MACQUARIE UNIVERSITY

Agglomeration Economies, Diseconomies, and City Size

The Case of Australian Cities

Muheed Jamaldeen

October 2015

Supervisor

Assoc. Prof. Tony Bryant

Co-Supervisor

Assoc. Prof. Roselyne Joyeux

Submitted to Macquarie University
In partial fulfilment of the requirements for the degree of

Master of Research (Economics)

October 2015

Abstract

This thesis examines the interplay between urban economies and diseconomies in Australian cities, and their relationship with productivity-maximising city size. Economic theory postulates the existence of an inverted U-shaped relationship between output per worker (productivity) and city size. Using the Au and Henderson (2006) (AH) model, we analyse this relationship for Australia. Given that the theoretical AH model imposes a monocentric circular urban structure, this thesis proposes a generalisation of the model that allows for non-circular polycentric urban structure. From a theoretical point of view, this extension illustrates the importance of transport efficiency, and housing affordability in improving urban-productivity outcomes.

Empirical results suggest that the generalised version of the AH model is supported by the data. In addition, we find that the classic Marshallian scale effect is dominant in Australian cities. It is argued that this is because the knowledge economy (services sector) plays an important role in an advanced economy such as Australia.

Empirical results are also used to estimate productivity-maximising city sizes for Australian Statistical Area 3 (SA3). This shows that though the vast majority of Australian SA3s are operating below their productivity-maximising peak sizes, the output loss due to lack of scale is moderate for major SA3s. However, of the major CBDs in Australia, output loss is more than 20% for North Canberra, Brisbane Inner - North and Adelaide City.

Acknowledgements

I would like to thank the Macquarie University for financial support through the course of this year, and SGS Economics and Planning for affording me the opportunity and support required to undertake this thesis. I would also like to offer my sincere gratitude to my two supervisors, Assoc. Prof. Tony Bryant and Assoc. Prof. Roselyne Joyeux, for their continual guidance, assistance and patience throughout the year.

In addition, I would like to, again, acknowledge Assoc. Prof. Tony Bryant in his capacity as the Master of Research co-ordinator. Our inspiring discussions have sustained my desire to seek intellectual exploration, and to pursue economic-research in the future. His belief in the capabilities of each student in this year's group has been a source of confidence to us all.

Special mentions also go to Julian Szafraniec, Armando Mazzei, and Torin Allen – colleagues at SGS, for their assistance with accessing data used in this thesis. I would also like to thank my parents and brother for putting up with my absence from their lives for most of this year. Their love, trust and confidence in me have been a constant source of inspiration. Last, but certainly not least, I would like to thank God. His grace, glory, and mercy have brought me to where I am. All those who have helped me throughout this year are a manifestation of His blessings in my life, and for that I am forever grateful.

TABLE OF CONTENTS

LIST OF FIGURES

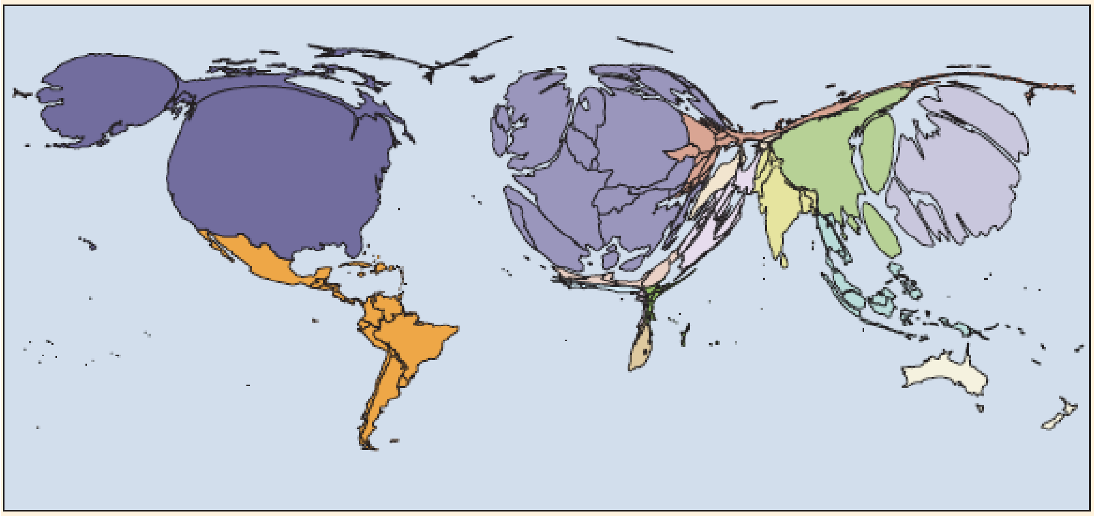
LIST OF TABLES

1 INTRODUCTION

1.1 Background

One of the most striking features about the organisation of economic activity is its unequal geographical distribution. High-income countries are almost entirely concentrated in a few temperate zones. Nearly 50 percent of World GDP is produced by 15 percent of the world's population, while 17 of the poorest 20 nations are located in tropical Africa (Henderson *et al*. 2001). The figure below, produced by the World Bank, shows this distribution visually. A country's size on this map shows the proportion of global GDP produced there. It is immediately apparent that there is an unequal distribution of economic activity across space, and that some large areas of land contribute proportionately much less to global GDP.

FIGURE 1. COUNTRY-SIZE AS A SHARE OF GLOBAL GDP



Source: WDR 2009 team using 2005 GDP (constant U.S. dollars).
Note: The cartogram was created using the method developed by Gastner and Newman (2004). This map shows the countries that have the most wealth when GDP is compared using currency exchange rates. This indicates international purchasing power—what someone's money is worth if spent in another country.

The uneven spatial distribution is apparent at the international level. However, concentrations of economic activity are also observed at a national level. As Quigley (1998) notes, scale economies are the historical rationale for the existence of cities, and it has long been argued that without the existence of scale economies in production, economic activity would be dispersed to save on transportation costs.

Since Krugman's *Geography and Trade* (1991), there has been a rise in the discussion of the spatial dimension in economics. However, the importance of geography in determining economic outcomes has been recognised for a long time. For instance, in 1776, Adam Smith observed that specialisation could occur through the opening up of markets, and the extent to which markets are open is partly determined—with the advent of transport innovations—by the geography of the economy (Smith, 1776). This simple, yet powerful, observation has potentially significant implications for spatial formulation of economic policy in many countries, including Australia.

There is a significant volume of theoretical and empirical work examining why firms tend to cluster (or agglomerate) in cities. Many of the empirical studies focus on disentangling the Marshallian 'trinity' of agglomeration economies (or benefits), namely a local pool of skilled labour, local supplier linkages, and local knowledge spillovers.

While there might be benefits from the agglomeration of economic activity, it follows intuitively that there might also be costs (diseconomies). Though not as extensive, there is a small strand of the literature that examines diseconomies[1]. For instance, congestion in the city might reduce the benefits of co-location (Krugman, 1998). The trade-off between the costs and benefits of agglomeration—or as Krugman (1998) puts it, the tension between 'centripetal' and 'centrifugal' forces—suggests that there exists a spatial equilibrium (for each location) that balances out these forces. It is plausible that the magnitude of these forces may also vary with the size of a city. That is, as a city gets larger, congestion costs might outweigh any increase in benefits from a larger city. Whilst the idea of optimal location is notably present in the theoretical literature (see for instance Starrett (1974)), there is limited theoretical and empirical work investigating 'optimal' city size.

There are a few studies that concurrently consider both economies and diseconomies, and their relationship with city scale. Au and Henderson (2006), one of the first contemporary studies to do so, formally models and then estimates net agglomeration economies for Chinese cities to determine 'optimal' city sizes where output per worker is maximised. However, the notion that there exists an 'optimal' city size (or sizes) can be traced back to Aristotle's *Politics*. As noted by Liansos (2010) (who formalises Aristotle's ideas)[2], Aristotle argues that the land-population ratio determines the welfare of a city. Interestingly, in his model, output per capita and public-private land mix ultimately determine the optimal land-population ratio.

The Australian literature on agglomeration, like the majority of international studies, is largely focussed on Marshallian economies[3]. This thesis utilises the Au and Henderson (2006) (AH) model to analyse the interplay between urban economies and diseconomies in Australian cities, and then models their net relationship with city size. The work presented here also extends the AH model to cover the cases of non-circular and polycentric urban structure observed in Australia. Using ABS Statistical Area 3 (SA3) data from 2011, we then empirically test both the AH model and its extension. This process enables the identification of key urban economies and diseconomies in the Australian data.

Following the AH approach, this thesis adopts a single criterion for 'optimal' city size. That is, 'peak' or 'optimal' city size is the scale (in terms of workforce) at which urban economies and diseconomies are equal, and output per worker is maximised. Using empirically derived estimates, we model the hypothesised inverted U-shaped relationship between city size and output per worker, and attempt to provide numerical values for 'optimal' city sizes.

1 See Rosenthal and Strange (2004), and Moretti (2004) for reviews.

2 The upper bound of the city size is such that it can be defended, and the lower bound is such that autarky is preserved.

3 See, for instance, Hensher et al (2012), which is a study that estimates agglomeration economies in the context of investments in transport.

1.2 Structure of the thesis

This thesis is structured as follows.

Chapter 1: Introduction	The present Chapter outlines the general background of the study, key contributions in the literature, and motivation for research.
Chapter 2: New Economic Geography, and location theory	This Chapter reviews the pertinent theoretical literature on spatial economics, with an emphasis on NEG models, and location theory.
Chapter 3: Agglomeration economies and the empirical literature	This Chapter reviews the empirical literature on agglomeration and city sizes. It also provides an overview of the nature, causes, and characterisation of agglomeration economies, as determined by available empirical studies.
Chapter 4: Theoretical model	This Chapter explains the theoretical framework used in this thesis. The AH model is described, and an extension of it is proposed. The implications of this extension are also discussed.
Chapter 5: Data and variables	This Chapter describes the data and variables used in the empirical analysis. We discuss the process of producing small-area data and variables, as well as the data sources used.
Chapter 6: Empirical results	This Chapter empirically tests the AH model, along with the extension proposed by us. Linearised versions of the AH model are also estimated. The Chapter concludes with a number of robustness checks on the validity of the obtained results.
Chapter 7: Productivity-maximising peak size	This Chapter estimates the hypothesised inverted-U relationship between city size and output per worker. Actual city sizes are compared with estimated peak sizes to derive output loss.
Chapter 8: Conclusion	This Chapter notes key findings, and the major contributions of the thesis. Policy implications and some potential areas for future research are identified.

2 NEW ECONOMIC GEOGRAPHY MODELS AND LOCATION THEORY

This Chapter provides a review of the role of space in contemporary economics. Two overlapping strands of the theoretical literature are examined in detail: New Economic Geography (NEG), and location theory. Where applicable, theoretical insights from this Chapter will be used and applied in this thesis.

2.1 Modelling space in economics

A good deal of economic theory has ignored the spatial dimension of economic activity. Specifically, economic activity is generally modelled as taking place in an aspatial context[4]. As Thisse and Walliser (1998) argue, it is not that economists have lacked interest in the spatial dimension of economic activity. Rather, the analytical tools employed in economics have often rendered economic analysis of space intractable. The following is a high-level classification of the types of theoretical models of space in the economic literature, as identified by Thisse and Walliser (1998).

2.1.1 Space as a physical substratum for economic activity

These models pre-suppose *physical* geographical constraints imposed on agent interaction. One of the very first types of these models comes from international trade theory – the Heckscher-Ohlin model (Ohlin, 1933). In such models, geographic variations impose constraints on international trade. Differences in factor endowments are modelled to explain different patterns of international and inter-regional trade.

2.1.2 Space as a location for activity

In this class of models, space is incorporated into economic analysis where agents' preferences and constraints determine locational choices. These models, in most cases, use transport cost as the mechanism to determine location choices, subject to preferences and constraints. Thisse and Walliser (1998) note that early works by Launhardt (1882) and Weber (1909) adopt this formulation, and model the location of a firm on the basis of a cost function in which distances to markets are weighted by the quantity of goods and transport costs.

[4] See, however, the Arrow-Debreu definition of a *commodity*, discussed overleaf.

2.1.3 Space as part of an economic good

Another way to incorporate space into economic models is to treat space as part of an economic good – more precisely, as a *characteristic* of a good, from which value is derived. This aspect of economic goods is incorporated into the seminal general equilibrium framework by Arrow and Debreu (1954), where space is explicitly incorporated into the very definition of a *commodity*. However, locational issues arise when the model attempts to guarantee the existence of a price system, since this depends on the assumptions of perfect competition, and convexity of preference and technology.

2.1.4 Space as a source of proximity effects

Increasing returns are often thought to be related to the existence of cities. In models that adopt this approach, spatial proximity is modelled to yield increasing returns due to indivisibilities across space. This idea can be traced back to Marshall (1890), and more formally, Koopmans (1957). German geographer Walter Chiristaller's 'central place theory' (Christaller, 1933; Lösch, 1940), which posits that settlements function as central locations providing services to surroundings, is also recognised as having a major influence on economic geography and its understanding of proximity effects (Eaton & Lipsey, 1977).

2.1.5 Space as a scarce resource

As Thisse and Walliser (1998) note, space can be privatised, unlike time. It follows then that property rights can be assigned to units of space. Such assignment of property rights has been at the heart of the democratic process of many nations. In models that treat space as a scarce resource, the interplay between the supply of, and demand for, land is modelled in a land market. The tension in a land market then ties into the competition between land uses; which in contemporary economies, is at heart of the tension between residential and commercial uses. This treatment of space can be traced back to von Thünen (1826), and subsequently Alsonso (1964), where the context was more urban.

In summary, the theoretical literature reveals the multi-faceted nature of 'space'. Each class of model seeks to explain specific properties of land, and the manner in which land interacts with economic activity.

2.2 Overview of the New Economic Geography literature

Endogenous growth theory helps explain how increasing returns may be achieved *via* specialisation and innovation (Romer, 1994). However, it does not account for spatial organisation and how this can contribute to productivity and economic growth. New Economic Geography (NEG) helps bridge this gap. NEG models[5] signify a revival of theoretical work in spatial economics, bringing it back into the mainstream, even though it borrows many concepts encapsulated in the treatments of space summarised above.

Seminal work by Krugman (1991) presents a simple model that shows how an economy could endogenously become differentiated into an industrialised 'core', and an agricultural 'periphery'. In the model, the manufacturing core is driven by firms trying to minimise transport costs, and maximise proximity to demand both from other manufacturing firms (intermediate goods), and consumers (final goods).

[5] Based on works by Krugman (1991, 1993, 1997, and 1998), and Fujita and Mori (1996).

Overall, the core-periphery pattern, or regional divergence of the Krugman model or NEG model, is driven by three parameters: transportation costs, economies of scale, and share of manufacturing in total income. The key assumptions in the model are summarised in Table 1 (Garretsen & Martin, 2010). It is clear from these assumptions that space, in the NEG framework, is modelled as a location of activity, and as a source of effects from proximity.

TABLE 1. KEY ASSUMPTIONS OF THE BASIC NEG MODEL

Consumption	Consumers are assumed to have preference for variety of goods, and to maximize their utility according to a constant elasticity of substitution (CES) function defined over all such varieties. Consumer preferences are assumed to be identical across space.
Producers	Producers are assumed to be atomistic, single-plant profit-maximising firms, with each plant producing only one (unique) good, and production with a single plant is cheaper than with multiple plants. Firms are assumed to move to be near markets (demand).
Workers	Labour markets are assumed to clear instantaneously. Workers are assumed to move between locations (regions) in response to spatial differences in real wages.
Market structure	A two sector economy is assumed: competitive traditional (agricultural) sector and manufacturing sector characterized by imperfect competition assumed to be of the Dixit-Stiglitz monopolistic type, with economies of scale both internal (fixed costs in production) and external (pecuniary or market size externalities) to manufacturing firms.
Trade (transport costs)	Assumed to be of 'iceberg' type. Indicates a preference for location in regions with large market access.
Geographical space	Assumes economic space consists of two hypothetical locations or regions, of equal size or extent. These regions can be of any scale: countries, regions of a country, cities, inner and outer areas of a city.

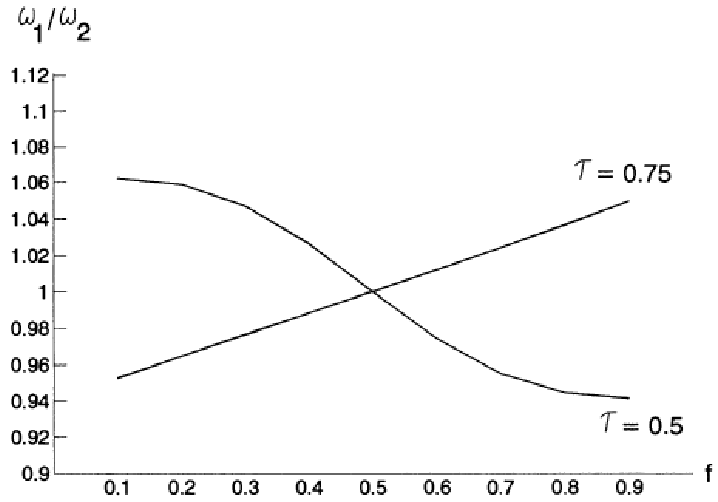
Source: Martin (2010); cited in Garrestan and Martin, (2010)

NEG models, as clearly indicated in the work of Krugman (1991), rely, to some extent, on the insights of Marshall (1890), though there are some differences. NEG models do not answer why a particular industry locates in a particular location; instead they are aimed at predicting the locational choices of homogeneous industries (Garretsen & Martin, 2010). The agglomeration economies associated with similar industries are known as *localisation economies*. The other difference is that NEG models focus on pure pecuniary benefits from agglomeration. So, non-measurable technological spillovers are not considered. The key point related to this is that in the presence of imperfect competition and increasing returns, pecuniary externalities matter (Krugman, 1991). Krugman argues that if one firm's actions affect the demand for a product of another firm, where price exceeds marginal cost, then pecuniary externalities have a real effect in the economy.

As Krugman (1991) notes, the story presented in NEG models is not necessarily new, and has been told previously by geographers. Circular causation or positive feedback will tend to drive concentrations of industries. That is, firms tend to locate close to markets, but the market will be large where there are clusters of firms (Arthur, 1990). However, NEG models are a tractable yet rigorous exposition of the ideas embodied in works such as Arthur (1990), and Marshall (1890).

The key result from the NEG model of Krugman (1991) is shown in Figure 2. Where *f* is the share of the manufacturing labour force in one region, and w_1/w_2 is the relative wage, the model shows two cases of stable equilibrium for a given transport cost of $(\tau = 0.75)$. If the relative wage decreases with *f*, workers will migrate out of the region, leading to convergence; conversely, if relative wage increases with *f*, then workers will migrate into the region, leading to divergence.

FIGURE 2. RELATIVE WAGE AND SHARE OF WORKERS



Source: Krugman (1991); page 492

Two central results, formally reached by Krugman (1991), are that, firstly, workers demand a larger wage premium in order to move from one region to another (forward linkage), and secondly, as the share of expenditure on manufactures increases, so does the relative market size of the region with manufacturing, producing a stronger home market effect (backward linkage). Another result is that higher transport costs reduce regional divergence, and encourage concentration of employment.

It is clear that the results derived here, given the parameters of the model, are entirely pecuniary. So, there are no explicit assumptions regarding technological spillovers, the geographic extent of externalities, or effects related to specific industries. As shall be discussed in Section 2.3, even though the economists considers these features, among other arguments, to be 'strengths' of the model, geographers consider them to be the weaknesses of the NEG framework.

Criticisms of, and extensions to, NEG models

Krugman (1991) acknowledges that the basic NEG model is an oversimplified view of the world with only localisation of *particular* industries, and the exclusion of non-pecuniary benefits. In addition, as noted by Garrestan and Martin (2010), there is no explicit consideration of geography or history. That is, geography enters the model through transport costs alone, and there is no consideration of initial spatial conditions. In addition, the inclusion of pure pecuniary benefits ignores the diversity and complexity of cities.

Specifically, the NEG model is a simple two-region model, which, by design, is for ease of tractability. As noted by Brakman *et al.* (2006), the unrealistic geography is problematic, when applied to a multi-region system. This is because the model is simply not capable of analysing multiple regions (Brakman, Garretsen, & Schramm, 2006).

The other major issue raised in the literature concerns the role of history. NEG theorists argue that, in the model, history is determined by a set of predetermined parameters that describe states of equilibrium. As argued by Brakman *et al.* (2006), a set of possible equilibrium economic landscapes are pre-given due to the structure and assumptions of the model, and so is the (unique) path that leads to the spatial distribution, and long-run equilibrium of economic activity. However, many empirical models of NEG go some way towards accounting for history and physical geography in a multi-region setting (see, for example, Brakman, Garretsen, & Schramm (2004), and Bosker, Brakman, Garretsen, & Schramm (2007)).

One approach to these criticisms has been empirical analysis of NEG models for a given spatial distribution of economic activity (Garretsen & Martin, 2010). This approach essentially translates to studies of market access (market potential). That is, tests of higher productivity or wages in regions near or within large markets, where the spatial distribution is taken as given. This renders the ideas of multiple equilibria, or path dependence irrelevant.

From a more theoretical point of view, however, the emerging research area known as Evolutionary Economic Geography (EEG) tries to address some of the issues in the simple NEG model. These EEG models use an evolutionary economic framework, which is still in its early stages of development. A non-mainstream approach to economic analysis, these models build on the notion that economic activity is not necessarily carried out in a state of equilibrium. The basic idea here is that history plays a part in evolving economic outcomes, and therefore path dependence and multiple equilibria feature strongly in economic geography (Garretsen & Martin, 2010). There is also due consideration of networks and network effects. Advocates of this approach argue that evolutionary theory is conducive for analysis because it has the potential to develop into a general theory in economic geography while being applicable empirically to specific processes in space and time (Boschma & Frenken, 2011).

Another criticism of the standard NEG model is its bias towards full agglomeration. The argument is that once transport costs get low enough, the inevitable outcome will be full agglomeration in either of the two regions in the model (Garretsen & Martin, 2010). There have been a number of NEG studies that attempt to overcome these criticisms, though ostensibly neglecting the issues of history, geography, and path dependence. Following Brakman *et al.* (2009), we now consider some of the various NEG models that expand on the initial Krugman model.

Puga (1999) modelled an extension to the NEG model, where both factor mobility and intermediate input linkages are incorporated. Similarly, Helpman (1998) assumed complete factor mobility, but introduced immobile amenities – such as housing, which acts as an additional dispersion force. In the Helpman NEG model, multiple industries may locate in multiple regions.

NEG theory has also been extended to include quasi-linear preferences (see, for instance, (Ottaviano, Takatoshi, & Thisse, 2002)). Though these models produce results similar to those of the simple NEG model, the analysis is heavily constrained to stylised settings with limited regions and industries (Redding, 2010). Other theoretical models have introduced greater analytical tractability to the core-periphery model in innovative ways. For instance, Baldwin, Rikard, Martin, Ottaviano, & Robert-Nicoud (2003) denominate fixed and variable costs in terms of different factors of production. These extensions attempt to create a more ‘realistic’ setting of the NEG world whilst providing additional insights.

2.3 Overview of location theory, and urban structure theory

In the Arrow-Debreu model of general economic equilibrium, location enters the definition of a *commodity*, but there is no explicit recognition of the dynamics of location choice, and where producers and consumers are located (Aizpurua, Benito, & Puértolas, 2003). In the Arrow-Debreu model, spatial outcomes are a passive by-product of preference and profit maximisation. In this Section, some specifics regarding theoretical models of location are examined. These aim to explain why agents locate where they do.

One of these is Starrett's marquee contribution - the 'spatial impossibility theorem' (Fujita & Thisse, 2002). Using a simple model of households and firms with transportation cost, Starrett (1978) demonstrates that if space is homogenous, transport is costly and preferences are locally non-satiated, and thus there is no *competitive equilibrium* involving transportation. Using firm profits and consumer income, he shows that under a wider set of circumstances, any market configuration of prices must offer an incentive for some agents to move. This is not possible under competitive equilibrium.

However, an extension to Starrett's work by Azipura *et al.* (2003) showed that it is possible to have a competitive equilibrium with transportation costs, when non-homogeneity of space (in the form of mobility costs) is introduced. They utilise a general-equilibrium framework to demonstrate the implications of non-homogenous space in a framework similar to Starrett (1978), but impose factor input transport costs. The authors do, however, acknowledge that Starrett (and others) were aware of the possibility of achieving a competitive equilibrium with transport costs.

Given the issues raised by the 'spatial impossibility theorem', Starrett (1974) identifies three approaches to location theory:

- diversity of the resource base
- partial-equilibrium models
- imperfect-competition models

Diversity of the resource base is at the core of Von Thunen's work, where farmers ship their product to a central marketplace and since shipping is costly, farmers will bid more for land closer to the marketplace (Arnott & McMillen, 2007), and Heckscher-Ohlin models of international trade.

Partial-equilibrium models, on the other hand, take CBD (or transport network) as given, and describe location choices assuming an exogenously determined spatial distribution (see for instance Alonso (1964) and Beckmann (1969)). For instance, in the Alonso monocentric city model, urban households receive utility from land, and a numeraire good. Land is homogenous except that sites closer to the CBD offer households lower commuting costs. In an attempt to avoid costly commutes, households are willing to pay a premium for sites closer to the CBD (Arnott & McMillen, 2007).

Another approach to location theory involves modelling imperfect competition (as noted by Fujita and Thisse (2002)). The basic premise here is one that derives from modelling space as a means to market access. As noted by Starrett (1974), firms have a distinct advantage selling to nearby consumers than firms farther away from those consumers. As a result, firms close to consumers do not act as perfect competitors with distant firms. This idea is embodied in the work of Muth and Mills, who extend the Alonso model to include housing. In this version, house prices, land rents, building heights, and population density all decline with distance from the CBD (Arnott & McMillen, 2007). Households receive utility from housing, which is built by housing producers. Imperfect competition, or increasing returns in the CBD drives up rent at the core, but declines as land use extends to the periphery.

In a similar vein, Starrett (1973) develops a general-equilibrium model with imperfect competition to show that optimal spatial distribution would be such that the benefits of increasing returns are balanced against increasing average transport costs. This work shows that optimal location is determined by three economic variables: the degree of increasing returns, the value of producer demand for transport, and the level of differential land rents (Starrett, 1973). This valuable insight points to the complex urban interactions between firms and their markets, and households and land rents.

More recently, Okamoto (2007) demonstrates that heterogeneity is another important determinant of the spatial distribution of households and firms. It is argued that heterogeneity of workers and local amenities is behind 'excess commuting' – which is the result of choosing locations that require above-average commute times. Brueckner *et al.* (1999) also examine the role of amenities in a theoretical framework, where amenity is dependent on the distance from the CBD. They divide urban amenities into three types: natural, historical, and modern; and posit that amenities are important in describing spatial distributions, and location choices.

Another approach in the literature is to treat location choice as a system-wide consideration utilising, in most cases, a general-equilibrium framework. Following the Alonso-Muth-Mills tradition, Rossi-Hansberg and Wright (2006) examine the relationship between urban structure and economic growth, in a general-equilibrium framework. The authors propose a mechanism to resolve the discrepancy between two empirical observations – increasing returns to scale in cities, and constant returns to scale in the aggregate economy. In their model, cities arise endogenously out of a trade-off between agglomeration benefits, and congestion costs. Exploiting this trade-off, they demonstrate that the size distribution of cities (and its evolution through the birth, growth, and death of cities) leads to the reconciliation of the two empirical observations. The urban structure of the economy prevents growth rates from diverging, which then ensures constant returns. In addition, the authors also demonstrate that city sizes increase with productivity shocks, which explains the empirical regularity observed by Zipf's Law – that the size distribution of cities is closely approximated by a Pareto distribution.

Conclusion

In summary, this theoretical literature review shows the multi-faceted nature of 'space', and the various types of models required to explain specific properties of land, and the manner in which it interacts with economic activity. The review also shows that theoretical models in spatial economics have evolved considerably, giving rise to NEG models, location theory, and urban structure models, which provide unique spatial insights into economic activity.

3 AGGLOMERATION IN THE EMPIRICAL LITERATURE

This Chapter firstly discusses the features, causes, and characterisation of agglomeration. This is followed by a review of the empirical literature on NEG models, location theory, and agglomeration theory. Australian empirical studies on agglomeration have also been identified.

3.1 Features of agglomeration

As noted by Puga (2010), Starrett's *spatial impossibility theorem* states that once we abstract from the heterogeneity of space, and without indivisibilities or increasing returns, any competitive equilibrium in the presence of transport costs will result in fully autarkic locations, where every good is produced at small scales (Starrett, 1978). It follows that agglomeration might be due to externalities (or increasing returns) arising from scale.

Before the forces behind agglomeration are examined in Section 3.2, two of its key features are described here – scale of agglomeration, and diversity in clustering. Whilst these are considered features of agglomeration, diversity, in particular is also identified as a cause of agglomeration (discussed subsequently).

The first and most obvious aspect of agglomeration is the magnitude of clustering. The size (or scale) of clustering is one of the features that leads to scale economies. Quigley (1998) notes that the hallmark of a 1950s study of economic activity in New York was its recognition of 'external economies' arising from scale.

The second, perhaps less obvious, feature of agglomeration is the heterogeneity of firms. Where inputs are sourced from a market that is monopolistically competitive, greater diversity in inputs is thought to improve production. In other words, variety in producer input types can yield external economies, in the presence of firms that earn normal profits (Quigley, 1998).

Similarly, Chinitz (1961) speculated that an urban environment with many firms producing heterogeneous output is more conducive to economic growth, because competitive conditions fostered in an environment with small firms, more entrepreneurial activity, and a more adaptable investment and banking infrastructure, is more conducive to economic growth than locations with homogenous firms (Chinitz, 1961).

3.2 Causes of agglomeration

As noted in Chapter 1, agglomeration and the importance of geography did not elude the keen eye of Adam Smith (1776). However, it is often argued that it was Alfred Marshall (1890) who identified the main causes of agglomeration. Marshall's suggested causes of agglomeration are still being debated and tested by contemporary economists (see Section 3.5 below). The so-called 'trinity' of agglomeration economies (Marshall, 1890) are: (i) a local pool of skilled labour; (ii) local supplier linkages; and (iii) local knowledge spillovers (see (Potter & Watts, 2014)). These classic Marshallian agglomeration economies, along with other possible causes of agglomeration, are discussed in this Section.

Sharing production inputs

There are three types of shared production inputs in the agglomeration literature: labour, infrastructure, and intermediate inputs. The sharing of these inputs is considered to be behind observed agglomeration. Broadly, these causes relate to the supply side in the economy.

From a firm's perspective, spatial clustering of firms enables specialisation in labour, since certain tasks outside the core business of a firm can then be outsourced locally. Additionally, the clustering of similar firms enables businesses to source workers locally. As noted by Quigley (1998), the sharing of production inputs is similar to the idea of 'economies of localised industry' put forward by Marshall (1890). The pure labour market description of sharing production inputs is also referred to as 'labour market pooling' (see, for instance, (Overman & Puga, 2010) or 'thick labour markets' (see for instance (Krugman, 1998). The idea here is that access to thick localised labour markets reduces the cost to firms, since there is a readily available supply of labour with specialised skills. This effect is due to transferability of skills in specialised industries. The treatment in Krugman (1993) draws on this point, by demonstrating that due to availability of workers, costs to businesses would be lower than with spatial dispersion.

Puga (2010) notes that the sharing of facilities related to the production process is another key reason for agglomeration. Once a large fixed cost (in most cases this is infrastructure) has been incurred, the larger the number of firms sharing the facility, the lower the cost per unit of output. However, the scale effect derived from sharing of a facility is diminished by congestion of the facility.

Seminal work by Abdel-Rahman & Fujita (1990) shows that firms cluster to share a large common base of input suppliers, generally observed in larger and more specialised cities. In their model, cities are subject to housing and commuting costs that increase with population. In equilibrium, aggregate production at the city-sector level exhibits increasing returns, despite constant returns to scale in perfectly competitive final production. The additional implication is that risk and uncertainity is reduced due to availability of suppliers. In the NEG literature, the sharing of intermediate inputs is referred to as forward market linkages (Krugman, 1998).

Market size effect

The market size effect relates to the fact that firms have a distinct advantage in selling to nearby consumers than firms farther away from those consumers. That is, the market potential or market size is determined by the consumer base accessible by a firm. In the NEG literature, Krugman (1998) refers to this scale effect as the home market effect (backward linkages). Several others (for instance (Becker & Henderson, 2000) and (Duranton, Labor Specialization, Transport Costs, and City Size, 1998)) also provide theoretical arguments to support the case that access to large markets leads to the clustering of firms.

Lowering transaction costs

Quigley (1998), notes that agglomeration of firms allows for more efficient labour market matching that reduces transaction costs. That is, the availability of workers (supply) and employers (demand) in geographical proximity enables more efficient job market matching outcomes. Helsley and Strange (1990) model equilibrium city sizes with search and match costs incurred in city labour markets. In their model, both workers and firms have heterogeneous skills and requirements. Their model demonstrates that a larger city allows for the skill space to be more densely covered by firms, which reduces the average cost of mismatches. Similarly, Coles and Smith (1998) use a stock-flow model in a search-match framework to derive increasing returns to scale from agglomeration. The idea here is that in a market with more job opportunities that can be explored simultaneously, it is less likely that none of the job opportunities would be successful (Puga, 2010).

Knowledge and information spillovers

Though not explicitly included in NEG models, Krugman (1998) refers to pure external economies as efficiencies owing to knowledge and information 'spillovers'. That is, workers have a higher probability of interacting with each other in agglomerated space, which has the potential to lead to knowledge and information sharing between and within industries. Glaeser (1999) adopts this idea in a model that has young workers migrating to urban areas to learn from experienced workers in dense city locations. The transmission of knowledge is attributed to the agglomerated nature of cities, since high-density employment-centres increases the probability of interaction between workers. Jacobs (1969) also posits that knowledge spillovers occur in cities, though she argues that cross-fertilisation between industries leads to knowledge spillovers. This is discussed in detail under 'industrial-scope' economies in Section 3.3.

Business cycle economies due to the 'law of large numbers'

As long as the purchase of inputs by firms are not completely synchronised, production inputs can be stabilised over the business cycle (Quigley (1998)). This is because, during fluctuations in output, some firms will be hiring whilst other will not. And similarly, some buyers will be buying outputs whilst other will not.

3.3 Characterisation of agglomeration economies

In this Section, we briefly discuss the *scope* of agglomeration economies. Rosenthal and Strange (2004) provide a good theoretical characterisation of the three types of scope that relate to economies of agglomeration. These are discussed below.

Industrial scope

This describes the degree to which external economies of agglomeration extend across industries, including and beyond its own. Economies of scale that arise from spatial concentration of activity within a given industry are known as *localisation economies*. The externalities that arise from the concentration of all economic activity (across multiple industries), or from city size itself, are known as *urbanisation economies*. The closer two firms are in industrial space (similar production processes), the greater the potential for industrial scope.

The classical Marshallian agglomeration economies of a local pool of skilled labour, local supplier linkages, and local knowledge spillovers are generally associated with *localisation economies* since they relate to firms operating in the same industry. Industrial scope is also particularly relevant to one of the features of agglomeration – diversity.

There are three key theoretical arguments at the heart of the urbanisation economies vs. localisation economies debate (see the Appendix for a more detailed discussion). Firstly, contributions by Marshall, Arrow, and Romer (MAR theory) posit that that local monopoly is better for growth than local competition. This is because local monopoly restricts the flow of ideas to others, and allows localisation economies to be internalised by the innovator (Glaesar, Kallal, Scheinkman, & Shleifer, 1992). Secondly, Porter (1990) argues that knowledge spillovers occurring in geographically concentrated are between highly competitive industries is better for growth. Even though competition reduces the returns to the innovator, it gives firms an incentive to innovate (Glaeser *et al*, 1992). Thirdly, Jacobs (1969) argues that diversity in industry and not specialisation is the key to economic growth in cities. This is because the most important knowledge transfers come from outside the core industry in which a business operates (Glaeser *et al*, 1992). The idea here is that variety of proximate industries (rather than industrial specialisation) promotes innovation and growth.

These arguments are all more or less valid depending on the context of the urban economy. The results from Glaeser *et al.* (1992) go some way towards demonstrating this fact.

Geographic scope

This describes the extent of agglomeration economies in terms of distance. That is, if agents are closer (in distance or time) there is greater likelihood for interaction between those agents, and from those interactions, external economies.

Temporal scope

This describes external economies arising from agents interacting temporally. It is plausible that an agent's interaction with another agent in a previous period could lead to productivity uplifts in the current period. For instance, inter-firm learning takes time, so interaction between a supplier and a firm may take time to yield efficiencies. This type of scope is also referred to as dynamic-externalities, and often relates to economic growth.

Given the nature of these three scope economies, these descriptions could apply to any of the causes of agglomeration. However, some scope economies might be more relevant than others.

3.4 Factors that reduce agglomeration economies

Krugman (1998) argues that spatial economic structure is a result of the 'interplay between two opposite forces': centripetal forces leading to spatial concentration, and centrifugal forces leading to dispersed activity. Table 2 summarises these forces. The attracting forces are of the classic Marshallian (1920) variety (discussed in detail in Section 3.2).

TABLE 2. FORCES AFFECTING GEOGRAPHICAL CONCENTRATION

Centripetal forces	Centrifugal forces
Market size effects (linkages)	Immobile factors
Thick labour markets	Land rents
Pure external economies	Pure external diseconomies

Source: Krugman 1998, p.8

Krugman (1998) argues that immobile factors such land and natural resource requirements act as constraints to agglomeration, because they prevent firms from clustering. Clustering in high-value CBD locations, driven by the desire to benefit from agglomeration economies, is likely to push up demand for land. This would drive up rents, and reduce agglomeration economies. Urban structure models that incorporate housing, such as Alonso (1960), Abdel-Rahman & Fujita (1990), Helpman (1998), also consider rent and commute costs as forces that reduce agglomeration. Lastly, clustering could lead to pure diseconomies such as congestion (Krugman, 1998) and increased crime (Quigley, 1998).

3.5 Review of the empirical literature

Each source of agglomeration economies identified above is consistent with a positive relationship between agglomeration and productivity. Many studies empirically observe this positive relationship, though the literature has struggled to isolate the sources of agglomeration (Maré & Graham, Agglomeration elasticities and firm heterogeneity, 2013). While some of the evidence is clear, other aspects remain opaque at best. In this Section, we review the empirical literature on agglomeration economies, diseconomies, and net agglomeration economies.

From an empirical point of view, the ideal way to estimate agglomeration economies is to estimate production functions for individual firms or spatial units (Rosenthal & Strange, 2004). Given the general lack of required data, many studies use employment growth, employment density, birth of new firms and their employment, wages, rents, or population density as an alternative. This Section highlights studies that use some of these measures of agglomeration, and others.

Many studies that test the empirical validity of agglomeration externalities utilise either a location theory or NEG framework to conceptualise the mechanism of effects. For instance, Coretz (2000) provides an empirical assessment of forward linkages that relate labour migrations to the geography of production through real-wage differentials. The results of this study support the notion of forward linkages in NEG models, and that migrants tend to follow market potential. For an extensive review of empirical work with NEG underpinnings, in addition to those identified in this Chapter, refer to Redding (2010).

Classical Marshallian economies

Given their prominence in the theoretical literature, it is not surprising that the Marshallian agglomeration economies of local pool of skilled labour, local supplier linkages, and local knowledge spillovers, have been thoroughly investigated in the empirical literature – albeit in most cases at an aggregate level. There are a few studies that test these three effects separately (Strange, Faggio, & Silva, 2014). Among others, Fallick *et al* (2006) and Almazan *et al* (2007) test labour market pooling and, in particular, job hopping, and among others Holmes (1999) tests input sharing. These studies show strong presence of Marshallian economies, though there it is still unclear which Marshallian effect dominates.

Studies such as Audretsch & Feldman (1996), and Rosenthal & Strange (2001) model the relationship between agglomeration and labour market pooling, input sharing, and knowledge spillovers. The latter study, which tests for the presence of knowledge spillovers, labour market pooling, and input sharing, uses the Ellison-Glaeser localisation index (a measure of agglomeration). For the US cities examined, they find that shipping-oriented attributes influence agglomeration at the state level, knowledge spillovers impact highly localised agglomeration, and that labour-pooling impacts agglomeration at all levels of geography.

One of the few studies to use firm level data, Henderson (2003a), also tests for the presence of Marshallian economies. In this paper, the author constructs a panel of plant level data (from the Longitudinal Research Database) including measures of capital stock, materials and labour. Interestingly, industrial scope is also controlled for in this study by differentiating between activities that take place within an industry, and those that do not. Temporal scope (or dynamic externalities) is also considered.

The author first uses a two-stage least squares (with local environment measures as instruments), then employs a Generalised Method of Moments (GMM) procedure, as the instruments in the 2SLS step were found to be weak. Since GMM also produces poor results, the author uses fixed effects, which shows evidence for Marshallian economies and their scope counterparts.

Strange *et al.* (2014) note that there is strong evidence in studies such as these (and many others) to support the empirical validity of Marshallian forces. Collectively, these studies validate the core-periphery theoretical proposition in NEG models.

Urbanisation and localisation economies

Classical Marshallian-forces are usually associated with *localisation* economies – that is, externalities between similar industries. Glaeser *et al.* (1992) is one of the first empirical tests of the effect of industrial diversity, or *urbanisation* economies. Using data on the employment growth of industries in 170 US cities between 1956 and 1987, the authors find that local competition and urban variety encourage employment-growth in industries. This suggests that knowledge spillovers might occur *between* rather than within industries. In a cross-section of city industries, they find that, as measured by employment, industries grow slower in cities in which they are more heavily overrepresented.

As noted by the authors, this result is in line with arguments advanced by Jacobs (1969), who suggested that ideas are stimulated in a diverse environment, and diversity leads to specialisation in inputs and outputs, which in turn leads to higher returns (Jacobs, 1969).

In a similar study, Henderson *et al* (1995) uses industry data from 1970 to 1987 to analyse the temporal aspect of urbanisation economies. They find that the extent of diversity in industry did not matter to more mature industries, though it did matter in attracting high-tech industries at the beginning of their life cycles.

Urban human capital and productivity

There is an extensive literature on human capital and economic growth (see for instance Barro (2001)), though much of the work is aspatial. If it is the case that agglomeration results in knowledge spillovers, it follows then that there should be spatial variation in human capital accumulation, and wages. That is, human-capital intensive cities should, in theory, experience higher wages due to agglomeration economies (Glaeser, 1999). Rauch (1993) tests this hypothesis with data on wages, and human capital (proxied by formal education and work experience) for Standard Metropolitan Statistical Areas in the US. The study finds that US cities with an additional year of education can be expected to raise total factor productivity by 2.8 percent.

Transport and agglomeration economies

Many Australian empirical studies on agglomeration economies (and urban structure) focus on the role of transport investments in enabling productivity enhancements. Some of these are identified below.

Daniels and Mulley (2011) use simple correlations to argue that, within strategic centres in Sydney, some industries were concentrated in different types of centres, indicating that they are more likely to benefit from agglomeration economies. They also find that in Sydney, there is a strong relationship between centre density and public-transport use for the journey to work, with public-transport use higher in high-density centres. Using survey data, the authors also find that public-transport users are slightly more likely than car users to report meeting and talking to people they know while travelling to work. The survey finds that transport enhances interactions which may lead to knowledge spillovers (Daniels & Mulley, 2011).

Hensher *et al.* (2012) measure the impact of transport investments on urban productivity. Their measure of agglomeration—effective employment density, a measure also used in Graham (2007) and Mare and Graham (2013)—reflects the area and distribution of jobs. They relate this to productivity estimates by industry. Consistent with other studies, they show that there is a positive relationship between agglomeration and productivity, and that this relationship varies by industry. In addition, the authors use these estimates of productivity-uplift and combine them with an integrated transport and location choice modelling system (TRESIS), and a spatial computable general-equilibrium model (SGEM) to compute the impacts of changes in transport infrastructure on the wider economy. The integrated model is then applied to the proposed North-West Rail Link project in Sydney. The study identifies a 17.6% mark-up over the conventional transport user benefit.

A similar study by *SGS Economics and Planning* (2012), a report for the Australian Government's COAG Reform Council, derives comparable estimates of the relationship between urban productivity and Effective Job Density (EJD). In this study EJD is agglomeration measured in terms of total transport (public and private) and accessibility to jobs. Using simple regressions for each industry, they produce elasticity estimates of the relationship between EJD and productivity for Statistical Local Areas in Sydney, Melbourne, and Adelaide.

Urban structure and productivity

One of the early studies to examine city scale, and productivity is Henderson (1986). In that study, the author uses 2-digit manufacturing industry data for urban areas in the US and Brazil to test whether agglomeration economies are more localised or urbanised. The author finds that localisation economies are strongest for industries in which cities tend to specialise, and that this effects diminishes as city size increases. In other words, manufacturing industry resources are not more productive in larger cities. This suggests that there are limits to agglomeration economies, which may be industry-specific.

The probable relationship between urbanisation and productivity naturally suggests the following question: Do urbanisation and urban structure impact on economic growth? Henderson (2003b) examines this question by looking at the relationship in a cross-country panel model. Given the wide variability in individual country-effects, and underlying mechanisms there-in, it is no easy task to establish the relationship between urbanisation and economic growth. The author concludes that urbanisation represents movements of industry sectors within an economy as development proceeds, but it does not necessarily stimulate economic growth. However, Henderson argues that the form taken by urbanisation, or the degree of urban concentration, strongly affects productivity growth.

An Australian study by Kulish *et al*. (2011) examines determinants of some aspects of the structure of Australian cities, including density, land prices, and housing prices. They calibrate a (monocentric) version of the Alonso-Muth-Mills model of urban structure to examine issues relating to the provision of infrastructure, zoning policies, frictions in housing supply, and population size. They find that the highly stylised aspects from the calibrated model are broadly consistent with empirical estimates in their study. Based on these findings, they argue that zoning restrictions alter the urban structure of a city, and put upward pressure on house prices.

Net urban economies and optimal city size

Though most studies reviewed in this Chapter focus on urban economies, there is a small strand of the literature that examines urban diseconomies (see Rosenthal and Strange (2004), and Moretti (2004) for reviews). There are even fewer studies that simultaneously consider both economies and diseconomies, and their relationship with city scale. Au and Henderson (2006), one of the first to do so, formally model[6] and then estimate net agglomeration economies for Chinese cities. They find that net urban-agglomeration benefits rise sharply with increases in city size from a low level. However, the net benefits level out near the peak scale, and then decline slowly past the peak. The authors also conclude that the majority of Chinese cities operate below their productivity-maximising peak size due to *hukou* system's restrictions on labour mobility.

Another approach to model the relationship between city scale and urban productivity is to estimate Zipf's law (see Chapter 7 for a detailed explanation). For instance, Li and Gibson (2015)—a study in response to shortcomings in Au and Henderson (2006) and Xu (2009)—use a more complete data set of Chinese workers to demonstrate that most Chinese cities are closer to the efficient size implied by Zipf's Law. In addition, using the Au and Henderson (2006) Generalised Leontief specification, they show that 80 percent of cities are close to the estimated productivity-maximizing scale, and that output losses from sub-optimal scale are typically below 10%.

However, not all studies of optimal size focus on urban productivity and workforce. Alonso (1971) presents an aggregative economic approach to the theory of city size. That is, city size is presented as a multidimensional phenomenon of both population and workforce. He argues that the population that minimises cost per capita is faulty criterion for optimal size. This study also argues that a more sensible objective of public policy would deal with the relation of outputs and inputs, rather than only with inputs.

A more recent study, Camagani *et al.* (2013), considers urban costs (malaise, rent-cost, sprawl), as well as urban benefits (urban amenity, diversity, density, functions, network effects), in their formal-model of optimal city size. Using data for 59 European cities, they estimate city-specific optimal sizes (Camagni, Capello, & Caragliu, 2013).

Conclusion

In summary, the literature reviewed in this Chapter shows that there is a substantial literature on agglomeration, and its nature and causes. The manner in which urban economic activity leads to economies and diseconomies has also been examined briefly. These findings will be drawn upon throughout the rest of this thesis.

[6] The Au and Henderson (2006) framework is used (and improved) in this thesis (see Chapter 4).

4 THEORETICAL MODEL

This Chapter outlines the formal model used for this thesis. Overall, the framework is based on Au and Henderson (2006), but modified in its application to Australian cities. The AH (2005) model is designed to examine urban economies and diseconomies within a single framework. The key improvement to the model is the relaxing of the monocentric-circular urban structure of effective labour, and the introduction of an alternative non-circular structure designed to reflect polycentric employment destinations and polygon-shaped residential locations (origins).

4.1 The Au and Henderson model

The theoretical framework used in this thesis is due to Au and Henderson (2006). The AH model is an appropriate foundation for this thesis because it accounts both for urban economies and diseconomies within the single framework, while at the same time accounting for differences in urban hierarchy. Agglomeration economies, which are at the heart of any model in spatial economics, are well-specified for the most familiar types of agglomeration benefits. The counterbalancing costs of agglomeration (so called 'urban diseconomies') are specified in a simple circular-monocentric framework. When economies and diseconomies operate within the one framework, it is possible that there is a unique peak point (optimum) where city size is such that urban economies and diseconomies are equated, and some socio-economic objective, such as output per worker, is optimised.

The original AH model, is described in detail below. Proposed modifications to the AH model, which form part of the foundation for our subsequent empirical work, are described in Section 4.2.

Final-good producers

There are two types of producers in the AH model – final-good producers and intermediate-good producers. The output of the latter group is used as an input to the production process of the former.

Net output of a final-good producer of good variety $y(i)$, for location (city) j is given in equation (1) below. Note that the subscript j is excluded since all variables are location-specific.

$$\tilde{y} = y - c_y = A(\cdot)\, k_y^{\alpha}\, l_y^{\beta} \left(\int_{s_x} x(i)^{\rho}\, di \right)^{\frac{\gamma}{\rho}} - c_y \qquad (1)$$

where $\alpha + \beta + \gamma = 1$, and $0 < \rho < 1$

In (1), final output (of variety i) is given by y, fixed cost of production is given by c_y, capital inputs is given by k_y^{α}, labour inputs is given by l_y^{β}, and the expression in brackets denotes the aggregation of different intermediate-good varieties (s_x) used to produce the final good. Note that $x(i)^{\rho}$ denotes a given intermediate-good variety. In all these expressions, the subscript y refers to the final-good variety being produced, and the superscript refers to the elasiticity of the input. Fixed costs are accounted for in units of the composite good $y(i)$. By design, the final-good production process incroporates the following widely acknowledged sources of agglomeration economies:

1. *Local external scale economies*

Also known as Marshallian economies, this effect is due to scale of the local labour force operating in the city. As noted in the Chapter 3, these classical economies relate to the local pool of skilled labour, local supplier linkages, and local knowledge spillovers. As the size of the city increases, economies generated by the larger labour force also increase. Following the typical specification in the literature, AH specify this as:

$$A(\cdot) = AL_g^{\varepsilon} \tag{2}$$

In this specification, ε relates to size of localised agglomeration economies, and sub-script g relates to the assumption that these Marhsallian scale externalities are internal to a given product – which is the common approach in the literature (in contrast to the ideas put forward by Jane Jacobs (1969), and tested in Glaesar (1992)). That is, an industry only gains from other final-good and intermediate-good suppliers operating in the same industry. The actual implementation of this assumption enters the model via the specification of urban hierarchy (discussed later).

2. *Number of local input varieties*

Intermediate-good varieties are given by s_x and as the number of input varieties available increases, so do the economies in the final-good production process. This could be thought of as diversity in intermediate inputs. As will be demonstrated later, this specification is key to incorporating urban hierarchy. AH show that for the symmetrical intermediate-input case, the production function reduces to the following:

$$y = A(\cdot)\, k_y^{\alpha}\, l_y^{\beta} (x s_x)^{\gamma} s_x^{\frac{\gamma(1-\rho)}{\rho}} \tag{3}$$

Here the term ${s_x}^{\gamma(1-\rho)/\rho}$represents the scale effect associated with increased input varieties (symmetric) that enter the production process.

3. *Implicit agglomeration benefits to consumers*

These benefits to consumers are a consequence of the inclusion of 'iceberg' type transport costs[7] in the specification of demand for final goods (see p. 23). The idea here is that consumers benefit from goods that are located close to their residential locations, which is consistent with the approach in New Economic Geography (NEG) models.

Intermediate-good producers

In the AH model, intermediate goods are of the non-traded services variety. That is, they are direct inputs to the production of final goods. Within the AH framework, the cost constraint, as shown below, is defined in terms of labour-units.

$$l_x = f_x + c_x X \tag{4}$$

where f_x is the fixed cost of production, and c_x is the marginal labour cost

In addition, AH alter the specification of the manner in which intermediate inputs enter the production of final goods, to include an urban-hierarchy parameter. Denoted by $(xs_x)^{\gamma} {s_x}^{\gamma(1-\rho)/\rho}$ in the symmetric case, γ is defined in terms of the ratio of manufacturing to services, which represents whether a city is specialised in manufacturing or services. This is described further in the 'urban hierarchy' discussion of this Chapter.

Consumer demand for final goods

Having specified the producer-agents operating in the AH model, we now turn to the demand side, by specifying consumer preference and implicit prices. Consumer preferences are given by:

$$U = \left(\int y(i)^{\frac{\sigma_y - 1}{\sigma_y}} \, di \right)^{\frac{\sigma_y}{\sigma_y - 1}} \tag{5}$$

where $\sigma_y > 1$

Notice that the subscripts imply that the preference parameter of consumption is the same for all product varieties of $y(i)$. This is altered slightly with the introduction of urban hierarchy.

In line with Starrett (1973), Thisse & Fujita (2002), and others, the AH model adopts the general conversion in the literature of monopolistic competition, which follows from the necessary condition of increasing returns in spatial models. Using standard results in Overman, Redding & Venables (2003), and Head & Mayer (2004), the price of final good $y(i)$ in location (city) j is given by:

$$p_{y,j} = \mathrm{MP}_j^{\frac{1}{\sigma_y}} (y - c_y)^{\frac{-1}{\sigma_y}} \tag{6}$$

where the own-price elasticity of demand is $\eta_y = -\sigma_y$

[7]. Common in NEG-models, iceberg cost factor discounts the value of a good based on the distance between production-origin, and consumption-destination of a good.

The price-elasticity of demand (given by σ_y) is a measure of the derived demand for intermediate inputs by final-good producers. The spatial variation in the elasticity of demand for inputs incorporates the spatially differentiated nature of the importance of intermediate inputs to the production process.

Market Potential (MP) for location (city) j is given by the following equation:

$$\mathrm{MP}_j = \sum_v \frac{E_v I_v}{\tau_{j,v}^{\sigma_y - 1}} \tag{7}$$

where, $I_v = \left[\sum_v s_{y,u}\left(p_{y,u}\tau_{v,u}\right)^{1-\sigma_y}\right]^{-1}$

As Harris (1954) argues, Market Potential (MP) relates the potential demand for final goods in a particular location to its proximity to consumer demand in other locations. In the specification of MP_j: E_v is total consumer expenditure in location v, I_v is a price index at location v, $\tau_{j,v}$ is the iceberg cost factor (common in NEG models) which represents a final good 'melting' as the distance between consumer and producer increases (given by locations j and v).

In the specification of the price index I_v: the summation is over all locations v, where $s_{y,u}$ represents the number of final-good varieties produced in location u, and $p_{y,u}\tau_{v,u}$ is the transport cost-discounted effective price of a good travelling from v to u. Given this specification, it is clear that all goods (of a given variety and location) sell at the same prices.

Urban hierarchy

Not every city location is homogenous in its industry structure. Some cities are more industrial than others, and conversely some cities have a much larger base of professional services. AH therefore suppose that all cities lie within a spectrum of product specialisation, because they argue that having two different product types in the same city would result in a city size that would be inefficient for at least one of the products. So, for instance, a city location producing textiles is more efficient if it specialises in its own industry and produces textile-related varieties of goods (both final and intermediate). Of course, in reality, it might be the case that the sectoral mix is more varied and less specialised. The authors point to the treatment of urban hierarchy in Black and Henderson (2003), Alexandersson (1959), and Bergsman, Greenston and Healy (1972) for more detailed work.

To include variations in urban hierarchy, the AH model defines γ in the following manner:

$$\gamma_g = \frac{1}{1 + \mathrm{MS}_g} \tag{8}$$

where MS_g is the ratio of manufacturing to services at a given location

The subscript g denotes that γ is product-specific, and is a consequence of the urban hierarchy of a given location. The relationship between γ and MS holds regardless of how other parameters vary across the urban hierarchy. For this reason, this specification is sufficient to introduce variation in urban hierarchy to the AH model. In addition, ignoring transport costs between cities, as MS rises, γ tends to fall, which leads to specialisation of cities—the structure desired by the authors.

Au and Henderson (2006) make the additional modification to consumer demand by specifying consumer preferences to vary by product g within a given variety of final good. Since there are now many products (within a given variety), consumption weights μ are also introduced.

$$U = \prod_g \left(\int y_g(i)^{\frac{\sigma_g - 1}{\sigma_g}} \, di \right)^{\frac{\mu_g \sigma_g}{\sigma_g - 1}} \tag{9}$$

However, Au and Henderson (2006) impose the common convention that $\sigma_g = \sigma_y$, and assume that only the consumption weights $\mu_g$vary by product. This modification impacts on the form of Market Potential (MP), though the authors do not pursue the issue further because variation in consumption weights cannot be identified in the data. Similarly, this cannot be identified in the Australian data, and for this thesis it is assumed that $\mu_g = 1$.

Effective labour in the city

This aspect of the model specifies the labour force in each city. The AH model assumes that the city is circular and monocentric in its structure, so that workers live around the city, and travel to the middle of the city for employment. This component of the AH model is crucial to modelling the inverted-U shape relating city scale to output per worker. Recognising that many cities are neither circular nor monocentric, we show in Section 4.2 how this part of the AH model can be modified to better reflect the polycentric non-circular structure of Australian cities studied in this thesis.

In the AH model, the CBD is assumed to be the centre point in a circle, and people live on lots of fixed size one, and each city is circular. The labour force is given by N and is uniformly distributed across the city. Since the city is circular, and all workers live on points in the circle (and travel to the centre for work), the area of the city is also given by N.

From Euclid, the area (A) of a circular city is $A = \pi R^2$. Based on the circular monocentric structure of the AH model's city, we can derive the radius (R) of the city as follows:

$$A = N$$
$$N = \pi R^2$$
$$R = N^{\frac{1}{2}} \pi^{-\frac{1}{2}} \tag{10}$$

People live at distance b from the city centre, where the maximum distance from the CBD centre is R. Each worker spends t amount of productive time (opportunity cost of travel) to commute a unit distance (there and back). This means that total commuting cost per worker is tb.

Given that a worker lives at distance b from the centre, and given that residential locations are a circle around the centre, the diameter of the circle at distance b from the centre is given by $2\pi b$. Since total commute-cost (or lost work-hours) per worker is tb, total commute cost (for all workers) at b is simply $2\pi b(tb)$. Adding up these concentric circles from the CBD centre, we get the following expression for Total Commute-cost (TC) for all resident workers in the monocentric-circular city of the AH model:

$$TC = \int_0^{N^{\frac{1}{2}} \pi^{-\frac{1}{2}}} 2\pi b(tb) \, db = 2\pi t \int_0^{N^{\frac{1}{2}} \pi^{-\frac{1}{2}}} b^2 \, db = 2\pi t [\frac{b^3}{3}]_0^{N^{\frac{1}{2}} \pi^{-\frac{1}{2}}}$$

Evaluating this integral:

$$TC = 2\pi t \frac{(N^{\frac{1}{2}}\pi^{-\frac{1}{2}})^3}{3} = \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}} \tag{11}$$

In the expression for TC, the first two terms are constants due to the assumed monocentric-circular structure. It is also clear that TC is linearly increasing in t, and non-linearly increasing in N (with an exponent of 1.5). Au and Henderson (2006) interpret the exponent on N as a congestion factor defined as z. Though a valid interpretation, strictly speaking, the fact that $z = 1.5$ is actually due to the circular structure of cities in the AH model. This thesis will generalise this restrictive aspect of the AH model at both the theoretical and empirical levels.

Since the expression of TC does not have units (monetary, distance, or time), the cost is effectively in units of N workers. In this thesis, we interchangeably refer to TC as "lost workers", since commute time is foregone *productive* worker-hours that would otherwise have been used in the production of final and intermediate goods[8].

The final definition in the AH model is the specification of the *effective* labour force (L) of a city, which is simply the difference between the gross labour force N, and lost workers described by TC.

$$L = N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}} \tag{12}$$

As mentioned earlier, the AH parameterisation of the model does not, strictly speaking, allow for congestion. From a theoretical perspective, Au and Henderson (2006) seek to model congestion as a form of urban diseconomy, which leads to a '$\cap -$shaped' function. To accommodate this, Au and Henderson (2006) alter the effective city labour specification to allow for a lower bound z (interpreted as a congestion parameter).

In other words, if $t(N)$ and $t'(N) > 0, L$ can be expressed as:

$$L = N - (\frac{2}{3}\pi^{-\frac{1}{2}}\tilde{t}) \times N^z \tag{13}$$

where $z \geq 1.5$

[8] Alternatively, the worktime could be used for leisure (non-work) activities. Though this is not explicitly considered in the AH model, it is implied. It would appear that the AH model assumes this to be constant. Consequently, there is little or no substitution between worktime saved and leisure. It is assumed that the 'saved' time would otherwise have been used for production of final or intermediate goods. This issue is considered in more detail in the empirical analysis of this thesis, though explicitly not considered in the AH model (2006).

The idea behind (13) is that Au and Henderson (2006) are trying to allow for the possibility that congestion (z) might increase with city size (N). Of course, the monocentric-circular nature of the specification requires that z has a lower bound of 1.5. In their empirical experiments, Au and Henderson (2006) find that increasing z above 1.5 is not satisfactory since the coefficient term $(a_0 = 1.5\pi^{-0.5}\tilde{t})$ gets smaller. In this thesis, we show that the monocentric-circular structure is not supported by Australian data, and an alternative specification is proposed, derived, and empirically validated.

Market-clearing conditions

There are five market-clearing conditions needed to solve the AH model for equilibrium employment allocations (between final and intermediate goods), and number of firms.

1. Final-good producers are profit-maximising firms that maximise production functions subject to linear cost functions.

$$\Pi = p_y A L^{\varepsilon} k_y^{\alpha} l_y^{\beta} \left(\int_{s_x} x(i)^{\rho} \, di \right)^{\frac{\gamma}{\rho}} - c_y - \int_{s_x} p_x(i) x(i) \, di - w l_y - r k_y \qquad (14)$$

where w is the local wage-rate, r is the cost of capital, p_y is the price of a given final-good variety, $p_x(i)$ is the local price of intermediate-input variety $x(i)$.

2. Intermediate-good producers are profit-maximising firms that maximise profit subject to their technology and linear cost functions.

3. The local (city) labour market clears with full employment.

$$s_x l_x + s_y l_y = L \qquad (15)$$

This means that the total number of local workers producing both final and intermediate goods must sum to the effective labour force. This implies that the city economy comprises only two types of production activities, and that there is full employment.

4. Intermediate-production is market-clearing.

$$X = s_y x \qquad (16)$$

This condition means that that intermediate-good supply (of any variety) must equal demand (from final-good producers), where X is the total supply of intermediate goods, and s_y (the number of final-good producers) purchase x of the intermediate good.

5. Final-good production is market-clearing.

This means that consumer demand for final goods equals the supply of final goods, under monopolistic competition market conditions.

Solving the model

The derivations required to solve the model are reported in the Appendix of Au and Henderson (2006), and are not fully repeated here for the sake of brevity. However, key results that close the model are shown below.

Maximising the representative final-good production function subject to the cost constraint yields the following gross-output function:

$$y = \sigma_y c_y \tag{17}$$

Note here that the expression of final-good prices (which includes Market potential) is used in the constrained optimisation of the objective function.

Maximising the representative intermediate-good production function subject to its cost constraint yields classic Dixit-Stiglitz results[9] where p_x is the price of intermediate goods, X is the total supply of intermediate goods, and l_x is the number of workers in the intermediate-good sector.

$$p_x = \frac{wc_x}{\rho} \tag{18}$$

$$X = \frac{f_x \rho}{(1-\rho)c_x} \tag{19}$$

$$l_x = \frac{f_x}{(1-\rho)} \tag{20}$$

After appropriate substitutions and associated manipulations, the model can be solved for the numbers of final goods (s_y) and intermediate goods (s_x). The results of these are as follows:

$$s_y = Q_0^{\frac{1}{(1-\alpha)}} MP^{\frac{\alpha}{\sigma_y(1-\alpha)}} r^{\frac{-\alpha}{(1-\alpha)}} A^{\frac{1}{(1-\alpha)}} L^{\frac{(\varepsilon+\gamma)/(\rho+\beta)}{(1-\alpha)}} \tag{21}$$

$$s_x = \frac{\gamma}{\gamma+\beta} \frac{(1-\rho)}{f_x} L \tag{22}$$

The criterion used to derive optimal city size is *net output per worker*, where net output is given by $(p\tilde{y} - rk_y)s_y$. According to the AH model, net output per worker can then be defined as follows.

Definition 1 (Net output per worker): *Net output per worker* is the disposable income per worker in the city, after capital rentals are paid.

[9] See Overman, Redding and Venables (2003); and Head and Mayer (2004), cited in Au and Henderson (2006).

Given this definition, and again doing the appropriate algebraic manipulations, we can derive the following expression for equilibrium Net Output Per Worker (NOPW) under market-clearing conditions:

$$NOPW = (p\tilde{y} - rk_y)s_y N^{-1} = Q_2 MP^{\frac{1}{\sigma_y(1-\alpha)}} r^{\frac{-\alpha}{(1-\alpha)}} A^{\frac{1}{(1-\alpha)}} (N - \frac{2}{3}\pi^{-\frac{1}{2}} t N^{\frac{3}{2}})^{\frac{(\varepsilon+\gamma)/(\rho+\beta)}{(1-\alpha)}} N^{-1} \quad (23)$$

where $Q_2 = Q_0^{\frac{1}{(1-\alpha)}} Q_1$ and Q_1 is a parameter cluster[10]

Productivity-maximising peak size

It should be noted that in the AH model, 'peak city size' is based on a single criterion, that of maximising labour productivity (specifically NOPW). There could, of course, be a multitude of other criteria that could (or should) determine 'optimal' size. For instance, Camagani *et al* (2013) consider urban costs (malaise, rent-cost, sprawl), as well as urban benefits (urban amenity, diversity, density, functions, network), in their model of 'optimal' city size. In contrast, the AH model uses a single-criterion framework.

Given the expression of NOPW in (23), we can maximise this with respect to number of workers (N) to derive optimal 'peak size' of a city. This peak size of course varies by urban hierarchy, and operates in a setting where there are many cities competing for mobile workers in national labour markets. The maximisation problem and derivation is laid out below.

$$\max_N (NOPW) = \mathrm{Q_2 MP}^{\frac{1}{(\sigma_y(1-\alpha))}} \mathrm{r}^{\frac{-\alpha}{(1-\alpha)}} \mathrm{A}^{\frac{1}{(1-\alpha)}} (\mathrm{N} - \mathrm{a_0 N}^{\frac{3}{2}})^{\chi} \mathrm{N}^{-1} \quad (24)$$

where $\chi \equiv \frac{\frac{\varepsilon+\gamma}{\rho+\beta}}{(1-\alpha)}$, and $a_0 = 1.5\pi^{-0.5}\tilde{t}$

Assuming enough differentiability on the relevant functions, maximising NOPW with respect to N gives us the following first-order necessary condition:

$$-1 \times \left(\mathrm{Q_2 MP}^{\frac{1}{(\sigma_y(1-\alpha))}} \mathrm{r}^{\frac{-\alpha}{(1-\alpha)}} \mathrm{A}^{\frac{1}{(1-\alpha)}} \right) \times \left(\mathrm{N} - \mathrm{a_0 N}^{\frac{3}{2}}\right)^{\chi} \times \mathrm{N}^{-2} +$$
$$\chi \times \left(\mathrm{Q_2 MP}^{\frac{1}{(\sigma_y(1-\alpha))}} \mathrm{r}^{\frac{-\alpha}{(1-\alpha)}} \mathrm{A}^{\frac{1}{(1-\alpha)}} \right) \times (\mathrm{N} - \mathrm{a_0 N}^{\frac{3}{2}})^{\chi-1} \times \mathrm{N}^{-1} \times \left(1 - \frac{3}{2}\mathrm{a_0 N}^{\frac{1}{2}}\right) = 0$$

[10] From Au and Henderson (2006); $Q_0 = \sigma_y(\sigma_y - 1)^{\alpha\left(1-\frac{1}{\sigma_y}\right)} c_y^{\alpha\left(1-\frac{1}{\sigma_y}\right)-1} \alpha^{\alpha}\rho^{\gamma} c_x^{-\gamma} \gamma^{\frac{\gamma}{\rho}} \beta^{\beta} (\gamma+\beta)^{-\left(\beta+\frac{\gamma}{\rho}\right)} \left(\frac{f_x}{1-\rho}\right)^{\gamma(1-\frac{1}{\rho})}$ and $Q_1 = (1-\alpha)(\sigma_y - 1)c_y)^{(\sigma_y-1)/\sigma_y}$

Rearranging this we get:

$$-\left(Q_2 MP^{\overline{\left(\sigma_y(1-\alpha)\right)}^{\,1}} r^{\frac{-\alpha}{(1-\alpha)}} A^{\frac{1}{(1-\alpha)}}\right)\left(N - a_0 N^{\frac{3}{2}}\right)^{\chi} N^{-2}$$

$$= -\chi\left(Q_2 MP^{\overline{\left(\sigma_y(1-\alpha)\right)}^{\,1}} r^{\frac{-\alpha}{(1-\alpha)}} A^{\frac{1}{(1-\alpha)}}\right)\left(N - a_0 N^{\frac{3}{2}}\right)^{\chi-1} N^{-1}\left(1 - \frac{3}{2} a_0 N^{\frac{1}{2}}\right)$$

$$\Rightarrow \left(N - a_0 N^{\frac{3}{2}}\right)^{\chi} \times N^{-2} = \chi \times \left(N - a_0 N^{\frac{3}{2}}\right)^{\chi-1} \times N^{-1} \times \left(1 - \frac{3}{2} a_0 N^{\frac{1}{2}}\right)$$

$$\Rightarrow \left(N - a_0 N^{\frac{3}{2}}\right)^{1} \times N^{-1} = \chi \times \left(1 - \frac{3}{2} a_0 N^{\frac{1}{2}}\right)$$

$$\Rightarrow N^{\frac{1}{2}} = \frac{\chi - 1}{a_0\left(\chi \frac{3}{2} - 1\right)}$$

$$\Rightarrow N^{*} = \left(\frac{\chi - 1}{a_0\left(\chi \frac{3}{2} - 1\right)}\right)^{2} \tag{25}$$

where $\chi \equiv \frac{\frac{\varepsilon+\gamma}{\rho+\beta}}{(1-\alpha)}$ ■

Equation (25) is referred to as the "AH peak size equation" throughout this thesis, and refers to the labour-productivity-maximising peak size of a city. As outlined in Au and Henderson (2006), it is apparent that the peak size equation has the following properties, which are consistent with mainstream agglomeration-theory:

1. Peak size rises with local (Marshallian) urban economies, so that $\partial N^*/\partial \varepsilon > 0$;

2. Peak size increases when the substitutability between intermediate inputs decreases. In other words, if there is specialisation (i.e., increasing the value of more differentiated variety) in intermediate inputs, peak size increases: $\partial N^*/\partial \rho < 0$;

3. If it is the case that $\beta(1-\rho) > \varepsilon\rho$, then it is clear that peak size is increasing with the importance of intermediate inputs $\partial N^*/\partial \gamma > 0$. This parametric restriction prevents a form of 'super' scale economies.

The NOPW equation (23) forms the basis for the empirical work of this thesis. However, we use its value-added output version in a log-linearised form. This is discussed in more detail in Chapter 6.

4.2 A Generalized AH model

In this Section, a modification to the AH model's effective-labour specification is proposed, and the resulting peak size equation is derived. Both the standard and modified versions of the AH model are tested using Australian data.

Modified effective-labour specification

As noted earlier, the AH model assumes a monocentric-circular urban structure, which is a potentially restrictive assumption, as many real-world cities do not conform to that structure. Empirical results in Au and Henderson (2006) suggest that, though restrictive, the model works well for Chinese cities in 1997. Perhaps this is because most workers live close to their destination of employment, and predominantly walk or ride a bike to work in that context[11]. This might also be influenced by the nature of Chinese land markets (Bertaud & Renaud, 1997).

By contrast, Australian journey-to-work data[12] in 2011 shows that workers' residential locations are varied, and are not based around a central node of employment. Moreover, an examination of the spatial distribution of employment data shows the polycentric nature of Australian cities. For instance, in New South Wales (NSW), the majority of employment is located in Sydney CBD, followed by Parramatta – which has been identified as 'Sydney's second CBD' in State Government policy documents[13].

To account for this difference in urban structure, the effective city labour component is modified to be unconstrained by the AH model's monocentric-circular structure. Instead, the specification below allows for non-circular polycentric urban structure. That is, workers can live at any location i, and travel to multiple employment centres (cities) j.

As in the AH model, time taken to travel to and from a place of employment is considered a loss of productive time that could otherwise have been used to engage in the production of final and intermediate goods.

Definition 2 (Commute-loss share): The share of total work hours lost due to commute time is specific to each employment destination (city j), and is called *commute-loss share* and denoted by λ_j.

Given that not all productive time would be 'lost' due to the commute, we make the following assumption about the range of λ_j:

Assumption 1: $0 < \lambda_j < 1$

In addition, we model this parameter as an exponent on N. This implies that the number of lost workers increase with city size, albeit at a decreasing rate. This is reasonable, since highly productive cities are likely to attract workers who reside in varied locations with high commute times.

[11] As documented in Bertaud and Melpazzi (1997), this notion is supported by the population density profiles in Chinese cities that decay quickly from the city centre.

[12] Sourced from ABS Census (2011).

[13] NSW Department of Planning and Environment, *A Plan for Growing Sydney*, December 2014.

The parameter λ_j captures multiple effects related to choice of residential location, quality and efficiency of transport infrastructure, impact of geographic constraints on the commute to work, and congestion-related effects. It is worth noting that the residential location choice effect reflects amenity of residence and housing market conditions (including affordability). However, the housing market is not fully specified in this model, so λ_j should be interpreted as the combination of these effects.

Notwithstanding the loss of productivity due to commuting, it is conceivable that some of the commute time might be used in productive activity, and not all of the time taken to travel from residential location i to employment location j would be lost. For instance, workers could engage in productive activity by using electronic devices (laptops, tablets, and mobile phones), reading, or even conducting meetings while travelling. Of course, this depends on the mode of transport, and the industry of employment. In addition, there is also the possibility that workers may try to reduce the commute time loss by increased productivity once the worker reaches employment destination j.

Definition 3 (Commute productivity): *Commute productivity* is the labour productivity during with the commute. Commute productivity will be denoted by θ.

Assumption 2: $\theta > 0$

Remark: Although we think it unlikely, we also allow for the possibility that $\theta > 1$, and so do not specify an explicit upper bound *a priori* on θ.

Having defined the two parameters λ_j and θ, we are able to specify our generalised effective-labour equation as follows:

$$L = N - \theta N^{\lambda_j} \tag{26}$$

where $0 < \lambda_j < 1$, and $\theta > 0$

In this more general specification, λ_j allows for more flexible urban structure. Workers can reside at any location i, and travel to city j for work, as long as the commute time does not exceed the number of productive work-hours in a day[14]. This implies that there is no longer a circular urban structure imposed. In addition, residents at a given location i could also travel to work at two different locations $(j \neq k)$, so there is also no longer a monocentric structure imposed.

Though not identical in the way it is used, the parameter θ is somewhat similar to the parameter q in what is known as Hensher's equation[15] [see (Hensher D. , 1977)], where q is defined as the productivity of work while travelling relative to the productivity of work at the employment destination.

[14] In our dataset, we exclude fly-in-fly-out workers (Pilbara in WA), and incidental work trips that exceed around 30% of an average workday.

[15] VBTTS = (1 – r – pq)MPL + MPF + (1 – r) VW + rVL; where VBTT = Value Of Savings In Business Travel Time, MP = Marginal Product of labour, MPF = value of extra output generated due to reduced (travel) fatigue, VL = the value to the employee of leisure time relative to travel time, VW = the value to the employee of work time at the workplace relative to travel time, r = proportion of travel time saved used for leisure purposes, p = proportion of travel time saved at the expense of work done while travelling, q = relative productivity of work done while travelling compared to the office (Fowkes, 2001).

In equation (26), there are three possibilities for the parameter θ:

1. $\theta \cong 1$, there is no productive activity during the commute and all N^{λ_j} workers are 'lost' due to the commute (and there is no post-commute productivity recovery either);

2. $\theta < 1$, some of the 'lost workers' are recovered due to productive activity during the commute (or increased productivity once at the workplace);

3. $\theta > 1$, more than all N^{λ_j} workers are lost during the commute.

It seems to us more likely that $0 < \theta \leq 1$. However, it is plausible that long commutes might increase fatigue, which might reduce productivity at the workplace, or require workers to stay longer hours which might incur additional costs (overtime, taxis, meals, or sick leave) and reduce output per worker. For this reason, an upper bound is not specified for θ, though it is unlikely that it would be far from 1.

Generalized peak size equation

Similar to the standard AH model, we are able to derive the productivity-maximising peak size equation for the generalised model with the modified effective-labour equation.

The NOPW equation for the modified AH model now becomes[16]:

$$NOPW = Q_2 MP^{\frac{1}{(\sigma_y(1-\alpha))}} r^{\frac{-\alpha}{(1-\alpha)}} A^{\frac{1}{(1-\alpha)}} (N - \theta N^{\lambda})^{\frac{\frac{\varepsilon+\gamma}{\rho+\beta}}{(1-\alpha)}} N^{-1} \tag{27}$$

As before, the maximisation problem, and derivation is laid out below.

$$\max_{N}(\mathrm{NOPW}) = \mathrm{Q_2 MP}^{\frac{1}{(\sigma_y(1-\alpha))}} \mathrm{r}^{\frac{-\alpha}{(1-\alpha)}} \mathrm{A}^{\frac{1}{(1-\alpha)}} (\mathrm{N} - \theta \mathrm{N}^{\lambda})^{\chi} \mathrm{N}^{-1} \tag{28}$$

Where $\chi \equiv \frac{\frac{\varepsilon+\gamma}{\rho+\beta}}{(1-\alpha)}$

Maximising NOPW with respect to N, assuming enough differentiability, we get the following first order necessary condition:

$$-1 \times \left(\mathrm{Q_2 MP}^{\frac{1}{\left(\sigma_y(1-\alpha)\right)}} \mathrm{r}^{\frac{-\alpha}{(1-\alpha)}} \mathrm{A}^{\frac{1}{(1-\alpha)}} \right) \times \left(\mathrm{N} - \theta \mathrm{N}^{\lambda}\right)^{\chi} \times \mathrm{N}^{-2} +$$

$$\chi \times \left(\mathrm{Q_2 MP}^{\frac{1}{\left(\sigma_y(1-\alpha)\right)}} \mathrm{r}^{\frac{-\alpha}{(1-\alpha)}} \mathrm{A}^{\frac{1}{(1-\alpha)}} \right) \times (\mathrm{N} - \theta \mathrm{N}^{\lambda})^{\chi-1} \times \mathrm{N}^{-1} \times \left(1 - \theta\lambda \mathrm{N}^{\lambda-1}\right) = 0$$

[16] Note that the subscript j is dropped since each location has a peak size, and all variables, and γ *and* λ are all defined as location-specific.

Rearranging this we get:

$$-\left(Q_2 MP^{\frac{1}{(\sigma_y(1-\alpha))}} r^{\frac{-\alpha}{(1-\alpha)}} A^{\frac{1}{(1-\alpha)}}\right) \times \left(N - \theta N^{\lambda}\right)^{\chi} \times N^{-2} =$$

$$-\chi \times \left(Q_2 MP^{\frac{1}{(\sigma_y(1-\alpha))}} r^{\frac{-\alpha}{(1-\alpha)}} A^{\frac{1}{(1-\alpha)}}\right) \times (N - \theta N^{\lambda})^{\chi-1} \times N^{-1} \times \left(1 - \theta\lambda N^{\lambda-1}\right)$$

$$\left(N - \theta N^{\lambda}\right)^{1} \times N^{-2} = \chi \times N^{-1} \times \left(1 - \theta\lambda N^{\lambda-1}\right)$$

$$\left(N - \theta N^{\lambda}\right)^{1} \times N^{-1} = \chi \times \left(1 - \theta\lambda N^{\lambda-1}\right)$$

$$N^{\lambda-1} = \frac{1-\chi}{\theta(1-\chi\lambda)} =$$

$$N^{*} = \left(\frac{1-\chi}{\theta(1-\chi\lambda)}\right)^{\frac{1}{\lambda-1}}$$

where $\chi \equiv \frac{\frac{\varepsilon+\gamma}{\rho+\beta}}{(1-\alpha)}$ ■

Throughout this thesis, equation (29) below is called 'the generalised peak size equation'.

$$N^{*} = \left(\frac{1-\chi}{\theta(1-\chi\lambda)}\right)^{\frac{1}{\lambda-1}} \quad (29)$$

The properties of agglomeration economies remain the same in the above equation. However, the modified specification of effective city labour has interesting implications on peak city size.

1. Peak size declines with commute productivity, so that $\partial N^*/\partial\theta < 0$. If workers can engage in more productive activity during the commute (low θ), then the loss of output per worker is lower, and the impact of commute on achieving peak size is minimal. Using the Hensher equation interpretation of θ, if workers are able to operate as (or more) productively during the commute than when at the workplace, the lost productivity during the commute would be low.

2. Though contingent on the values of θ and χ, it can be shown[17] that peak size declines if lost workers are high $\partial N^*/\partial\lambda < 0$. In other words, if commute times to an employment destination are high, then the city would operate below its productivity-maximising peak size.

[17] Since $\frac{\partial N^*}{\partial\lambda} = \left(\frac{(\chi-1)(\chi\lambda-1)}{\theta}\right)^{\frac{1}{\lambda-1}} \left(\frac{\chi}{(\lambda-1)(\chi\lambda-1)} - ln\frac{\frac{(\chi-1)(\chi\lambda-1)}{\theta}}{(\lambda-1)^2}\right)$ we have that $\partial N^*/\partial\lambda < 0$.

The fact that $\partial N^*/\partial\lambda < 0$ has two important policy-implications.

Firstly, *ceteris paribus*, improvements to transport infrastructure (or its operations) which lead to reduced commute times allow a city to reach its productivity-maximising city size.

Secondly, holding variation in amenity fixed, improvements in housing affordability (such that workers can live closer to their place of employment) would lead to an increase in the productivity-maximising city size. But, of course, it should be noted that the housing market is not fully specified in this model, and we leave that for future research.

There are a number of innovations that arise from the generalised AH model. Firstly, the model proposed a simple way to incorporate polycentric city structure into the model. Secondly, the model incorporates commute productivity – an increasingly important aspect of modern economic activity, often ignored in most urban economic models. Thirdly, the extension incorporates exogenously determined housing location choices, and points to the importance of housing affordability in achieving optimal city size.

Conclusion

Overall, our generalisation of the standard AH model makes a number of key contributions. The modified specification of effective labour allows for non-circular polycentric urban structure in λ, and it also captures labour productivity during the commute (θ). Both aspects cannot be derived from the standard AH framework. Importantly, this modification leads to two key results that have potentially important policy implications.

Having developed the theoretical framework for this thesis, both the standard AH model and our generalised version of the AH model are estimated and tested empirically in subsequent Chapters.

5 DATA AND VARIABLES

This Chapter outlines data and variables used in the empirical analysis component of this thesis. The data used is mainly sourced from the ABS Census from 2011, and is aimed at measuring the variables in the theoretical models described in Chapter 4. This includes both the standard AH model, and its modified versions. Descriptive statistics of the data are reported in the Appendix.

5.1 Spatial unit – functional areas

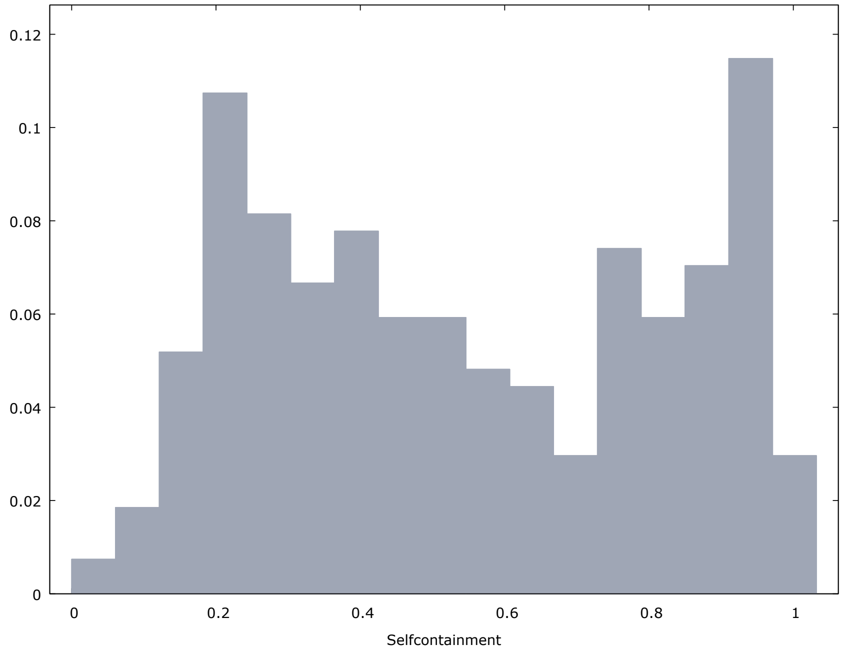
The main spatial unit used in this thesis is Statistical Area 3 (SA3). This is referred to as a 'city', 'city location', or 'functional area' throughout the thesis. An SA3 is a geographic unit defined by the Australian Burea of Statistics (ABS) that represents functional areas of regional cities, and large urban transport and service hubs. The ABS classifies SA3 as a regional breakup of Australia aimed at creating a standard framework for the analysis of ABS data at the regional level through clustering groups of smaller spatial units (SA2s) that have similar regional characteristics[18].

Camagani *et al.* (2013), in a study looking at optimal city sizes in Europe, use a similar spatial unit Functional Urban Area (FUA), where the 'city' is defined in terms daily commuting flows between one or more core areas. Studies such as Fox and Kumar (1965), CLG SAU and Coombes (2010), and Coombes (2009) argue that economic analysis is best undertaken at the spatial level at which the relevant economic market operates. They argue that economic analysis (and policy design) at a spatial level that represents functional regions ensures that most of the impacts of economic policy will be contained. Undertaking analysis, and designing economic policy based on administrative boundaries (for instance Local Government Areas) would be less efficient since the effects are more likely to spill over to other administrative locations and lead to perverse outcomes.

For these reasons, SA3 – which is a functional area definition produced by the ABS – is considered an appropriate spatial unit for analysis. In this thesis, SA3s are considered functional areas because they capture a large proportion of the spatial dynamics (labour and housing markets, transport nodes), and hence represent the spatial economy. As an illustration of spatial labour markets, Figure 3 below shows the distribution of self-containment rates—persons residing and working in the same SA3 as share of total workers—for all SA3s in the dataset. The mean in the data is 54 percent, which suggests that a large majority of workers live and work in the same SA3. In addition, nearly 26 percent of SA3s in the dataset have self-containment rates above 80 percent.

[18]See http://www.abs.gov.au/websitedbs/d3310114.nsf/home/australian+statistical+geography+standard+(asgs) for a full description of each spatial unit in the Australian Statistical Geography Standard (ASGC).

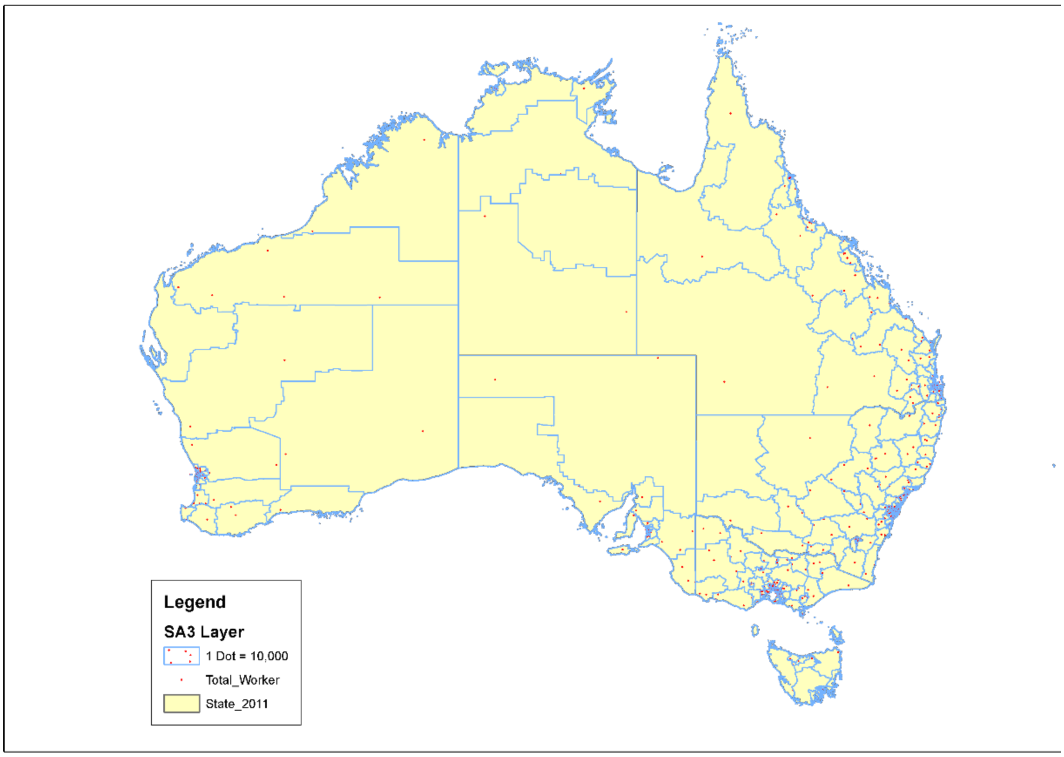
FIGURE 3. DISTRIBUTION OF SA3 WORKER SELF-CONTAINMENT



Source: Author calculations, using ABS Census (2011) data from Table Builder.
Notes: Sample is restricted SA3s with more 10,000 workers. Number of bins = 31, mean = 0.546237, sd = 0.279221

Figure 4 shows the SA3s in Australia, where each red dot on the map represents 10,000 workers in an SA3. It is apparent that the 351 Australian SA3s are geographically large spatial units that are designed to contain the vast majority of the spatial economy, and that the vast majority of employment is concentrated along the east coast.

FIGURE 4. STATISTICAL AREA 3 (SA3) IN AUSTRALIA



Source: Author, using ABS data layer, and ABS Census (2011).

5.2 Production function variables

Workers by ANZSIC industry

ABS Census 2011 records data on the number persons working at each SA3 by 19 Australia New Zealand Industry Classification (ANZSIC) (1 digit) groups[19]. This ANZSIC (1 digit) industry classification is the highest level of aggregatation available, and captures the major industry of operation. Following the approach in Au and Henderson (2006), and Li and Gibson (2015) Agriculture and Mining employment is excluded from the dataset, because these industries are strongly influenced by natural-resource deposits, and not by agglomeration forces.

However, the ABS records susbtantial employment in these industries in highly dense locations such as Sydney CBD – which means that these workers are 'office workers' who belong to mining and agricultural companies. To overcome this limitation, using spatial data from Geoscience Australia,[20] this thesis deems those SA3s in which mines and mineral processing plants operate as employing mining workers. Mining workers in the remaining SA3s are re-classified as professional service workers, and retained in the dataset. Excluding these workers (and associated GVA) would exclude high-value 'knowledge-workers' in the mining sector.

Gross Value Add (GVA) by ANZSIC industry $(\tilde{y}_j)$

The ABS does not collect data on small-area GDP or GVA by industry. The lowest geographical level of National Accounts data available is for States and Territories. Using State and Territory National Accouts, and ABS Census (2011) income data, we estimate GVA by ANZSIC industry for each SA3 in Australia. The estimation process is described below.

The small-area GVA estimation method uses a top-down approach, where State-level GVA is distributed to SA3s in the State using income data. The 2011 ABS Census records the total income a person recieves annually, and this information is coded for the place of work (SA3), and ANZSIC industry (1 digit). These self-reported income figures include both wages and non-wage income for employed persons. The spatial distribution of estimated SA3 income (as a share of total State Census-income) is used to allocate total State GVA (June 2011) by ANZSIC industry. This is shown in the equation below.

$$GVA_{j,a} = \frac{Income_{j,a}}{\sum_{j \in S} Income_{j,a}} \times GVA_{j \in S,a} \quad (30)$$

where, j denotes an SA3 in a given State or Territory (S), and a is the ANZSIC industry of employment

[19] Agriculture, Forestry and Fishing, Mining, Manufacturing, Electricity, Gas, Water and Waste Services, Construction, Wholesale Trade, Retail Trade, Accommodation and Food Services, Transport, Postal and Warehousing, Information Media and Telecommunications, Financial and Insurance Services, Rental, Hiring and Real Estate Services, Professional, Scientific and Technical Services, Administrative and Support Services, Public Administration and Safety, Education and Training, Health Care and Social Assistance, Arts and Recreation Services, Other Services.

[20] Sourced from Geosciences Australia's OZMIN database at http://www.australianminesatlas.gov.au/mapping/downloads.html

Equation (30) assumes that income is a reasonable proxy to allocate GVA by industry. One would imagine that a location which has high GVA would also generate high incomes for its workers[21]. This approach of estimating small-area GVA is similar to that used by the US Bureau of Economic Analysis (see Panek *et al*, 2007), and others (see SGS Economics and Planning, 2012) who also have to deal with the lack of available data.

One of the limitations of this income share approach is that wages and rents *may* differ from value-add in locations where profits are high due to the presence of knowledge workers in the legal branch of large corporations. Whilst we note this limitation, we do not have additional information to tackle this problem. In any case, this issue is likely to be a seldom occurrence in our dataset.

Capital stock by ANZSIC industry($\boldsymbol{K_j}$)

Similar to GVA, the ABS does not collect data on small area capital stock. Again, a top-down appraoch is used to derive estimates of capital stock for SA3s in Australia. This process uses a number of data sources, and is somewhat different to GVA estimation. The difficulty being that whilst the ABS collects data on aggregate capital expenditure at a State-level (Cat. 5220.0), capital stock data (by ANZSIC industry) is only available at the National level (Cat. 5204.0). Therefore, the first step is to estimate State level capital stock using available capital expenditure data. Subsequently, the capital stock estimates are produced for SA3s in each State.

There are two key assumptions involved in this process. Firstly, it is assumed that the five-year present value of capital expenditure is an accurate representation of the spatial distribution of capital stock between States. Secondly, it is assumed that (except for mining and manufacturing for four States) the National industry composition of capital stock is a plausible representation of the relative capital intensity in each State's ANZSIC industry. In other words, it is assumed that the manner in which different industries utilise capital is relatively similar across Australian States. This assumption is not strong since industry estimates (exlcuding agriculuture and mining) are aggregated for our emprical analysis.

The literature on capital stock estimation, notes that in a two-period world, capital accumulation can be described as follows.

$$K_t = I_t + (1 - \delta)K_{t-1} \tag{31}$$

where, K_t denotes capital stock at a given period, I_t is the capital expenditure in the same period, and δ is the depreciation rate

As noted by Stevens (1989) and Mueller (2008), there are certainly issues in estimating capital stock with investment expendtiure. Fortunately, the ABS collect data on capital stock at a National level. The innovation in this thesis is to use a top-down approach to produce small area capital stock estimates. In other words, capital expenditure is only used to estimate the spatial distribution of capital stock.

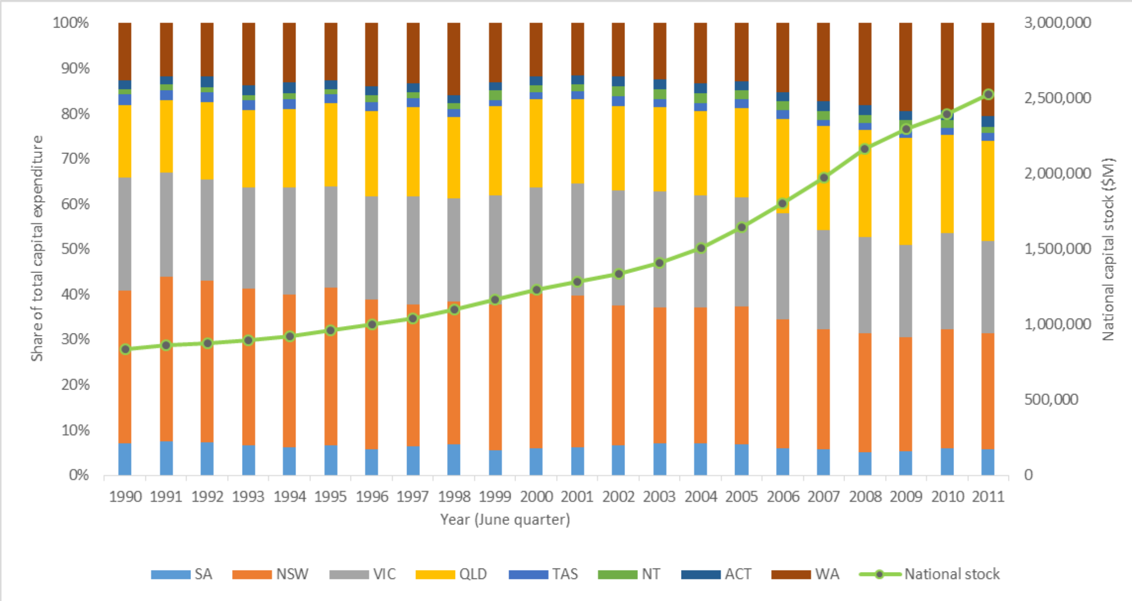
The graph below shows that whilst aggregate real capital stock (right axis) has been increasing at the National level, the spatial distribuition of aggregate capital expenditure (left axis) has been relatively stable. This might be expected since many spatial processess are cummulative, and are spatially and temporally correlated. The only noteworthy, break in the series is for WA – where

[21] This is also observed in the correlation between State-level GVA and State-level Census-income.

after year 2000 this State has a larger share in expenditure, reflecting the increase in mining related expenditure. This break has been accounted for (using data in Cat. 5625.0), in the estimation process described below.

Overall, given the stability in capital expenditure shares, and because capital stock is a spatially cumulative process (as suggested by stable expenditure over long periods), we argue that that the use of present-value of capital expenditure is an accurate representation of the spatial distribution of capital stock between States.

FIGURE 5. TOTAL CAPITAL STOCK AND EXPENDITURE SHARES



Source: Author calculations, using ABS Cat. 5204.0, and ABS Cat. 5220.0

ABS State Accounts data on total gross fixed capital formation (public and private, but excluding dwellings) (Cat. 5220.0) is used to calculate the present value of estimated capital stock at June 2011[22]. It is assumed that capital expendtiure flows form previous years accumulate to form captial stock at a point in time, but a large proportion of the previous years' stock is likely to depreciate. To account for this, we use a discount rate of 20 percent[23], to calculate the present value of expenditure over five years, and derive an estimate of June 2011 total capital stock for each State. These estimates are then used to caculate *initial* spatial distributions between States.

The ABS also collects data on actual capital expenditure (Cat. 5625.0) for two capital intensive industries – mining and manufacturing. All other industries are lumped into an 'other' catgeory However, the data is only available for the more capital intensive States of NSW, VIC, QLD, WA[24]. Following the same present value appraoch, we estimate 2011 capital stock for mining, manufacturing, and 'other' industries for these four States. As before, 20 percent discount rate is used for mining and manufactring, however, 10 percent is used for the 'other' industry category since capital expenditure is more likely to last longer[25].

[22]June 2011 was chosen since it aligns with the month of the ABS Census in 2011.

[23]This is informed by Australian Tax Office (ATO) estimates of capital stock depreciation, and verified against studies such as Stevens (1989).

[24]Tasmania was available, but was not used because there were no ABS data for mining and manufacturing.

[25]Given that we work with shares (for the spatial distributions), the estimates are not sensitive to the choice of discount rate.

According to the intial spatial distribution estimates (using Cat. 5220.0), around 89 percent of total capital stock is located in the four States of NSW, VIC, QLD, WA. The distribution of total capital stock *between* these four States is estimated using the second set of capital stock estimates (Cat. 5625.0). And for these four States, we also know the industry composition of total estimated capital stock for mining, manufacturing, and ‘other’ industries. The decomposition of ‘other’ into the remaining 17 industries is based on composition of National net capital stock data for these industries (Cat. 5204.0).

For the remaining States and Territories, the intial spatial distribution estimates (Cat. 5220.0) for these areas is used. That is, we use the five-year present value total capital expenditure to estimate the spatial distribution of total capital stock between the remaining States. Since no data on the industry composition is available, National net capital stock (Cat. 5204.0) data is used to derive the industry composition.

The calculated spatial distributions between States is applied to the National net capital stock data (Cat. 5204.0) to derive the total capital stock accumulated in each State and Territory at June 2011. Finally, the calculated industry composition within each State (largely based on the National composition, except for the identified four States where data is available) is applied to total estimated capital stock for each State to derive stock estimates by industry. Then, each State’s capital stock by industry is allocated to its SA3s.

For identified major capital city SA3s,[26] which are labour-intensive CBD locations, the spatial distribution of the difference between Census income, and estimated GVA is used to distribute State capital stock by ANZSIC industry. The idea here is that where the difference between Census income and estimated GVA is large, it is more like that non-labour input is important to the production process in these locations. Conversely where wage income (which is a large component of Census reported income) is close to estimated GVA, then capital stock is less likely to play a major role in GVA production. For all other SA3s, the spatial distribution of GVA is used to distribute State capital stock by ANZSIC industry, since a similar *a priori* identification is not possible. Extensive experimention showed that these two approaches produce reasonable capital stocks estimates at the SA3 level.

[26] Sydney Inner City, Melbourne City, Brisbane Inner, Brisbane Inner – North, Adelaide City, Perth City, Hobart – North East, Hobart – North West, Hobart Inner, Darwin City, North Canberra, and South Canberra. These capital-city SA3s are highly labour-intensive, and less likely to require substantial capital input.

5.3 Distance, time, and market potential

Market potential ($\mathbf{MP}_j$)

Recall that from Chapter 4, Market Potential (MP) for location (city) j is given by the following equation.

$$\mathrm{MP}_j = \sum_v \frac{E_v I_v}{\tau_{j,v}^{\sigma_y - 1}} \tag{32}$$

where $I_v = \left[\sum_v s_{y,u}\left(p_{y,u}\tau_{v,u}\right)^{1-\sigma_y}\right]^{-1}$

In this specification: E_v is total consumer expenditure in location v; I_v is a price index for location v, where producers in each location face the same prices. With urban hierarchy in the AH model, location-specific price information is product-specific. Since we do not have location- and product-specific price information, we follow the approach by Au and Henderson (2006) and normalise $I_v = 1$. Following Head and Mayer (2004), and Au and Henderson (2006) we refer to this as *nominal* Market Potential.

The other components of MP are calculated differently to Au and Henderson (2006). They measure expenditure (E_v) with total GDP in the whole prefecture and not just the urbanised area (the spatial unit in their study). As the authors note, using GDP captures only consumer and not producer markets for inter-city traded goods. In contrast, this thesis uses GVA by ANZSIC industry (for each SA3) since GDP data is not available by ANZSIC industry at a State level, and therefore SA3 estimates cannot be derived by industry. We argue that each SA3's GVA primarily reflects producer markets (since the spatial variation in production is captured), but to a lesser extent also reflects consumer demand potential (since wages are a large component in GVA). Ideally, small-area consumer expenditure data (based on residential location) ought to be used. Unfortunately, this data is not available to us.

To discount expenditure for distance ('iceberg' cost-factor), Au and Henderson (2006) follow the approach by Hummels (2004), and calculate $\tau_{j,v} = Ad_{j,v}^{\delta}$, where $d_{j,v}^{\delta}$ is the distance from the centre of locality j to that of v. Since better data is available, the time taken to travel from one SA3 to another is used to calculate $\tau_{j,v}$ in this thesis (see next Section). The advantage of this approach is that there is no longer the need for assumptions regarding δ, which is an elasticity parameter set to 0.82 (for all Chinese cities) in Au and Henderson (2006). In contrast, we prefer a data-driven approach where an *a priori* and spatially uniform assumption regarding travel time is not required.

These components are brought together to calculate the travel time-discounted GVA for each SA3. That is, the GVA of each SA3 (v) is travel time-discounted, and then summed to derive the *nominal* market potential index for a given SA3 (j). It is clear that this variable captures both transport infrastructure efficiency (in terms of time), and each location's accessibility to the market (as captured by GVA). For instance, a location (j) which is poorly connected to other SA3s will have a low index score, since $\tau_{j,v}$ is large (as travel times are high).

Commute time calculation

Commute time data is not used in Au and Henderson (2006), possibly due to lack of availability. Studies such as Okamoto (2007), consider commute time in a theoretical framework, though there is very little applied econometric work that uses travel time. Most studies that use estimated travel time tend to be in transport modelling frameworks. For instance, Hensher *et al* (2012) use an integrated transport and location choice modelling system (TRESIS) and a Spatial-computable General Equilibrium Model (SGEM) to measure the employment agglomeration impact of transport investments.

The travel time matrix used in this study, provided and calculated by *SGS Economics and Planning* (2015) shows the time taken to commute by car from one SA2 (i) to every SA2 (j) in Australia. There are 2193 SA2s in Australia, resulting in a 2193 x 2193 matrix of travel times.

The travel time calculations are based on centroids of an SA2. That is, the centre of each SA2 is used as the reference point where a commute begins and ends. The time taken to travel from the edge of an SA2 to its centroid is used to measure intra-SA2 travel. The calculations use Route finder 3.74, an extension for *MapInfo* (GIS software), where the road network for Australia is sourced from *Road Net Lite* - a high-level network for Australia, which includes both main roads and highways, but excludes smaller roads. The *Road Net Lite* network is disaggregated into eight categories, which are assigned different average speeds[27] ranging from 40 km per hour on main roads, to highways at 80 km per hour. These average speeds are set such that they closely match travel times from Google Maps API[28]. Though there are no explicit assumptions regarding congestion, given the calibration of speeds to match Google Maps travel times, average speeds are set to reflect current waiting times at traffic lights, and peak hour traffic conditions.

[27]Category codes 300 to 302 were assigned 80 km/h; Category codes 303 and 304 were assigned 60 km/h, and Category codes 305, 306, and 312 were assigned 40 km/h.

[28]See https://developers.google.com/maps/documentation/distance-matrix/intro?hl=en

As noted earlier, the spatial unit used in this thesis is SA3 (which is composed of smaller SA2s), since it reflects functional regions, and the spatial economy. Though the travel time matrix has data for every SA2, workers do not uniformly reside in every SA2. In other words, worker locations are not homogenous, circular, and monocentric. To account for this, we use ABS origin-destination place-of-work data[29] (2011), which records the place of residence (origin), and place of employment (destination) of each worker. Using this data, we are able to identify the origin of workers (SA2 i) employed at each destination-location (SA2 j). Commute times used in this thesis are based on travel times to the SA2 with the biggest share of employment within its SA3. This is to reduce any bias due to centres located in SA2s that border the major SA2 of employment. Using this procedure, we are able to calculate private car commute times to and from each SA3 in Australia.

Since travel times reflect both locational choice of residence and efficiency of transport infrastructure (including peak time traffic), it is not possible to disentangle the two effects. As such, all interpretations relating to commute times in this thesis consider both effects. In addition, since the travel times are based on private car, the data do not fully account for public-transport travel. Having said that, private-car commute times closely mirror bus commute times and to some extent, also likely to capture the *relative* train times (since effective geographic constraints such as distance and terrain are captured in the data).

On-road distance calculation

The distance matrix, also provided and calculated by SGS Economics and Planning (2015), shows the on-road distance from one SA2 (i) to every SA2 (j) in Australia. Again, this is a departure from the approach taken by Au and Henderson (2006), and many other papers that use linear (crow-flies) distances. As before, the distance calculations are computed using Route finder 3.74, an extension for *MapInfo* (GIS software), and the road network for Australia is sourced from *Road Net Lite*, a high-level network for Australia. These non-linear distances, also capture residential location choices, and effective-geography constraints, though previous caveats still apply.

Though not provided in the *SGS Economics and Planning* distance matrix, intra-SA2 distance is approximated by the radius of the area of each SA2. Using the procedure described earlier, we are able to calculate on-road distance to and from each SA3 in Australia.

[29] Sourced from ABS Table Builder (2011).

5.4 Derived variables

In this Section, three derived variables are discussed. These use data described in Sections 5.2 and 5.3. Of these, the Manufacturing to Services (MS) ratio is used in all estimated models since it captures urban hierarchy. Time per unit of distance travelled is used in the 'unconstrained' version of the monocentric-circular AH model (Model 2), and commute loss is a derived variable used in the estimation of the generalised non-circular polycentric model (Model 3).

Manufacturing to Services ratio($\boldsymbol{MS}$)

Recall that the AH model defines γ in the following manner.

$$\gamma_g = \frac{1}{1 + \mathrm{MS}_g} \tag{33}$$

where MS_g is the ratio of manufacturing to services GVA in a given location

Using ABS ANZSIC 1-digit industry classifications, we define MS as the ratio of 'Manufacturing' to 'Professional, Scientific and Technical Services'. The idea here is that these ABS definitions of industry groups sufficiently capture the extent of spatial industry specialisation in each SA3. For instance, Sydney CBD has high concentrations of professional services, as does Macquarie Park (a technology-cluster), implying that spatial specialisation is well-captured by this definition.

We experimented with alternative definitions of manufacturing (including transport and warehousing), and services (including financial services, and all other ANZSIC groups), but these definitions of MS produced similar empirical results.

Commute time per unit of distance travelled ($\boldsymbol{t_j}$)

This variable measures t in the AH model's specification of effective city labour. Each worker spends t amount of time to commute a unit distance, and total commuting cost per worker is tb.

Au and Henderson (2006) econometrically estimate an average t for all cities by imposing a monocentric circular structure on the data. In contrast, we are able to use ABS origin-destination place-of-work data, travel time (minutes), and on-road distance data (KM), to calculate t for each SA3 in Australia. Since all workers residing in a given SA2 face the same commute time and distance, we calculate these for every SA2-resident who works at a given SA3. We then take the average of these time per distance values as a measure of t for each SA3 (j). This is summarised in the equation below.

$$t_j = \frac{1}{j} \times \sum_{i=1}^{j} \frac{travel\ time_{i,j}}{distance_{i,j}}, N_{i,j} \neq \emptyset, \tag{34}$$

where i is an origin SA2, j is the destination SA2[30], and the sum is over all i where $N_{i,j} \neq 0$ (that is, origin SA2s with no workers are excluded).

Commute-loss share ($\boldsymbol{\lambda_j}$)

Recall that in the generalised AH model, we introduce a new specification of effective labour. The key variable to model non-circular polycentric urban structure is the commute loss share λ_j,

[30]Recall that the destination SA2 is the major employment area in each SA3.

which is defined as the share of total work-hours lost due to commute time. This is specific to each employment destination SA3 j, and captures residential location choice as well efficiency of transport infrastructure (which would be a function of congestion).

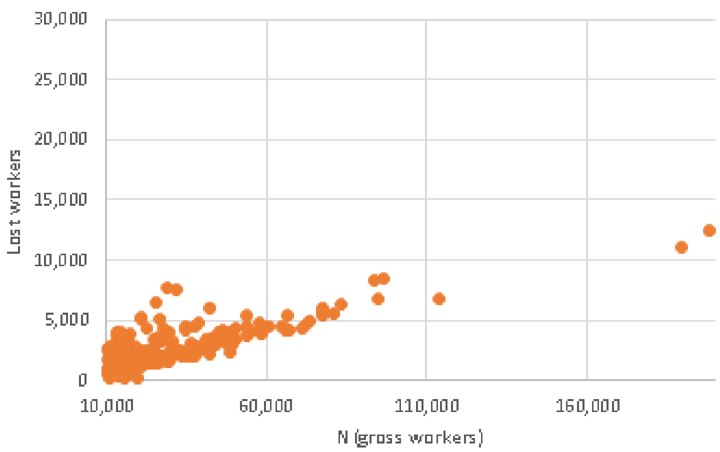
Using ABS Census (2011) data on the number of hours worked, we are able to calculate the average number of hours per day spent working (per worker) at each destination SA3 j[31]. Since we also know the commute time (using the travel time matrix) from each worker's origin SA2 i, to the destination SA3 j (and back), we are able to directly calculate the share of average productive-hours (or workers) 'lost' due to the commute. This is shown in the equation below.

$$Lost\ workers_j = \sum_{i=1}^{j} N_{i,j} \times \frac{2 \times travel\ time_{i,j}}{average\ hours\ spent\ working_j}\ , N_{i,j} \neq \emptyset, \qquad (35)$$

where i is an origin SA2, j is the destination SA2[32], and the sum is over all i where $N_{i,j} \neq 0$ (that is, origin SA2s with no workers are excluded).

In the above equation, note that the scalar 2 is to account for two-trips per day (to and from work), and that intra-SA2 commute times are used for cases where $i = j$. Recall also that in the generalised AH model, we specify λ_j as an exponent on N. This implies that the number of $Lost\ workers_j$ increases with city size (N), albeit at a decreasing rate since $\lambda_j < 1$. This is reasonable, since highly productive SA3 locations are likely to attract workers who reside in varied locations with high commute times. The assertion is supported in the graph below which plots 'lost workers' and city size. Broadly, this shows that as city size (N) increases, 'lost workers' increase at a decreasing rate.

FIGURE 6. LOST WORKERS AND CITY SIZE



Source: Author calculations, using ABS Census (2011), and SGS Economics and Planning (2015) travel time matrix
Notes: for ease of visual representation, the sample is restricted to 10,000 < N <200,000.

[31] As is the case with the rest of the analysis in this thesis – Agriculture and (actual) Mining hours are excluded. All other industry hours and workers are aggregated for this calculation.

[32] Recall that the destination SA2 is the major employment area in each SA3.

We are now able to estimate the 'commute-loss' share (λ_j) variable for each SA3. Given the assumed functional form N^{λ_j}, the following equation holds true.

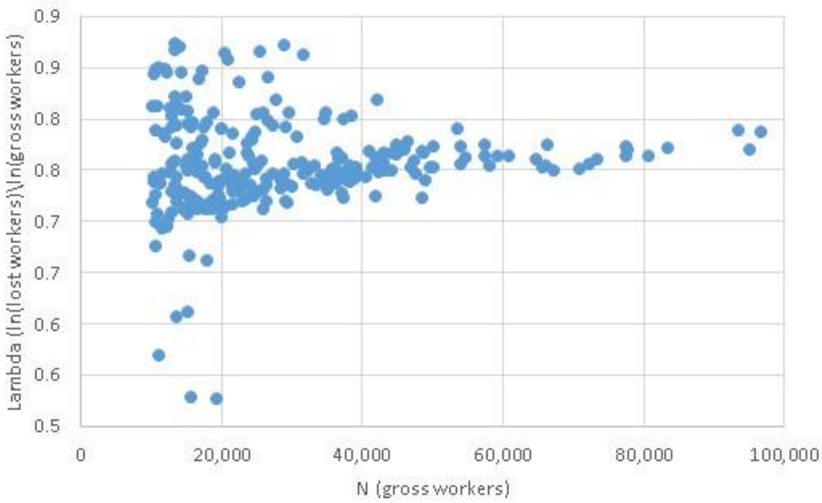
$$Lost\ workers_j \equiv \ N^{\lambda_j}$$

It is then straightforward to solve for λ_j as follows[33].

$$\lambda_j = \frac{\ln(Lost\ workers_j)}{\ln(N_j)} \qquad (36)$$

The empirical relationship between λ_j and city size is displayed in the graph below. Though not controlling for other factors, it is clear that λ_j rises and then stabilises with city size. This pattern is apparent since both N, and $Lost\ workers$ are location-specific, which enables us to calculate λ_j for each SA3 (destination of employment) in the data set.

FIGURE 7. LOG SHARE OF GROSS WORKERS LOST AND CITY SIZE



Source: Author calculations, using ABS Census (2011), and SGS Economics and Planning (2015) travel time matrix
Notes: for ease of visual representation, the sample is restricted to 10,000 < N <100,000.

[33] Note that since both numerator and denominator have N, they can be cancelled. This implies that λ_j is simply the ratio of (log) commute time to (log) average hours worked at j. We opt for the representation in (36) for ease of interpretation.

5.5 Detailed information on data sources

The following table provides additional detail on data sources used in this thesis. All data is for Main Statistical Area Structure (Main ASGS) Statistical Area 3 (SA3) unless specified otherwise. Where applicable, we note data that has been constructed for this thesis.

TABLE 3. DATA SOURCES USED IN ESTIMATIONS

Variable	**Data source(s) used**
Workers by ANZSIC industry	ABS, Census, (Table Builder Pro), (Place of Work) by INDP Industry of Employment, 2011
GVA output by ANZSIC industry *(constructed)*	ABS, 5220.0 Australian National Accounts: State Accounts (Table 5), 2011
	ABS, Census (Table Builder Pro), (Place of Work) INDP Industry of Employment and INCP Total Personal Income, 2011
Capital stock by ANZSIC industry *(constructed)*	ABS, 5220.0 Australian National Accounts: State Accounts (Table 5), 2011
	ABS, 5625.0 - Private New Capital Expenditure and Expected Expenditure, Australia (Table 7), 2011
	ABS, 5204.0 Australian System of National Accounts (Table 63), 2011
	ABS, Census (Table Builder Pro), (Place of Work) INDP Industry of Employment and INCP Total Personal Income, 2011
	GVA output by ANZSIC industry (*estimated*)
Market Potential *(constructed)*	GVA output by ANZSIC industry (*estimated*)
	SGS Economics and Planning, Private-car travel time matrix, 2015
Commute time per unit of distance *(constructed)*	SGS Economics and Planning, Private-car travel time matrix, 2015
	SGS Economics and Planning, On-road distance matrix, 2015
	ABS, Census (Table Builder Pro), (Place of Work), Journey-to-work data, 2011
Commute loss share *(constructed)*	SGS Economics and Planning, Private-car travel time matrix, 2015
	ABS, Census (Table Builder Pro), (Place of Work), Journey-to-work data, 2011
	ABS, Census (Table Builder Pro), (Place of Work), INDP Industry of Employment and HRSP Hours Worked, 2011
Share of total workers with university education	ABS, Census (Table Builder Pro), (Place of Work), by QALLP - 1 Digit Level, 2011
SA3 Area (square KM)	ABS, Statistical Area Level 3 (SA3) ASGS Ed 2011 Digital Boundaries in ESRI Shape file Format, 2012

Notes: Net capital stock data (Table 63) was chosen because the data here was after depreciation.

It should be noted that the data used in this thesis is not estimated by regression models. That is, the variables used are not Generated Regressors[34], and as such the issues relating to Generated Regressors do not apply. Instead, this thesis uses "constructed" variables, which are derived from ABS (2011) Census data.

[34] See Pagan (1984), which identifies issues that arise from regressions that are derived as functions of output from another regression.

6 ECONOMETRIC ANALYSIS

This Chapter tests the empirical validity of the AH model for Australian data. The results suggest that the model's monocentric-circular urban structure does not hold for Australian SA3-cities. The generalised version of the AH model is also tested. The alternative theoretical specification produces robust results. Lastly, Taylor-series and Generalised Leontief expansions of the models are tested. All non-linear models are estimated using Non-Linear Least Squares (NLS), while their linearised versions use Ordinary Least Squares (OLS). Robustness tests suggest that these estimates are valid.

6.1 Testing the Au and Henderson model in the Australian context

In this Section, we test two versions of the AH model. The first follows Au and Henderson (2006) and constrains the urban area—and hence the spatial structure from which the city's effective labour is drawn—to be circular and monocentric. The second is an 'unconstrained' version where the requirements of circularity and monocentricity are relaxed by including location-specific commute time in the data. Before describing the estimation process and results, we firstly lay out the relationship between output per worker and value-added per worker in the basic AH framework.

In the AH model, the criterion for 'optimal' size of each city is *output per worker*. However, we cannot directly estimate the AH model's output per worker (equation (23)) since this requires data on location-specific capital rent (r). Moreover, we do not have small-area GDP data. For these reasons, we follow Au and Henderson (2006), and use location-specific Total Gross Value Added (Total GVA) output instead.

Following Au and Henderson (2006)[35], various expressions for Total GVA may be developed as follows. Total GVA is defined as $p_y(y - c_y)s_y$, where p_y is unit price of final-good variety y, y is the quantity produced (output of the production function) of the final-good variety, c_y is the fixed cost of production, and s_y is the number of final-good varieties.

Recall from equation (6) that $p_{y,j} = \mathrm{MP}_j^{1/\sigma_y}(y - c_y)^{-1/\sigma_y}$. Using this, we are able to derive an expression for GVA as follows.

[35] See derivation in page 573.

$$\text{Total GVA} = p_y(y - c_y)s_y = (\text{MP}_j^{\frac{1}{\sigma_y}}(y - c_y)^{\frac{-1}{\sigma_y}}) \times (y - c_y) \times s_y$$

$$\text{Total GVA} = (\text{MP}_j^{\frac{1}{\sigma_y}}(y - c_y)^{\frac{\sigma_y - 1}{\sigma_y}})s_y$$

Using the result $s_y = Q_0^{\frac{1}{(1-\alpha)}} MP^{\frac{\alpha}{\sigma_y(1-\alpha)}} r^{\frac{-\alpha}{(1-\alpha)}} A^{\frac{1}{(1-\alpha)}} L^{\frac{(\varepsilon+\gamma)/(\rho+\beta)}{(1-\alpha)}}$, and a number of other substitutions, we get the following equation.

$$\text{Total GVA} = Q_3 MP^{\frac{1}{\sigma_y}} AK^{\alpha}(N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}})^{\varepsilon+\beta+\frac{\gamma}{\rho}} \quad (37)$$

where $Q_3 = Q_0\alpha^{-\alpha}\left(c_y(\sigma_y - 1)\right)^{\frac{(1-\alpha)(\sigma_y-1)}{\sigma_y}}$ ■

The derivation of the per worker equivalent of equation (37), GVA per worker (GVAPW) is given below. The purpose of this derivation is to show that, holding K/N constant, the AH model peak size equation (25) is valid for both NOPW and GVAPW.

$$\text{GVAPW} = Q_3 MP^{\frac{1}{\sigma_y}} AK^{\alpha}(N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}})^{\varepsilon+\beta+\frac{\gamma}{\rho}} N^{-1}$$

Since $\alpha + \beta + \gamma = 1$, we can use the relation $\beta = 1 - \gamma - \alpha$ below

$$\text{GVAPW} = Q_3 MP^{\frac{1}{\sigma_y}} AK^{\alpha}(N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}})^{\varepsilon+(1-\gamma-\alpha)+\frac{\gamma}{\rho}} N^{-1}$$

$$\text{GVAPW} = Q_3 MP^{\frac{1}{\sigma_y}} AK^{\alpha}(N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}})^{1-\alpha+\varepsilon} (N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}})^{\frac{\gamma(1-\rho)}{\rho}} N^{-1}$$ ■

To get GVAPW in terms of a capital per worker relation, we can rewrite:

$$\text{GVAPW} = Q_3 MP^{\frac{1}{\sigma_y}} A\left(\frac{K}{N}\right)^{\alpha} (N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}})^{1-\alpha+\varepsilon+\frac{\gamma(1-\rho)}{\rho}} N^{(\alpha-1)} \quad (38)$$

In the relation $\alpha + \beta + \gamma = 1$, where γ varies by location, either α or β (or both) may change across the urban hierarchy. It is more like that capital elasticity (α) is relatively stable across the hierarchy. Extensive experimentation in Au and Henderson (2006)[36]; showed that α is invariant across space, and we adopt the same convention for this thesis. This implies that as γ rises (when MS ratio falls), the role of internal labour (β) declines. That is, final good producers switch from internal labour usage to local outsourcing.

[36] See page 556.

Monocentric circular urban structure

To estimate the GVAPW equation, we re-write equation (38) in logs as follows.

$$\ln(\frac{\text{GVA}}{N}) = \ln(Q_3) + \frac{1}{\sigma_y}\ln(MP) + \ln(A) + \alpha\ln(\frac{K}{N})$$
$$+(1-\alpha+\varepsilon)\ln(N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}}) + \frac{\gamma(1-\rho)}{\rho}\ln(N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}}) - (1-\alpha)\ln(N) \quad (39)$$

Equation (38) and some variants of it underpin all the empirical work in this thesis. Using the data and variables described in Chapter 5, we are able to estimate equation (39) for Australian SA3s. We use Non-linear Least Squares (NLS), since the estimation equation is non-linear in both variables, and parameters.

The reason for estimating (39) is to identify the parameters $\sigma_y, \alpha, \varepsilon, \rho$, and a_0 ($= 2/3\,\pi^{-1/2}tN^{3/2}$) . The parameter-cluster [37] (Q_3) and the technology parameter (A) are assumed lumped in the constant term (β_0) of the regression.

We term this model 'constrained' because the 'congestion parameter' z is set to 1.5, which is a result of the monocentric urban structure imposed. That is, the effective-labour component of the AH model is such that $a_0 = 0.376 \times tN^z$, where $z = 1.5$. In the standard AH framework, the model freely estimates a_0, whilst restricting z. Extensive experimentation[38] in Au and Henderson (2006) suggests that this restriction is appropriate for Chinese cities in 1997.

In addition to the data described in Chapter 5, we also include a number of controls. These are: (i) fixed effects for each State and Territory in Australia – to capture regional specificities; (ii) share of university-educated[39] workers in each SA3 – to capture the impact of higher-education on high-value output; and (iii) area of SA3 (square KM) – to control for variation in the sizes of spatial units (and employment density).

The dependent variable of the model is GVA per worker, and as described in Chapter 5, all data is for 2011, and excludes agriculture and mining industries. The sample is restricted to SA3s with at least 10,000 workers. This is to exclude SA3s that are too small to have a meaningful interpretation in the context of the functional regions of the majority of the dataset.

As can be seen in Table 4, the NLS estimation of the AH model produces mixed results. On the one hand, σ_y, ρ, and α are significant, and have theory-consistent signs. For instance, the market potential elasticity is 0.12, and the scale effect associated with diversity of intermediate inputs is 0.04. These are discussed in detail and compared with other models in Section 6.2.

[37] Given the AH model's definition of the cluster, it is possible that the parameter may vary across the urban hierarchy. Similar to Au and Henderson (2006), identifying f_x and c_x is beyond the scope of the Australian SA3 data, so we assume it is lumped in the constant term.

[38] As part of their experiments, the authors increase the exponent z to accelerate how commuting costs rise with city size. They find that a_0 falls significantly such that the proportion of time spent commuting in cities declines as the exponent rises. For example, when z= 1·7, the a_0-coefficient falls from 0·0347 to 0·00957.

[39] This is sourced from ABS Census (2011), Table Builder SA3 (Place of Work) data. University education is defined as workers with Graduate Diploma and Graduate Certificate Level, Bachelor Degree Level, and Postgraduate Degree Level qualifications.

On the other hand, ε and a_0 are not significant, and the former has the 'wrong' sign. These parameters are estimated by the effective-labour component of the model. It is likely that the imposed monocentric-circular structure is not appropriate for Australian SA3 data. Given the varied residential locations of workers, it is appears that setting $z = 1.5$ is not supported by the data.

TABLE 4. MODEL 1 - CONSTRAINED MONOCENTRIC-CIRCULAR AH MODEL

Parameter	Estimated coefficient	
β_0	8.583	
	(5.9591)	
σ_y	7.975	***
	(1.9731)	
α	0.122	***
	(0.03756)	
ε	−0.0392	
	(0.8542)	
ρ	0.985	***
	(0.0078)	
a_0	−0.00381	
	(0.0392)	
Control variables		
% university-educated workers	−0.033	
	(0.0631)	
VIC	−0.1320	***
	(0.0199)	
QLD	0.044	
	(0.0296)	
SA	0.079	**
	(0.0334)	
WA	0.180	***
	(0.0472)	
TAS	0.218	***
	(0.0493)	
NT	0.310	**
	(0.1239)	
ACT	0.234	***
	(0.0524)	
Area of SA3	0.021	***
	(0.0075)	
R-squared	0.56	
n (sample)	264	
Minimum number of workers in an SA3	10,0000	

Dependent variable of the model is (log) GVA per worker. All variables, except 'Area' enter the model in logs. Standard-errors are in parentheses.
* indicates significance at the 10 percent level; ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level
Analytical derivatives are used to solve the model.
Estimation of the model uses a modified version of the Levenberg–Marquardt algorithm, and is implemented in Gretl.
NLS covariance matrix and standard-errors are calculated as described in Davidson and MacKinnon, *Econometric Theory and Methods*, 2004.

Relatively unconstrained model with location-specific commute time

In this Section, we further explore the implications of the monocentric-circular structure implied by setting $z = 1.5$. Recall that when $a_0 = 0.376 \times tN^z$, the two unknowns are time per unit of distance travelled (t), and the exponent (z) on N which is interpreted as a 'congestion parameter' by Au and Henderson (2006). For Chinese cities, the authors estimate that $a_0 N^{1.5} = 2/3\,\pi^{-1/2} tN^{1.5} = 0.035N^{1.5}$. Given that $2/3\,\pi^{-1/2} = 0.376$, it must be the case that $t = 0.093$; for all Chinese cities. If interpreted in minutes and metres, that is around 9 minutes per 100 metres, or 90 minutes per 1000 metres, which suggests a very slow commute to work (possibly due to congestion). The authors note that this estimate implies that around 25 percent of workforce hours are 'lost' due to commuting.

The approach used in Au and Henderson (2006) is to estimate a single t for all Chinese cities. However, it is reasonable to suppose that this varies by location – due to geographic constraints, and transport-efficiency. In this thesis, we instead use estimated average commute times (minutes) per (on-road) KM travelled for each SA3 (t_j). This variable captures residential location choice, and transport efficiency (which is a function of congestion). Importantly, we are able to capture SA3-specific variations in t. Since we have t_j, we are then able to freely estimate the remaining unknown parameter z to test the validity of the imposed monocentric-circular urban structure. It is for this reason that we term this model 'relatively unconstrained'.

Using SA3 city-specific t produces better results than Model 1. Estimates of Model 2 in Table 5 below shows that ε is significant and of the right sign, though the other parameter estimates are similar. The key variable of interest in this estimation z, is statistically significant, and is estimated to be 0.90. This is substantially less than the imposed value of $z = 1.5$, suggesting that the monocentric-circular urban structure (which follows from the theoretical framework of the model), is not supported by the Australian data[40].

It is reasonable to suppose that this is driven by the fact that Australian urban settlement patterns are more complex than suggested by a circular-monocentric structure. The data has a mean self-containment rate of 54 percent. That is, more than one in two residents live and work in the same SA3. In addition, the data also shows that there are varied commute times within an SA3. These facts suggest that residents seek amenity and cheaper housing, and are willing to accept longer intra SA3 (implying non-circularity), or inter SA3 (implying polycentric structure) commutes.

[40]TSNLS produced a similar result $z = 0.93$ for Model 2 (reported in the Appendix).

TABLE 5. MODEL 2 – RELATIVELY UNCONSTRAINED MODEL WITH COMMUTE TIME

Parameter	Estimated coefficient	
β_0	8.192	***
	(0.5267)	
σ_y	11.044	***
	(2.3220)	
α	0.091	***
	(0.0351)	
ε	0.127	***
	(0.0196)	
ρ	0.982	***
	(0.0070)	
z	0.901	***
	(0.0922)	
Control variables		
% university-educated workers	-0.075	
	(0.0634)	
VIC	-0.129	***
	(0.0214)	
QLD	0.020	
	(0.0248)	
SA	0.034	
	(0.0337)	
WA	0.126	**
	(0.0501)	
TAS	0.177	***
	(0.0439)	
NT	0.215	
	(0.1463)	
ACT	0.230	***
	(0.0606)	
Area of SA3	0.000	***
	(1.69E-07)	
R-squared	0.57	
n (sample)	264	
Minimum number of workers in an SA3	10,0000	

Dependent variable of the model is (log) GVA per worker. All variables, except 'Area' enter the model in logs. Standard-errors are in parentheses.
* indicates significance at the 10 percent level; ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level
Analytical derivatives are used to solve the model.
Estimation of the model uses a modified version of the Levenberg–Marquardt algorithm, and is implemented in Gretl.
NLS covariance matrix and standard-errors are calculated as described in Davidson and MacKinnon, *Econometric Theory and Methods*, 2004.

6.2 Applying the generalised AH model to Australian data

In this Section, we test the empirical validity of a generalised version of the AH model. In particular, the effective city labour component of the model is modified to include non-circular and polycentric urban structure. The extended model also allows for productivity during commute as a means to offset 'lost' productive time. The parameter λ_j is SA3-specific, so it accounts for both residential location choice, and transport infrastructure efficiency, whilst θ is the average productivity during commute for all locations. It is plausible that the latter varies by both industry, and mode of transport, however, given the objective of this thesis industries are aggregated. As before, the sample is restricted to SA3s with a minimum of 10,000 workers, and additionally, the new specification requires that we only include SA3s where $\lambda_j > 0$.

The estimation equation for Model 3 is shown in equation (40) below.

$$\ln(\frac{\text{GVA}}{N}) = \ln(Q_3) + \frac{1}{\sigma_y}\ln(MP) + \ln(A) + \alpha\ln(\frac{K}{N})$$
$$+(1-\alpha+\varepsilon)\ln(N - \frac{2}{3}\pi^{-\frac{1}{2}}tN^{\frac{3}{2}}) + \frac{\gamma(1-\rho)}{\rho}\ln(N - \theta N^{\lambda_j}) - (1-\alpha)\ln(N) \qquad (40)$$

Interestingly, as can be seen in Table 6, the parameter estimates of urban economies are relatively similar to Models 1 and 2. Though small in magnitude, the only noteworthy difference is in σ_y, which is slightly lower (implying a higher market potential elasticity).

The urban diseconomy parameter estimate is significant and has interesting implications. Since λ_j is location-specific, we are able to freely estimate $\theta = 0.69$. Firstly, it is interesting to note that the statistically significant result of $\theta < 1$ implies that there is some form of productive activity during the commute. This is perhaps not surprising since workers in labour-intensive industries (the majority of the dataset) are able to read, conduct phone meetings, or work on digital devices during (public transport, or non-self-driven) commutes. Secondly, work during the commute 'recovers' around 30 percent of lost productivity, on average. For a commute of 60 minutes, this is around 18 minutes of recovered output. This appears to be reasonable since not *all* time during the commute can be used productively or effectively.

As noted in Chapter 4, though not directly comparable, a similar concept to θ is the q parameter in the Hensher equation, which measures commute productivity relative to productivity at the destination. Fowkes (2001), in a review of the Hensher equation and valuations of travel time savings, notes that studies such as Ramjerdi *et al* (1996) find that when q is restricted to 1, productivity during the commute in Norway is around 70 to 80 percent (depending on the mode of travel) lower than at the workplace.

TABLE 6. MODEL 3 – GENERALISED POLYCENTRIC NON-CIRCULAR AH MODEL

Parameter	Estimated coefficient	
β_0	7.571	***
	(0.4504)	
σ_y	7.745	***
	(1.0990)	
α	0.090	***
	(0.0331)	
ε	0.135	***
	(0.0188)	
ρ	0.983	***
	(0.0066)	
θ	-0.693	***
	(0.2399)	
Control variables		
% university-educated workers	-0.084	
	(0.0587)	
VIC	-0.131	***
	(0.0200)	
QLD	0.045	*
	(0.0244)	
SA	0.055	*
	(0.0312)	
WA	0.130	***
	(0.0475)	
TAS	0.208	***
	(0.0375)	
NT	0.290	**
	(0.1387)	
ACT	0.235	***
	(0.0527)	
Area of SA3	0.000	***
	(1.35E-07)	
R-squared	0.58	
n (sample)	263	
Minimum number of workers in an SA3	10,0000	

Dependent variable of the model is GVA per worker. All data is for 2011. Standard-errors are in parentheses.
Heteroskedasticity-robust standard errors, variant HC1 are used.
* indicates significance at the 10 percent level; ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level
All variables, except fixed effects, and area are in natural logs. Analytical derivatives are used to solve the model.
Estimation of the model uses a modified version of the Levenberg–Marquardt algorithm, and is implemented in Gretl.
NLS covariance matrix and standard-errors are calculated as described in Davidson and MacKinnon*, Econometric Theory and Methods*, 2004.

Comparison of urban economies

As shown in Table 7, all three NLS models produce similar parameters, suggesting that the specification of urban economies is quite stable across the models.

TABLE 7. COMPARISON OF PARAMETER ESTIMATES

Parameter	Model 1		Model 2		Model 3	
β_0	8.949		8.192	***	7.571	***
σ_y	10.701	***	11.044	***	7.745	***
α	0.112	***	0.091	***	0.090	***
ε	-0.031		0.127	***	0.135	***
ρ	0.983	***	0.982	***	0.983	***

Using these parameter estimates, we are able to calculate the urban economies (scale effects) implied by the AH model. These are reported in Table 8 below, along with the production function elasticity estimates. All estimates are in percentages for a 10 percent increase in the variable to which it relates. For instance, the Marshallian scale effect from Model 2 should be interpreted as a 1.27% increase in GVA per worker (urban productivity) due to a 10% increase in total SA3 labour. It should also be noted that average values for urban hierarchy are used (γ_{Ave}) in the relevant calculations.

TABLE 8. URBAN ECONOMIES AND ELASTICITIES (%) FOR 10% INCREASES

	Specification	Model 1	Model 2	Model 3
Increase in number of intermediate-input varieties (s_x)				
Diversity scale effect	$\gamma_{Ave}(1-\rho)/\rho$	0.07	0.08	0.07
Increase in effective SA3 labour (L)				
Marshallian total effective city labour scale effect	ε	n.a.	1.27	1.35
Increase in market potential (MP)				
Market potential effect	$1/\sigma$	0.93	0.91	1.29
Production function elasticities (10% increase in each input)				
Capital (K)	α	1.12	0.91	0.90
Labour (used in final-good production) (l_y)	$\beta = 1 - \alpha - \gamma_{Ave}$	4.88	5.09	5.10
Intermediate inputs (x)	ρ	4.06	4.07	4.07

Notes: all calculations are converted to 10 percent increases for ease of comparison. Average values for urban hierarchy are used ($\gamma_{Ave} = 0.4$).

Labour inputs to the production process have the highest elasticity (4.8 to 5.1%), followed by intermediate inputs (4.06 to 4.07%). The capital elasticity estimates are lower at around 0.9 to 1.12%. These seem reasonable, given that most industries in the data are labour-intensive.

The scale effects are also relative similar across the models. Au and Henderson (2006) estimate that the diversity effect is 1.8% (for a 10% increase) – which, as they note, is quite high. In contrast, their estimate of the Marshallian scale effect (also for a 10% increase) is much lower at 0.33%. This means that diversity of production inputs is the largest effect for Chinese cities at the top end of the urban hierarchy.

In contrast, we find that the classic Marshallian scale effect (1.27 to 1.35%) is dominant in Australian cities. This seems reasonable since the knowledge economy (services sector) plays a much larger role in a developed economy such as Australia. Moreover, these estimates are in line with other Australian and New Zealand agglomeration studies[41]. The diversity effect is smaller in comparison (0.07 to 0.08%). This implies, for instance, that an increase in the number of firms providing tax accounting services (input varieties (s_x)) only has a small effect on urban productivity. We postulate that the diversity effect might be low because Australian firms are more likely to have 'preferred' contractor relationships with intermediate-good producers, so an increase in input variety only has a limited effect.

Interestingly, the market potential effect also plays an important role (0.9 to 1.3%) in urban productivity increases. Recall that MP is measured in terms of commute time-discounted GVA. So, the elasticity captures both 'expenditure' in other locations, and the extent to which those markets can be accessed by a firm in a given location. Consistent with NEG models, it follows that improving access to markets (by improving transport connections) increases market potential.

6.3 Flexible functional-form models

Since they are non-linear in both parameters and variables, the models described in Sections 6.1 and 6.2, were estimated using NLS. In this Section, we estimate linearised equivalents of the AH models. The main reason for doing this is to derive productivity-maximising 'peak size' estimates, a concept discussed at length in Chapter 4.

Specifically, following, Au and Henderson (2006), and Li and Gibson (2015) we estimate two linearised 'flexible' functional-form versions of the AH model. The first is a second-order *Regular Taylor Series* expansion, while the second is its *Generalised Leontief* version.

To get the first linearised model, recall equation (39) that underlies all AH model estimations:

$$\ln\left(\frac{\mathrm{GVA}}{N}\right) = \frac{1}{\sigma_y}\ln(MP) + \ln(A) + \alpha\ln\left(\frac{K}{N}\right) + [\mathbf{ln}(\boldsymbol{Q_3})$$
$$+(\mathbf{1} - \boldsymbol{\alpha} + \boldsymbol{\varepsilon})\,\mathbf{ln}(\boldsymbol{N} - \frac{\mathbf{2}}{\mathbf{3}}\boldsymbol{\pi}^{-\frac{1}{2}}\boldsymbol{t}\boldsymbol{N}^{\frac{3}{2}}) + \frac{\boldsymbol{\gamma}(\mathbf{1}-\boldsymbol{\rho})}{\boldsymbol{\rho}}\mathbf{ln}(\boldsymbol{N} - \frac{\mathbf{2}}{\mathbf{3}}\boldsymbol{\pi}^{-\frac{1}{2}}\boldsymbol{t}\boldsymbol{N}^{\frac{3}{2}}) - (\mathbf{1}-\boldsymbol{\alpha})\mathbf{ln}(\boldsymbol{N})]$$

It is possible that ε could vary across the urban hierarchy, and that urban structure may be more complex than this specification allows. For these reasons we follow Au and Henderson (2006), and Li and Gibson (2015) and approximate the terms in the square brackets (in bold) using a second-order Taylor Series expansion in MS and N.

[41] See for instance Hensher *et al* (2012); and SGS Economics and Planning (2013). Additionally, Mare and Graham (2009) estimate 1.2 % for New Zealand.

We are able to write the second-order Taylor Series expansion shown in equation (41), using the fact that $\gamma = 1/(1 + MS)$. Higher order expansions were considered, but not modelled due to strong multicollinearity among the terms in the resulting equations.

$$\ln\left(\frac{\text{GVA}}{N}\right) = \frac{1}{\sigma_y}\ln(MP) + \ln(A) + \alpha\ln\left(\frac{K}{N}\right) + [\mathbf{a_1N - a_2N^2 - a_3N \times MS + a_4MS + a_5MS^2}] \quad (41)$$

The second 'flexible' functional form is a Generalised Leontief version, where the second-order expansion is in square-roots (see equation (42)). Similar to Au and Henderson (2006), we find that this functional form produces better results, and is more useful in the estimation of 'peak sizes' in the Chapter 7.

$$\ln\left(\frac{\text{GVA}}{N}\right) = \frac{1}{\sigma_y}\ln(MP) + \ln(A) + \alpha\ln\left(\frac{K}{N}\right) + [\mathbf{a_1N^{\frac{1}{2}} - a_2N^1 - a_3N^{\frac{1}{2}} \times MS^{\frac{1}{2}} + a_4MS^{\frac{1}{2}} + a_5MS^1}] \quad (42)$$

Table 9 shows the results of OLS estimations of (41) and (42). The controls used in the non-linear estimations of Models 1 to 3 are used here with one minor adjustment: we used the number of workers per square KM of SA3 area (job density), as opposed to area alone. We find that this variable is particularly useful to strengthen the Generalised Leontief estimates. The other difference is that we restrict the sample to only include SA3s with a minimum of 15,000 workers, since this produces marginally better results. However, we also report Generalised Leontief estimates for other samples (see Section 6.4).

Interestingly, both models produce similar estimates of capital elasticity, though higher than Models 1 to 3. The elasticity on Market potential is also higher, though not significant in either model. The coefficients on the expansions of MS and N have the right signs in both models (such that there is an inverted U-shaped relationship with GVA per worker). Of particular interest are the coefficients on N (a_1 to a_3), since they will be used to estimate 'peak size' in Chapter 7. Given that the Generalised Leontief expansion produces statistically significant results for these coefficients (and non-negative peak sizes) we use Model 5 in our peak size estimations.

TABLE 9. FLEXIBILE FUNCTIONAL FORMS: MODELS 4 AND 5

Parameter	Model 4: Taylor Series		Model 5: Generalised Leontief	
β_0	8.400	**	8.000	**
	(0.57)		(0.53)	
$1/\sigma_y$	0.034		0.032	
	(0.034)		(0.034)	
α	0.210	**	0.210	**
	(0.051)		(0.049)	
$\mathbf{a_1}$	5.40E-06	**	0.0037	**
	(1.00E-06)		(0.00082)	
$\mathbf{a_2}$	-9.50E-12	**	-0.0000028	**
	(4.00E-12)		(0.0000014)	
$\mathbf{a_3}$	-4.80E-07	**	-0.00088	**
	(1.40E-07)		(0.00025)	
$\mathbf{a_4}$	0.016		0.047	
	(0.0099)		(0.059)	
$\mathbf{a_5}$	0.001		0.035	**
	(0.00044)		(0.01)	
Control variables				
% university-educated	0.042		0.014	
	(0.056)		(0.054)	
VIC	-0.130	**	-0.130	**
	(0.017)		(0.017)	
QLD	0.007		0.004	
	(0.029)		(0.028)	
SA	0.036		0.032	
	(0.03)		(0.03)	
WA	0.170	**	0.150	**
	(0.044)		(0.043)	
TAS	0.110	*	0.095	
	(0.066)		(0.064)	
NT	0.200	*	0.210	*
	(0.12)		(0.11)	
ACT	0.230	**	0.200	**
	(0.044)		(0.053)	
$N/Area$	0.00005	**	4.9e-05	**
	(0.000025)		(0.000024)	
Adjusted R-squared	0.71		0.71	
n (sample)	209		209	
Min. of workers in SA3	15,000		15,000	

Dependent variable of the model is (log) GVA per worker. All data is for 2011. Standard-errors are in parentheses. Italicised variables enter in logs.
* indicates significance at the 10 percent level; ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level
Heteroskedasticity-robust standard errors, variant HC1 are used.

6.4 Robustness tests

In this Section, we test the robustness of the empirical results primarily focusing on the issues of endogeneity of labour, and minimum workers in an SA3. Issues relating to the spatial unit, MS ratio, and spatial dependence are also discussed.

Endogeneity of labour

Studies of urban agglomeration are often bedevilled by concerns about the direction of causation – opening the way for potentially damaging simultaneity bias. For instance, geographic concentration might enhance productivity, but it also might be the case that productive locations tend to attract more firms. This endogeneity has the potential to bias estimates upwards, but in some cases the bias can be small or negligible[42]. The general approach in the literature to deal with this is to use Two-Stage Least Squares (TSLS) estimators with Instrumental Variables (IV). In the 'optimal' city size literature, both Au and Henderson (2006) and Li and Gibson (2015) follow a similar approach. Both these studies find that the IV estimates produce similar results for all models tested.

Given that effective labour is both potentially spatially *and* temporally correlated, many studies use historical variables as instruments for current labour. For instance, past labour is likely to be correlated with current labour, but not correlated with current levels of productivity. Au and Henderson (2006) utilise a complex design of planning[43] and amenity[44] variables to instrument for all time-varying covariates in the model. These include lagged workers, and city area. Similarly, Li and Gibson (2015) use the square root of the (log) number of non-agricultural *hukou* in 2000—a measure of lagged workers—as their instrument.

In this thesis, we use ABS Estimated Resident Population (ERP) in 2001[45] per square km of area (population density) to instrument for workers in 2011. The assumption here is that population growth 10 years ago is strongly correlated with 2011 economic activity and effective labour. Given that spatial processes are temporally correlated, historical growth in population is likely to be associated with economic activity in the present day. However, population growth from 10 years ago is not likely to be related to current urban productivity, and land area is also exogenous.

Our objective is to test whether $Cov(X, e) = 0$, where X is the suspected endogenous variable being instrumented, and e is the error term of the model. Following the specification in Li and Gibson (2015), we test for the endogeneity of labour in the linearised versions of the AH model using N, and $\sqrt{N}$. Similar to Li and Gibson (2015) we find that higher-order terms of N are insignificant when instrumented. The results of the TSLS estimations are shown in Table 10 for samples with minimums of 10,000 and 15,000 workers in an SA3.

[42]See for instance (Melo & Graham, 2009), (Lin & Truong, 2012), and (Graham D. , 2009).

[43]The 1990 capital to labour ratio, percentage of population with over 6 years of high school education, spatial area of the central business district (and that interacted with the manufacturing to service ratio), agriculture to other sector ratio, FDI to labour force, sales of independent accounting units to all enterprises, and whether a city had FDI are used as instruments,

[44]The 1990 population of the rural area is used as an exogenous IV since they argue that this population is the base for much of the migration into the nearby city determining 1997 labour force. In addition, per capita library books, doctors, telephones, and roads in 1990 are used to instrument for amenity in 1997.

[45]ERP from previous years produce similar results suggesting that OLS estimates are valid. We choose ERP 2001 because it more closely reflects developments associated with the Australian mining boom.

TABLE 10. TESTING FOR ENDOGNIETY OF LABOUR – TSLS ESTIMATES

Variable	Model A1		Model A2		Model B1		Model B2	
constant	8.100	**	8.400	**	8.100	**	8.400	**
	(0.63)		(0.57)		(0.69)		(0.66)	
MP	0.027		0.041	*	0.043	**	0.060	*
	(0.019)		(0.025)		(0.022)		(0.031)	
K/N	0.230	**	0.190	**	0.240	**	0.200	**
	(0.055)		(0.054)		(0.059)		(0.059)	
$\sqrt{\mathbf{N}}$	0.002	**	0.002	**				
	(0.00078)		(0.00086)					
N					0.000003	**	0.000003	**
					(0.0000013)		(0.0000015)	
Control variables								
$\%\ of\ uni\ educated$	-0.001		0.013		0.016		0.031	
	(0.039)		(0.045)		(0.041)		(0.048)	
VIC	-0.140	**	-0.140	**	-0.140	**	-0.140	**
	(0.019)		(0.02)		(0.018)		(0.02)	
QLD	-0.004		0.011		-0.004		0.013	
	(0.029)		(0.033)		(0.03)		(0.034)	
SA	0.049		0.058	*	0.058		0.067	
	(0.031)		(0.034)		(0.036)		(0.041)	
WA	0.170	**	0.160	**	0.170	**	0.180	**
	(0.051)		(0.055)		(0.05)		(0.053)	
TAS	0.180	**	0.180	**	0.190	**	0.200	**
	(0.068)		(0.081)		(0.07)		(0.083)	
NT	0.160		0.220	*	0.160		0.210	*
	(0.11)		(0.12)		(0.12)		(0.12)	
ACT	0.250	**	0.240	**	0.250	**	0.230	**
	(0.039)		(0.043)		(0.033)		(0.04)	
$\mathrm{N/Area}$	0.000	*	0.000		0.000		0.000	
	(0.000027)		(0.00003)		(0.000031)		(0.000036)	
Adjusted R-squared	264		209		264		209	
n (sample)	0.62		0.66		0.62		0.64	
Min. workers in SA3	10,000		15,000		10,000		15,000	
Hausman Test								
Asymptotic test statistic (χ^2)	0.686		0.937		0.571		1.120	
p-value	0.407		0.332		0.449		0.289	
Weak instrument test								
First-stage F-statistic	20.38		26.31		17.26		16.55	

Dependent variable of the model is (log) GVA per worker. All data is for 2011. Standard-errors are in parentheses. Italicised variables enter in logs.
* indicates significance at the 10 percent level; ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level
Heteroskedasticity-robust standard errors, variant HC1 are used.
Note that for the weak instrument test, F-statistics < 10 may indicate weak instruments; and for the Hausman-Test H_0= OLS is consistent

The TSLS estimations using N (Models A), and $\sqrt{N}$ (Models B) produce similar results to those of the flexible form Models 4 and 5, in that elasticity estimates on capital and market potential are similar. Both sets of models produce significant coefficients on the N variable. Crucially, the Hausman tests for all models suggest that OLS is consistent (the null hypothesis), and the Weak-Instrument tests also suggest that ERP 2001 per square KM is a good instrument for SA3 labour in 2011.

These results provide robust evidentiary support for the use of flexible functional form Models 4 and 5. Consequently, we proceed on this basis to estimate productivity-maximising peak sizes. However, we do acknowledge that these results do not necessarily apply to the NLS estimates (due to non-linearity). For the sake of comparison, we also produce Two-Stage NLS (TSNLS) estimates for Models 2 (see Appendix for details). Similar to the results in Table 10, we find that TSNLS estimates are largely similar to NLS, though some estimates are not significant. Moreover, an informal examination of the first-stage estimates of TSNLS suggests the appropriateness of the instrument. For these reasons, similar to Au and Henderson (2006), and Li and Gibson (2015), we focus on NLS and OLS results in our analysis.

Number of minimum workers in an SA3

To test the stability of the estimates across the sample, we estimate Model 5 (Generalised Leontief) across samples with varying minimum number of workers, ranging from 10,000 to 30,000.

The results in Table 11 show that the estimated coefficients a_i are largely similar across the different unit-sizes. We find that including job density (workers per square KM) improves the significance of the coefficients of interest. We suspect that this is due to the variance in the size of each SA3. The only notable exception is the sample with a minimum of 10,000 workers where a_2 and a_3 are not significant.

Other issues considered

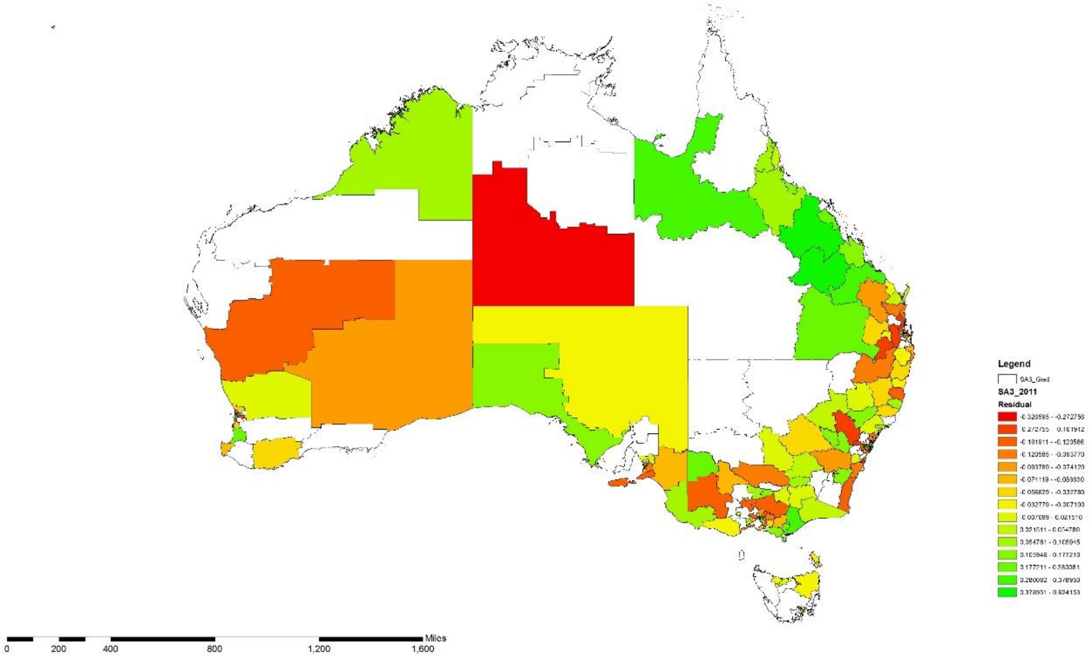
The spatial unit used in this thesis is SA3 – an ABS definition of functional regions in Australia. As argued in Chapter 5, functional regions capture the spatial economy and its functional dynamics. In determining productivity-maximising peak size, it follows that the peak size estimates relate to *functional regions*. We therefore argue that the relationship between GVA per worker and city scale in Australian SA3 data is a valid interpretation of productivity-maximising 'peak size' that should be interpreted as being appropriate for Australian functional regions. In other words, whilst we acknowledge that different spatial-unit definitions may produce different results, we argue that peak size associated with functional regions is an appropriate definition since the obvious alternative—administrative boundaries—are arbitrary and do not represent spatial dynamics of a 'city'.

The MS ratio used in this thesis represents the urban hierarchy of each location. We experimented with alternative definitions of manufacturing (including transport and warehousing), and services (including financial services, and all other ANZSIC groups), but these definitions of MS produced broadly similar results to Models 1 to 3. This suggests that the results are robust to alternative definitions of MS. The choice of urban hierarchy definition is based on the Au and Henderson (2006) model, which forms the analytical backbone of this thesis. No doubt the use of industrial data for the urban-hierarchy classification could miss a much richer picture of unique properties of different locations ignoring the immense diversity of place. We consider alternative definitions of urban hierarchy an area for future research.

The possibility of spatially correlated errors was also considered. That is, unobserved effects shared by spatially proximate SA3s that might cause the error term to be spatially correlated. *A priori* we believed that this is less likely to be an issue, since SA3s are geographically large units. Studies that use smaller spatial units, such as postcodes or suburbs (see, for instance, Abelson *et al.*, 2012) tend to require spatial lag or spatial error model specifications. However, we conducted both informal spatial-inspection of the errors, and formal Moran's I global spatial autocorrelation tests.

The visual inspection (Figure below) shows that the errors are largely random (mix of green and red shades) along the East Coast SA3s, though there is some spatial autocorrelation between SA3s in North Queensland.

FIGURE 8. SPATIAL DISTRIBUTION OF MODEL 3 RESIDUALS



Source: Author calculations, using residuals from estimated Model 3.
Notes: The white SA3s are those that are excluded from the sample. Green shading refers to positive residuals, while red shading is negatives.

Since economic activity only takes place within a small proportion of each SA3, the distance between each spatial unit is not very meaningful. For instance, it might be the case that there is correlation between units 400 km apart. But this has no structural interpretation because economic activity within each SA3 is largely self-contained.

As Fischer and Getis (2010) note, since Moran's I is distributed normally, outliers in distance or regression residuals will yield meaningless results. For this reason, we use 'zone of indifference' in our calculation of Global Moran's I. This simply means that SA3s within a threshold distance (default being 7.2 degrees or around 700 km) of an SA3 receive a weight of 1, and influence computations for a given SA3. Once the threshold distance is exceeded, weights (and the influence of each a neighbouring SA3s) diminish with distance. Calculations of Moran's I with both the default threshold (p-value 0.53), and threshold of 5 degrees (p-value 0.08) do not reject the null of no spatial autocorrelation (reported in the Appendix).

TABLE 11. GENERALISED LEONTIEF ESTIMATES BY SAMPLE

Parameter	Model 5		Model 5A		Model 5B		Model 5C		Model 5D	
β_0	8.000	**	7.900	**	8.000	**	8.100	**	7.600	**
	(0.53)		(0.6)		(0.6)		(0.69)		(0.58)	
$1/\sigma_y$	0.032		0.036		0.027		0.013		0.110	**
	(0.034)		(0.024)		(0.05)		(0.06)		(0.039)	
α	0.210	**	0.240	**	0.210	**	0.210	**	0.160	**
	(0.049)		(0.054)		(0.057)		(0.063)		(0.046)	
$\mathbf{a_1}$	0.004	**	0.002	**	0.004	**	0.004	**	0.005	**
	(0.00082)		(0.00077)		(0.001)		(0.0013)		(0.0015)	
$\mathbf{a_2}$	0.000	**	0.000		0.000	**	0.000	*	0.000	**
	(0.0000014)		(0.0000014)		(0.0000016)		(0.0000018)		(0.0000019)	
$\mathbf{a_3}$	-0.001	**	0.000		-0.001	**	-0.001	**	-0.001	**
	(0.00025)		(0.00024)		(0.0003)		(0.00041)		(0.00047)	
$\mathbf{a_4}$	0.047		-0.021		0.065		0.180		0.170	
	(0.059)		(0.064)		(0.079)		(0.12)		(0.14)	
$\mathbf{a_5}$	0.035	**	0.021	*	0.033	**	0.023	*	0.026	
	(0.01)		(0.013)		(0.011)		(0.013)		(0.018)	
Controls										
% Uni-edu.	0.014		-0.027		-0.011		0.069		-0.030	
	(0.054)		(0.055)		(0.071)		(0.084)		(0.084)	
VIC	-0.130	**	-0.130	**	-0.130	**	-0.130	**	-0.120	**
	(0.017)		(0.016)		(0.02)		(0.022)		(0.025)	
QLD	0.004		-0.013		0.001		0.030		0.100	*
	(0.028)		(0.027)		(0.038)		(0.051)		(0.053)	
SA	0.032		0.034		0.011		0.052		0.110	**
	(0.03)		(0.027)		(0.033)		(0.047)		(0.053)	
WA	0.150	**	0.160	**	0.150	**	0.210	**	0.170	**
	(0.043)		(0.046)		(0.058)		(0.069)		(0.061)	
TAS	0.095		0.120	*	0.150		0.110		0.210	**
	(0.064)		(0.069)		(0.095)		(0.12)		(0.091)	
NT	0.210	*	0.170		0.280	**	0.420	**		
	(0.11)		(0.11)		(0.13)		(0.082)			
ACT	0.200	**	0.220	**	0.210	**	0.220	**	0.250	**
	(0.053)		(0.044)		(0.065)		(0.081)		(0.12)	
$N/Area$	4.9e-05	**	6.2e-05	**	5.1e-05	*	4.8e-05		1.9e-05	
	(0.000024)		(0.000024)		(0.000028)		(0.000029)		(0.000024)	
Adj. R^2	0.69		0.63		0.67		0.65		0.73	
n (sample)	209		264		160		121		93	
Min. workers	15,000		10,000		20,000		25,000		30,000	

Dependent variable of the model is (log) GVA per worker. All data is for 2011. Standard-errors are in parentheses. Italicised variables are in logs.
* indicates significance at the 10 percent level; ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level
Heteroskedasticity-robust standard errors, variant HC1 are used.

7 PRODUCTIVITY-MAXIMISING PEAK SIZE

In this Chapter, we use the *Generalised Leontief* flexible functional-form model to estimate productivity-maximising 'peak sizes', and compare these with actual SA3-sizes in the 2011 data. Since the AH model's 'optimal' size is based on a single criterion, we also estimate Zipf's Law for our data as an additional measure of 'efficient' city size. Using the results of the productivity-maximising peak sizes, we also make some general observations regarding theoretical propositions.

7.1 Flexible functional-form peak size

All econometric estimations in Chapter 6 were based on GVA Per worker (GVAPW). However, to estimate the inverted-U relationship between city scale and output per worker we convert GVA per worker to *Net Output Per Worker* (NOPW). Following Au and Henderson (2006)[46], if we assume that on the LHS capital-per worker is fixed, and on the RHS capital-rent is fixed, the following equation holds:

$$\delta ln\left(\frac{GVA}{N}\right) = (1-\alpha) \times \delta ln\left(\frac{Net\ output}{N}\right) \quad (43)$$

Given equation (43), GVAPW and NOPW are equivalent up to a proportionality constant α. We use this property when presenting empirical results on the inverted-U function for NOPW against city size. Note though, as was demonstrated in Chapter 6, the peak size is the same for both GVAPW and NOPW. Although, the structural model (Models 3), and its Taylor Expansion (Model 4) produce similar results, we focus our peak size analysis on the Generalised Leontief (Model 5) estimates. This is because, similar to the results in Au and Henderson (2006) and Li and Gibson (2015), we find that these estimates produce non-negative peak sizes.

The basic equation underlying our output per worker estimates is the following:

$$\ln\left(\frac{Net\ output}{N}\right) = \frac{1}{(1-\alpha)} \times \ln\left(\frac{GVA}{N}\right) = [\mathbf{a_1 N^{\frac{1}{2}} - a_2 N^{1} - a_3 N^{\frac{1}{2}} \times MS^{\frac{1}{2}}}] \quad (44)$$

Figure 9 below shows the inverted-U relationship between urban productivity and SA3-scale, which is calculated using equation (44), holding remaining terms in equation (42) fixed. The reason for the inverted-U shape in the relationship between productivity and city size is that, as the number of workers increases (from a low base), agglomeration economies outweigh diseconomies, leading to productivity increases.

[46] See p. 565.

However, as city size continues to increase, diseconomies begin to outweigh economies, leading to productivity decline. The point at which these economies and diseconomies are equal—the 'top' of the inverted-U—will be referred to as *peak size*.

We empirically model the inverted U-function for the following three SA3s on the urban hierarchy spectrum: (i) Inner-Sydney – a CBD location with highly dense professional-services employment; (ii) Newcastle – historically a mining and steel manufacturing hub, now transitioning into the service sector; and (iii) Lower Hunter – an SA3 with high manufacturing.

Before delving into the specifics of the estimation, some general properties of the U-function presented here should be noted.

Firstly, as city scale rises, we account only for the 'internal' returns to city scale. That is, the full general-equilibrium effects are not accounted for. Consistent with Au and Henderson (2006), there are two potential 'external' scale effects at play. They are: (i) as a city's productivity increases, so would its own market potential. This then impacts on other locations (or shifts their U-functions upward); (ii) as a city expands from a low worker base (where economies outweigh diseconomies), induced demand for output from other locations also increases. Given that city scale increases output, and ergo increases market potential, we are likely to be underestimating the full productivity increase associated with rising city scale.

Secondly, given that a city could, in principle, sit anywhere along the urban-hierarchy spectrum, there is not a single 'optimal-scale' for all cities. Instead, there are multiple peak sizes, subject to the unique conditions of each location.

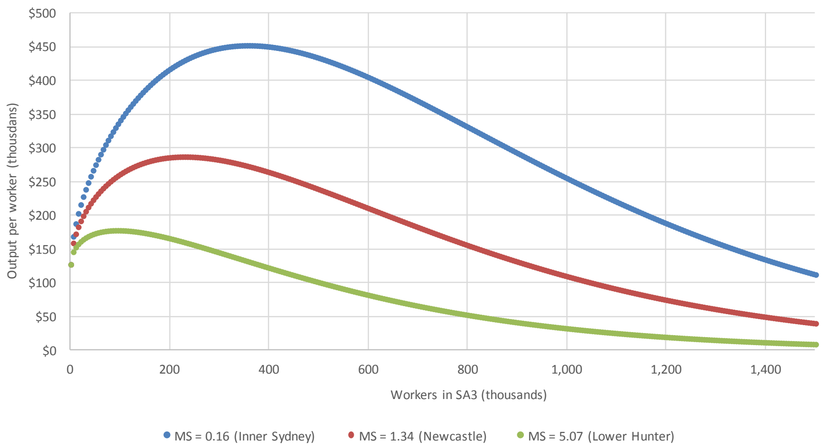
Thirdly, since all control variables (including market potential) are held constant, it follows that a change in any of these could shift the inverted U-function. So, the 'peak size' on the horizontal axis stays the same, but the peak-value for (NOPW) on the vertical axis is altered. For instance, shocks to the production technology (captured by the constant in Model 5) would shift the function upwards, increasing urban productivity without increasing scale.

Fourthly, as noted by both the AH (25), and Generalised (29) peak size equations, any shocks to urban-economy or diseconomy parameters would alter a given peak size. For instance, a reduction in λ, by improving either transport infrastructure efficiency or housing affordability, would increase a city's peak size. This also implies an increase in output per worker at the higher peak size.

Turning now to the details of Figure 9, we note that cities with high concentrations of industrial-activity reach the peak size much faster than service-oriented locations. This is plausibly because manufacturing industries require large land areas with few workers, reflecting the capital- (and land-) intensive nature of manufacturing. In contrast, the services sector has a much higher peak size because office-workers require small workspaces (small land area). It is also apparent that industrial SA3s lose productivity much faster than the service sector SA3s, since agglomeration economies are much stronger for services than manufacturing[47]. The other apparent feature is that the vast majority of productivity benefits are generated when a city reaches around 50% of its peak scale. It follows that operating at a scale more than 50% below the peak has a large impact on productivity.

[47] See for instance Hensher *et al* (2012) for agglomeration economies by ANZSIC industry.

FIGURE 9. INVERTED-U FUNCTION AND PEAK SIZES



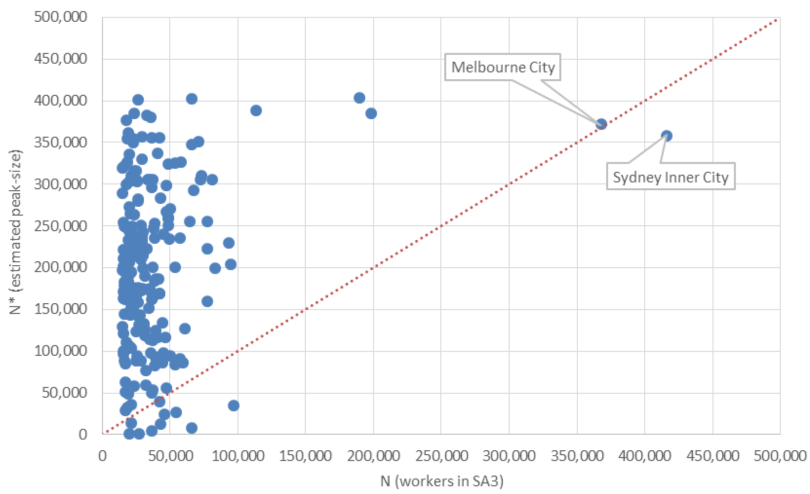
Source: Author calculations, using Model 5 estimates. Note that output per worker is calculated as the average Y/N in the data where MS <6.0. This is around $127,500, and is multiplied by equation (42) at each N.

We next compare the actual number of workers from the 2011 Census, with Model 5 predictions of productivity-maximising peak city size. Holding the capital-per worker ratio constant (along with other controls), and maximising equation (42) with respect to N, we get the Generalised Leontief peak size equation (45).

$$N^* = \left(\frac{a_1 - a_3 MS^{\frac{1}{2}}}{2a_2}\right)^2 \tag{45}$$

Points along the dashed line from the origin ($N = N^*$) show where an SA3 is at its theoretical NOPW maximising peak size. The blue dots indicate actual city sizes in 2011. From this plot we see that the vast majority of Australian SA3s are operating below their productivity-maximising peaks. This is plausible because Australia only has a few major cities, and because Australia is a sparsely populated landmass. The notable exceptions are Melbourne City, and Sydney Inner City which are at, and slightly above their peak sizes, respectively. There are also a few SA3s (less than 100,000 in size) that operate above their peak.

FIGURE 10. ACTUAL-SIZE AND ESTIMATED PEAK SIZE



Source: Author calculations, using Model 5 estimates using a sample of SA3s with minimum 15,000 workers. Note that output per worker is calculated as the average Y/N in the data where MS <6.0. This is around $127,500, and is multiplied by equation (45).

To examine peak size relationships more finely, we calculate the percentage of productivity-loss (using equation (46)) for SA3s with more than 50,000 workers in 2011.

$$\ln\left(\frac{Net\ output}{N}\right)^{*} - \ln\left(\frac{Net\ output}{N}\right) = \frac{1}{(1-\alpha)}\left[\,(\mathbf{a_1} - \mathbf{a_3}\mathrm{MS}^{\frac{1}{2}})(\sqrt{\mathrm{N}^{*}} - \sqrt{\mathrm{N}}) - \mathbf{a_2}(\mathrm{N}^{*} - \mathrm{N})\right] \quad (46)$$

Of the 32 major SA3s identified in Table 12, Sydney Inner City, Dandenong, Tullamarine–Broadmeadows, and Bankstown are above peak size. Interestingly, all except Sydney Inner City specialise in manufacturing. Even though these locations are above peak in size, the maximum output loss is only 10% (Tullamarine-Broadmeadows).

The remaining SA3s in Table 12 operate (at or) below the estimated peak city scale -the vast majority of which specialise in services (MS < 1). Within these locations, the average output loss is around 20%. North Canberra (50%), Adelaide City (29%), Perth City (14%), and Brisbane Inner – North (39%) are major CBDs with output loss due to lack of scale.

TABLE 12. PRODUCTIVITY LOSS FOR MAJOR SA3 LOCATIONS

Major SA3	MS ratio	N	GVA loss %	Output loss %
Sydney Inner City	0.16	415,810	1%	1%
Melbourne City	0.11	367,716	0%	0%
Brisbane Inner	0.06	198,393	9%	11%
Perth City	0.03	189,514	11%	14%
Adelaide City	0.06	113,762	23%	29%
Dandenong	9.04	96,765	4%	5%
Monash	1.76	95,051	6%	7%
Newcastle	1.34	93,582	8%	11%
Parramatta	1.86	83,546	7%	9%
Port Phillip	0.47	80,823	20%	26%
Ryde - Hunters Hill	0.97	77,758	14%	18%
Townsville	1.45	77,700	10%	13%
Geelong	2.75	77,544	4%	5%
Chatswood - Lane Cove	0.44	73,430	23%	29%
Yarra	0.47	72,385	23%	29%
North Sydney - Mosman	0.19	71,035	30%	38%
Stirling	0.58	67,273	22%	28%
Tullamarine - Broadmeadows	13.22	66,353	8%	10%
North Canberra	0.03	66,199	40%	50%
Boroondara	0.21	65,651	31%	39%
Canning	0.98	64,717	18%	22%
Knox	3.76	60,857	3%	4%
Kingston	5.43	59,432	1%	1%
Brisbane Inner - North	0.32	58,176	31%	39%
Botany	5.22	57,595	1%	1%
Toowoomba	1.24	57,576	17%	21%
Bankstown	10.03	54,727	1%	2%
Gosford	1.84	54,041	13%	16%
Belmont - Victoria Park	0.33	54,005	32%	41%
Nundah	5.60	53,695	1%	1%
Strathfield - Burwood - Ashfield	0.80	50,243	25%	31%
Auburn	5.07	50,233	2%	2%

Source: Author calculations. Note that all locations exclude agriculture and mining workers.

Overall, though the show that the majority of Australian SA3s operate below their productivity-maximising peak scales (Figure 10), the output loss for 'under-sized' major SA3s (Table 12) is around 20%. However, there are a few major CBDs with output loss greater than 20%. This might be explained by the sparsely populated nature of Australian States and Territories. Major CBD locations such as Sydney and Melbourne operate near peak scale, whilst, other major CBDs are well below their peaks.

Duranton and Puga (2004) argue that under conditions of free mobility (within country), if a city is not at its peak, then it is likely to be beyond its peak. This is due to either stability conditions in labour markets, or conditions on Nash-equilibrium migration decisions. It is not immediately apparent if this argument could apply to Australia, since it only has a few major CBDs, and is sparsely populated. Conceivably, the Duranton and Puga (2004) theory would apply to similar cities that have achieved a threshold city scale. In our data set, this might explain why Melbourne and Sydney sit on, and to the right of their respective peaks. In contrast, other major CBDs such as Perth, Brisbane, and Adelaide sit to the left of their peaks. Though we have briefly speculated about the reasons for this potential 'puzzle', since we do not explicitly model migration, or cultural factors that might influence migration-decisions, we consider this a question for future research.

7.2 Zipf's Law

An alternative benchmark of 'efficient' city size, common in the literature[48], is Zipf's Law. The idea that city size distributions can be approximated by a Pareto distribution, was first proposed by Auerbach (1913), and later refined by Zipf (1949). Zipf's refinement was that in addition to following a Pareto distribution, city sizes have a Pareto parameter[49] of $\alpha = 1$.

Zipf's Law can be stated as follows.

$$y = Ax^{-\alpha} \text{ or}$$

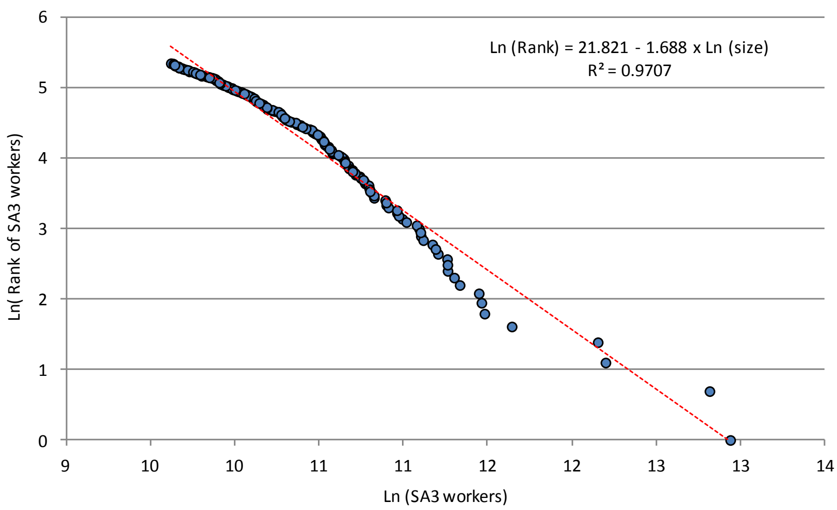
$$\ln(y) = \ln(A) - \alpha \times \ln(x) \qquad (47)$$

For the application we have in mind, in (47), x is the size of a city, y is the number of cities with size greater than x, and A and α are constants. Zipf's idea was that $\alpha = 1$, where A is the largest city. In other words, Zipf's Law is a log-log relationship of city size, and city rank (relative to the largest city).

A cross-country study by Rosen and Resnick (1980) examined the Pareto exponent (α) for populations of 44 countries. Since parameter estimates for 32 out of 44 countries exceeded unity (mean 1.14), Soo (2005) argues that populations in most countries are more evenly distributed than predicted by the rank size rule. In addition, as noted by Soo (2005), studies such as Guerin-Pace's (1995) find that the evidence in the literature is mixed, and that results are highly sample-specific. Some studies find that $\alpha > 1$, while others observe unity.

Similar to Li and Gibson (2015), in addition to using the Generalised Leontief AH model, we use Zipf's Law as an alternative measure of 'efficient' city size. Figure 11 shows the rank size relationship for Australian SA3s.

FIGURE 11. ZIPF'S LAW FOR AUSTRALIAN SA3 IN 2011



Source: Author calculations, using ABS Census 2011 SA3 data excluding workers in agriculture and mining. Sample is restricted SA3s with a minimum of 15,000 workers.

[48] See for instance Rosen and Resnick (1980), and Soo (2004).

[49] See Soo (2004) for a detailed review of the evolution of Zipf's Law.

The data suggests that the Pareto-parameter $\alpha = 1.68$, and that the relationship is a good fit (R-squared of 0.97). Since the departure from the estimated linear-fit is at the upper and lower tails, we also estimated the model after restricting the sample to only include SA3s with $\ln(Rank) > 3$, and found the results to be largely similar ($\alpha = 1.59$).

Given this result, we note that the city size distribution is relatively even between Australian SA3s, with exceptions at the tails. Though not directly comparable with Figure 10, the SA3s diverging from the estimated Zipf's Law linear fit (above and below 'efficient' size) are similar in both analyses. As noted earlier, this is most likely explained by the fact that Australia has few major CBDs, and is sparsely populated.

8 CONCLUSION

The theoretical literature reviewed in this thesis pointed out the contributions made to spatial economics by New Economic Geography (NEG) and location theory-inspired models. We also surveyed some of the vast empirical literature on agglomeration economies, their causes, nature, characterisation and consequences. It was noted that there are only a few studies that concurrently consider both economies and diseconomies of agglomeration (both of which impact on urban productivity), and their relationship with city size. An influential strand of the literature, represented by Au and Henderson (2006), postulates the existence of an inverted U-shaped relationship between output per worker and city size. The productivity-maximising 'peak' of this function is considered the 'optimal' city size, where the criterion for optimal city size is maximising output per worker.

This thesis applied the AH model to Australian cities, to analyse the interplay between urban economies and diseconomies in Australia, and to derive productivity-maximising 'peak' or 'optimal' city sizes. Given that the AH model imposes a strict monocentric circular urban structure, this thesis proposed a generalisation of the model that allows for non-circular polycentric urban structure.

The generalised version of the AH model introduced two parameters: (1) commute loss share – which captures the share of work hours 'lost' due to the commute to work; (2) commute productivity – labour productivity during the commute that offsets the work-hours lost during the commute. The inclusion of these parameters produces two interesting theoretical results related to productivity-maximising city size:

1. *Peak size declines with commute productivity*: if workers can engage in more productive activity during the commute, then the loss of output per worker can be reduced.

2. *Peak size declines if lost work-hours are high*: if commute times to an employment destination are high, then the loss of output per worker is also high.

The second result has two potentially important policy implications.

Firstly, *ceteris paribus*, improvements to transport infrastructure (or its operations), which lead to reduced commute times allow a city to increase its productivity-maximising city size, and therefore output per worker.

Secondly, holding amenity-preferences fixed, improvements in housing affordability (such that workers can live closer to their place of employment) could lead to an increase in the productivity-maximising city size, and therefore output per worker. However, it should be noted that the housing market is not fully specified in this model, and we leave that for future research.

To test both the AH model and the extended version of it proposed here, a small-area dataset was developed for Statistical Area 3 (SA3) – an ABS definition of the functional regions of Australia, designed to capture the functional dynamics of its cities and regions. Since the ABS does not produce National Accounts data below the State level, this thesis developed Gross Value Added (GVA) output, and capital stock estimates for SA3s. Whilst the estimation method for GVA is relatively common, the capital stock estimation procedure is an innovation proposed in this thesis. The location-specific 'market potential' of an SA3 is also calculated in an innovative manner, where GVA of all locations is discounted by commute times instead of linear distance.

The dataset of Australian SA3s was then used to test the AH model, and its extension using Non-linear Least Squares (NLS). The results obtained by doing this show that the monocentric-circular city structure imposed by the AH model is not supported by the data. We argue that this is driven by the fact that Australian urban-settlement patterns are more complex than suggested by a circular-monocentric city-structure. Moreover, the spatial distribution of employment data shows the polycentric nature of Australian cities. For instance, in New South Wales (NSW), the majority of employment is located in Sydney CBD, followed by Parramatta – which has been identified as 'Sydney's second CBD' in State Government policy documents[50]

In contrast, the generalised version of the AH model proposed in this thesis produced far better results. With SA3-specific 'commute loss' shares, this model estimates that productive activity during the commute 'recovers' around 30 percent of lost productivity, on average. For a commute of 60 minutes, this is around 18 minutes of recovered output.

There are three agglomeration (centripetal) forces at play in both versions of the AH model – Marshallian scale effect, diversity of inputs scale effect, and market potential effect. For Chinese cities, Au and Henderson (2006) find that the second scale effect dominates. In contrast, we find that the classic Marshallian scale effect (1.27 to 1.35% for a 10% increase in city scale) is dominant in Australian cities. This seems reasonable since the knowledge economy (services sector) plays a much larger role in a developed economy such as Australia. In addition, the market potential effect also plays an important role (0.9 to 1.3% for a 10% increase in market potential) in urban productivity increases. That is, improving the efficiency of transport infrastructure (access to markets) leads to an increase in output per worker.

With these observations as motivation, this thesis follows Au and Henderson (2006), and Li and Gibson (2015) by estimating two linearised 'flexible' functional-form versions of the AH model. The first is a *Regular Taylor Series* expansion, and the second is its *Generalised Leontief* version. Estimates from the latter are used to derive both the postulated inverted U-shaped function (relating output per worker and city scale), and implied productivity-maximising 'peak sizes'.

[50] NSW Department of Planning and Environment, *A Plan for Growing Sydney*, December 2014.

The estimated inverted-U function has a number of interesting properties, and produces the following useful insights.

1. Given that a city could, in principle, sit anywhere along the urban-hierarchy spectrum, there is not a single 'optimal scale' for all cities. Instead, there are multiple peak sizes, subject to the unique conditions of each location.

2. Cities with high concentrations of industrial-activity reach the peak size much faster than service-oriented locations. It is argued that this is because manufacturing industries require large land areas with few workers, reflecting the capital- (and land-) intensive nature of manufacturing.

3. Conversely, once past the productivity-maximising peak, heavy-industry based locations lose productivity much faster than service-sector locations. This is because agglomeration economies are much stronger for services than manufacturing.

4. The vast majority of productivity-benefits are generated when a city reaches around 50% of its peak scale. In other words, urban economies are substantially stronger than diseconomies for around 50% of peak size. It follows that operating at a scale more than 50% below the peak, results in large losses in productivity.

Comparing actual city sizes (from the 2011 Census), with estimated 'peak' city sizes shows that the vast majority of Australian SA3s are operating below their productivity-maximising peaks. It is argued that this is because Australia only has a few major CBDs, and because Australia is a sparsely populated land-mass. The notable exceptions are Melbourne City, and Sydney Inner-City which are at, and slightly above their peak sizes, respectively.

However, calculations of output losses for the 32 major SA3s (more than 50,000 workers) shows that, of the SA3s above 'peak size' (Sydney Inner City, Dandenong, Tullamarine–Broadmeadows, and Bankstown) the maximum output loss is only 10% (Tullamarine–Broadmeadows). The remaining 28 major SA3s operate (at or) below 'peak size' resulting in an average output loss of around 20%. North Canberra (50%), Adelaide City (29%), Perth City (14%), and Brisbane Inner – North (39%) are major CBDs with output loss due to lack of scale.

Duranton and Puga (2004) argue that under conditions of free mobility (within country), if a city is not at its peak, then it is likely to be beyond its peak. It is not immediately apparent that this argument could apply to Australia, since it only has a few major CBDs, and is sparsely populated. Since we do not explicitly model migration, or cultural factors that might influence migration-decisions and city scale, we consider this a question for future research.

Overall, this thesis has made both theoretical and empirical contributions to the literature on urban economies, diseconomies, and their relationship with city scale. A number of interesting policy implications relating to transport efficiency, housing affordability, urban productivity, and city scale arise from what has been uncovered. However, it should be noted that 'optimal' size here is based on the single criterion of the maximisation of output per worker. It might be that social, environmental, and ecological considerations should also influence 'optimal' city size. As such, this thesis ought to be considered as a contribution to one piece of a much larger spatio-temporal general-equilibrium puzzle.

9 APPENDIX

Extended discussion on industrial scope

Economies of scale that arise from spatial concentration of activity within a given industry are known as *localisation economies*. The externalities that arise from the concentration of all economic activity (across multiple industries), or from city size itself, are known as *urbanisation economies*. The closer two firms are in industrial space (similar production processes), the greater the potential for industrial scope.

The classical Marshallian agglomeration economies of a local pool of skilled labour, local supplier linkages, and local knowledge spillovers are generally associated with *localisation economies* since they relate to firms operating in the same industry. Industrial scope is also particularly relevant to one of the features of agglomeration – diversity.

There are three key theoretical arguments at the heart of the urbanisation economies vs. localisation economies debate.

The first of these is due to independent contributions by Marshall, Arrow, and Romer. Known as MAR theory, the collective ideas from their individual theories posits that that local monopoly is better for growth than local competition, because local monopoly restricts the flow of ideas to others and so allows externalities to be internalised by the innovator (Glaesar, Kallal, Scheinkman, & Shleifer, 1992). It follows that the ability to internalise externalities enables faster growth. MAR theory thus argues that local concentration (or localisation economies), *and* monopolisation is better for growth, and that specialisation or less industrial diversity is key to this.

A second theory that relates to this debate is due to Porter (1990), who argues that knowledge spillovers in specialised geographically concentrated industries are better for growth. Unlike MAR, Porter (1990) argues that that local competition, as opposed to local monopoly, drives innovation. This is because, even though competition reduces the returns to the innovator, it also incentivises firms to innovate (Glaeser *et al*, 1992). It follows that firms that are unable to progress by innovation will not survive in the face of strong competition from similar firms.

A third theory that relates to this debate is due to Jacobs (1969). Unlike MAR and Porter, she argues that diversity in industry, and not specialisation is key to economic growth in cities. This is because the most important knowledge transfers come from outside the core industry in which a business operates (Glaeser *et al*, 1992). In essence, the argument here is that the process of 'cross fertilisation' enables better growth outcomes. Variety of proximate industries rather than industrial specialisation of proximate industries promote innovation, and growth. One example is the brassiere industry, which grew out of dressmakers' innovations rather than the lingerie industry (Glaeser, 1999). In addition, Jane Jacobs also argues that local competition, as opposed to local monopoly, drive innovation and growth.

This is because like Porter, she argues that competition speeds up innovation and the adoption of technology (Glaeser *et al*, 1992). It follows that industries located in areas that are highly industrially diversified should grow faster, owing to the cross-fertilisation between industries in a highly competitive environment.

The theoretical considerations here are equally valid, and dependent on the *economic context* of the urban economy. As a result, empirical analysis is required to resolve the relative significance of each of the factors identified in the theory. Among others, the results from Glaeser *et al.* (1992), which go some way towards this, are discussed in the empirical literature review.

Two Stage Non-linear Least Squares (TSNLS) estimates for Model 2

The table below compares NLS and TSNLS estimates for Model 2. The instrument used in TSNLS is Estimated Resident Population (ERP) in 1981. Estimates with earlier years of ERP produce similar results, though we find that ε is negative, and in some cases significant.

Crucially, Table 13 shows that $z = 0.93$, which is very similar to the NLS estimate. In addition, estimates of σ_y and ρ are also largely similar. This test shows that both NLS and TSNLS produce similar results, and importantly that both sets of results support the notion that the monocentric-circular urban structure (imposed by the AH model) is not supported by the Australian data.

TABLE 13. MODEL 2 – NLS AND TSNLS COMPARISON

Parameter	NLS estimates		TSNLS estimates	
β_0	8.192	***	10.352	***
	(0.5267)		(0.84950)	
σ_y	11.044	***	7.159	***
	(2.3220)		(1.4578)	
α	0.091	***	0.038	
	(0.0351)		(0.0515)	
ε	0.127	***	-0.045	
	(0.0196)		(0.0463)	
ρ	0.982	***	0.989	***
	(0.0070)		(0.0079)	
z	0.901	***	0.938	***
	(0.0922)		(0.0720)	
Control variables				
% university-educated workers	-0.075		0.004	
	(0.0634)		(0.0715)	
VIC	-0.129	***	-0.111	***
	(0.0214)		(0.0260)	
QLD	0.020		0.032	
	(0.0248)		(0.0278)	
SA	0.034		0.033	
	(0.0337)		(0.0379)	
WA	0.126	**	0.183	***
	(0.0501)		(0.0559)	
TAS	0.177	***	0.177	***
	(0.0439)		(0.0337)	
NT	0.215		0.272	*
	(0.1463)		(0.1546)	
ACT	0.230	***	0.241	***
	(0.0606)		(0.0581)	
Area of SA3	0.000	***	0.000	***
	(1.69E-07)		(2.013E-07)	
R-squared	0.57		0.46	
n (sample)	264		264	
Minimum number of workers in an SA3	10,0000		10,0000	

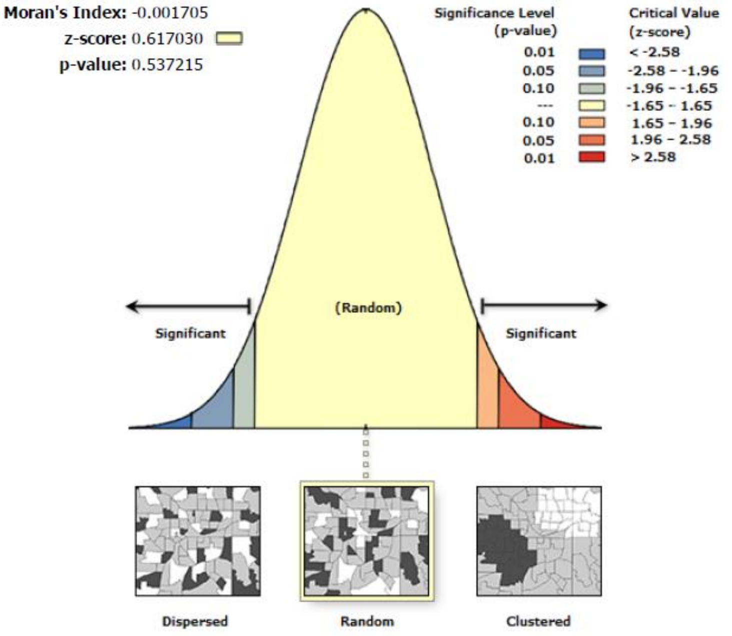
Dependent variable of the model is (log) GVA per worker. All variables, except 'Area' enter the model in logs. Standard-errors are in parentheses.
* indicates significance at the 10 percent level; ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level
Analytical derivatives are used to solve the model.
Estimation of the model uses a modified version of the Levenberg–Marquardt algorithm, and is implemented in Gretl.
NLS covariance matrix and standard-errors are calculated as described in Davidson and MacKinnon, *Econometric Theory and Methods*, 2004.

Similar TSNLS estimations for Models 1 and 3 were attempted (with ERP 1981 as the instrument). However, both models failed to estimate beyond the first stage. Alternative instruments (varying ERP years and including Area) were attempted without any success. It would appear that the specification of Models 1 and 3 may require a more complex instrument structure to account for the non-linearity inherent in the effective-labour component of the model.

However, given the similarity of the NLS and TSNLS results in Table 13, along with the results of the endogeneity-tests in Table 10, we focus our analysis and discussion on NLS (Models 1 to 3), and OLS (Models 4 and 5) estimations.

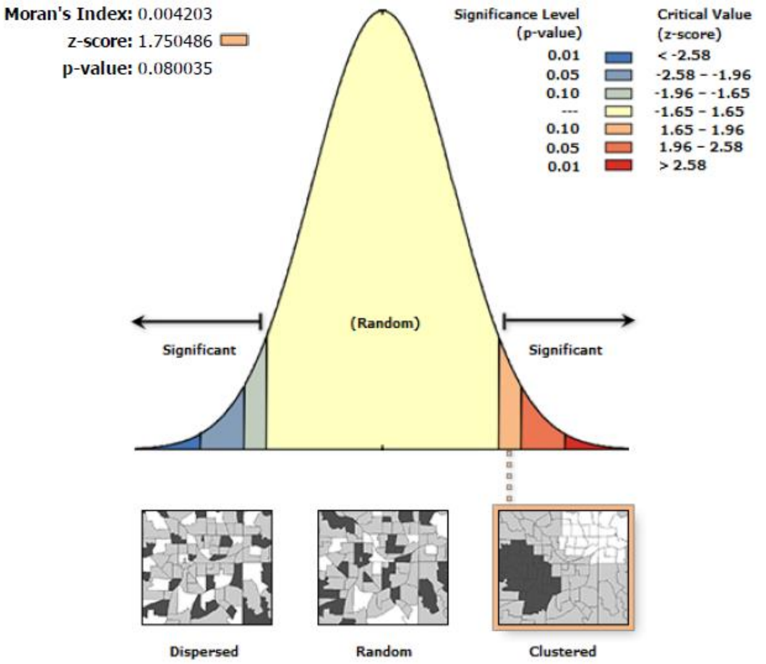
Moran's Index tests for spatial autocorrelation

FIGURE 12. MORAN'S USING DEFAULT DISTANCE THRESHOLD (7.2 DEGREES)



Source: Calculated in ARC GIS 10.2

FIGURE 13. MORAN'S USING DEFAULT DISTANCE THRESHOLD (5 DEGREES)



Source: Calculated in ARC GIS 10.2

Descriptive statistics of data

The following table shows descriptive statistics for SA3 data used. The statistics below are SA3s with a minimum of 10,000 workers (n=264).

TABLE 14. DESCRIPTIVE STATISTICS OF SA3 DATA

Variable	Mean	Median	Min.	Max.	Std. Dev.
Total GVA ($M)	4,093.90	2,413.20	781.51	86,055.00	7,530.10
Total GVA excluding Agriculture and mining ($M)	3,793.50	2,121.00	774.98	86,008.00	7,440.70
Total workers	34,095.00	24,809.00	10,394.00	416,000.00	38,746.00
Total workers excluding Agriculture and mining	32,999.00	23,776.00	10,194.00	415,810.00	38,992.00
Total capital ($M)	9,063.20	6,438.40	897.62	106,060.00	9,002.10
Total capital excluding Agriculture and mining ($M)	7,484.80	5,094.80	794.21	33,794.00	6,110.20
GVA per worker excluding Agriculture and mining ($)	98,921.00	92,495.00	70,004.0	240,680.00	25,125.00
Capital per worker excluding Agriculture and mining ($)	247,090.0	233,510.00	11,672.0	629,560.00	84,131.00
Market Potential index (commute time-discounted)	9,422.60	6,267.10	485.65	72,818.00	9,646.10
Share of university-educated workers (%)	0.23	0.20	0.10	0.54	0.10
Ratio of Manufacturing to Services GVA (MS ratio)	2.89	1.85	0.03	26.09	3.33
Share of workers (or work-hours) lost due to commute (%)	0.09	0.07	0.00	0.30	0.05
Area of an SA3 (square KM)	19,750.00	162.16	10.64	714,830.0	87,490.00
Average minutes per KM taken to commute to work	0.88	0.87	0.00	1.39	0.13
Estimate Resident Population at 2001	68,366.00	57,543.00	1,420.00	178,030.0	34,966.00

Notes: SA3s with average minutes per KM zero is for SA3s with commute times more than 30% of an average work-day. This captures fly-in-fly-out workers in SA3s such as Pilbara.

Software acknowledgments

It should be noted that the MS Word template used to write this thesis is a modified version of the MS Word template provided by *SGS Economics and Planning*. Similarly, market potential calculations also used an MS Excel-tool provided by *SGS Economics and Planning*. All spatial calculations and analysis was conducted on ARC GIS 10.2, and all econometric calculations used Gretl 1.10.1.

10 REFERENCES

Abdel-Rahman, H. M., & Fujita, M. (1990). Product Variety, Marshallian Externalities, and City Sizes. *Journal of Regional Science, 30*(2), 165-183.

Abelson, P., Joyeux, R., & Mahuteau, S. (2012). Modelling House Prices across Sydney with Estimates for Access, Property Size, Public Transport, Urban Density and Crime. *NILS Working Paper Series - Flinders University, 181*.

Aizpurua, J., Benito, J., & Puértolas, J. (2003). Market allocations of location choice: An example. *Universidad Pública de Navarra Department of Economics Discussion paper, 1*(4).

Alexandersson, G. (1956). *The Industrial Structure of American Cities: a geographic study of urban economy in the United States.* Stockholm : Almqvist & Wiksell.

Almazan, A., De Motta, A., & Titman, S. (2007). Firm location and the creation and utilization of human capital. *Review of Economic Studies, 74*(4), 1305-1327.

Alonso, W. (1964). *Location and Land Use.* Cambridge (MA): Harvard University Press.

Alonso, W. (1971). The economics of urban size. *Papers in Regional Science, 26*(1), 67–83.

Arnott, R., & McMillen, D. (2007). *A Companion to Urban Economics.* Malden, MA: Blackwell Publishing.

Arrow, K. J., & Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, 265–290.

Arthur, W. B. (1990). Positive feedbacks in the economy. *Scientific American, 262*(2), 92-99.

Au, C., & Henderson, V. J. (2005). Are Chinese Cities Too Small? *The Review of Economic Studies, 73*(3), 549-576.

Audretsch, D., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review, 86*(3), 630-640.

Auerbach, F. (1913). Das gesetz der bevolkerungskoncentration. *Petermanns Geographische Mitteilungen*, pp. 74– 76.

Baldwin, R., Rikard, F., Martin, P., Ottaviano, G., & Robert-Nicoud, F. (2003). *Economic Geography and Public Policy.* Princeton, NJ: Princeton University Press.

Barro, R. J. (2001). Human Capital and Growth. *American Economic Review, 91*(2), 12-17.

Becker, R., & Henderson, J. V. (2000). Intra-Industry Specialization and Urban Development. In J.-M., & H. J.-F. Thisse, *Economics of Cities: Theoretical Perspectives* (pp. 138-166). Cambridge: Cambridge University Press.

Beckmann, M. J. (1969). On the distribution of urban rent and residential density. *Journal of Economic Theory, 1*(1), 60-68.

Bergsman, J., Greenston, P., & Healy, R. (1972). The Agglomeration Process in Urban Growth. *Urban Stubies, 9*(3), 263–288.

Bertaud, A., & Renaud, B. (1997). Socialist cities without land markets. *Journal of Urban Economics, 41*(1), 137-151.

Black, D., & Henderson, V. (2003). Urban Evolution in the USA. *Journal of Economic Geography, 3*(4), 343–373.

Boschma, R., & Frenken, K. (2011). The emerging empirics of evolutionary economic geography. *Journal Of Economic Geography, 11*(2), 295-307.

Bosker, M., Brakman, S., Garretsen, H., & Schramm, M. (2007). Looking for multiple equilibria when geography matters: German city growth and the WWII shock. *Journal of Urban Economics, 61*(1), 152-169.

Bosker, M., Brakman, S., Garretsen, H., & Schramm, M. (2012). Relaxing Hukou: Increased labor mobility and China's Economic Geography. *Journal of Urban Economics, 72*(2), 252-266.

Brakman, S., & Garretsen, H. (2009). Trade and geography: Paul Krugman and the 2008 Nobel Prize for Economics. *Spatial Economic Analysis, 4*(1), 5-23.

Brakman, S., Garretsen, H., & Schramm, M. (2004). The strategic bombing of German cities during WWII and its impact on city growth. *Journal of Economic Geography, 4*(2), 201-218.

Brakman, S., Garretsen, H., & Schramm, M. (2006). Putting new economic geography to the test: free-ness of trade and agglomeration in the EU regions. *Regional Science and Urban Economics, 36*(5), 613-636.

Brueckner, J., Thisse, J.-F., & Zenou, Y. (1999). Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review, 43*(1), 91-107.

Camagni, R., Capello, R., & Caragliu, A. (2013). One or infinite optimal city sizes? In search of an equilibrium size for cities. *The Annals of Regional Science, 51*(2), 309-341.

Chinitz, B. J. (1961). Contrasts in Agglomeration: New York and Pittsburgh. *American Economic Review, 51*(2), 279-289.

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland.* Iena, DE: Fischer.

Coles, M. G., & Smith, E. (1998). Marketplaces and Matching. *International Economic Review, 39*(1), 239-254.

Coombes, M. (2009). Functional economic areas. In WMRO, *Refreshing the West Midlands evidence base: a state of the region thematic report* (pp. 12-18). Birmingham: West Midlands Regional Observatory.

Crozet, M. (2004). Do migrants follow market potentials? An estimation of a new economic geography model. *Journal of Economic Geography, 4*(4), 439-458.

Daniels, R., & Mulley, C. (2011). Exploring the role of public transport in agglomeration economies and centres. *Australasian Transport Research Forum (ATRF) Proceedings.* Adelaide: Australasian Transport Research Forum.

Duranton, G. (1998). Labor Specialization, Transport Costs, and City Size. *Journal of Regional Science, 38*(4), 553-573.

Duranton, G., & Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. In J. V. Henderson, & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2063–2117). Amsterdam: North-Holland.

Eaton, B., & Lipsey, R. G. (1977). The introduction of space into the neoclassical model of value theory. In M. Artis, & A. Nobay, *Studies in Modern Economies* (pp. 59-96). Oxford: Basil Blackwell.

Fallick, B., Fleischman, C., & Rebitzer, J. B. (2006). Job hopping in the Silicon Valley: The microfoundations of an industry cluster. *Review of Economics and Statistics, 88*(3), 472-481.

Fischer, M. M., & Getis, A. (2009). *Handbook of Applied Spatial Analysis.* Berlin Heidelberg: Springer.

Fowkes, T. (2001). Principles of Valuing Business Travel Time Savings. *University of Leeds Institute of Transport Studies Working Paper*.

Fox, K., & Kumar, T. (1965). The Functional Economic Area: Delineation and Implications for Economic Analysis and Policy. *Papers in Regional Science, 15*(1), 57-85.

Fujita, M., & Mori, T. (1996). The Role of Ports in the Making of Major Cities: Self-Agglomeration and Hub-Effect. *Journal of Development Economics, 49*(1), 93-120.

Fujita, M., & Thisse, J.-F. (2002). *Economics of Agglomeration: Cities, Regions, and International Trade.* Cambridge, MA: MIT Press.

Garretsen, H., & Martin, R. (2010). Rethinking (New) Economic Geography Models: Taking Geography and History More Seriously. *Spatial Economic Analysis, 5*(2), 127-160.

Glaesar, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, 1126-1152.

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics, 46*(2), 254-277.

Graham, D. (2009). Identifying Urbanisation and Localisation Externalities in Manufacturing and Service Industries. *Papers in Regional Science, 88*(1), 63-84.

Graham, D. J. (2007). Agglomeration, productivity and transport investment. *Journal of Transport Economics and Policy, 41*(3), 317-343.

Guerin-Pace, F. (1995). Rank-size distribution and the process of urban growth. *Urban Studies, 32*(3), 551– 562.

Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers, 44*(4), 315-348.

Head, K., & Mayer, T. (2004). The Empirics of Agglomeration and Trade. In J. V. Henderson, & J.-F. Thisse, *Handbook of Regional and Urban Economics* (pp. 2609–2669). Amsterdam: North-Holland.

Helpman, E. (1998). The Size of Regions. In D. Pines, E. Sadka, & I. Zilcha, *Topics in Public Economics: Theoretical and Applied Analysis* (pp. 33-54). Cambridge: Cambridge Universtiy Press.

Helpman, E. (n.d.). The size of regions. In D. Pines, I. Sadka, & I. Zilcha, *Topic in public economics. Theoretical and empirical analysis* (pp. 33-54). Cambridge University Press.

Helsley, R. W., & Strange, W. (1990). Matching and Agglomeration Economies in a System of Cities. *Regional Science and Urban Economics, 20*(2), 189-212.

Henderson, J. V. (1986). Efficiency of resource usage and city size. *Journal of Urban Economics, 19*(1), 47-70.

Henderson, J. V. (2003). Marshall's Scale Economies. *Journal of Urban Economics, 53*(1), 1-28.

Henderson, J. V. (2003). The Urbanization Process and Economic Growth: The So- What Question. *Journal of Economic Growth, 8*(1), 47-71.

Henderson, J. V., Kuncoro, A., & Turner, M. (1995). Industrial Development in Cities. *Journal of Political Economy, 103*(5), 1067-1090.

Henderson, J. V., Zmarak, S., & Venables, A. (2001). Geography and development. *Journal of Economic Geography, 1*(1), 81-105.

Hensher, D. (1977). *Value of Business Travel Time.* Elmsford, NY, USA: Pergamon Press.

Hensher, D. A., Truong, T. P., Mulley, C., & Ellison, R. (2012). Assessing the wider economy impacts of transport infrastructure investment with an illustrative application to the north-west rail link project in Sydney, Australia. *Journal of Transport Geography, 24*, 292-305.

Holmes, T. J. (1999). Localization of industry and vertical disintegration. *Review of Economics and Statistics, 81*(2), 314-325.

Hummels, D. (2004). *Towards a Geography of Trade Costs.* Mimeo: Purdue University.

Jacobs, J. (1969). *The Economy of Cities.* New York, NY, USA: Random House.

Koopmans, T. C., & Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica, 25*(1), 53-76.

Krugman, P. (1991). *Geography and Trade.* Cambridge, MA: MIT Press.

Krugman, P. (1991). Increasing returns and economic geography. *The Journal of Political Economy, 99*(3), 483-499.

Krugman, P. (1993). First Nature, Second Nature, and Metropolitan Location. *Journal of Regional Science, 33*(2), 129-144.

Krugman, P. (1995). *Development, Geography and Economic theory.* Cambridge, MA: The MIT Press.

Krugman, P. (1998). The Role of Geography in Development. *Annual World Bank Conference on Development Economics.* Washington, D.C.: World Bank.

Krugman, P. (1998). What's new about the new economic geography? *Oxford Review of Economic Policy, 14*(2), 7-17.

Krugman, P. (2010). The New Economic Geography, now middle-aged. *Prepared for presentation to the Association of American Geographers*. Princeton University.

Kulish, M., Richards, A., & Gillitzer, C. (2011). Urban Structure and Housing Prices: Some Evidence from Australian Cities. *RBA Research Discussion Paper*.

Launhardt, W. (1882). Die Bestimmung des zwekmässigsten Standorts einer gewerblichen Anlage. *Zeitschrift des Vereins Deutscher Ingenieure*, 106-115.

Lee, M. (2012, January). New Economic Geography and Tax Competition in the PRC: A Firm-Level Data Analysis with Policy Implications. *ADB Economics Working Paper Series, 297*.

Li, C., & Gibson, J. (2015). City scale and productivity in China. *Economics Letters, 131*, 86-90.

Lianos, P. T. (n.d.). Aristotle's Macroeconomic model of the city-state. In Balfoussias S., *Essays in economics: Applied studies on the Greek economy* (pp. 675–687). Athens: Planning and Economic Research.

Lin, T., & Truong, T. P. (2012). Transport improvement, agglomeration effect and urban productivity: The case of Chinese cities. *Institute of Transport and Logistics (Working Paper)*.

Lösch, A. (1954). *The economics of location.* New Haven, CT: Yale University Press.

Maré, D. C., & Graham, D. J. (2009). Agglomeration Elasticities in New Zealand. *Motu Working Paper, 9*(6).

Maré, D. C., & Graham, D. J. (2013). Agglomeration elasticities and firm heterogeneity. *Journal of Urban Economics, 75*, 44-56.

Marshall, A. (1890). *Principles of Economics.* London: Macmillan.

Melo, P., & Graham, D. (2009). Agglomeration Economies and Labour Productivity: Evidence from longitudinal worker data for GB's Travel-to-Work areas. *SERC Discussion Paper, 31*.

Moretti, E. (2004). Human Capital Externalities in Cities. In J. V. Henderson, & J. F. Thisse, *Handbook of Urban and Regional Economics* (pp. 2243–2291). Amsterdam: North-Holland.

Mueller, S. (2008). Capital stock approximation using firm level panel data - A modified perpetual inventory approach. *Jahrbücher für Nationalökonomie und Statistik, 228*(4), 357-371.

Ohlin, B. (1933). *Interregional and International Trade.* Cambridge (Mass.): Harvard University Press.

Okamoto, R. (2007). Location choices of firms and workers in an urban model with heterogeneities in skills and preferences. *Regional Science and Urban Economics, 37*(6), 670-687.

Ottaviano, G., Takatoshi, T., & Thisse, J. (2002). Agglomeration and Trade revisited. *International Economic Review, 43*, 409-436.

Overman, H. G., & Puga, D. (2010). Labour pooling as a source of agglomeration: An empirical investigation. In E. L. Glaeser, *Agglomeration Economics* (pp. 133-150). Chicago, IL: University of Chicago Press.

Overman, H. G., Redding, S., & Venables, A. J. (2004). The Economic Geography of Trade, Production, and Income: A Survey of Empirics. In J. Harrington, & K. Choi, *LSE Handbook of International Trade* (pp. 353–387). Malden, MA: Basil Blackwell.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review, 25*(1), 221-247.

Panek, S. D., Baumgardner, F. T., & McCormick, M. J. (2007). *Introducing New Measures of the Metropolitan Economy Prototype GDP-by-Metropolitan-Area Estimates for 2001–2005.* U.S. Bureau of Economic Analysis.

Porter, M. E. (1990). *The Competitive Advantage of Nations.* New York: Free Press.

Potter, A., & Watts, D. (2014). Revisiting Marshall's Agglomeration Economies: Technological Relatedness and the Evolution of the Sheffield Metals Cluster. *Regional Studies, 48*(4), 603-623.

Puga, D. (1999). The Rise and Fall of Regional Inequalities. *European Economic Review, 43*(2), 303-334.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of regional science, 50*(1), 203-219.

Quigley, J. M. (1998). Urban Diversity and Economic Growth. *Journal of Economic Perspectives, 12*(2), 127-138.

Ramjerdi, F., Rand, L., & Soelensminde, K. (1996). The Norwegian Value of Time Study - Some Preliminary Results.

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics, 34*(3), 380-400.

Redding, S. (2010). The empirics of new economic geography. *Journal of Regional Science, 50*(1), 297-313.

Romer, P. M. (1994). The Origins of Endogenous Growth. *The Journal of Economic Perspectives, 8*(1), 3–22.

Rosen, K., & Resnick, M. (1980). The size distribution of cities: an examination of the Pareto law and primacy. *Journal of Urban Economics, 8*(2), 165–186.

Rosenthal, S. S., & Strange, W. C. (2001). The determinants of agglomeration. *Journal of Urban Economics, 50*(2), 191-229.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In G. D., H. V., & S. W., *Handbook of Urban And Regional Economics* (Vol. 4, pp. 2119-2171). Elsevier.

Rossi-Hansberg, E., & Wright, M. L. (2007). Urban Structure and Growth. *Review of Economic Studies, 74*(2), 597-624.

SAU, C., & Coombes, M. (2010). *Functional Economic Market Areas: an economic note.* London: Communities and Local Government.

SGS Economics and Planning. (2012). *Productivity and Agglomeration Benefits in Australian Capital Cities.* Sydney: COAG Reform Council.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.* Edinburgh, Scotland.

Soo, K. (2005). Zipf's Law for cities: a cross-country investigation. *Regional Science and Urban Economics, 35*(3), 239–263.

Starrett, D. A. (1974). Principles of optimal location in a large homogenous area. *Journal of Economic Theory, 9*(4), 418-448.

Starrett, D. A. (1978). Market Allocations of Location Choice in a Model with Free Mobility. *Journal of Economic Theory, 17*(1), 21-37.

Stevens, G. V. (1989). *A substitute for the capital stock variable in investment functions.* Washington, DC: Board of Governors of the Federal Reserve System.

Strange, W. C., Faggio, G., & Silva, O. (2014). Heterogeneous Agglomeration. *SERC Discussion Paper*.

Thisse, J., & Walliser, B. (1998). Is space a neglected topic in mainstream economics? *Louvain Economic Review, 64*(1), 11-22.

Venables, A. J. (2010). Economic Geography and African Development. *Papers in Regional Science, 89*(3), 469-483.

von Thünen, J. H. (1826). *The Isolated State.* Hamburg: Perthes.

Weber, A. (1909). *Über den Standort der Industrien.* Tübingen: J.C.B. Mohr.

Weinstein, D. R. (2002). Bones, Bombs, and Break Points: The Geography of Economic Activity. *The American Economic Review, 92*(5), 1269-1289.

World Bank. (2009). *World Development Report.*

Xu, Z. (2009). Productivity and agglomeration economies in Chinese cities. *Comparative Economic Studies, 51*(3), 284-301.

Zipf, G. (1949). *Human Behaviour and the Principle of Least Effort.* Reading, MA: Addison-Wesley.