the individuals and, thus, identified two distinct market segments with different preferences for travel mode and airline choice. The LC model not only outperformed the MNL model but also provided segment-specific parameter estimates and membership, as well as the posterior individual-specific segment probabilities. The outputs from these models can provide a better understanding and comparison of the preference differences between the latent segments and also compute the general demographic, economic and trip characteristics of each segment for exploring the possible relationships between the preference difference and these observed characteristics. In general, the modelling results provide evidence that the key determinants affecting regional travellers' mode choice among car, bus and airlines are travel cost, journey time, seat comfort and service frequency, while the importance of access time may depend on the market groups or study areas. This research was especially interested in the identified latent segments of the travellers and, thus, the results of direct elasticity and willingness to pay for each of the two segments are reported. As discussed before, in contrast to segment 1, segment 2 mainly consists of travellers with middle to high income, middle to old age and business trip purposes. These travellers are more sensitive to the service-quality related factors including travel time, access time, seat comfort and service frequency and, thus, are willing to pay more for the improvement of these factors.

The next chapter is the final chapter of this thesis. It concisely summarises the key findings and achievements of the whole work including limitations regarding data collection and the modelling methods as well as recommendations and ideas for future research.

# CHAPTER 10 EVALUATION AND CONCLUSIONS

## 10. 1 Introduction

Although the Western Australian government has proposed a number of policies to tackle the aviation issues and provide a better air service for regional businesses and local communities, (e.g. regulation/deregulation of air routes), there remains an urgent need for policymakers and the airline industry to better understand the market. This thesis proposed a series of linked studies investigating the key parameters influencing air travel demand (Chapter 4), the characteristics of the regional aviation market (Chapters 6 and 7), the travel behaviour of regional travellers, and the key factors that affect travel mode and airline choice (Chapters 8 and 9).

This chapter summaries the major findings and empirical results from this thesis, and discusses possible limitations in the data collection and modelling methods. Some recommendations are then proposed, that may be of benefit to and direct future studies. Finally, there is a discussion on how the research objectives in this thesis have been met.

## 10. 2 Summary of Research Findings

Air travel demand is the inherent motivating force behind decisions to invest in airport infrastructure, the provision of affordable air services and improving service quality. Nevertheless, accurate forecasts of passenger movements are not available to policy makers due to the lack of relevant air passenger movement information (Regional Aviation Association of Australia, 2013). Therefore, this research collected data on the total seats available on Western Australian domestic RPT air routes, (i.e. within the state), to represent air travel demand. Adapted gravity models were then applied to forecast air travel demand, (air passenger seat numbers), as gravity models are an effective tool to help understand spatial structures and interactions (Nijkamp, 1997).

In this thesis, a survey using efficient SP experimental design principles was constructed to investigate the travel mode and airline choices of regional travellers. A realistic and statistically efficient SP survey can best reflect participants' choice-making behaviours and therefore maximise the reliability of modelling results. Setting up constraints in an SP experimental design can assist in improving the realism level of the survey questions. However, this process is tedious and time consuming. Although an increasing number of previous researchers has focused on the methodology of constructing SP experimental designs, limited research has been done regarding design constraints and few of these studies have attempted to automate the process. In order to address this gap, this research developed a semi-systematic method for generating an efficient SC design by extending the Modified Federov Algorithm, that can help to more easily specify all required constraints and effectively ensure appropriate realism and statistical efficiency of the choice questions.

A major challenge identified in the thesis is that airlines cannot easily balance the preferences of different group of passengers while maintaining a sound commercial and economic status (Shaw, 2016), as they have various preferences for, or needs from, the same air services factors (Kotler, 2009). In this thesis, the market segmentation analysis using an EM algorithm estimator was applied to identify and investigate the regional aviation market segments in terms of both airport and non-airport respondent samples. It used a mixture model-based market segmentation approach that can effectively identify the distinct market segments (Fraley and Raftery, 2002; Jacques et al., 2013). The EM algorithm is an efficient tool for estimating the mixture model parameters and producing reliable segmentation/clustering results (Kishor and Venkateswarlu, 2016; Neal and Hinton, 1998; Xu and Wunsch, 2005).

As indicated in the literature chapter, airlines are not only facing competition within the industry, but also experiencing increasingly fierce competition from non-air travel modes. Thus, it is necessary to investigate regional travellers' mode choice, as well as how the key factors such as travel cost, journey time and service quality affect the choice. Although a number of studies has explored travel mode and airline choice between cities, states and/or countries, no studies have investigated travel mode and airline choice in regional Western Australia. Most of the studies conducted elsewhere applied discrete choice analysis based on SP data collected at airports or train stations, that may be valuable for certain questions but are subject to selection bias and may not generalise to the rest of the population, especially potential air travellers. Therefore, this thesis addressed these gaps by the face-to-face collection of SP data in both regional airports and other settings likely to involve those who do not frequently choose to fly, and applied a range of statistical analysis techniques to more comprehensively analyse travel mode and airline choice (see Chapters 8 & 9).

### 10.2.1 Domestic air travel demand forecasting

The objective of Chapter 4 was to estimate domestic air passenger seat numbers between airport-pairs based on modified gravity models. Particularly, it aimed at investigating the impact of distance, airfare, catchment areas, population, tourism and the mining sector on forecasting air passenger seat numbers in order to inform and guide policy making. This research collected appropriate data and produced valid models that can predict the numbers of air passenger seat offered on RPT air services in regional Western Australia. The models considered both geographic and service-related variables, such as the catchment areas of airports, and population and number of tourists within those catchment areas. Two kinds of airport catchment areas were generated in this study, based on Thiessen polygons and 2.5 hour driving distances. The Thiessen polygon catchment areas covered the whole of Western Australia, while the 2.5 hour driving catchment areas only covered 32 percent of Western Australia. The size of the catchment area can affect the magnitude of the factors and therefore influence the modelling results. When deciding the catchment area for the study, it was important to take the spatial distribution of factors into consideration.

For both the Thiessen polygon and 2.5 hour driving distance catchment areas, the model results illustrated that distance between airports, airfare of the flight, population of the origin airport's catchment area, and the number of operating mine sites and tourists within the destination airport's catchment area are significantly correlated with domestic air travel seat capacity provided. For Western Australia, the mining sector was found to have more influence than the tourism sector on total seat availability. The findings indicate that facilitating mining development and stimulating tourism growth would be effective ways to increase air passenger movements. One noteworthy point is that the airlines and government should prepare countermeasures to respond to fluctuations in the mining industry in Western Australia, as mining upturns and downturns can both significantly influence air travel demand. These results improve the understanding of the key parameters of regional passenger aviation services and help guide policy makers to better implement airport investment.

### 10.2.2 Stated preference efficient experimental design

A realistic and statistically efficient SP experiment is a crucial step in understanding regional travellers' travel behaviour. In this thesis, a novel, efficient and practical semi-automatic constraint setting method of EMFA was developed for constructing experimental designs

(Chapter 5). The proposed EMFA was implemented to create a D-efficient SP survey for the case of Western Australia, in order to understand those travellers' travel mode and airline choice behaviour. The constructed SP survey confirmed that the proposed method performs well in finding and specifying the constraints, and the survey not only maintained a relatively high statistical efficiency but also an appropriate behavioural plausibility and realism. A random simulation test was applied to examine the SP design's modelling performance, which indicated that the generated SP survey was appropriate for collecting the data and modelling travel mode and airline choice.

This chapter deliberately developed the computer analytical codes using open source software, with the intention that these codes could be accessed by the broader research community for easier and better experimental choice design. This also remains an area for future research; the EMFA method could be further improved by developing and expanding an open source library for accessing basic constraints regarding different choice study areas.

### 10.2.3 Regional aviation market segmentation

Chapter 7 provided an approach for identifying and investigating existing and potential aviation markets, to assist local governments and airlines in developing more efficient marketing strategies. The mixture model-based market segmentation approach using an EM algorithm estimator was applied to identify corresponding market segments in terms of socio-demographics, (e.g., age and income), trip characteristics and stated preferences among air and non-air travel modes, (car and bus). Three distinct segments were identified in both the air and non-air passenger samples. For the airport respondent sample, one segment mainly consisted of male, older and middle-income travellers who were relatively less likely to choose air travel than the other two segments. This segment, consisting of 18% of the total sample, chose air travel in 52.5% of choice tasks, compared to the pooled other two segments, who consisted of 82% of the sample, and chose airlines 95.3% of the time. Conversely, in the non-airport respondent sample, two of the three segments typically comprised of relatively younger, lower income and non-business purpose travellers who were less likely to choose air transport than the other segment. These two segments, consisting of 45% of the total sample, chose air travel in 35.7% of choice tasks, compared to the pooled other segments, who consisted of 55% of the sample, and chose airlines 87.8% of the time. Therefore, a highly individualised marketing strategy could be developed by the regional airlines and/or government to target these market segments, as it is relatively easy to attract those travellers who are moderately likely to choose air transport. Through the proposed approaches, the study revealed potential aviation markets and their characteristics and travel preferences. This provides a clear insight into the air transport strategy, where the aim is to draw more travellers into air travel. Two identified potential aviation market segments, for instance, were that of relatively young travellers from low or middle income groups. These were more likely to use non-air transport and, especially, on a non-business trip. Therefore, for the air transport department and the regional aviation industry, policies and or strategies targeting these market segments could focus on the provision of airport and airplane amenities that appeal to the younger population of travellers, reducing airfares for lower income groups, and offering extra incentives for leisure travellers, such as the free travel strategy, extra coupons for some of the restaurants and hotels that relatively closer to the tourist places within the destination cities, and the small gift of the tourism souvenir. The suggestion of importance to the regional aviation industry is that, for the travellers with low to middle incomes and a non-business travel purpose, reducing airfares (especially for leisure travel) and introducing low-cost carriers could be an effective way to further increase the aviation market share. Overall, these findings, showing heterogeneous groups in both populations, provide a more nuanced view of the aviation market than has been developed to date, and allow key stakeholders in the aviation sector, (including airlines and government), to better predict the consequences of the various policy levers at their disposal.

#### 10.2.4 Travel mode and airline choice estimation

Chapter 8 investigated the travel mode and airline choice of air and non-air travellers in regional Western Australia using discrete logit models. MNL and NL models were used as key analysis tools for the SP survey data. Both assume homogenous preferences across respondents for a particular sample group. In this chapter, the SP data were divided into four subsets, air travel business and non-business groups, and non-air travel business and non-business groups. The modelling analysis was applied to each of the four groups. The results indicated that travel cost, journey time, service frequency and seat comfort played important roles in affecting travellers' regional travel mode choices. For business trips, air passengers were willing to pay more to reduce journey time and increase seat comfort and service frequency compared to non-air passengers, while for non-business trips, these differences were much smaller. The air and non-air passengers had statistically significantly different scales of sensitivity to the travel mode choice related factors. Apart from that, business travellers were more time sensitive and less price sensitive compared to the non-business travellers. The findings suggest that reducing airfares is an efficient way to attract more air passengers, while reducing road journey time may encourage more travellers to use road travel modes, such as the introduction of express bus and or reasonably increase speed limit of some sections of the regional roads and highways that may reduce the car journey time. This guidance for policy makers and airlines provides important insights into understanding people's travel mode choice behaviour.

These findings are valuable and it is important to extend them by considering how preference heterogeneity may influence results in each of the subgroups. Therefore, Chapter 9 accommodated such potential preference heterogeneity and subsequently investigated the travel mode choice behaviour within, and among, passenger segments. A market segmentation approach, using LC modelling, was applied to identify latent passenger segments. The LC model effectively accommodated unobserved preference heterogeneity by assuming a discrete distribution of parameters across individuals to accommodate for the heterogeneity in the sample population, identifying two differentiated market segments and outperforming the logit model in estimating the mode choice. The results show a distinct difference between segments in terms of demographics, economics and trip characteristics. One segment, (60.1% of the sample population), comprising travellers with relatively high income, older age and travelling more frequently on business, was sensitive to service-quality factors. The other, comprising personal travellers and those on relatively lower incomes, placed relatively more importance on price. These findings are consistent with the MNL and NL modelling results obtained in Chapter 8, but also provide a more comprehensive understanding of the regional travellers' travel behaviour, as well as their sensitivities to the key factors. Additionally, as shown in Figure 6-7, the stacked bar chart indicated that the regional air travellers ranked journey time and travel cost as the most important factors in affecting travel mode choice, and access time as the least important factor. This ranking outcome was confirmed by the NL and LC modelling results. For example, the direct elasticity results imply that a unit percent increase of journey time and travel cost (e.g., airfare) for a particular travel mode would cause the largest proportional reduction in sales/usage.

Five key aviation policy suggestions can be obtained based on the choice modelling outcomes. Firstly, reducing airfares would be an effective approach to improve air transport patronage especially for young, low to middle-income, female and leisure travellers. Secondly, reducing journey time could also help to attract travellers, especially for the non-air travel modes. For example, within an acceptable and safe range, increasing the flight speed or the speed limit of the regional roads and highways would impact on passenger demand. Thirdly, improving the service quality factors such as seat comfort and service frequency would impact most on demand from older and high-income passengers, and those travelling for business. Fourthly, access time to reach the airport is found to be only significant to business and high-income travellers, which suggests that the government could consider upgrades to the roads to the airports for those mining or industry towns. Lastly, cross elasticity statistics (Chapters 8 & 9) have revealed the competition between airlines, between air and ground travel modes, and the factors affecting the competition. The airfare was found to be the most effective factor on influencing the competition between airlines, while a unit increase of a given airline's airfare would significantly reduce its competitiveness against the other airline and the non-air travel modes. However, for the non-air travel modes, increasing travel cost and/or journey time would significantly reduce their competitiveness against air transport. As a whole, the findings highlight the importance of understanding mode choice behaviour based on market segmentation and provide insights to policy-makers and airlines for developing more effective polices and marketing strategies and, hence, to better tackle the existing aviation issues.

## 10. 3 Limitations of the Research

This research could provide contributions to the government and aviation industries in more comprehensively understanding regional aviation market and competition, through a series of spatial and statistical modelling analyses. However, there are limitations that should be acknowledged and are discussed in this section.

### 10.3.1 Limitations of data collection

- Flight data collection

Flight data including real-time flights and seats information were collected from the websites www.Flightradar24.com and https://planefinder.net. These data have slight differences/errors in comparison with the available government-published flight data, (4% to 6% monthly difference), as discussed in section 4.5.4. These differences may add some uncertainty to the air travel demand modelling analysis. Additionally, historical data on air passenger numbers were not available due to commercial and confidentiality reasons. Therefore, total seats on the flights on the RPT routes in regional Western Australia were used as a proxy for modelling air travel demand, which may slightly reduce the accuracy, subject to the actual aircraft load factors.

- SP survey data collection

The SP survey data used to identify the regional aviation market segments, as well as to model individual travel mode and airline choice, were collected in four selected regional towns, (and their airports), in 2018. Although the four regional towns were chosen in collaboration with government transportation agency partners, met the research selection criteria, and deliberately represented a variety of locations and socio-economic characteristics, Western Australia had 26 RPT airports, (located in 26 regional towns), servicing the public in 2018. Hence this thesis cannot with certainty assert that the survey had captured the opinions of all regional travellers across the whole state.

### 10.3.2 Limitations of modelling methods

- Gravity model based air travel demand forecasting

Several variables, including distance, population, airfare and number of tourists and operating mine sites within the airport catchment area, were accounted for in the modified gravity model to forecast the air travel demand. However, the network structure of flights between airports was not calibrated into the model, (i.e., hub flight locations may have an extra impact on improved the flight frequencies), and, as mentioned in the literature chapter, it may also have an impact on travel demand. Additionally, for consistency and simplicity, only fully flexible airfares were considered in the gravity model, which may result in some discrepancies between the assumed airfares and those actually paid by the passengers. This limitation may lead to some variations in the estimated parameter/scale of the airfare factor. Therefore, more accurate information on airfares could further improve the accuracy of the air travel modelling results.

- Mixture model based aviation market segmentation

The main limitations of the mixture model based market segmentation approach are twofold. Firstly, individual preferences for air and non-air transport are represented as the mean probabilities calculated based on his/her travel mode choices across the SP mode choice questions. Although the SP questions were designed to be reflective of actual choices faced in the real world, they are still hypothetical scenarios and therefore may lose some capacity in representing respondent's real experiences. Secondly, while applying an EM algorithm to estimate the mixture model parameters, the numeric (mean probabilities) and nominal attributes (i.e., age and gender) were assumed to be independently distributed. However, there may be some correlation between the attributes. As recognised by Witten et al. (2016), it is a quite complex task to quantify/accommodate such potential correlations, but achieving this may further improve the reliability of the modelling results.

- EMFA for constructing efficient stated preference survey

Each iteration in the EMFA method requires the researcher to find the undesirable choice tasks manually, and then to determine the constraints for rejecting these choice tasks. The number of iterations required is subject to the focus of the specific choice problem that may be relatively large for complex problems, with relatively more alternatives, attributes and attribute-levels. In these cases, the manual process for recognising problem choice tasks and specifying the corresponding constraints could be substantial and potentially result in an unreasonable workload for the researcher. Researchers who use other approaches may need some proficiency in coding to transform the pseudocodes provided in Chapter 5.

- Discrete choice model based travel mode and airline choice estimation

There are three major limitations associated with the application of discrete choice modelling. The first relates to the omitted factors, as the model did not account for egress times or the time required for check-in and bag collection. This may add an extra hour, or more, to the overall airline travel time and, thus, may reduce the choice probability/utility of air travel. The second limitation is that, for simplicity, the travel cost of car did not include any accommodation or other costs that might accrue, especially on longer trips, and would increase the real cost of car travel. The primary reason for omitting this as a separate variable was that including accommodation and other costs would require a strong assumption about how car travellers would undertake their journey across multiple days. Secondly, it is potentially true that flying rather than driving may not reduce the total accommodation cost as extra nights at the final destination may be spent in paid accommodation. Finally, leg room distance was used as the proxy for seat comfort level in this study. However, the value of this proxy measure might vary depending upon the length of the journey; for example, respondents may be more likely to value a seat that can be fold down than the long leg room if their journey included overnight travel, This may change the estimated impact (estimated parameter) of seat comfort on the individual travel mode and airline choices.

### 10. 4 Future Research Directions

Regarding air travel demand estimation, further research could be conducted to improve the demand forecasting accuracy, including consideration of other potential factors such as the air travel network and seasonal variations in demand and service provision. It is also suggested that a time series analysis of air passenger trips in Western Australia would be valuable in investigating how changes over time in the key drivers of demand, such as mining and tourism, impact on air passenger trips. It could not only contribute to offering more reliable air travel demand forecasts, but also provide insight to the government and airlines who are seeking to more effectively develop policies and strategies for improving air transport usage and service quality.

For the market segmentation analysis, an advantage is that this study explored the characteristics of non-air passenger respondents while investigating the aviation market. Future aviation market studies could also possibly include these respondents in addition to the air passengers to achieve more reliable results, especially of the potential airline market. The mixture model-based market segmentation approach allows statistical analysis to make a formal inference for further exploration and understanding of the market. Therefore, future studies could explicitly apply further statistical analysis, such as factor analysis, that may improve the significance of the results and provide a deeper understanding of the aviation markets and traveller behaviour. Additionally, more effectively handling the potential correlations between numeric and categorical attributes of the vector sample data in the segmentation analysis may further enhance the modelling performance.

There are also future directions that would be worthwhile considering in the context of the discrete choice modelling analysis. Firstly, future research is recommended to explicitly consider the omitted parameters, such as the egress time and schedule convenience, and interactions between specific variables such as the potential effect of the interaction of price and schedule convenience. More than that, with sufficient funds, researchers may try to collect the data in more diverse areas and invest more resources on the data collection phase to better inform respondents about the choices they face and to understand how respondents complete the survey. For example, future work could consider employing eye-tracking technology or video surveys, in order to filter out the small segment of respondent outliers represented by the residual segment, (approximately 7% of respondents), who either did not, or could not, meaningfully engage with the survey. Finally, it would be worthwhile comparing regional travel mode choice behaviour for different regions in order to gain a deeper understanding of how distance, history, culture, dominant industries and populations influence the regional travel behaviour, especially mode choice behaviour.

### 10. 5 Achievement of research objectives

This thesis has investigated the regional aviation market and its interactions with its competitors for the regional travel market in Western Australia, over space. Four main objectives were set up in the introductory chapter (Chapter 1). In the subsequent literature review chapter (Chapter 2), the findings provided possible explanations relating to issues raised in the background and problems of the regional aviation market, as well as the theoretical analysis that can explore the aviation market from various representative statistical viewpoints. The methodology for the analysis and forecasting of air travel demand, constructing efficient and reliable SP experimental design, exploring regional aviation market characteristics and estimating travel mode choice and behaviours, was then established and demonstrated in the research framework and methodology chapter (Chapter 3). The four key objectives were achieved in the subsequent chapters. Chapter 4 developed a set of modified gravity models that accomplished the first objective in terms of forecasting bilateral air travel demand on RPT airport-pairs in regional Western Australia for a given time period, as well as identifying the significant key factors that could affect the aggregate travel demand. Chapter 5 proposed and implemented a semi-automatic experimental design procedure that addressed the second objective, relating to more effectively constructing statistically efficient SP experiments while maintaining an appropriate realism level of the choice questions. The third main objective sought to more explicitly explore the regional aviation market characteristics. Thus, Chapter 6 generated a series of visualisations, based on the air travel information data of regional air passenger respondents, and provided a basic understanding of the air passenger characteristics. Chapter 7 developed a mixture model-based market segmentation approach using an EM algorithm estimator to identify the existing and potential aviation market segments in regional Western Australia, and explored the prominent characteristics of each distinct segment, respectively. The final objective was to estimate individual passenger travel mode and airline choice for regional trips in Western Australia and identify the key factors that drive passenger choice decisions. This objective is of crucial importance to the transport government and airlines as it can not only reveal the internal, (between peer airlines), and external, (between air and non-air transportation), competition but also quantify the passengers' sensitivities to key factors such as travel cost, time and service frequency. In order to achieve this goal, Chapters 8 and 9 estimated travel mode and airline choice from two different statistical perspectives. Chapter 8 focused on considering the potential correlation or substitution between regional airlines using an NL modelling approach, whereas Chapter 9 concentrated particularly on accommodating potential preference heterogeneity across individuals through an LC modelling technique.

### 10. 6 Conclusions

This thesis has proposed a realistic, rigorous and easily computed modified gravity model, a mixture clustering model and travel mode and airline choice models to identify the determinants affecting air travel demand and the potential aviation markets. It has also explained the travel mode and airline choices of regional traveller and revealed the competition in the regional aviation market.

The case study was conducted in Western Australia, where the innovative exploration and modelling methods were carried out. The thesis has applied the modified gravity models to forecast regional air travel demand between airports, where the scales of the identified key drivers have been revealed. EMFA was developed subsequently to more easily and effectively generate the optimal SP experimental design (SP survey). Based on the face-to-face collected SP and air travel survey data, the study then deliberately identified and investigated the existing and potential aviation markets through the model-based market segmentation approach. Finally, this thesis employed MNL, NL and LC models to more comprehensively estimate regional traveller's travel mode and airline choice, as well as the competition between not only airlines, but also air and non-air travel modes. The modelling has quantified individuals' sensitivities to the key factors such as travel cost, travel time and seat comfort, which were found to be consistent with the average rankings they placed in the survey for these factors.

These findings can shed light on the regional aviation market, especially for the government transport agencies and the regional airlines who are seeking to more effectively respond to the market (as outlined in section 10.2). Therefore, Western Australian current aviation issues could be better tackled, such as the high regional airfares, the decline in air passenger movements and the lack of regulation and airport investment policy-making guidance. The findings may also be generalised and applied to other regions, (with some caveats), that have similar geographical pattern and characteristics to Western Australia. Some algorithm or modelling approaches established in this thesis could also be applied to other studies. For example, the proposed EMFA could be used for generating optimal SP surveys for other research areas, such as urban development, retailing and healthy. Similarly, the established aviation market segmentation approach may also be utilised to identify the potential segments of other different type of markets.

## REFERENCES


Agresti, A., & Kateri, M. (2011). *Categorical data analysis*: Springer.

Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Weaven, S. (2019). Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. International Journal of Hospitality Management, 80, 52-77. doi:https://doi.org/10.1016/j.ijhm.2019.01.003

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle *Selected papers of hirotugu akaike* (pp. 199-213): Springer.

Albany (Harry Riggs) Regional Airport. (2018). City of Albany: Albany (Harry Riggs) Regional Airport. Retrieved from http://www.albany.wa.gov.au/utility/venues/albany-harry-riggs-regional-airport/

Alhussein, S. N. (2011). Analysis of ground access modes choice King Khaled International Airport, Riyadh, Saudi Arabia. *Journal of Transport Geography, 19*(6), 1361-1367. doi:10.1016/j.jtrangeo.2011.07.007

Anderson, J. E. (2010). The Gravity Model. *National Bureau of Economic Research Working Paper Series, No. 16576*. doi:10.3386/w16576

Anderson, S. P., De Palma, A., & Thisse, J.-F. (1992). *Discrete choice theory of product differentiation*: MIT press.

Angulo, F., Pergelova, A., & Rialp, J. (2010). A market segmentation approach for higher education based on rational and emotional factors. *Journal of Marketing for Higher Education, 20*(1), 1-17.

Anwar, S., & Nguyen, L. P. (2011). Foreign direct investment and trade: The case of Vietnam. *Research in International Business and Finance, 25*(1), 39-52.

Asmael, N., & Waheed, M. (2018). Demand estimation of bus as a public transport based on gravity model. Paper presented at the MATEC Web of Conferences.

Australian Bureau of Statistics. (2015a). 6202.0 - Labour Force. Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/mf/6202.0

Australian Bureau of Statistics. (2015b). Population of Western Australia, 2015. Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/mf/6202.0

Australian Bureau of Statistics. (2016). Population of Western Australia, 2016. Retrieved from http://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/5?opendocument

Äyrämö, S., & Kärkkäinen, T. (2006). Introduction to partitioning-based clustering methods with a robust example. *Reports of the Department of Mathematical Information Technology. Series C, Software engineering and computational intelligence*(1/2006).

Bain, J. H. (1976). *Activity choice analysis time allocation and disaggregate travel demand modeling*. Massachusetts Institute of Technology.

Baker, D., Merkert, R., & Kamruzzaman, M. (2015). Regional aviation and economic growth: cointegration and causality analysis in Australia. *Journal of Transport Geography, 43*, 140-150.

Balcombe, K., Fraser, I., & Harris, L. (2009). Consumer willingness to pay for in-flight service and comfort levels: A choice experiment. *Journal of Air Transport Management, 15*(5), 221-226.

Barry, J., & Weinstein, A. (2009). Business psychographics revisited: from segmentation theory to successful marketing practice. *Journal of Marketing Management, 25*(3-4), 315-340.

Becken, S., & Carmignani, F. (2020). Are the current expectations for growing air travel demand realistic? Annals of Tourism Research, 80, 102840. doi:https://doi.org/10.1016/j.annals.2019.102840

Beirão, G., & Cabral, J. S. (2008). Market segmentation analysis using attitudes toward transportation: exploring the differences between men and women. *Transportation research record, 2067*(1), 56-64.

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions *Handbook of transportation science* (pp. 5-33): Springer.

Ben-Akiva, M. E. (1973). *Structure of passenger travel demand models.* Massachusetts Institute of Technology.

Ben-Akiva, M. E., Lerman, S. R., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9): MIT press.

Bergkvist, E., & Westin, L. (1998). *Forecasting interregional freight flows by gravity models: Utilising OLS-, NLS-estimations and Poisson-, Neural Network-specifications.* Paper presented at the 38th Congress of the European Regional Science Association.

Bhat, C. R. (1997). An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science, 31*(1), 34-48. Retrieved from http://www.jstor.org/stable/25768750

Binova, H. (2015). MODIFIED METHOD OF GRAVITY MODEL APPLICATION FOR TRANSATLANTIC AIR TRANSPORTATION. *Neural Network World, 25*(2), 203. Retrieved from http://search.proquest.com/docview/1701082682?accountid=10382

Bínová, H. (2015). Modified method of gravity model application for TransAtlantic air transportation. *Neural Network World, 25*(2), 203.

Bliemer, M. C., & Collins, A. T. (2016). On determining priors for the generation of efficient stated choice experimental designs. *Journal of Choice Modelling, 21*, 10-14.

Bliemer, M. C., & Rose, J. M. (2005). Efficiency and sample size requirements for stated choice studies.

Bliemer, M. C., & Rose, J. M. (2006). *Designing stated choice experiments: state of the art.* Emerald.

Bliemer, M. C., & Rose, J. M. (2011). Experimental design influences on stated choice outputs: an empirical study in air travel choice. *Transportation Research Part A: Policy and Practice, 45*(1), 63-79.

Bliemer, M. C., Rose, J. M., & Hensher, D. A. (2007). Constructing efficient stated choice experiments allowing for differences in error variances across subsets of alternatives. *Transportation Research B.*

Bliemer, M. C., Rose, J. M., & Hensher, D. A. (2009). Efficient stated choice experiments for estimating nested logit models. *Transportation Research Part B: Methodological, 43*(1), 19-35.

Bliemer, M. C. J., & Rose, J. M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B-Methodological*, *44*(6), 720-734. Retrieved from <Go to ISI>://WOS:000278601100002

Bolduc, D. (1999). A practical technique to estimate multinomial probit models in transportation. *Transportation Research Part B: Methodological, 33*(1), 63-79. doi:https://doi.org/10.1016/S0191-2615(98)00028-9

Boonekamp, T., Zuidberg, J., & Burghouwt, G. (2018). Determinants of air travel demand: The role of low-cost carriers, ethnic links and aviation-dependent employment. Transportation Research Part A: Policy and Practice, 112, 18-28. doi:https://doi.org/10.1016/j.tra.2018.01.004

Boxall, P. C., & Adamowicz, W. (1999). Understanding heterogeneous preferences in random utility models: The use of latent class analysis.

Boxall, P. C., & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and resource economics, 23*(4), 421-446.

Broome International Airport. (2018). About. Retrieved from http://broomeair.com.au/passenger-info/bia-flight-information/weekly-schedules/

Bruning, E. R., Kovacic, M. L., & Oberdick, L. E. (1985). Segmentation analysis of domestic airline passenger markets. *Journal of the Academy of Marketing Science, 13*(1-2), 17-31.

Buraga, A. M., & Rusu, A. (2014). Estimating the territorial autocorrelation of air passengers flows in europe using a multivariate gravity model. *Analele Stiintifice ale Universitatii" Al. I. Cuza" din Iasi. Serie Noua. Geografie, 60*(2).

Bureau of Infrastructure Transport and Regional Economics. (2007). *Australian Domestic Airline Activity 2007*. Retrieved from Department of Infrastructure, Regional Development and Cities: https://bitre.gov.au/publications/ongoing/files/domestic_airline_activity_2007.pdf

Bureau of Infrastructure Transport and Regional Economics. (2012). *Australian Domestic Airline Activity 2012*. Retrieved from Department of Infrastructure, Regional Development and Cities: https://www.bitre.gov.au/publications/ongoing/files/domestic_airline_activity_2012.pdf

Bureau of Infrastructure Transport and Regional Economics. (2018). *Australian Domestic Airline Activity 2018*. Retrieved from Department of Infrastructure, Regional Development and Cities: https://www.bitre.gov.au/publications/ongoing/files/domestic_airline_activity_2018.pdf

Bureau of Transport and Communications Economics. (1994). *Transport and Communications Indicators: A Quarterly Review of Transport and Communications Activity*. Retrieved from

Burgess, L., Knox, S. A., Street, D. J., & Norman, R. (2015). Comparing designs constructed with and without priors for choice experiments: A case study. *Journal of Statistical Theory and Practice*, *9*(2), 330-360.

Burrough, P. A., McDonnell, R., McDonnell, R. A., & Lloyd, C. D. (2015). *Principles of* *geographical information systems*: Oxford University Press.

Button, K., Martini, G., Scotti, D., & Volta, N. (2019). Airline regulation and common markets in Sub-Saharan Africa. *Transportation Research Part E: Logistics and Transportation Review, 129*, 81-91.

Cahill, D. J. (1997). Target marketing and segmentation: valid and useful tools for marketing. *Management Decision, 35*(1), 10-13.

Carlsson, F., & Martinsson, P. (2003). Design techniques for stated preference methods in health economics. *Health economics, 12*(4), 281-294.

Çekyay, B., Kabak, Ö., Ülengin, F., Ulengin, B., Toktaş Palut, P., & Özaydın, Ö. (2020). A multi-commodity network flow and gravity model integration for analyzing impact of road transport quotas on international trade. Research in Transportation Economics, 80, 100816. doi:https://doi.org/10.1016/j.retrec.2020.100816

Cerwick, D. M., Gkritza, K., Shaheed, M. S., & Hans, Z. (2014). A comparison of the mixed logit and latent class methods for crash severity analysis. *Analytic Methods in Accident Research, 3*, 11-27.

Chang, L.-Y. (2012). International air passenger flows between pairs of APEC countries: A non-parametric regression tree approach. *Journal of Air Transport Management, 20*, 4-6. doi:http://doi.org/10.1016/j.jairtraman.2011.04.001

Chang, L.-Y. (2014). Analysis of bilateral air passenger flows: A non-parametric multivariate adaptive regression spline approach. *Journal of Air Transport Management, 34*, 123-130. doi:http://doi.org/10.1016/j.jairtraman.2013.09.003

Chang, L.-Y., & Sun, P.-Y. (2012). Stated-choice analysis of willingness to pay for low cost carrier services. *Journal of Air Transport Management, 20*, 15-17.

Chen, H.-T., & Chao, C.-C. (2015). Airline choice by passengers from Taiwan and China: A case study of outgoing passengers from Kaohsiung International Airport. *Journal of Air Transport Management, 49*, 53-63. doi:10.1016/j.jairtraman.2015.08.002

Cherchi, E., & Hensher, D. A. (2015). Workshop synthesis: Stated preference surveys and experimental design, an audit of the journey so far and future research perspectives. *Transportation Research Procedia, 11*, 154-164.

ChoiceMetrics. (2018). *Ngene 1.2 USER MANUAL & REFERENCE GUIDE*. Retrieved from http://www.choice-metrics.com/

Christian, W., & Braden, W. (1966). Rural Migration and the Gravity Model. *Rural Sociology, 31*(1), 73.

Cohen, M. F., Chen, S. E., Wallace, J. R., & Greenberg, D. P. (1988). A progressive refinement approach to fast radiosity image generation. *ACM SIGGRAPH computer graphics, 22*(4), 75-84.

Collins, A. T., Bliemer, M. C., & Rose, J. M. (2014). *Constrained stated choice experimental designs*. Retrieved from University of Sydney, Australia: http://hdl.handle.net/2123/19274

Congdon, P. (2000). A Bayesian Approach to Prediction Using the Gravity Model, with an Application to Patient Flow Modeling. *Geographical Analysis, 32*(3), 205-224. doi:10.1111/j.1538-4632.2000.tb00425.x

Cook, R. D., & Nachtrheim, C. J. (1980). A comparison of algorithms for constructing exact

D-optimal designs. *Technometrics, 22*(3), 315-324.

Croley, T. E., & Hartmann, H. C. (1985). Resolving Thiessen polygons. *Journal of Hydrology, 76*(3), 363-379. doi:http://dx.doi.org/10.1016/0022-1694(85)90143-X

D'Alfonso, T., Jiang, C., & Bracaglia, V. (2015). Would competition between air transport and high-speed rail benefit environment and social welfare? *Transportation Research Part B: Methodological, 74*, 118-137.

De Vany, A. (1974). The revealed value of time in air travel. *The Review of Economics and Statistics*, 77-82.

Deloitte Access Economics. (2014a). *2015 - 2025 Western Australian Resources Sector Outlook*. Retrieved from

Deloitte Access Economics. (2014b) Western Australian Resources Sector Outlook.

Department of Infrastructure and Regional Development. (2017). *Australian Domestic Aviation Activity*. Retrieved from https://bitre.gov.au/publications/ongoing/domestic_airline_activity-monthly_publications.aspx

Department of Infrastructure and Regional Development. (2003). *Regional Public Transport In Australia: Long-Distance Services, Trends And Projections* (1440-9707). Retrieved from https://bitre.gov.au/publications/2003/files/wp_051.pdf

Department of Infrastructure and Regional Development. (2017). Airport traffic data. Retrieved from https://bitre.gov.au/publications/ongoing/airport_traffic_data.aspx

Department of Infrastructure Regional Development and Cities. (2018). *Australian Domestic Aviation Activity Monthly Publications*. Retrieved from Canberra:

Department of Mines and Petroleum. (2017). Mines and mineral deposits (MINEDEX) Retrieved from http://www.dmp.wa.gov.au/Mines-and-mineral-deposits-1502.aspx. Retrieved March 2017 http://www.dmp.wa.gov.au/Mines-and-mineral-deposits-1502.aspx

Department of the Prime Minister and Cabinet. (2013). *Best practice regulation handbook*: Australian Government, Canberra.

Department of Transport. (2014). *Review of Regulated Regular Public Transport Air Routes in Western Australia Position paper – for public comment*. Retrieved from https://www.transport.wa.gov.au/mediaFiles/aviation/AV_P_RPT_PositionPaper_AirRoutes.pdf

Department of Transport. (2015a). *Review of Regulated Regular Public Transport Air Routes in Western Australia*. Retrieved from http://www.transport.wa.gov.au/mediaFiles/aviation/AV_P_ReviewRegulatedAirRoutesRPT.pdf

Department of Transport. (2015b). Review of Regulated Regular Public Transport Air Routes in Western Australia Final Public Report 2015. Retrieved from http://www.transport.wa.gov.au/mediaFiles/aviation/AV_P_RevRegRPTAirRoutesWAFinalRpt2015.pdf

Department of Transport. (2015c). *Western Australian State Aviation Strategy*. Retrieved from Perth: http://www.transport.wa.gov.au/mediaFiles/aviation/AV_P_State_Aviation_Strategy2015.pdf

Department of Transport. (2015d). *Western Australian State Aviation Strategy*. Retrieved from http://www.transport.wa.gov.au/mediaFiles/aviation/AV_P_State_Aviation_Strategy2015.pdf

Dibb, S. (1998). Market segmentation: strategies for success. *Marketing Intelligence & Planning, 16*(7), 394-406.

Ding, C., Wang, D., Liu, C., Zhang, Y., & Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice, 100*, 65-80. doi:http://doi.org/10.1016/j.tra.2017.04.008

Dobruszkes, F., Lennert, M., & Van Hamme, G. (2011). An analysis of the determinants of air traffic volume for European metropolitan areas. *Journal of Transport Geography, 19*(4), 755-762. doi:http://dx.doi.org/10.1016/j.jtrangeo.2010.09.003

Doganis, R. (2009). *Flying off course: airline economics and marketing*: Routledge.

Douglas, E. J. (1993). *Airline competition and strategy in Australia*: Bond University, School of Business.

Drakos, K., Kyriazidou, E., & Polycarpou, I. (2014). A dynamic gravity model for global bilateral investment holdings *Macroeconomic Analysis and International Finance* (pp. 125-152): Emerald Group Publishing Limited.

Ekinci, Y., Uray, N., Ülengin, F., & Duran, C. J. T. (2018). A segmentation based analysis for measuring customer satisfaction in maritime transportation. 33(1), 104-118.

Endo, N. (2007). International trade in air transport services: Penetration of foreign airlines into Japan under the bilateral aviation policies of the US and Japan. *Journal of Air Transport Management, 13*(5), 285-292.

Espino, R., Martín, J. C., & Román, C. (2008). Analyzing the effect of preference heterogeneity on willingness to pay for improving service quality in an airline choice context. *Transportation Research Part E: Logistics and Transportation Review, 44*(4), 593-606. doi:https://doi.org/10.1016/j.tre.2007.05.007

Esri. (2017). NLOGIT reference guide : version 5.0: Redlands, CA: Esri. Retrieved from http://resources.arcgis.com/en/help/main/10.2/index.html

Ferrini, S., & Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of environmental economics and management, 53*(3), 342-363.

Forsyth, P. (2002). Privatisation and regulation of Australian and New Zealand airports. *Journal of Air Transport Management, 8*(1), 19-28. doi:http://dx.doi.org/10.1016/S0969-6997(01)00036-9

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association, 97*(458), 611-631.

Freathy, P., & O'Connell, F. (2012). Spending time, spending money: passenger segmentation in an international airport. *The International Review of Retail, Distribution and Consumer Research, 22*(4), 397-416.

Freathy, P., & O'Connell, F. (2000). Market segmentation in the European airport sector. *Marketing Intelligence & Planning, 18*(3), 102-111.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 1-67.

Gaggero, A. A., & Piga, C. A. (2010). Airline competition in the British Isles. *Transportation Research Part E: Logistics and Transportation Review, 46*(2), 270-279. doi:http://dx.doi.org/10.1016/j.tre.2009.08.009

Garrow, L. A. (2010). *Discrete Choice Modelling and Air Travel Demand : Theory and Applications*. Farnham, UNITED KINGDOM: Routledge.

Gauss. (2011). Gauss 8.0. Black Diamond: Aptech Systems, Inc.,.

Gautam, G. (2002). Competition and Regulation in the Airline Industry. Retrieved from http://www.frbsf.org/economic-research/publications/economic-letter/2002/january/competition-and-regulation-in-the-airline-industry

George, K., Han, E.-H., & Kumar, V. (1999). CHAMELEON: a hierarchical clustering algorithm using dynamic modeling. *IEEE computer, 27*(3), 329-341.

Geraldton Airport. (2018). City of Greater Geraldton, Geraldton Airport. Retrieved from https://airport.cgg.wa.gov.au/

Ghobrial, A. (1992). An aggregate demand model for domestic airlines. *Journal of Advanced Transportation, 26*(2), 105-119.

Givoni, M., & Rietveld, P. (2009). Airline's choice of aircraft size – Explanations and implications. *Transportation Research Part A: Policy and Practice, 43*(5), 500-510. doi:http://dx.doi.org/10.1016/j.tra.2009.01.001

Green, P. E. (1977). A new approach to market segmentation. *Business Horizons, 20*(1), 61-73.

Greene, W., Harris, M. N., Hollingsworth, B., & Maitra, P. (2014). A latent class model for obesity. *Economics Letters, 123*(1), 1-5. doi:https://doi.org/10.1016/j.econlet.2014.01.004

Greene, W. H. (2012). NLOGIT reference guide : version 5.0: Plainview, NY : Econometric Software, Inc., 2002. Retrieved from https://search.library.wisc.edu/catalog/999935175002121

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological, 37*(8), 681-698.

Greene, W. H., & Hensher, D. A. (2013). Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics, 45*(14), 1897-1902.

Grosche, T., Rothlauf, F., & Heinzl, A. (2007). Gravity models for airline passenger volume estimation. *Journal of Air Transport Management, 13*(4), 175-183. doi:http://doi.org/10.1016/j.jairtraman.2007.02.001

Grylls, B. (2016). *Air route connectivity and north west of Western Australia* Retrieved from

Guha, S., Rastogi, R., & Shim, K. (1998). *CURE: an efficient clustering algorithm for large databases.* Paper presented at the ACM Sigmod Record.

Guimerà, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America, 102*(22), 7794-7799. doi:10.1073/pnas.0407994102

Gumbel, E. J. (2012). *Statistics of extremes*: Courier Corporation.

Günster, A., & van Dijk, M. (2016). The impact of European antitrust policy: Evidence from the stock market. *International Review of Law and Economics, 46*, 20-33. doi:http://dx.doi.org/10.1016/j.irle.2015.12.001

Hall, E. (2015). Tourism can be the new gold mine in WA. Retrieved from https://au.news.yahoo.com/thewest/a/26805171/tourism-can-be-the-new-gold-mine-in-wa/#page1

Harrison, A., Popovic, V., & Kraal, B. (2015). A new model for airport passenger segmentation. *Journal of Vacation Marketing, 21*(3), 237-250.

Hassan, S. S., & Craft, S. H. (2005). Linking global market segmentation decisions with strategic positioning options. *Journal of Consumer Marketing, 22*(2), 81-89.

Hazledine, T. (2009). Border effects for domestic and international Canadian passenger air travel. *Journal of Air Transport Management, 15*(1), 7-13. doi:http://doi.org/10.1016/j.jairtraman.2008.09.007

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, 271-320.

Hensher, D. A. (1982). Functional measurement, individual preference and discrete-choice modelling: theory and application. *Journal of Economic Psychology, 2*(4), 323-335.

Hensher, D. A. (1994). Stated preference analysis of travel choices: the state of practice. *Transportation, 21*(2), 107-133.

Hensher, D. A., & Barnard, P. O. (1988). The orthogonality issue in stated choice designs.

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation, 30*(2), 133-176.

Hensher, D. A., & Louviere, J. J. (1983). Identifying individual preferences for international air fares: an application of functional measurement theory. *Journal of transport economics and policy*, 225-245.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015a). *Applied Choice Analysis* (2 ed.). Cambridge: Cambridge University Press.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015b). Experimental design and choice experiments *Applied Choice Analysis* (2 ed., pp. 189-319). Cambridge: Cambridge University Press.

Hess, S., Adler, T., & Polak, J. W. (2007). Modelling airport and airline choice behaviour with the use of stated preference survey data. *Transportation Research Part E: Logistics and Transportation Review, 43*(3), 221-233. doi:https://doi.org/10.1016/j.tre.2006.10.002

Hess, S., Ben-Akiva, M., Gopinath, D., & Walker, J. (2009). *Advantages of latent class choice models over continuous mixed logit models.* Paper presented at the Proceedings of the 12th International Conference on Travel Behaviour Research, Jaipur, India. International Association for Travel Behaviour Research, Washington, DC, USA.

Hess, S., & Polak, J. W. (2006a). Airport, airline and access mode choice in the San Francisco Bay area. *Papers in Regional Science, 85*(4), 543-567.

Hess, S., & Polak, J. W. (2006b). Exploring the potential for cross-nesting structures in airport-choice analysis: A case-study of the Greater London area. *Transportation Research Part E: Logistics and Transportation Review, 42*(2), 63-81.

doi:https://doi.org/10.1016/j.tre.2005.09.001

Hollywood, L., Armstrong, G., & Durkin, M. (2007). Using behavioural and motivational thinking in food segmentation. *International Journal of Retail & Distribution Management, 35*(9), 691-702.

Horowitz, J. L. (1991). Reconsidering the multinomial probit model. *Transportation Research Part B: Methodological, 25*(6), 433-438.

Hruschka, H., Fettes, W., & Probst, M. (2004). Market segmentation by maximum likelihood clustering using choice elasticities. *European Journal of Operational Research, 154*(3), 779-786. doi:https://doi.org/10.1016/S0377-2217(02)00807-X

Hsiao, C.-Y., & Hansen, M. (2011). A passenger demand model for air transportation in a hub-and-spoke network. *Transportation Research Part E: Logistics and Transportation Review, 47*(6), 1112-1125. doi:10.1016/j.tre.2011.05.012

Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of marketing research, 33*(3), 307-317.

Hwang, C.-C., & Shiao, G.-C. (2011). Analyzing air cargo flows of international routes: an empirical study of Taiwan Taoyuan International Airport. *Journal of Transport Geography, 19*(4), 738-744.

Iles, R. A., & Rose, J. M. (2014). Stated Choice design comparison in a developing country: recall and attribute nonattendance. *Health economics review, 4*(1), 25.

Inoue, G., Ono, M., Uehara, K., & Isono, F. (2015). Stated-preference analysis to estimate the domestic transport demand following the future entry of LCCs and the inauguration of the Linear Chuo Shinkansen in Japan. *Journal of Air Transport Management, 47*, 199-217.

Jacques, C., Manaugh, K., & El-Geneidy, A. M. (2013). Rescuing the captive [mode] user: an alternative approach to transport market segmentation. *Transportation, 40*(3), 625-645.

Jiang, C., & Zhang, A. (2016a). Airline network choice and market coverage under high-speed rail competition. *Transportation Research Part A: Policy and Practice*. doi:http://dx.doi.org/10.1016/j.tra.2016.06.008

Jiang, C., & Zhang, A. (2016b). Airline network choice and market coverage under high-speed rail competition. *Transportation Research Part A: Policy and Practice, 92*, 248-260.

Jiao, P., Liu, M., & Guo, J. (2015). Integrated model of joint residence-workplace location choice and commute behavior using latent class and mixed logit methods. *Discrete Dynamics in Nature and Society, 2015*.

Johansson, M. V., Heldt, T., & Johansson, P. (2006). The effects of attitudes and personality traits on mode choice. *Transportation Research Part A: Policy and Practice, 40*(6), 507-525.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., . . . Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health, 16*(1), 3-13.

Jorge-Calderón, J. D. (1997). A demand model for scheduled airline services on international European routes. *Journal of Air Transport Management, 3*(1), 23-35. doi:http://dx.doi.org/10.1016/S0969-6997(97)82789-5

Jovicic, G., & Hansen, C. O. (2003). A passenger travel demand model for Copenhagen. *Transportation Research Part A: Policy and Practice, 37*(4), 333-349. doi:https://doi.org/10.1016/S0965-8564(02)00019-8

Jung, S.-Y., & Yoo, K.-E. (2014). Passenger airline choice behavior for domestic short-haul travel in South Korea. *Journal of Air Transport Management, 38*, 43-47. doi:10.1016/j.jairtraman.2013.12.017

Kabir, M., & Salim, R. (2016). Is trade in electrical and electronic products sensitive to IPR protection? Evidence from China's exports. *Applied Economics, 48*(21), 1991-2005.

Kabir, M., Salim, R., & Al-Mawali, N. (2017). The gravity model and trade flows: Recent developments in econometric modeling and empirical evidence. *Economic Analysis and Policy, 56*, 60-71. doi:https://doi.org/10.1016/j.eap.2017.08.005

Kamakura, W. A., & Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of marketing research, 26*(4), 379-390.

Kamthania, D., Pawa, A., & Madhavan, S. S. (2018). Market Segmentation Analysis and Visualization Using K-Mode Clustering Algorithm for E-Commerce Business. *Journal of computing and information technology, 26*(1), 57-68.

Kanafani, A. (1983). *Transportation demand analysis*: McGraw-Hill College.

Karatha Airport. (2018). Visitor Guide. Retrieved from http://www.karrathaairport.com.au/visitor-guide/

Karemera, D., Oguledo, V. I., & Davis, B. (2000). A gravity model analysis of international migration to North America. *Applied Economics, 32*(13), 1745-1755. doi:10.1080/000368400421093

Kaufman, L., & Rdusseeun, P. J. (1987). Clustering by means of medoids.

Keane, M. P. (1997). Current Issues in Discrete Choice Modeling. *Marketing Letters, 8*(3), 307-322. Retrieved from http://www.jstor.org/stable/40216456

Kessels, R., Goos, P., & Vandebroek, M. (2006). A comparison of criteria to design efficient choice experiments. *Journal of marketing research, 43*(3), 409-419.

Kieu, L. M., Ou, Y., & Cai, C. (2018). Large-scale transit market segmentation with spatial-behavioural features. Transportation Research Part C: Emerging Technologies, 90, 97-113. doi:https://doi.org/10.1016/j.trc.2018.03.003

Kim, H.-C., Nicholson, A., & Kusumastuti, D. (2017). Analysing freight shippers' mode choice preference heterogeneity using latent class modelling. *Transportation Research Procedia, 25*, 1109-1125.

Kishor, D. R., & Venkateswarlu, N. (2016). A novel hybridization of expectation-maximization and K-means algorithms for better clustering performance. *International Journal of Ambient Computing and Intelligence (IJACI), 7*(2), 47-74.

Klojgaard, M. E., Bech, M., & Sogaard, R. (2012). Designing a Stated Choice Experiment: The Value of a Qualitative Process. *Journal of Choice Modelling, 5*(2), 1-18. Retrieved from <Go to ISI>://WOS:000209157600001

Koo, T. T., Wu, C.-L. R., & Dwyer, L. (2010). Ground travel mode choices of air arrivals at regional destinations: The significance of tourism attributes and destination contexts. *Research in Transportation Economics, 26*(1), 44-53.

Kotler, P. (2009). *Marketing management: A south Asian perspective*: Pearson Education India.

Kuhfeld, W. F. (2010). *Marketing Research Methods in SAS*: Citeseer.

Lai, K., Twine, N., Aidan, O. B., & Guo, Y. (2018). Artificial Intelligence and Machine Learning in Bioinformatics. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, 55*, 272.

Lancsar, E., & Louviere, J. (2006). Deleting 'irrational'responses from discrete choice experiments: a case of investigating or imposing preferences? *Health economics, 15*(8), 797-811.

Lancsar, E., & Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics, 26*(8), 661-677.

Lapparent, M. d., Frei, A., & Axhausen, K. W. (2009). Choice of mode for long distance travel: current SP-based models from three European countries. *Arbeitsberichte Verkehrs-und Raumplanung, 579*.

Laughlin, J. L., & Taylor, C. R. (1991). An approach to industrial market segmentation. *Industrial Marketing Management, 20*(2), 127-136.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*: Houghton Mifflin Co.

Lee, B. J., Fujiwara, A., Zhang, J., & Sugie, Y. (2003). *Analysis of mode choice behaviours based on latent class models.* Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne.

Lee, J.-K., Yoo, K.-E., & Song, K.-H. (2016). A study on travelers' transport mode choice behavior using the mixed logit model: A case study of the Seoul-Jeju route. *Journal of Air Transport Management, 56*, 131-137. doi:https://doi.org/10.1016/j.jairtraman.2016.04.020

Leung, S. C., & Lai, K. (2002). Multiple objective decision-making in the mode choice problem: A goal-programming approach. *International Journal of Systems Science, 33*(1), 35-43.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory, 28*(2), 129-137.

Long, W. H. (1970). AIR TRAVEL, SPATIAL STRUCTURE, AND GRAVITY MODELS. *Annals of Regional Science, 4*(2), 97. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=5820944&site=ehost-live

Long, W. H., & Uris, R. B. (1970). Distance, intervening opportunities, city hierarchy and air travel. *The Annals of Regional Science, 5*(1), 152-161. doi:10.1007/bf01288120

Louviere, J. J., & Hensher, D. A. (1982). On the design and analysis of simulated choice or allocation experiments in travel choice modelling. *Transportation research record, 890*(1982), 11-17.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and applications*: Cambridge university press.

Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of marketing research*, 350-367.

Lu, J., Lv, C., Yang, Z., & Hansen, M. J. S. (2019). Market Segmentation of New Gateway Airports Incorporating Passengers' Curiosity. 11(24), 7042.

Ma, T.-Y., Gerber, P., Carpentier, S., & Klein, S. (2015). Mode choice with latent preference heterogeneity: a case study for employees of the EU institutions in Luxembourg. *Transportmetrica A: transport science*, *11*(5), 441-463.

Maimon, O., & Rokach, L. (2005). Data mining and knowledge discovery handbook.

Marrocu, E., & Paci, R. (2013). Different tourists to different destinations. Evidence from spatial interaction models. *Tourism Management, 39*, 71-83.

Marshall, D., Bridges, J. F., Hauber, B., Cameron, R., Donnalley, L., Fyie, K., & Johnson, F. R. (2010). Conjoint analysis applications in health—how are studies being designed and reported? *The Patient: Patient-Centered Outcomes Research*, *3*(4), 249-256.

Mason, K. J., & Gray, R. (1995). Short haul business travel in the European Union: a segmentation profile. *Journal of Air Transport Management, 2*(3-4), 197-205.

Massiani, J., Danielis, R., & Marcucci, E. (2007). THE HETEROGENEITY IN SHIPPER'S VALUE OF TIME, RESULTS FROM AN SP EXPERIMENT USING MIXED LOGIT AND LATENT CLASS. *Pomorstvo/Journal of Maritime Studies, 21*(2).

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of applied Econometrics*, *15*(5), 447-470.

McLachlan, G., & Chang, S. (2004). Mixture modelling for cluster analysis. *Statistical methods in medical research, 13*(5), 347-361.

Mehdizadeh, M., Nordfjaern, T., & Mamdoohi, A. J. T. (2018). The role of socio-economic, built environment and psychological factors in parental mode choice for their children in an Iranian setting. 45(2), 523-543.

Melnykov, V., & Maitra, R. (2010). Finite mixture models and model-based clustering. *Statistics Surveys, 4*, 80-116.

Mikkonen, K., & Luoma, M. (1999). The parameters of the gravity model are changing – how and why? *Journal of Transport Geography, 7*(4), 277-283. doi:http://dx.doi.org/10.1016/S0966-6923(99)00024-1

Molesworth, B. R., & Koo, T. T. (2016). The influence of attitude towards individuals' choice for a remotely piloted commercial flight: A latent class logit approach. *Transportation Research Part C: Emerging Technologies, 71*, 51-62.

Monchambert, G. (2020). Why do (or don't) people carpool for long distance trips? A discrete choice experiment in France. *Transportation Research Part A: Policy and Practice, 132*, 911-931. doi:https://doi.org/10.1016/j.tra.2019.12.033

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., & Foley, J. A. (2012). Closing yield gaps through nutrient and water management. *Nature, 490*(7419), 254-257.

Neal, R. M., & Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants *Learning in graphical models* (pp. 355-368): Springer.

Nicolas, G., Apolloni, A., Coste, C., Wint, G. W., Lancelot, R., & Gilbert, M. J. P. o. (2018). Predictive gravity models of livestock mobility in Mauritania: The effects of supply, demand and cultural factors. 13(7), e0199547.

Nijkamp, P. (1997). Gravity models of spatial interaction behavior (Book). *Journal of Economic Literature, 35*(3), 1435-1436. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9710242806&site=

ehost-live

Norusis, M. J. (2010). *PASW statistics 18 guide to data analysis*: Prentice Hall Press.

O'Kelly, M. E., Song, W., & Shen, G. (1995). New Estimates of Gravitational Attraction by Linear Programming. *Geographical Analysis, 27*(4), 271-285. doi:10.1111/j.1538-4632.1995.tb00911.x

Orme, B. (1998). Sample size issues for conjoint analysis studies. *Sawthooth Software Research paper Series Squim, WA, USA: Sawthooth Software Inc*.

Pearmain, D., & Kroes, E. P. (1990). Stated preference techniques: a guide to practice.

Pels, E., Nijkamp, P., & Rietveld, P. (2000). Airport and Airline Competition for Passengers Departing from a Large Metropolitan Area. *Journal of Urban Economics, 48*(1), 29-45. doi:http://dx.doi.org/10.1006/juec.1999.2156

Pels, E., Nijkamp, P., & Rietveld, P. (2003). Access to and competition between airports: a case study for the San Francisco Bay area. *Transportation Research Part A: Policy and Practice, 37*(1), 71-83. doi:http://dx.doi.org/10.1016/S0965-8564(02)00007-1

Pels, E., Njegovan, N., & Behrens, C. (2009). Low-cost airlines and airport competition. *Transportation Research Part E: Logistics and Transportation Review, 45*(2), 335-344. doi:10.1016/j.tre.2008.09.005

Peyhardi, D. J. (2020). Robustness of Student link function in multinomial choice models. *Journal of Choice Modelling*, 100228. doi:https://doi.org/10.1016/j.jocm.2020.100228

Poot, J., Alimi, O., Cameron, M. P., & Maré, D. C. (2016). The gravity model of migration: the successful comeback of an ageing superstar in regional science.

Psaraki, V., & Abacoumkin, C. (2002). Access mode choice for relocated airports: the new Athens International Airport. *Journal of Air Transport Management, 8*(2), 89-98.

Puckett, S. M., & Hensher, D. A. (2009). Revealing the extent of process heterogeneity in choice analysis: An empirical assessment. *Transportation Research Part a-Policy and Practice, 43*(2), 117-126. Retrieved from <Go to ISI>://WOS:000263219600002

Qiao, Y., Huang, Y., Yang, F., Zhang, M., & Chen, L. (2016). Empirical study of travel mode forecasting improvement for the combined revealed preference/stated preference data–based discrete choice model. *Advances in Mechanical Engineering, 8*(1), 1687814015624836.

Quiggin, J. (1997). Evaluating airline deregulation in Australia. *Australian Economic Review, 30*(1), 45-56.

Quinn, L., Hines, T., & Bennison, D. (2007). Making sense of market segmentation: a fashion retailing case. *European Journal of Marketing, 41*(5/6), 439-465.

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org/

Rao, C. P., & Wang, Z. (1995). Evaluating alternative segmentation strategies in standard industrial markets. *European Journal of Marketing, 29*(2), 58-75.

Regional Aviation Association of Australia. (2013). *Western Australia Draft State Aviation Strategy*. Retrieved from https://www.raaa.com.au/wp-content/uploads/2016/12/RAAA-Response-WA-Aviation-Strategy-Jan2013.pdf

Reid, R. D., & Bojanic, D. C. (2009). *Hospitality marketing management*: John Wiley and Sons.

Rengaraju, V., & Arasan, V. T. (1992). Modeling for air travel demand. *Journal of Transportation Engineering, 118*(3), 371-380.

Revelt, D., & Train, K. (1998). Incentives for appliance efficiency in a competitive energy environment: random parameters logit models of households' choices. *Review of Economics and Statistics, 80*(4), 647-657.

Rodrigue, J.-P., Comtois, C., & Slack, B. (2013). *The geography of transport systems*: Routledge.

Román, C., Arencibia, A. I., & Feo-Valero, M. (2017). A latent class model with attribute cut-offs to analyze modal choice for freight transport. *Transportation Research Part A: Policy and Practice, 102*, 212-227.

Rose, J., Collins, A., Bliemer, M., & Hensher, D. (2009). Ngene 1.0 stated choice experiment design software. *University of Sydney*.

Rose, J. M., & Bliemer, M. C. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews, 29*(5), 587-617.

Rose, J. M., & Bliemer, M. C. (2013). Sample size requirements for stated choice experiments. *Transportation, 40*(5), 1021-1041.

Rose, J. M., Bliemer, M. C., Hensher, D. A., & Collins, A. T. (2008). Designing efficient stated choice experiments in the presence of reference alternatives. *Transportation Research Part B: Methodological, 42*(4), 395-406.

Ruto, E., Garrod, G., & Scarpa, R. (2008). Valuing animal genetic resources: a choice modeling application to indigenous cattle in Kenya. *Agricultural economics, 38*(1), 89-98.

Sánchez-Fernández, R., Iniesta-Bonillo, M. Á., & Cervera-Taulet, A. (2019). Exploring the concept of perceived sustainability at tourist destinations: a market segmentation approach. *Journal of Travel & Tourism Marketing, 36*(2), 176-190. doi:10.1080/10548408.2018.1505579

Sandell, D. D. (1978). Deregulation-Has it Finally Arrived-The Airline Deregulation Act of 1978. *J. Air L. & Com., 44*, 799.

Sandor, Z., & Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of marketing research, 38*(4), 430-444.

Sándor, Z., & Wedel, M. (2002). Profile construction in experimental choice designs for mixed logit models. *Marketing Science, 21*(4), 455-475.

Sarabia, F. J. (1996). Model for market segments evaluation and selection. *European Journal of Marketing, 30*(4), 58-74.

SAS Institute. (2011). The SAS system for Windows: SAS Institute Inc.

Scarpa, R., & Thiene, M. (2005). Destination choice models for rock climbing in the Northeastern Alps: a latent-class approach based on intensity of preferences. *Land economics, 81*(3), 426-444.

Scarpel, R. A. (2013). Forecasting air passengers at São Paulo International Airport using a mixture of local experts model. *Journal of Air Transport Management, 26*, 35-39.

Schipper, Y., Rietveld, P., & Nijkamp, P. (2003). Airline deregulation and external costs: a welfare analysis. *Transportation Research Part B: Methodological, 37*(8), 699-718. doi:http://dx.doi.org/10.1016/S0191-2615(02)00047-4

Seelhorst, M., & Liu, Y. (2015). Latent air travel preferences: Understanding the role of frequent flyer programs on itinerary choice. *Transportation Research Part A: Policy and Practice, 80*, 49-61.

Sener, I. N., Eluru, N., & Bhat, C. R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation, 36*(5), 511-539. Retrieved from <Go to ISI>://WOS:000268257200003

Sepp, P., & Aiello, T. (2018). *Airline Deregulation At 40*. National Taxpayers Union Retrieved from https://www.ntu.org/publications/detail/airline-deregulation-at-40.

Shang, B., & Zhang, X. N. (2013). *Study of Travel Mode Choice Behavior Based on Nested Logit Model*. Paper presented at the Applied Mechanics and Materials.

Sharma, N., Bajpai, A., & Litoriya, M. R. (2012). Comparison the various clustering algorithms of weka tools. *facilities, 4*(7), 78-80.

Shaw, S. (2016). *Airline marketing and management*: Routledge.

Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics, 41*(22), 2915-2924.

Shen, H., Zou, B., Lin, J., & Liu, P. (2020). Modeling travel mode choice of young people with differentiated E-hailing ride services in Nanjing China. Transportation Research Part D: Transport and Environment, 78, 102216. doi:https://doi.org/10.1016/j.trd.2019.102216

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of marketing, 21*(1), 3-8.

Srisaeng, P., Baxter, G. S., & Wild, G. (2015). Forecasting demand for low cost carriers in Australia using an artificial neural network approach. *Aviation, 19*(2), 90-103.

Starkie, D. (2002). Airport regulation and competition. *Journal of Air Transport Management, 8*(1), 63-72. doi:http://dx.doi.org/10.1016/S0969-6997(01)00015-1

Tan, P.-N. (2018). *Introduction to data mining*: Pearson Education India.

Teichert, T., Shehu, E., & von Wartburg, I. (2008). Customer segmentation revisited: The case of the airline industry. *Transportation Research Part A: Policy and Practice, 42*(1), 227-242.

Theis, G., Adler, T., Clarke, J.-P., & Ben-Akiva, M. (2006). Risk aversion to short connections in airline itinerary choice. *Transportation research record, 1951*(1), 28-36.

Tourism Research Australia. (2017). TRA online.

Tourism Western Australia. (2012). STATE GOVERNMENT STRATEGY FOR TOURISM IN WESTERN AUSTRALIA 2020. Retrieved from http://www.tourism.wa.gov.au/Publications%20Library/About%20Us/State%20Government%20Strategy%20for%20Tourism%20in%20WA%202020%20-%20Detailed.pdf

Train, K. E. (2009). *Discrete choice methods with simulation*: Cambridge university press.

Valdes, V. (2015). Determinants of air travel demand in Middle Income Countries. *Journal of Air Transport Management, 42*, 75-84.

Van Can, V. (2013). Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model. *Transportation Research Part A: Policy and Practice, 49*, 149-159.

Van Rossum, G., & Drake Jr, F. L. (1995). *Python tutorial*: Centrum voor Wiskunde en Informatica Amsterdam.

Verleger Jr, P. K. (1972). Models of the demand for air transportation. *Bell Journal of Economics, 3*(2), 437-457.

Vij, A., Carrel, A., & Walker, J. L. (2013). Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transportation Research Part A: Policy and Practice*, *54*, 164-178. doi:https://doi.org/10.1016/j.tra.2013.07.008

Walker, J. L., & Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems, 9*(1), 77-101.

Walker, J. L., Wang, Y., Thorhauge, M., & Ben-Akiva, M. (2018). D-efficient or deficient? A robustness analysis of stated choice experimental designs. *Theory and Decision, 84*(2), 215-238. doi:10.1007/s11238-017-9647-3

Wang, Y., Li, L., Wang, L., Moore, A., Staley, S., & Li, Z. (2014). Modeling traveler mode choice behavior of a new high-speed rail corridor in China. *Transportation Planning and Technology, 37*(5), 466-483.

Wedel, M., & Kamakura, W. A. (2002). Introduction to the special issue on market segmentation. *Intern. J. of Research in Marketing, 19*, 181-183.

Wedel, M., & Kamakura, W. A. (2012). *Market segmentation: Conceptual and methodological foundations* (Vol. 8): Springer Science & Business Media.

Wei, W., & Hansen, M. (2006). An aggregate demand model for air passenger traffic in the hub-and-spoke network. *Transportation Research Part A: Policy and Practice, 40*(10), 841-851. doi:http://doi.org/10.1016/j.tra.2005.12.012

Wei, Y. (2014). The Network Effects of Air-Travel Demand. *PIER Working Paper Archive 14-027*. Retrieved from https://ideas.repec.org/p/pen/papers/14-027.html

Wen, C.-H., & Lai, S.-C. (2010). Latent class models of international air carrier choice. *Transportation Research Part E: Logistics and Transportation Review, 46*(2), 211-221.

Wen, C.-H., Lai, S.-C., & Yeh, W.-Y. (2008). Segmentation and positioning analysis for international air travel market: Taipei-to-Tokyo route. *Transportation research record, 2052*(1), 46-53.

Wen, C.-H., Wang, W.-C., & Fu, C. (2012). Latent class nested logit model for analyzing high-speed rail access mode choice. *Transportation Research Part E: Logistics and Transportation Review*, *48*(2), 545-554.

Williams, C. (2015). *Australia's commuting distance: cities and regions*. Retrieved from Canberra Department of Infrastructure and Regional Development:

Wind, Y. (1978). Issues and advances in segmentation research. *Journal of marketing research, 15*(3), 317-337.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Wu, C.-L., & So, T. H. (2018). On the flight choice behaviour of business-purpose passengers in the Australian domestic air market. *Journal of Air Transport Management, 72*, 56-67.

Wu, J., Yang, M., Sun, S., & Zhao, J. (2018). Modeling travel mode choices in connection to metro stations by mixed logit models: a case study in Nanjing, China. *30*(5), 549-561.

Wu, R.-S., & Chou, P.-H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications, 10*(3), 331-341.

Xia, J. C., Evans, F. H., Spilsbury, K., Ciesielski, V., Arrowsmith, C., & Wright, G. (2010). Market segments based on the dominant movement patterns of tourists. *Tourism Management, 31*(4), 464-469.

Xu, R., & Wunsch, D. C. (2005). Survey of clustering algorithms.

Yang, Y., Wang, C., Liu, W., & Zhou, P. (2018). Understanding the determinants of travel mode choice of residents and its carbon mitigation potential. *Energy Policy, 115*, 486-493. doi:https://doi.org/10.1016/j.enpol.2018.01.033

Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E., & Ruzzo, W. L. (2001). Model-based clustering and data transformations for gene expression data. *Bioinformatics, 17*(10), 977-987.

Zhang, Y., Sampaio, B., Fu, X., & Huang, Z. (2018). Pricing dynamics between airline groups with dual-brand services: The case of the Australian domestic market. *Transportation Research Part A: Policy and Practice, 112*, 46-59.

Zhang, Y., & Zhang, A. (2016). Determinants of Air Passenger Flows in China and Gravity Model: Deregulation, LCCs, and High-speed Rail. *Journal of Transport Economics and Policy (JTEP), 50*(3), 287-303. Retrieved from http://www.ingentaconnect.com/content/lse/jtep/2016/00000050/00000003/art00005

Zhou, H., Xia, J., Luo, Q., Nikolova, G., Sun, J., Hughes, B., . . . Falkmer, T. (2018). Investigating the impact of catchment areas of airports on estimating air travel demand: A case study of regional Western Australia. *Journal of Air Transport Management, 70*, 91-103. doi:https://doi.org/10.1016/j.jairtraman.2018.05.001

Zhou, H., Xia, J., Norman, R., Hughes, B., Nikolova, G., Kelobonye, K., . . . Falkmer, T. (2019). Do air passengers behave differently to other regional travellers?: A travel mode choice model investigation. *Journal of Air Transport Management, 79*, 101682. doi:https://doi.org/10.1016/j.jairtraman.2019.101682

# APPENDIX A PERMISSION FROM CO-AUTHORS

## PUBLICATIONS BASED ON THIS THESIS

**PhD student name: Heng Zhou** **Student ID: 17287764**

Some of the work presented in this thesis has already been published or submitted by the thesis author and the paper co-authors.

Journal papers (peer-reviewed and fully refereed)

1. Zhou, H., Xia, J., Luo, Q., Nikolova, G., Sun, J., Hughes, B., . . . Falkmer, T. (2018). Investigating the impact of catchment areas of airports on estimating air travel demand: A case study of regional Western Australia. Journal of Air Transport Management, 70, 91-103. doi:https://doi.org/10.1016/j.jairtraman.2018.05.001 *[Published]*
2. Zhou, H., Norman, R., Kelobonye, K., Xia, J., Hughes, B., Nikolova, G., . . . Falkmer, T. (2020). Market Segmentation Approach to Investigate Existing and Potential Aviation Markets. *Manuscript, [Submitted to Journal of Transport Policy, revision requested]*
3. Zhou, H., Xia, J., Norman, R., Hughes, B., Nikolova, G., Kelobonye, K., . . . Falkmer, T. (2019). Do air passengers behave differently to other regional travellers?: A travel mode choice model investigation. Journal of Air Transport Management, 79, 101682. doi:https://doi.org/10.1016/j.jairtraman.2019.101682 *[Published]*
4. Zhou, H., Norman, R., Xia, J., Hughes, B., Kelobonye, K., Nikolova, G., & Falkmer, T. (2020). Analysing travel mode and airline choice using latent class modelling: A case study in Western Australia. Transportation Research Part A: Policy and Practice, 137, 187-205. doi:https://doi.org/10.1016/j.tra.2020.04.020 *[Published]*

**Thesis attribution of papers published or being submitted**

### Author: Heng Zhou

| **Papers** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| Conceptualization | ✓ | ✓ | ✓ | ✓ |
| Data curation | ✓ | ✓ | ✓ | ✓ |
| Methodology | ✓ | ✓ | ✓ | ✓ |
| Software | ✓ | ✓ | ✓ | ✓ |
| Formal analysis | ✓ | ✓ | ✓ | ✓ |
| Interpretation and discussion | ✓ | ✓ | ✓ | ✓ |
| Writing – Original Manuscript | ✓ | ✓ | ✓ | ✓ |
| Writing – Review and Editing | ✓ | ✓ | ✓ | ✓ |

**I acknowledge that these represent my contribution to the above research output:**

## Author: Jianhong Cecilia Xia

| Papers | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Conceptualization | ✓ | ✓ | ✓ | ✓ |
| Data curation | ✓ | ✓ | ✓ | ✓ |
| Methodology | ✓ | ✓ | ✓ | ✓ |
| Software | ✓ | | | |
| Formal analysis | | | | |
| Interpretation and discussion | ✓ | ✓ | ✓ | ✓ |
| Writing – Original Manuscript | | | | |
| Writing – Review and Editing | ✓ | | ✓ | |

**I acknowledge that these represent my contribution to the above research output:**

## Author: Richard Norman

| Papers | 2 | 3 | 4 |
|---|---|---|---|
| Conceptualization | ✓ | ✓ | ✓ |
| Data curation | | | |
| Methodology | ✓ | ✓ | ✓ |
| Software | | ✓ | ✓ |
| Formal analysis | | | |
| Interpretation and discussion | ✓ | ✓ | ✓ |
| Writing – Original Manuscript | | | |
| Writing – Review and Editing | ✓ | ✓ | ✓ |

**I acknowledge that these represent my contribution to the above research output:**

### Author: Brett Hughes

| Papers | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Conceptualization | | | | |
| Data curation | ✓ | ✓ | ✓ | ✓ |
| Methodology | | | | |
| Software | | | | |
| Formal analysis | | | | |
| Interpretation and discussion | ✓ | ✓ | ✓ | ✓ |
| Writing – Original Manuscript | | | | |
| Writing – Review and Editing | ✓ | ✓ | ✓ | ✓ |

**I acknowledge that these represent my contribution to the above research output:**

### Author: Torbjorn Falkmer

| Papers | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Conceptualization | | | | |
| Data curation | | | | |
| Methodology | | | | |
| Software | | | | |
| Formal analysis | | | | |
| Interpretation and discussion | ✓ | ✓ | ✓ | ✓ |
| Writing – Original Manuscript | | | | |
| Writing – Review and Editing | ✓ | ✓ | ✓ | ✓ |

**I acknowledge that these represent my contribution to the above research output:**

### Author: Gabi Nikolova

| Papers | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Conceptualization | | | | |
| Data curation | ✓ | ✓ | ✓ | ✓ |
| Methodology | | | | |
| Software | | | | |
| Formal analysis | | | | |
| Interpretation and discussion | | | | |
| Writing – Original Manuscript | | | | |
| Writing – Review and Editing | ✓ | ✓ | ✓ | ✓ |

**I acknowledge that these represent my contribution to the above research output:**

### Author: Keone Kelobonye

| Papers | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Conceptualization | | | | |
| Data curation | | | | |
| Methodology | | ✓ | | ✓ |
| Software | | | | |
| Formal analysis | | | | |
| Interpretation and discussion | ✓ | ✓ | ✓ | ✓ |
| Writing – Original Manuscript | | | | |
| Writing – Review and Editing | ✓ | ✓ | ✓ | ✓ |

**I acknowledge that these represent my contribution to the above research output:**

## Author: Jie Sun

| **Papers** | **1** |
|---|---|
| Conceptualization | ✓ |
| Data curation | |
| Methodology | |
| Software | |
| Formal analysis | |
| Interpretation and discussion | ✓ |
| Writing – Original Manuscript | |
| Writing – Review and Editing | ✓ |

**I acknowledge that these represent my contribution to the above research output:**

## Author: Qingzhou Luo

| **Papers** | **1** |
|---|---|
| Conceptualization | |
| Data curation | |
| Methodology | ✓ |
| Software | ✓ |
| Formal analysis | |
| Interpretation and discussion | |
| Writing – Original Manuscript | |
| Writing – Review and Editing | |

**I acknowledge that these represent my contribution to the above research output:**

## Author: Hui Wang

| **Papers** | **1** |
|---|---|
| Conceptualization | |
| Data curation | ✓ |
| Methodology | |
| Software | |
| Formal analysis | |
| Interpretation and discussion | ✓ |
| Writing – Original Manuscript | |
| Writing – Review and Editing | |

**I acknowledge that these represent my contribution to the above research output:**

## Author: Kai Du

| **Papers** | **3** |
|---|---|
| Conceptualization | |
| Data curation | ✓ |
| Methodology | |
| Software | |
| Formal analysis | |
| Interpretation and discussion | ✓ |
| Writing – Original Manuscript | |
| Writing – Review and Editing | |

**I acknowledge that these represent my contribution to the above research output:**

## APPENDIX B INFORMATION SHEET

Curtin University | Department of Transport

Curtin University Human Research Ethics Committee (HREC) has approved this study.

***Project Title:*** Deregulation and Competition: Comparison of Regional Aviation Markets in Western Australia

Our study which is seeking to investigate passenger's travel mode and airline choice for traveling. The aim of the research is to get a better understanding of how passengers react to airfare, travel distance, comfortability, specifically how these factors influence the demand to WA air services and the competition between regional airports. The expected results may provide insights to civic aviation policymakers and airlines about air travel demand to WA regional areas and how some key parameters, such as airfare, flight frequency and accessibility to airports, affect people's travel mode and airline choices. It is funded by the Department of Transport (DoT) WA.

***Passenger Information will be captured through questionnaire survey***

We would like you to help us understand the passengers' travel mode and airline choice for their domestic travelling better by doing three things for us.

The three things we would like you to do are these:

- Tell us a little about you and your travel.
- Record your travel mode and airline choice for your recent travel.
- Make a choice among bus, car or airlines if you want travel to a given destination.

***Contacts:***

A/Prof Cecilia Xia
08 9266 7563
c.xia@curtin.edu.au

Heng Zhou
0416749997
Heng.zhou@postgrad.curtin.edu.au

# APPENDIX C ETHIC APPROVAL

Curtin University

Office of Research and Development

GPO Box U1987
Perth Western Australia 6845

**Telephone** +61 8 9266 7863
**Facsimile** +61 8 9266 3793
**Web** research.curtin.edu.au

21-Nov-2017

Name: Cecilia Xia
Department/School: Department of Spatial Sciences
Email: C.Xia@curtin.edu.au

Dear Cecilia Xia

**RE: Ethics Office approval**
**Approval number: HRE2017-0815**

Thank you for submitting your application to the Human Research Ethics Office for the project **Deregulation and Competition: Comparison of Regional Aviation Markets in Western Australia**.

Your application was reviewed through the Curtin University Negligible risk review process.

The review outcome is: **Approved**.

Your proposal meets the requirements described in the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research (2007)*.

Approval is granted for a period of one year from **21-Nov-2017** to **20-Nov-2018**. Continuation of approval will be granted on an annual basis following submission of an annual report.

Personnel authorised to work on this project:

| Name | Role |
|---|---|
| Xia, Cecilia | CI |
| Norman, Richard | Co-Inv |
| Zhou, Heng | Student |

Approved documents:

| Document |
|---|

**Standard conditions of approval**

1. Research must be conducted according to the approved proposal
2. Report in a timely manner anything that might warrant review of ethical approval of the project including:

- proposed changes to the approved proposal or conduct of the study
- unanticipated problems that might affect continued ethical acceptability of the project
- major deviations from the approved proposal and/or regulatory guidelines
- serious adverse events

3. Amendments to the proposal must be approved by the Human Research Ethics Office before they are implemented (except where an amendment is undertaken to eliminate an immediate risk to participants)
4. An annual progress report must be submitted to the Human Research Ethics Office on or before the anniversary of approval and a completion report submitted on completion of the project
5. Personnel working on this project must be adequately qualified by education, training and experience for their role, or supervised
6. Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, that bears on this project
7. Changes to personnel working on this project must be reported to the Human Research Ethics Office
8. Data and primary materials must be retained and stored in accordance with the Western Australian University Sector Disposal Authority (WAUSDA) and the Curtin University Research Data and Primary Materials policy
9. Where practicable, results of the research should be made available to the research participants in a timely and clear manner
10. Unless prohibited by contractual obligations, results of the research should be disseminated in a manner that will allow public scrutiny; the Human Research Ethics Office must be informed of any constraints on publication
11. Approval is dependent upon ongoing compliance of the research with the Australian Code for the Responsible Conduct of Research, the National Statement on Ethical Conduct in Human Research, applicable legal requirements, and with Curtin University policies, procedures and governance requirements
12. The Human Research Ethics Office may conduct audits on a portion of approved projects.

**Special Conditions of Approval**

None

**This letter constitutes low risk/negligible risk approval only**. This project may not proceed until you have met all of the Curtin University research governance requirements.

Should you have any queries regarding consideration of your project, please contact the Ethics Support Officer for your faculty or the Ethics Office at hrec@curtin.edu.au or on 9266 2784.

Yours sincerely

[redacted]

Amy Bowater
Acting Manager, Research Integrity

# APPENDIX D INTERCEPT SURVEY QUESTIONNAIRES

## Appendix D1 Air Travel Information

Curtin University

Department of Transport

Curtin University Human Research Ethics Committee (HREC) has approved this study.

*This survey by the School of Earth and Planetary Sciences, Curtin University is designed to get a better understanding of travel methods from a regional airport to other places and how some key parameters, such as airfare and flight frequency, affect people's travel mode and airline choices. This is an independent study from Curtin University, with assistance and support from the Department of Transport in WA.*

## Air Travel Survey

*Thank you for taking the time to undertake this survey.*

**Flight number:** ______________ **Date:** _____ /_____ / 2018

1. **Where did you start your trip to the airport today? Please tick one.**
   1) Your own home
   2) Someone else's home
   3) A place of business / Workplace
   4) Accommodation: hotel, motel, inn, bed & breakfast, Airbnb, backpackers
   5) A restaurant
   6) A school, college, or university
   7) A tourist attraction
   8) Somewhere else (please specify) ______________

2. **What was the suburb or postcode from which you started your trip to the airport today?**
   This may not necessarily be your usual residence if you are a visitor. If you don't know the suburb or postcode, what was the name of the place or street from which you left?

   ______________________________

3. **If you will transit through Perth, what is your final destination?**
   If Perth is your final destination, please write Perth.

   ______________________________

4. **How did you travel to the airport today? Please tick more than one if applicable.**
   1) Taxi
   2) Ridesharing, such as, Uber
   3) Company car
   4) Private car
   5) Rental car
   6) Public Transport (e.g. Bus)
   7) Motorcycling
   8) Cycling
   9) Lift by friend/s, colleague/s, relative/s, or someone else
   10) Other (please specify) ______________

Department of Transport

Curtin University Human Research Ethics Committee (HREC) has approved this study.

**5. Who is travelling with you today? Please tick more than one, if applicable.**
1) Alone
2) Partner/spouse
3) Friend/s
4) Family
5) Business associate/s or colleague/s
6) Special interest group/s
7) Other (please specify) ______________

**6. What was your main reason for taking this trip? Please tick one.**
1) Holiday / Leisure
2) Visiting Friends or Relatives
3) Education
4) Work – Government (Commonwealth, State or Local public servant)
5) Work – Fly in or fly out (regular mining operational staff)
6) Work – Other Business
7) Medical or health reasons
8) Other reason (please specify) __________________

**7. Why did you choose to fly rather than travel by car or coach? Please tick one.**
1) Distance is too long to drive by car
2) The airfare was cheap / affordable
3) Convenience / more time efficient
4) It is an emergency trip
5) Flying is safer
6) Flying is more comfortable
7) Other (please specify) __________________

**8. How much did you personally pay for your flight one-way (divide by two if you have a return flight)? Please tick one.**
1) Nothing: my employer paid
2) Nothing: someone else paid (e.g. family, friend/s)
3) Nothing: flight was paid through the Patient Assisted Travel Scheme (PATS)
4) $0 - $199
5) $200 - $299
6) $300 - $399
7) $400 - $499
8) Over $500

**9. In last 12 months, how many times, including this trip, did you fly to this destination? Please tick one.**
1) Once (this trip)
2) Twice
3) 3-5 times
4) 6 or more times

## Appendix D2 Travel Mode and Airline Choice Block 1

Curtin University

Department of Transport

Curtin University Human Research Ethics Committee (HREC) has approved this study.

# Business Travel Block 1

Below are six hypothetical trips for travelling for business in regional WA. Each trip has 4 travel mode options (car, bus, and two different airlines) and some trip attributes such as cost and time. **For each trip**, evaluate the different attributes of each travel option and **please tick one option that appeals to you most overall (one option only)**.

| Trip 1 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$150 | A$175 | A$350 | A$200 |
| Time to bus-station or airport-terminal | N/A | 15 mins | 30 mins | 15 mins |
| Journey or travel time | 12 hrs | 15 hrs | 2 hrs | 2 hrs |
| Service frequency (weekly) | Any time | 30 buses/coaches a week | 16 flights a week | 2 flights a week |
| Seat comfort level | High (leg room 90 cm) | Medium (leg room 80 cm) | Medium (leg room 80 cm) | Low (leg room 70 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

| Trip 2 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$400 | A$425 | A$500 | A$650 |
| Time to bus-station or airport-terminal | N/A | 30 mins | 60 mins | 60 mins |
| Journey or travel time | 30 hrs | 35 hrs | 4 hrs | 4 hrs |
| Service frequency (weekly) | Any time | 16 buses/coaches a week | 30 flights a week | 44 flights a week |
| Seat comfort level | High (leg room 90 cm) | High (leg room 90 cm) | Low (leg room 70 cm) | Medium (leg room 80 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

| Trip 3 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$275 | A$300 | A$500 | A$350 |
| Time to bus-station or airport-terminal | N/A | 60 mins | 60 mins | 45 mins |
| Journey or travel time | 21 hrs | 25 hrs | 3 hrs | 4 hrs |
| Service frequency (weekly) | Any time | 44 buses/coaches a week | 30 flights a week | 2 flights a week |
| Seat comfort level | High (leg room 90 cm) | Low (leg room 70 cm) | Medium (leg room 80 cm) | Medium (leg room 80 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

Curtin University

Department of Transport

Curtin University Human Research Ethics Committee (HREC) has approved this study.

## Trip 4

| | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$150 | A$50 | A$200 | A$200 |
| Time to bus-station or airport-terminal | N/A | 45 mins | 60 mins | 30 mins |
| Journey or travel time | 21 hrs | 25 hrs | 3 hrs | 3 hrs |
| Service frequency (weekly) | Any time | 2 buses/coaches a week | 30 flights a week | 30 flights a week |
| Seat comfort level | Medium (leg room 80 cm) | Medium (leg room 80 cm) | High (leg room 90 cm) | Low (leg room 70 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

## Trip 5

| | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$275 | A$300 | A$350 | A$350 |
| Time to bus-station or airport-terminal | N/A | 30 mins | 15 mins | 60 mins |
| Journey or travel time | 30 hrs | 35 hrs | 4 hrs | 4 hrs |
| Service frequency (weekly) | Any time | 16 buses/coaches a week | 2 flights a week | 44 flights a week |
| Seat comfort level | High (leg room 90 cm) | High (leg room 90 cm) | Low (leg room 70 cm) | Medium (leg room 80 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

## Trip 6

| | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$25 | A$50 | A$200 | A$200 |
| Time to bus-station or airport-terminal | N/A | 60 mins | 15 mins | 30 mins |
| Journey or travel time | 3 hrs | 5 hrs | 1 hrs | 1 hrs |
| Service frequency (weekly) | Any time | 2 buses/coaches a week | 44 flights a week | 16 flights a week |
| Seat comfort level | Medium (leg room 80 cm) | Low (leg room 70 cm) | Medium (leg room 80 cm) | High (leg room 90 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

## Appendix D3 Travel Mode and Airline Choice Block 2

Curtin University

Department of Transport

Curtin University Human Research Ethics Committee (HREC) has approved this study.

# Business Travel Block 2

Below are six hypothetical trips for travelling for business in regional WA. Each trip has 4 travel mode options (car, bus, and two different airlines) and some trip attributes such as cost and time. **For each trip**, evaluate the different attributes of each travel option and **please tick one option that appeals to you most overall (one option only).**

| Trip 1 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$150 | A$175 | A$200 | A$350 |
| Time to bus-station or airport-terminal | N/A | 15 mins | 60 mins | 45 mins |
| Journey or travel time | 12 hrs | 15 hrs | 2 hrs | 2 hrs |
| Service frequency (weekly) | Any time | 44 buses/coaches a week | 30 flights a week | 2 flights a week |
| Seat comfort level | High (leg room 90 cm) | Low (leg room 70 cm) | High (leg room 90 cm) | Medium (leg room 80 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

| Trip 2 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$275 | A$300 | A$650 | A$650 |
| Time to bus-station or airport-terminal | N/A | 60 mins | 60 mins | 45 mins |
| Journey or travel time | 30 hrs | 35 hrs | 4 hrs | 4 hrs |
| Service frequency (weekly) | Any time | 30 buses/coaches a week | 30 flights a week | 44 flights a week |
| Seat comfort level | Medium (leg room 80 cm) | Low (leg room 70 cm) | High (leg room 90 cm) | Low (leg room 70 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

| Trip 3 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$150 | A$175 | A$350 | A$350 |
| Time to bus-station or airport-terminal | N/A | 15 mins | 30 mins | 60 mins |
| Journey or travel time | 12 hrs | 15 hrs | 2 hrs | 2 hrs |
| Service frequency (weekly) | Any time | 44 buses/coaches a week | 16 flights a week | 16 flights a week |
| Seat comfort level | Medium (leg room 80 cm) | High (leg room 90 cm) | Low (leg room 70 cm) | High (leg room 90 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

Curtin University

Department of Transport

Curtin University Human Research Ethics Committee (HREC) has approved this study.

| Trip 4 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$25 | A$50 | A$200 | A$200 |
| Time to bus-station or airport-terminal | N/A | 45 mins | 45 mins | 30 mins |
| Journey or travel time | 3 hrs | 5 hrs | 1 hrs | 1 hrs |
| Service frequency (weekly) | Any time | 2 buses/coaches a week | 2 flights a week | 44 flights a week |
| Seat comfort level | High (leg room 90 cm) | High (leg room 90 cm) | Medium (leg room 80 cm) | Low (leg room 70 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

| Trip 5 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$275 | A$300 | A$350 | A$500 |
| Time to bus-station or airport-terminal | N/A | 60 mins | 15 mins | 30 mins |
| Journey or travel time | 21 hrs | 25 hrs | 4 hrs | 3 hrs |
| Service frequency (weekly) | Any time | 2 buses/coaches a week | 44 flights a week | 30 flights a week |
| Seat comfort level | Medium (leg room 80 cm) | Medium (leg room 80 cm) | Low (leg room 70 cm) | High (leg room 90 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

| Trip 6 | Option1: Car | Option2: Bus | Option3: Airline1 | Option4: Airline2 |
|---|---|---|---|---|
| Ticket fare or driving cost | A$400 | A$425 | A$650 | A$500 |
| Time to bus-station or airport-terminal | N/A | 15 mins | 45 mins | 15 mins |
| Journey or travel time | 30 hrs | 35 hrs | 4 hrs | 4 hrs |
| Service frequency (weekly) | Any time | 2 buses/coaches a week | 44 flights a week | 16 flights a week |
| Seat comfort level | Medium (leg room 80 cm) | Medium (leg room 80 cm) | Low (leg room 70 cm) | High (leg room 90 cm) |
| Which one would you choose for your trip? | ○ | ○ | ○ | ○ |

# Appendix D4 Demographic Information

Curtin University Human Research Ethics Committee (HREC) has approved this study.

## Demographic information

**10. Please rank the following factors affecting your choice of transport mode (1 is the most important and 5 is the least important).**

| Factors | Rank (1(most) – 5(least)) |
|---|---|
| Ticket fare or driving cost | |
| Time to bus-station or airport-terminal | |
| Journey / travel time (form terminal to destination) | |
| Service frequency | |
| Seat comfort | |

**11. Please choose your gender. Please tick one.**
1) Male
2) Female

**12. Please choose your age category. Please tick one.**
1) Under 18
2) 18-24
3) 25-34
4) 35-44
5) 45-54
6) 55-64
7) 65 or older

**13. What is the highest level of education have you completed? Please tick one.**
1) Primary/some secondary
2) Senior high school
3) Vocational/Technical
4) Diploma
5) University (undergraduate)
6) University (postgraduate)

**14. What is your current employment? Please tick one.**
1) Employed by a small company (less than 50 employees)
2) Employed by a large company (more than 50 employees)
3) Self-employed
4) Currently looking for work or not in paid employment
5) Retired
6) Student
7) Government (local, state or federal)
8) Other ___________________

Curtin University

Department of Transport

Curtin University Human Research Ethics Committee (HREC) has approved this study.

**15. What is the place of your usual residence?**

1) Australia Postcode or Suburb

2) Overseas Country

**16. What is the total personal income you usually receive before tax per month (in Australian dollars)? Please tick one.**

1) Nil
2) $1 - $1,749
3) $1,750 - $3,499
4) $3,500 - $5,499
5) $5,500 - $6,499
6) $6,500 - $8,699
7) $8,700 or more

**17. If you have travelled by air to a regional destination in WA in the last two years, how far in advance did you book a flight? Please tick one.**

1) Less than 24 hours
2) Less than two days
3) Between two days and one week
4) Between one week and one month
5) 1 – 3 months
6) 3 – 6 months
7) 6 months – 1 year
8) Over 1 year
9) Not applicable

**This is the end of the survey. Thank you very much for your help.**

*This survey by the School of Earth and Planetary Sciences, Curtin University is designed to get a better understanding of air travel demand to WA regional areas and how some key parameters, such as airfare and flight frequency, affect people's travel mode and airline choices. This is an independent study from Curtin University, with assistance and support from the Department of Transport in WA. If you have any queries about the survey please contact Mr Heng Zhou or A/Prof Jianhong (Cecilia) Xia at on (08) 9266 7563, heng.zhou@postgrad.curtin.edu.au or c.xia@curtin.edu.au. Mail address: GPO Box U1987 Perth Western Australia 6845.*

# APPENDIX E PEARSON CORRELATION MATRICES

## Appendix E1 Pearson Correlation Between Independent Variables By Classifying Catchment Area Using Thiessen Polygons

| | | ln_Mine_sites_origin | ln_Mine_sites_destination | ln_Travel_time | ln_Population_origin | ln_Population_destination | ln_Driving_distance | ln_Average_income_origin | ln_Average_income_destination | ln_ catchment_area_origin | ln_catchment_area_destination | ln_Tourists_destination | ln_Airfare |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ln_Total_seats | Pearson Correlation | .302** | .282** | .986** | .416** | .390** | -.163** | -.090 | -.092* | .052 | .044 | .360** | .987** |
| | P-value | .000 | .000 | 0.000 | .000 | .000 | .000 | .054 | .047 | .267 | .342 | .000 | 0.000 |
| ln_Mine_sites_origin | Pearson Correlation | 1 | -.048 | .301** | .570** | -.027 | -.041 | .196** | -.009 | .128** | -.006 | -.015 | .284** |
| | P-value | | .307 | .000 | .000 | .561 | .385 | .000 | .841 | .006 | .896 | .746 | .000 |
| ln_Mine_sites_destination | Pearson Correlation | -.048 | 1 | .274** | -.027 | .570** | -.080 | -.009 | .196** | -.006 | .128** | .317** | .258** |
| | P-value | .307 | | .000 | .561 | .000 | .087 | .841 | .000 | .896 | .006 | .000 | .000 |
| ln_Travel_time | Pearson Correlation | .301** | .274** | 1 | .412** | .381** | -.160** | -.094* | -.101* | .056 | .052 | .357** | .996** |
| | P-value | .000 | .000 | | .000 | .000 | .001 | .043 | .031 | .234 | .263 | .000 | 0.000 |
| ln_Population_origin | Pearson Correlation | .570** | -.027 | .412** | 1 | -.048 | .105* | -.286** | .014 | .111* | -.005 | -.040 | .399** |
| | P-value | .000 | .561 | .000 | | .307 | .024 | .000 | .770 | .017 | .910 | .397 | .000 |
| ln_Population_destination | Pearson Correlation | -.027 | .570** | .381** | -.048 | 1 | .078 | .014 | -.286** | -.005 | .111* | .830** | .366** |
| | P-value | .561 | .000 | .000 | .307 | | .094 | .770 | .000 | .910 | .017 | .000 | .000 |
| ln_Driving_distance | Pearson Correlation | -.041 | -.080 | -.160** | .105* | .078 | 1 | -.124** | -.129** | .242** | .114* | .089 | -.182** |
| | P-value | .385 | .087 | .001 | .024 | .094 | | .008 | .006 | .000 | .015 | .055 | .000 |
| ln_Average_income_origin | Pearson Correlation | .196** | -.009 | -.094* | -.286** | .014 | -.124** | 1 | -.048 | -.041 | .002 | .020 | -.094* |
| | P-value | .000 | .841 | .043 | .000 | .770 | .008 | | .307 | .380 | .967 | .668 | .044 |
| ln_Average_income_destination | Pearson Correlation | -.009 | .196** | -.101* | .014 | -.286** | -.129** | -.048 | 1 | .002 | -.041 | -.420** | -.099* |
| | P-value | .841 | .000 | .031 | .770 | .000 | .006 | .307 | | .967 | .380 | .000 | .033 |
| ln_ catchment _area_origin | Pearson Correlation | .128** | -.006 | .056 | .111* | -.005 | .242** | -.041 | .002 | 1 | -.048 | .000 | .063 |
| | P-value | .006 | .896 | .234 | .017 | .910 | .000 | .380 | .967 | | .307 | .999 | .175 |
| ln_ catchment _area_destination | Pearson Correlation | -.006 | .128** | .052 | -.005 | .111* | .114* | .002 | -.041 | -.048 | 1 | -.001 | .057 |
| | P-value | .896 | .006 | .263 | .910 | .017 | .015 | .967 | .380 | .307 | | .980 | .222 |
| ln_Tourists_destination | Pearson Correlation | -.015 | .317** | .357** | -.040 | .830** | .089 | .020 | -.420** | .000 | -.001 | 1 | .345** |
| | P-value | .746 | .000 | .000 | .397 | .000 | .055 | .668 | .000 | .999 | .980 | | .000 |
| ln_Airfare | Pearson Correlation | .284** | .258** | .996** | .399** | .366** | -.182** | -.094* | -.099* | .063 | .057 | .345** | 1 |
| | P-value | .000 | .000 | 0.000 | .000 | .000 | .000 | .044 | .033 | .175 | .222 | .000 | |

*Significant at the 5% level
**Significant at the 1% level

# Appendix E2 Pearson Correlation Between Independent Variables By Classifying Catchment Area Using 2.5 Hour Drive Distance

| | | ln_Mine_sites_origin | ln_Mine_sites_destination | ln_Travel_time | ln_Population_origin | ln_Population_destination | ln_Driving_distance | ln_Average_income_origin | ln_Average_income_destination | ln_ catchment_area_origin | ln_catchment_area_destination | ln_Tourists_destination | ln_Airfare |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ln_Total_seats | Pearson Correlation | .250** | .235** | .986** | .398** | .377** | -.163** | -.064 | -.066 | .052 | .047 | .334** | .987** |
| | P-value | .000 | .000 | 0.000 | .000 | .000 | .000 | .170 | .154 | .261 | .310 | .000 | 0.000 |
| ln_Mine_sites_origin | Pearson Correlation | 1 | -.048 | .250** | .542** | -.026 | -.065 | .458** | -.022 | .629** | -.030 | -.012 | .234** |
| | P-value | | .307 | .000 | .000 | .580 | .163 | .000 | .640 | .000 | .521 | .804 | .000 |
| ln_Mine_sites_destination | Pearson Correlation | -.048 | 1 | .228** | -.026 | .542** | -.109* | -.022 | .458** | -.030 | .629** | .243** | .214** |
| | P-value | .307 | | .000 | .580 | .000 | .019 | .640 | .000 | .521 | .000 | .000 | .000 |
| ln_Travel_time | Pearson Correlation | .250** | .228** | 1 | .397** | .366** | -.160** | -.070 | -.073 | .060 | .045 | .324** | .996** |
| | P-value | .000 | .000 | | .000 | .000 | .001 | .134 | .117 | .198 | .330 | .000 | 0.000 |
| ln_Population_origin | Pearson Correlation | .542** | -.026 | .397** | 1 | -.048 | -.036 | -.071 | .003 | .187** | -.009 | -.040 | .379** |
| | P-value | .000 | .580 | .000 | | .307 | .436 | .128 | .942 | .000 | .849 | .397 | .000 |
| ln_Population_destination | Pearson Correlation | -.026 | .542** | .366** | -.048 | 1 | -.002 | .003 | -.071 | -.009 | .187** | .830** | .350** |
| | P-value | .580 | .000 | .000 | .307 | | .962 | .942 | .128 | .849 | .000 | .000 | .000 |
| ln_Driving_distance | Pearson Correlation | -.065 | -.109* | -.160** | -.036 | -.002 | 1 | -.087 | -.158** | -.025 | -.050 | .037 | -.182** |
| | P-value | .163 | .019 | .001 | .436 | .962 | | .060 | .001 | .596 | .282 | .426 | .000 |
| ln_Average_income_origin | Pearson Correlation | .458** | -.022 | -.070 | -.071 | .003 | -.087 | 1 | -.048 | .241** | -.011 | .013 | -.073 |
| | P-value | .000 | .640 | .134 | .128 | .942 | .060 | | .307 | .000 | .805 | .776 | .115 |
| ln_Average_income_destination | Pearson Correlation | -.022 | .458** | -.073 | .003 | -.071 | -.158** | -.048 | 1 | -.011 | .241** | -.279** | -.077 |
| | P-value | .640 | .000 | .117 | .942 | .128 | .001 | .307 | | .805 | .000 | .000 | .099 |
| ln_ catchment _area_origin | Pearson Correlation | .629** | -.030 | .060 | .187** | -.009 | -.025 | .241** | -.011 | 1 | -.048 | .004 | .057 |
| | P-value | .000 | .521 | .198 | .000 | .849 | .596 | .000 | .805 | | .307 | .925 | .220 |
| ln_ catchment _area_destination | Pearson Correlation | -.030 | .629** | .045 | -.009 | .187** | -.050 | -.011 | .241** | -.048 | 1 | -.092* | .047 |
| | P-value | .521 | .000 | .330 | .849 | .000 | .282 | .805 | .000 | .307 | | .049 | .317 |
| ln_Tourists_destination | Pearson Correlation | -.012 | .243** | .324** | -.040 | .830** | .037 | .013 | -.279** | .004 | -.092* | 1 | .312** |
| | P-value | .804 | .000 | .000 | .397 | .000 | .426 | .776 | .000 | .925 | .049 | | .000 |
| ln_Airfare | Pearson Correlation | .234** | .214** | .996** | .379** | .350** | -.182** | -.073 | -.077 | .057 | .047 | .312** | 1 |
| | P-value | .000 | .000 | 0.000 | .000 | .000 | .000 | .115 | .099 | .220 | .317 | .000 | |

*Significant at the 5% level
**Significant at the 1% level

# APPENDIX F AIR PASSENGER MARKET SEGMENTATION

## Appendix F1 Information Criteria

| No. of segments | N | Log likelihood (LL) | K | AIC | Segment size |
|---|---|---|---|---|---|
| 1 | 476 | -1,141.01 | 13 | 2308.03 | 100% |
| 2 | 476 | 959.05 | 27 | -1864.11 | 78%,22% |
| **3** | **476** | **3,000.33** | **41** | **-5918.66** | **53%,21%,26%** |
| 4 | 476 | 507.87 | 55 | -905.75 | 16%,45%,21%,18% |
| 5 | 476 | 156.65 | 69 | -175.30 | 10%,30%,20%,14%,27% |

## Appendix F2 Air Passenger Market Segments Block Two

| **Characteristics** | **Air passenger Block 2** | **Segment a1** | **Segment a2** | **Segment a3** |
|---|---|---|---|---|
| Segment size | Proportion of sample | **53%** | 21% | 26% |
| Car probability | Mean | 0.044 | 0.373 | 0.173 |
| | Std. dev. | 0.077 | 0.304 | 0.110 |
| Bus probability | Mean | 0.000 | 0.120 | 0.001 |
| | Std. dev. | 0.000 | 0.178 | 0.004 |
| Airline probability | Mean | **0.950** | **0.508** | **0.827** |
| | Std. dev. | 0.077 | 0.272 | 0.110 |
| Trip purpose | Business | **65%** | 46% | **80%** |
| | Non-business | 35% | **54%** | 20% |
| Gender | Female | 43% | 45% | 34% |
| | Male | **57%** | **55%** | **66%** |
| Age | Under 25 | 7% | 15% | 4% |
| | 25 to 44 | 29% | 29% | **82%** |
| | 45 or more | **64%** | **55%** | 14% |
| Education background | Basic education | 43% | 32% | 12% |
| | Tertiary education | **57%** | **68%** | **88%** |
| Income | Low income | 8% | 22% | 2% |
| | Middle income | 24% | 26% | 30% |
| | High income | **68%** | **52%** | **68%** |

# APPENDIX G NON-AIR PASSENGER MARKET SEGMENTATION

## Appendix G1 Information Criteria

| No. of segments | N | Log likelihood (LL) | K | AIC | Segment size |
|---|---|---|---|---|---|
| 1 | 422 | -1,704.48 | 13 | 3,434.96 | 100% |
| 2 | 422 | 890.54 | 27 | -1,727.08 | 30%,70% |
| **3** | **422** | **1,144.24** | **41** | **-2,206.48** | **57%,25%,18%** |
| 4 | 422 | -674.39 | 55 | 1,458.78 | 25%,17%,39%,19% |
| 5 | 422 | -775.27 | 69 | 1,688.53 | 22%,16%,29%,19%,15% |

## Appendix G2 Non-air Passenger Market Segments Block Two

| **Characteristics** | **Non-air passenger Block 2** | **Segment n1** | **Segment n2** | **Segment n3** |
|---|---|---|---|---|
| Segment size | Proportion of sample | **57%** | 25% | 18% |
| Car probability | Mean | 0.197 | **0.654** | 0.331 |
| | Std. dev. | 0.199 | 0.312 | 0.253 |
| Bus probability | Mean | 0.000 | 0.037 | 0.233 |
| | Std. dev. | 0.000 | 0.068 | 0.246 |
| Airline probability | Mean | **0.803** | 0.309 | **0.436** |
| | Std. dev. | 0.199 | 0.286 | 0.266 |
| Trip purpose | Business | 36% | 23% | 32% |
| | Non-business | **64%** | **77%** | **68%** |
| Gender | Female | **53%** | **65%** | **53%** |
| | Male | 47% | 35% | 47% |
| Age | Under 25 | 16% | 7% | **65%** |
| | 25 to 44 | **44%** | **62%** | 10% |
| | 45 or more | 41% | 32% | 25% |
| Education background | Basic education | 25% | 25% | 73% |
| | Tertiary education | **75%** | **75%** | **27%** |
| Income | Low income | 21% | 13% | 59% |
| | Middle income | **44%** | **72%** | 36% |
| | High income | 35% | 15% | **5%** |