CHAPTER 1 INTRODUCTION

1.1 Defining the research question

This study addresses the question of delineating the spatial dimensions of the residential urban real estate submarket structure without reference to any \textit{a priori} definition. The real estate market may be viewed as consisting of a series of spatial submarkets within which may be simultaneously nested structural submarkets (Adair et al. 1996b; Tu 1997; Watkins 2001). The research question focuses on the spatial dimensions of the submarket structure, recognising that this structure is a complex bundle of both spatial and structural components that form the real estate geography of a study area. Importantly, it is this geography that consumers purchase when satisfying their need for housing (Evans 1995, p.6) and therefore needs to be quantified and analysed in terms of the market place in order to gain an understanding of the spatial submarket structure.

Wide-ranging data are collected at the individual property level describing the real estate geography across the entire study area. It is the spatial variability of the market price of this real estate geography that alone defines the spatial dimensions of that market. These dimensions are defined, in this study, as geospatial submarkets. The study is undertaken against a background of a residential real estate market that is unlike any other commodity market in terms of its complexity and dimensions. The property market is complex because it is heterogeneous in nature. Each property is unique, immobile, modifiable and expensive to both buy and sell (Galster 1996, p.1798). The property market is inefficient because it violates many of the classical assumptions of an efficient market in that it is not made up
of a homogenous product with many buyers and sellers all with a full knowledge of the marketplace (Evans 1995, p.16). Also the property market is unlike many other markets in terms of its dimensions. Although it does have the classical price quantity equilibrium it also has a spatial equilibrium (price accessibility) dimension as well (Thrall 2002, p.38).

The real estate market can be viewed over a wide spectrum of scales ranging from the macro national level to the micro urban level (Meen 2001, p.1). This thesis addresses only the urban-level end of the residential real estate market spectrum and the spatial delineation of the submarkets that may exist therein. Due to the complex heterogeneous nature of the market as a whole, urban land professionals have adopted the view that the market may be perceived as a series of interconnected submarkets in which more homogeneous market behaviour may be observed (Galster 1996, pp.1803–4; Grigsby 1963, p.34).

However, to date there has been difficulty in defining and delineating these submarket boundaries. Even at the broadest level there is much discussion as to whether they should be expressed in terms of spatial or structural attributes or indeed both (Watkins 2001, p.2235). Structural attributes include the physical descriptors of the dwelling including the wall and roof construction, the size, condition, and age of the dwelling (Bourassa et al. 2007; Grether & Mieszkowski 1974). The spatial attributes describe the amenity of the location, including environmental and socio economic characteristics as well as the proximity to various services such as health, schools and shopping (Chhetri et al. 2006; Jackson et al. 2007; Kestens et al. 2004). The exact list of such attributes may vary from one study area to another depending on the importance the different markets may place on the different attributes. Orford (1999, p.50) argues the structural attributes and location are related. This makes them both important in delineating the spatial dimensions of the submarket structure. Although the
number of structural attributes is limited and well researched, this is not the case with spatial attributes, whose numbers are almost limitless (Orford 1999, p.63). Hedonic modelling has been widely used for a long time to express property value in terms of its constituent components and this literature is well reviewed (Kauko 2000). Accounting for such a wide ranging number of attributes in hedonic modelling can be problematic. Historically, hedonic modelling has incorporated up to 20 or more spatial and structural variables (Orford 1999, p.64, Song and Knaap 2003, p.227). Problems of interpretability and multicollinearity encountered in models of this size are of concern as discussed by Orford (1999, p.64). As more spatial attributes are identified (in terms of their contribution to market behaviour) and measured there will be a need to manage these data in hedonic modelling from the point of view of both statistical validity and interpretability of the results. The use of principal component analysis (PCA) may be an appropriate technique that quantifies appropriate data in a form that can be used in further analysis by reducing the large data sets to a meaningful and manageable number of dimensions. This also satisfies the identified need in the submarket literature to base submarket analysis on an understanding of the complete underlying residential real estate structure rather than select components thereof (Maclennan & Tu 1996, p.389).

Anecdotally, ‘location’ is understood to be an important component in real estate market behaviour. A three bedroom brick dwelling in Adelaide is in a different market to a three bedroom brick dwelling located on Sydney or New York. The same is true within a metropolitan area; one location may be a different market to another and the further they are apart the more likely this is to be so. This may be viewed as an adaptation of Tobler’s First Law (Tobler 1970). As Thrall (2002, p.38) points out ‘location matters, geography matters’. It is therefore important to understand the locational aspects of the real estate
submarket structure as real estate is considered to be a ‘key driver’ of the urban economy (Gibb & Hoesli 2003, p.887).

This study focuses on delineating the spatial aspects of the submarket structure in terms of the market’s interpretation of the underlying residential living structure across geographic space, and defines these spatial aspects to be geospatial residential submarkets. The research question is, that given these geospatial submarkets exist, how are they to be found?

1.2 Objective of the study

The objective of the study is to develop a methodology for deriving the spatial dimensions of urban residential real estate submarkets based on the behaviour of the marketplace with respect to the underlying dimensions of the urban residential real estate geography, without reference to existing *a priori* spatial definitions. The efficacy of the derived methodology is to be assessed by applying it to a study area containing approximately 440 000 urban residential properties.

The objective is achieved in two stages:

1 Derive the underlying dimensions of the urban residential real estate geography using variables describing both the spatial and structural attributes of all the residential properties in the study area. These dimensions will be called the residential living structure (RLS) of the study area and describes its underlying residential real estate geography.

2 Derive the spatial dimensions of the urban residential submarkets through the relationship of the marketplace, as represented by housing transaction prices, and the RLS (stage 1) at a particular point in time.
1.3 Nature of submarkets

In order to appreciate the context of the research question and the resulting objectives of the study it is helpful to understand the nature and importance of residential real estate submarkets and studies already undertaken to define them. There is general support in the literature for the theoretical existence of residential real estate submarkets, and for their importance (Goodman & Thibodeau 2003, p.182; McCluskey et al. 2002, p.2), although there is very little consensus as to what the model of a submarket should look like (Adair et al. 1996b, p.67). The literature contains examples of many submarket identifiers that are used at the local urban level in terms of spatial or structural characteristics or indeed both (Bourassa et al. 2003. p.13). A comprehensive summary of the studies covering the different identification types is given by Watkins (2001, pp.2238, 2242) who outlines suggested reasons for the failure to develop submarket models (p.2235). The principal reasons, in his opinion, include a lack of submarket definition and identification procedures.

The structural market characteristics include property attributes such as property type, number of rooms, dwelling size and condition, number of stories, block size and shape (Watkins 1999, pp.161–3). The number and type of these attributes may vary between study areas but most areas have similar core characteristics. The spatial characteristics include attributes that describe the location of the property and its neighbourhood. These include socio economic, amenity, environmental, school zones, crime rate, real estate agent influence, and ethnicity measures as well as accessibility to various services (retail, health and education) and other data that may be considered as influencing a buyer’s choice of housing (Kestens et al. 2004; Orford 1999; Palm 1978). In reality both the spatial and structural submarket attributes influence the buyer’s choice of housing and they should be
considered simultaneously when defining submarkets (Adair et al. 1996b; Maclellan & Tu 1996; Watkins 2001). This is the adopted approach in this study.

There have been a number of different approaches taken to determine spatial submarket delineation. For example, there is a body of literature in which the spatial components of submarkets are thought to be important but whose delineation is assumed to conform with existing administrative boundaries such as suburbs, postcode districts, local government areas, electoral boundaries and other forms of *a priori* definitions proposed by real estate agents and other land professionals (Adair et al. 1996b; Maclellan & Tu 1996; Watkins 2001). In studies based on *a priori* definitions, separate submarket status is established by the now accepted test (Watkins 2001) for submarket difference laid down by Schnare and Struyk, (1976) as generating significantly different hedonic house price models based on structural property attribute data. Although this approach claims to result in different submarkets, the more exacting goal of optimising the spatial submarket definition still remains unanswered.

Other researchers adopt an approach where the data are used to determine the spatial submarket boundary (Bourassa et al. 1999; Bourassa et al. 2003; Maclellan & Tu 1996; Watkins 2001). These studies approach the identification of submarkets by recognising firstly that *a priori* spatial segmentation does not necessarily give an optimal solution, and secondly, that submarkets constitute more than individual property characteristics and therefore need to be based on the whole underlying residential structure of the study area. It is critical when determining submarket boundaries that the complete (including both spatial and structural criteria) urban structure is used and not just individual dwelling characteristics. Watkins (2001, p.2241) refers to microeconomic theory, and states that:
Chapter 1 – Introduction

... housing submarkets exist where the interaction between segmented demand, characterised by consumer groups, and segmented supply, characterised by product groups, generate price differences for some hypothetical standardised dwelling.

Watkins stresses two important points. Firstly, both product groups (representing the segmented supply characteristics) and consumer groups (representing the segmented demand characteristics) need to be reflected in the model and, secondly, such a model has to be market based as submarkets are clearly a market entity. This second point is vital and supported by Pryce, (2004b, p.2–10) who reports that the use of non-economic criteria to delineate an economic entity 'contains an essential flaw'. Therefore, if submarket boundaries are to be identified, not only is it important to express the residential structure in terms of both product and consumer groups, but also in terms of the marketplace so the resulting spatial submarket structures can be recognised as economic entities.

This study considers an important element in the nature of submarkets to be their structure as a continuous surface rather than a series of chloroplethic steps across geographic space. This is a view expressed by McCluskey et al. (1999, p.4) and supported by Grigsby (1963, p.34) who describes the interdependence of submarkets as a continuum.

There are few Australian examples of submarket studies. Bourassa et al. (1999), using a sample of properties from the metropolitan areas of Melbourne and Sydney adopted principal component analysis (PCA) to extract a reduced set of independent components from structural and local government data for a sample of residential properties to describe the underlying structure and then used cluster analysis to appropriately group these components into submarkets. Adopting a different approach Costello (2004) used property transaction data for Perth to establish market segmentation. Costello took a priori postcode districts and examined price variation across these to build a price-location model of
market segregation. An interesting spatial aspect emerges showing his price-location model submarkets are not necessarily contiguous, a finding supported by Rothenberg et al. (1991, pp.63–4). Although not specifically interested in submarket delineation, Reed (2001) investigated the link between social constructs and established house prices in Brisbane between 1976 to 1996. This work used PCA to group socio economic variables into a number of independent components that were related to house prices. Reed concluded that demographic data had a distinct role to play in the analysis of established housing markets, and that income levels and house value had a close relationship. While the Australian case studies are few there is not the coherent submarket definition and identification procedure suggested by Watkins (2001, p.2235) as being necessary for the development of submarket models.

Urban economists have modelled property markets in terms of the trade-off between the price of a location and its proximity to a centre of interest such as the CBD. However, modern cities have many centres of interest, making the modelling of the resulting house price gradients from these centres quite complex (Orford 1999, p.157). Consequently, this leads to the questioning of this type of modelling approach as being appropriate in accounting for the complexity that exists in modern real estate markets (Badcock 1984, p.18; Badcock 2002, p.157). Regardless of the complexity, the importance for understanding submarket structure is present, and growing, in the Australian context as it is elsewhere.

### 1.3.1 The importance of understanding submarket structure

People live in household units that form the basis of our society. The importance of people’s welfare and happiness through the ability to satisfy their housing needs in the marketplace is of fundamental importance to the wellbeing of that society (Rothenberg et al. 1991, p.1) and
hence, the importance of understanding the underlying residential living structure and its relationship to the marketplace.

The literature highlights the importance of submarkets in the context of an overall need to understand the dynamics and structure of the residential real estate market and how this understanding may be applied (Rothenberg et al. 1991, pp.3–4). It is important for housing policy-makers to understand market structure to improve their decision-making process and not rely on a simplified market model based on assumptions that may not be valid (Maclennan & Tu 1996, p.389; Meen & Meen 2003, p.90). Galster (1996, p.1797) reinforces the view that housing policy in a market-dominated economy should be based on a market view of the housing sector. This is supported in the recommendations of the Review of Housing Supply (Barker 2004) carried out in the U.K., which suggests that market indicators should be used to determine when planning intervention may be appropriate in the development process. A better understanding of the impact of the planning process can be seen if such processes are related to the submarket structure rather than an irrelevant *a priori* spatial proxy. Bates (2006, p.5), in supporting this view, says:

> Analyzing the urban spatial structure as an array of housing submarkets can inform planners in better anticipating the overall effects of policy interventions as caused by housing submarket dynamics and by spatial processes.

Can and Megbolugbe, (1997, p.203) report on the importance of creating housing price indices and explain how these are used in a diversity of applications. These include:

- the ability to monitor and assess risk in the housing and mortgage markets
- measuring housing demand
- establishing price trends
- formulating and designing housing and mortgage policies.
All of which rely on an understanding of the structure of residential submarkets. As Pryce and Evans (2007, p.207) comment, a lot of the political and administrative process (including policy matters) is fundamentally spatial in nature. People are represented and administered in spatial units and therefore an understanding of the spatial aspects of real estate markets will help better align policy administration with marketplace structures.

The concept of submarkets is also important in understanding the market value of property, especially in the field of mass appraisal. Mass appraisal is the art and science of assessing the market value of real property as a basis for determining property tax. Mass appraisal is the derivation of market value at a specific time for an entire jurisdiction, often extending to many hundreds of thousands of properties (Eckert 1990, p.303). A good fiscal cadastre (the end product of a mass appraisal process) should provide comparability of values across the entire jurisdiction and the ability to update property values whenever property taxes are levied. An important feature in the mass appraisal process is the incorporation of the locational aspects of property in modelling its price. Allowing for local market variation in property price enables more accurate and reliable statistical models for the prediction of individual property values to be developed (Figueroa 1999; Fletcher et al. 2000; Gallimore et al. 1996; McCluskey et al. 1999; McCluskey et al. 2002; Watkins 1999). Current mass appraisal modelling does not easily incorporate the spatial structure into mass appraisal models and consequently, the formulation of market based submarket groups may assist this by allowing separate models for separate submarkets (Adair et al. 1996b, p.69; Goodman & Thibodeau 2007, p.209).

The uses to which an understanding of the submarket structure may be put are many and varied. Such an understanding is as important in the Australia economy as in any other
economy. However, before the submarket delineation process can be undertaken, it is important to understand the issues involved.

1.3.2 Identified Issues in defining and delineating submarkets

Case studies reported in the literature identify a number of fundamental issues that need addressing to achieve submarket delineation. Principally, there needs to be:

- An understanding of the underlying structure and dimensions of the marketplace in which submarkets reside (Maclennan & Tu 1996, p.389).
- A recognition of the overemphasis in existing studies on the individual dwelling characteristics rather than the underlying residential structure to determine submarket boundaries (Bourassa et al. 2003, p.14; Orford 2002, p.108) and the reliance on what Orford (1999, p.48) describes as ‘location insensitive’ models that rely mainly on structural characteristics.
- A recognition that submarket behaviour is about the market and hence needs to be considered as a market entity (Pryce 2004a, p.2–10).
- Less reliance on the multitude of different submarket identifiers and more focus on both structural and spatial attributes considered simultaneously (Watkins 2001, p.2235).

1.4 Choice of submarket definition

There are many ways of defining residential submarkets and this is one of the identified problems with current submarket research (Adair et al. 1996b, p.69). However, there appears to be an emerging sense that a workable definition of submarkets should be based on:

- Data analysis and not on a priori knowledge (Bourassa et al. 1999, p.161).
- A recognition that a submarket may be composed of both structural and spatial components simultaneously (Adair et al. 1996a, p.22; Watkins 2001, p.2235).
The underlying residential structure of the study area (Maclennan & Tu 1996, p.389).

The concept of substitutability constrained by price (Bourassa et al. 2003, p.14; Grigsby 1963, p.34). Submarkets are an economic entity and therefore should be defined with reference to the marketplace (Pryce 2004b, pp.2–10).

The definition of a residential spatial submarket adopted in this study accommodates all these elements, thus adhering to residential submarket theory, and is termed a *geospatial residential submarket*. It has been defined for this study as:

A geographic area within which the market prices of the individual components of the underlying residential living structure have a predefined homogeneous pattern.

The submarket identifier is price. In particular it is the price pattern of the continuously changing individual hedonic parameter estimates of the underlying residential living structure components across geographical space that is the basis for the definition. The definition of a particular set of boundaries delineated along that continuous surface is defined by the user for a particular purpose as Bourassa et al. (2003, p.12) point out:

We maintain that the appropriate definition of submarkets depends on the use to which they will be put.

In this study the definition is the degree of homogeneity within the geospatial submarket boundaries of the price patterns required by the user.

The resulting geospatial residential submarkets are not subjected to the test introduced by Schnare and Struyk, (1976) to establish submarket difference, namely, a significant statistical difference in hedonic pricing of contributing components must be observed between submarkets if they are to be deemed separate submarkets (Watkins 2001, p.2241). It is impractical to do so in this study as there are potentially many hundreds of submarkets across the study area. However, while it is recognised this submarket definition may be
Chapter 1 – Introduction

unacceptable to some practitioners, the submarket identifier is price homogeneity although it is not defined at a particular level of significance. This study is not an attempt to redefine the submarket term and so for the sake of clarity has been called a geospatial residential submarket. This is discussed in more detail in chapter 2.

1.5 Achieving the objective

It was an important aspect in achieving the objective of this study that the resulting spatial submarket dimensions were easily understood. They had to be explainable to the end users, and through them the general public, in terms of the main market influences that shaped the geospatial submarket structure. This promotes confidence and acceptance of results because they can be seen as having been based on housing market analysis and an understanding of the submarket structure. The efficacy of the two stage methodology was carried out through its application to a defined study area.

The study area is basically the metropolitan area of Adelaide, South Australia containing approximately 1.1 million people and approximately 440,000 residential dwellings (excluding flats as they represent an investment market) as detailed in chapter 3. It is recognised that the commodity traded in the market place is more than just a dwelling, but rather a piece of real estate geography, which is a complex bundle of many different features (Evans 1995, p.6). Therefore, the data used has been gathered from a variety of data custodians and include socio economic data from the quinquennial 2001 Australian population census, property accessibility data from The University of Adelaide (GISCA), structural property characteristic, sales transaction price data and the property cadastre from the South Australian Valuer-General and remotely sensed amenity data from the South Australian Government’s Department of Environment and Heritage. The date
adopted for this study is August 2001, the month in which the 2001 census was carried out. In Australia, a complete household census is carried out every five years. Other data were collected as near as practicable to the same date. At the time of undertaking this study the Australian census data for 2006 was unavailable.

In order to address the research question fully the process has been divided into two stages so as to recognise the important issues identified in submarket analysis. The first two issues, discussed earlier, namely recognising the underlying residential structure and the use of a wide variety of attribute data to describe that underlying structure are addressed in the first stage. The last two issues are addressed in the second stage by relating the market behaviour to the underlying structure through the adoption of price as the submarket identifier.

1.5.1 Stage 1 – The creation of the residential living structure (RLS)

This study proposes a residential living structure (RLS) as a vehicle for understanding and quantifying the underlying structure and dimensions of the urban real estate geography in which submarkets are embedded. Principal component analysis (PCA) is used to reduce a large number of spatial and structural property attributes into a smaller number of components that describe the underlying dimensions of that data. They are adopted as the dimensions of the RLS. As discussed more fully in Chapter 4, PCA has been extensively used in the real estate literature (including Australian studies) to describe underlying data structures (Bourassa et al. 1999; Maclennan & Tu 1996; Watkins 1999). This study builds upon existing research and extends it in the following ways by:

- Including a broad range of relevant property attributes describing both the spatial and structural aspects of the property (in this study 50 variables are described in chapter 3).
Chapter 1 – Introduction

- Making no assumption as to the spatial pattern of the resulting RLS, as such attribute 
data are collected at the individual property level.

- Including all properties in the study area (in this case, approximately 440,000 
residential properties).

The contribution of the RLS as a basis for the determination of submarket boundaries is 
significant in that the resulting components provide a more complete understanding of the 
underlying structure as it is based on all residential properties in the study area and not just 
a sample. The RLS contributes to the first issue identified earlier in this chapter that states 
submarket delineation should be based on the residential structure rather than individual 
characteristics.

Further, it contributes to the second issue by calculating component scores for each 
property in the study area and adopting them as surrogate property characteristics 
removing the overemphasis on the individual dwelling characteristics.

It therefore provides a comprehensive and independent set of surrogate property 
characteristics that can be used in stage 2 of the project.

1.5.2 Stage 2 – The relationship of the marketplace to the RLS

The aim of stage 2 is to address the remaining two identified issues concerning submarket 
identification namely the relating of the RLS to the marketplace and the question as to a 
submarket identifier. In stage 1 a number of surrogate property characteristics were derived 
for each of the approximate 440,000 residential properties in the study area. A subset of 
these properties that sold, thus describing the market at the study date (August 2001), was 
extracted together with their surrogate property characteristics and selling price and used to 
relate the marketplace with the RLS using ‘price’ as the submarket identifier.
This relationship between the market place and the surrogate property characteristics is achieved using a geographically weighted regression (GWR) model (Fotheringham et al. 2002). GWR is a technique that differs from normal regression models in that it does not assume that independent variables' contribution to the dependent variable is the same at different geographic locations. It does not assume spatial stationarity and can be conceptualised as shown in figure 1.1. A number of different regression points (X in figure 1.1) are taken across geographic space using surrounding sales data each weighted according to their distance from the regression point pursuant to a prescribed weighting function (Wij in figure 1.1).

Figure 1.1 GWR conceptual model

Note: The Wij is the weighting function attributed to each of the sales data. It portrays the Gaussian decay nature of the function across space.

Source: (Fotheringham et al., 2002, p.45)

The model assigns the transaction price of a property, at a given point in time, as the dependent variable and the residential living structure principal components (surrogate property characteristics) as the independent variables. This is a point of difference with the current literature in that the selling price of the property can be related to their surrogate property characteristics knowing they reflect the dimension of that property in relation to
the structure of the whole study area from which they were derived. The hedonic GWR model takes the general form:

**Equation 1.1 General GWR hedonic model**

\[
\text{Sale Price} = b_0 (u,v) + b_1 (F_1)(u,v) + \ldots + b_n (F_n)(u,v) + \text{error}(u,v)
\]

where:
- \((u,v)\) are the location of the sale
- \(b_0\) to \(b_n\) are parameter estimates
- \(F_1 \ldots F_n\) are the surrogate property characteristics.

The essence of this thesis, and the basis of the definition of geospatial submarket delineation, is that it is the difference in the price pattern of the parameter estimates (\(b_0\) to \(b_n\) in equation 1.1) of the independent variables as the GWR model moves across the study area will serve as a proxy indicator of a submarket change and hence the presence of a submarket boundary.

At each regression point, as the GWR model moves across space, hedonic parameter estimates (\(b_1\) to \(b_n\) in the model shown in equation 1.1) are calculated for each of the independent variables. Rothenberg et al. (1991) adopted the hedonic index (HI) as an indicator of housing quality and used it to rank individual market segments. In summary, the use of hedonic indexes is about viewing the value of what Rothenberg refers to as the housing unit quality implicitly defined by the marginal prices of a particular bundle of characteristics. The single quality index is then used to define submarkets by partitioning the array of hedonic indexes (HI) into a number of groups of housing units having similar hedonic values. In this study the individual parameter estimates are derived from the hedonic geographically weighted regression models generated across the study area and clustered into groups of similar price patterns using Ward’s hierarchical clustering algorithm. By considering the HI rather than the individual parameter estimates the variation in the different combinations of these parameter estimates that constitute the same HI at each regression point can be missed. This is important as, in this study, it is the variation in the price patterns
of the parameter estimates that denoted a change in the market’s interpretation of the RLS and hence a change in submarket. Therefore, groups of similar patterns of parameter estimates are used to denote similar submarkets and used as a definition of a geospatial submarket.

1.6 Conclusions

The residential real estate market is a complex market, which needs to be understood in terms of its spatial structure in order to support the two main identified users namely urban policy and planning, and property tax administration functions. The literature highlights the important aspects of residential submarket theory that should be preserved in order to successfully delineate the residential submarket structure and to a lesser extent the analytical techniques capable of achieving this. The residential submarket theory may be best summarised as being a recognition that it is the interpretation by the marketplace of the real estate geography in which a dwelling resides that is fundamental to its’ spatial submarket classification. The wide range of variables describing the real estate geography collected for the total population of the study area may be reduced to a manageable and statistically independent number of surrogate property characteristics. The change in the price patterns of these surrogate property characteristics in a hedonic geographically weighted regression model may be defined as the change in geospatial submarket boundaries. This approach addresses the research question of establishing a methodology to delineate the spatial component of what may or may not be a nested submarket structure. It preserves the theoretical concepts of spatial submarkets in a simple and transparent way providing an understandable structure in which to pursue any submarkets that may be nested within the geospatial submarket framework.
The focus of the thesis is predominately concerned with the derivation of a methodology, rather than a definitive set of property attributes, that may be used to define geospatial submarket boundaries and to test the efficacy of that methodology within a defined study area. An important component of such a methodology is its ability to accommodate the wide range of data that is needed to explain the complex nature of the real estate geography. If the methodology is to be useful across different study areas it can not be data specific as residential real estate markets may be influenced by different property attributes in different study areas. Therefore, although it is important to select appropriate data, deriving the best possible data is considered a separate and complex problem in its own right. The focus of this study is therefore about equipping the real estate professional with a consistent and transparent process that can accommodate a wide range of property attributes and explain the resulting boundaries in terms of those attributes without any reference to proxy a priori spatial boundaries. The data used in this study were taken from secondary sources that were readily available in terms of the imposed time and cost restraints of the study. It is anticipated that the application of the resulting methodology can highlight data deficiencies and thereby contribute to future data collection projects necessary to better understand the residential real estate submarket structure.

1.6.1 Contribution of the thesis

This thesis contributes to the overall understanding of residential real estate submarket delineation in a number of ways.

- The methodology developed, from issues identified in the literature, recognises that the spatial submarket structure is inexorably linked with the real estate geography of the study area.
- Improved land management techniques allow an increasing number of attributes about each property to be gathered and formed into a quantifiable geography for the
study area. The RLS contributes to this by allowing these extra data to be included as they become available and are found to be relevant in a manageable and statistically meaningful manner.

- Powerful computer-based statistical techniques allow the researcher to unravel this multidimensional geography that exists in data space and express it in terms of the marketplace across geographic space as delineating two dimensional residential real estate submarket boundaries. The clustering of parameter estimates identified from the GWR model contributes to this.

### 1.7 Structure of the thesis

In chapter 2 the current submarket literature is reviewed. The literature spanning many years has been well summarised (Watkins 2001, pp.2238, 2242). Therefore this chapter focuses on identifying the theory and major issues in which a contribution to a better understanding of residential submarkets may be made. Chapter 3 describes the study area and data used in assessing the efficacy of the methodology. Chapter 4 is concerned with developing a methodology for the derivation of the RLS and the associated parameter estimate price patterns, including appropriate data and analytical techniques, to address the issues uncovered in the literature review. It discusses analytical techniques used in previous studies and identifies areas in which previously used techniques can be extended and how techniques not previously used can be adapted to make a contribution to the thesis. These methods are then described for each of the two stages in more detail in chapters 5 and 6 respectively. The results are presented and discussed for each of the two stages in chapters 7 and 8 with application to both the planning and development process as well as to enhancing mass appraisal techniques that support the property taxation system. Chapter 9 draws conclusions from the study and outlines areas for further research that may contribute to further understanding of the residential real estate submarket structure.
CHAPTER 2 THE NATURE OF THE RESIDENTIAL REAL ESTATE MARKET AND ITS SUBMARKETS

2.1 Introduction

Land is of fundamental importance to people all over the world. It provides food, shelter and the raw materials of production that account for a significant portion of an economy’s GDP (Gibb & Hoesli 2003, p.888). Urban residential land provides the housing and living environment for the urbanised population and to manage its development and affordability requires informed government policy based on a sound understanding of the housing market structure and its behaviour. As Rothenberg et al. (1991, p.1) state:

To understand the urban housing market is … to understand that which fundamentally influences the wellbeing of most citizens.

The importance of people’s stability and welfare through the ability to satisfy their housing needs in the marketplace requires an understanding of the local housing market structure. This will assist in understanding submarket delineation and have direct input into the resolution of many housing policy issues.

The real estate market can be viewed over a wide spectrum of scales ranging from the national level at the broadest end of the spectrum to the urban level at the other (Meen 2001, p.1). This thesis addresses only the urban level of the residential real estate market and the spatial delineation of the submarkets that may exist therein. The real estate market is complex and unique. The commodity traded is the single most expensive investment a person is likely to make and the consequence of a bad decision can affect not only the
individual concerned, but the community as a whole, through the loss of money that may take many years to recover from.

The aim of this chapter is to investigate the current understanding of the residential real estate market and identify opportunities to contribute to the mapping of the spatial dimensions of its submarket structure.

2.2 The general real estate market

A general land market theory was described by David Ricardo and Johann Heinrich von Thünen in the eighteenth century. As discussed by Thrall (2002, p.61), Ricardo argued, from an economic perspective, that the more productive a parcel of land is the more valuable it must be and the comparability of land parcels could be based on their relative productivity. The difference in cost to make two parcels of land equally productive was the difference in value between the two parcels. Again as discussed by Thrall (2002, p.62) it was von Thünen who contributed to the economic concept of land value by adding the spatial component and arguing that the location of property could be quantified in terms of the economic cost of getting the produce from the land to the marketplace. These simple models help us to understand that the land market is economically driven and even though the variables are more complex today, both in terms of their number and their nature, the underlying economic principles remain relevant. Today real property is a prime component of the urban economy and has been suggested that it accounts for as much as 15 to 20% of GDP (Gibb & Hoesli 2003, p.888). Understanding such a significant portion of the economy is vital to the well being of society as a whole. To gain such an understanding a useful starting point is the general theory of economic market modelling.
Classic economic theory simplifies the concept of the general market by making a series of assumptions to generate an understandable economic model. For the market to be efficient, then assumptions must prevail. Evans (1995, p.16) and Orford (1999, p.6) suggest that these include:

- Market participants have full knowledge of the market in which they are participating.
- There is full and free flow of information of the market amongst its participants.
- There are enough transactions generated that any one does not dominate the market.
- The product traded is homogeneous.
- The market is in equilibrium (a price/quantity equilibrium).

Although necessary and useful, these assumptions can be misleading or simply inapplicable when applied to the real estate market and can result in false conclusions being drawn, namely submarkets not existing when in fact they do and existing when they do not. These assumptions are not necessarily valid for the real estate market (Evans 1995, p.16; Maclennan & Tu 1996, p.388; Meen & Meen 2003, p.90). The housing market is different to other commodity markets. As Galster (1996, p.1798) says:

> Housing is a spatially immobile, highly durable, highly expensive, multidimensionally heterogeneous, and physically modifiable commodity. These characteristics shape attitudes and behaviours towards housing and, in turn, influence neighbourhood characteristics, mortgage markets, urban growth and decline and national housing policies.

Examination of the relevance of each of these assumptions can give an overview of the unique nature of real estate as a marketplace commodity. From the market participant’s perspective, the information flow appears to be controlled by the local land agent, who generally specializes in a relatively small geographic area. Various government authorities
who are custodians of certain real estate information gives them a ‘gatekeepers’ role in making available valuable information in the real estate market process. The control by government is usually the search fee which, when included with transaction costs and liquidity constraints, can contribute to ‘trading frictions’ causing inefficiencies in the market as seen by the rise in the number of transactions before a movement in price (Meen 2001, p.64). These transaction costs may be divided into various categories including search, legal and administrative, adjustment and financing costs (Quigley 2003, pp.59–61). The control by the agent lies in the ability to selectively give information that understandably may put their client (normally the seller) in the best light. These controls naturally restrict the flow of information so that the assumption that buyers and sellers are fully informed as regards the market as a whole may not necessarily be true.

The real estate market is heterogeneous. No two properties can be exactly the same. Their different location, no matter how close, will always make each property unique. Even two houses built on adjacent parcels of land by the same builder to the same plan, would be different once the individual owners had made their own modifications in terms of level of maintenance, landscaping and decorative schemes.

Because of the heterogeneous nature of property the validity of the sale price as an indicator of the market in general, or for another particular property, will require adjustments to be made for those differences. Making these adjustments is not an exact science and can lead to differences in the price offered in the marketplace. Information about properties that have sold is not always known to the whole market and therefore some sections of the market may have unfair advantages in negotiating a sale price.
Evans (1995, p.16) says that the real estate market is imperfect and inefficient when he concludes:

It is not true that in every property market there is a homogenous product with many buyers, many sellers, and a high degree of knowledge. This conclusion seems almost too obvious to state, but even though most would recognize that the property market is not like this, few seem to have been willing to accept the consequent conclusion – that the property market is not a perfect market and so is an imperfect market and inefficient.

He emphasizes two consequences of such a market. Firstly, the buyer/seller psychology can have an effect on the final price paid and secondly, the market can be influenced by the negotiating skill of individual buyers and sellers. As a result, he argues, the land agent plays an important role in determining the final price. The individual psychology can be seen in the time a property may spend on the market and the resistance that may build up to a property that has been on the market for a long time.

The housing market has many spatial scales. Meen (2001, p.1) suggests that the spatial scale spectrum ranges from the international to the regional to the urban market scale. At the international end of the scale there is no mention of spatial analysis whereas at the urban market end of the spectrum spatial analysis is commonplace. Arguably, perhaps the most important difference between the housing market, particularly at the urban level, and other commodity markets lies in the nature of their equilibrium. The market equilibrium for the housing market is more than just the classical equilibrium between price and
quantity. The housing market has an extra equilibrium of price and geographical (spatial) position. Housing is unique and fixed in space and as Thrall, (2002, p.38) explains:

[T]he real estate market analyst must consider locational relationships, in addition to price – quantity relationships. Because location matters, geography matters, and we must therefore bring geographic equilibrium into real estate market analysis.

Geographic (spatial) equilibrium describes the trade off between price of property and its accessibility to various nodes of interests. Generally the lower priced properties are further away from nodes of interest such as employment and shopping centres and the more expensive properties are closer. Equilibrium occurs where people are indifferent to alternative locations. In addition, Thrall (2002, p.39) argues that there is more to spatial equilibrium than just the forces attracting to, and repulsing from, certain locations. It is also concerned with:

… the trajectory of movement of land values at a particular location as a result of a change in market forces. If the land values are below spatial equilibrium, then households will move in, bringing the price of land up until successive households are no longer willing to make the move and pay the price required for that location. Land values adjust.

A logical consequence of having a large, imperfect and inefficient single market might be to develop the ability to objectively identify smaller submarkets in which the classic market assumptions may be better observed. The existence of real estate submarkets is generally accepted as a valid concept (Adair et al. 2000, p.1080; Grigsby 1963, p.34), and it is the spatial identification of these submarkets that is important and the focus of this thesis.
2.3 The importance of identifying submarket boundaries

The literature talks about the importance of submarket groups in the context of an overall need to understand the dynamics and structure of the residential real estate market (Maclennan & Tu 1996, p.387; Meen 2001, ch.5). Implied in this are the applications to which such an understanding can be put.

Table 2.1 contains a summary of various functions that are performed by both the public and private sectors and appear to be related to either the formulation of planning policy and strategic planning issues or real estate market value derivation and monitoring.

<table>
<thead>
<tr>
<th>The need (community need)</th>
<th>Organisation responsible for fulfilling the need</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulation of market-based housing policy</td>
<td>Government planning bodies</td>
</tr>
<tr>
<td>Land use policy formulation and administration</td>
<td>State and local government bodies</td>
</tr>
<tr>
<td>Understanding the local housing system (Meen 2001)</td>
<td>Policy planners</td>
</tr>
<tr>
<td>Urban renewal location and programmes (Berry et al. 2002)</td>
<td>State government planning bodies</td>
</tr>
<tr>
<td>Definition of marketing areas (Thrall 2002)</td>
<td>Urban developers</td>
</tr>
<tr>
<td>Identification of deprived and affluent regions (Randolph 2004)</td>
<td>Real estate agents</td>
</tr>
<tr>
<td>Understanding the effect of location – the relative nature of one area with another (Orford 2000, 2002) (Pryce &amp; Evans 2007)</td>
<td>Marketing organisations</td>
</tr>
<tr>
<td>Mass appraisal – more equitable distribution of property tax burden (McCluskey et al. 2002)</td>
<td>State and local government bodies</td>
</tr>
<tr>
<td>Mass appraisal – more equitable distribution of property tax burden (McCluskey et al. 2002)</td>
<td>Property taxation authorities (state and Commonwealth Government)</td>
</tr>
<tr>
<td>Real estate industry functions</td>
<td>Real estate community</td>
</tr>
<tr>
<td>Monitoring housing and mortgage markets (Can and Megbolugbe 1997)</td>
<td>Commonwealth Grants Commission</td>
</tr>
<tr>
<td>Strategic planning (Adair et al. 2000)</td>
<td>Marketing, state and local government planning bodies</td>
</tr>
<tr>
<td>Administrative boundary formulation and maintenance</td>
<td>Electoral commission, Local government ward boundaries</td>
</tr>
</tbody>
</table>

Planning policy addresses the questions of urban renewal, land use policy and strategic planning, whereas real estate market analysis includes concern with marketing, mass appraisal and property and mortgage portfolio risk management.
The importance of an understanding of market structure for an overall housing policy is summarized by Galster (1996, p.1804) when he says:

The ultimate (as opposed to transitory) efficiency and equity of any public or private housing policy is crucially affected by the form of the inter-submarket repercussions.

From a market analysis perspective the need for submarket delineation is expressed by Adair et al. (1996b, p.67) and McCluskey et al. (2002) when referring to the need for the delineation of submarket boundaries in order to accurately reflect the influence of location in property valuation modelling. Watkins (1999, p.161) further notes that global mass appraisal models may suffer from aggregation bias that may be negated by recognized submarkets and adopting separate models for each.

The literature identifies two broad areas of application in which a knowledge of submarket structure can make significant contribution.

### 2.3.1 Housing policy

Housing policy embraces a wide range of issues, from affordable community housing to sustainable urban development. A major challenge in the management of the urban housing market is to understand its structure, its behaviour and thus the ability to satisfy the housing needs of the community. Traditionally, government housing policy, implemented mainly through planning controls and the taxation structure, has ignored an understanding of the marketplace in which the policies have had to operate. In general terms, the overreaching need is for housing policy formulation to be based on the real estate market structure (Maclellnan & Tu 1996, p.389). This point has been taken up in the Review of Housing Supply (Barker 2004), commissioned in the United Kingdom to review the lack of responsiveness of housing policy to the needs of the housing market. The previous recommendations specifically refer to the need to incorporate market signals into
the development process as a guide for planning intervention. Inherent in this must be an understanding of the submarket system of the local urban market structure. This is a view supported by Bates (2006, p.5) who observes:

When considering interventions for urban revitalization, planners often target neighborhoods defined by historical convention, codifying these into administrative boundaries. However, when predefined neighborhoods do not align with behavior of consumers in the housing market, it will be difficult to predict accurately how the area is likely to respond to policies.

The importance of the relationship between the property market and planning policy is emphasised by White and Allmendinger, (2003, pp.969–70) as Government interventions can have a ‘substantial’ impact on the availability of land for housing and can lead to higher house prices and housing densities.

In the broader social geography literature there is much discussion as to the changing nature and structure of the urban environment (Forster 2004; Randolph 2004). It has also been suggested through this body of literature that new approaches to metropolitan planning are needed if desirable social outcomes from ever more complex urban living environments are to be achieved. This is summed up by Randolph (2004, p.491):

Moreover, the longer term social outcomes of these developments, or the way housing markets actually work, are poorly understood or acknowledged in planning circles.

In general policy-making decisions, Pryce and Evans (2007, p.9) points out that policymakers often have a spatial perspective in their decision-making process:

… the political and administrative process in the United Kingdom and most other countries is fundamentally territorial. Members of Parliament, local authority councillors, and Members of the European Parliament are all elected at particular
spatial scales. As a result, their obligation is to a particular constituency or ward and the allocation of resources is ultimately and inevitably decided along spatial lines.

Submarket structures indicate to policy-makers spatial patterns of desirability and, if based on the underlying real estate structure, the market determinants of that spatial desirability or lack thereof. An understanding of these spatial patterns allows policy-makers to allocate and administer funding knowing the spatial structure of the housing market determinants of their constituents.

The need for a market understanding has been further extended by Can and Megbolugbe, (1997, p.203) when creating housing price indices. They point out these indexes:

... constitute a critical input for the measurement of housing demand; comparative analysis of price trends locally, regionally and nationally; the evaluation of real estate investment decisions; assessment of new mortgage products as well as a risk/default assessment of existing ones; and formulation and design of housing/mortgage policies, programs, and products.

They go on further to say these indices are dependent on the ‘operation of local housing markets’ (p.220), and therefore depend spatially on the underlying urban market structure. Although it may be logical to suppose that any housing policy should include an understanding of the workings of the housing market, it is another challenge entirely to do it.

The geospatial submarkets may fulfil this role by not only delineating the geospatial submarket breaks but also describing their nature in terms of the interaction of the underlying real estate geography and the marketplace.
The second broad area of application lies in the need to understand the market structure in conducting market analysis, particularly in the field of property taxation policy.

### 2.3.2 Property taxation

Mass appraisal is the art and science of assessing the market value of real property as a basis on which a property tax can be levied. It is the derivation of market value for an entire jurisdiction, at a specific time, often containing many hundreds of thousands of properties (Eckert 1990, p.35). The hallmarks of a good fiscal cadastre (the end product of a mass appraisal process) are its comparability of values across the entire jurisdiction and the ability of the custodian to update it whenever property taxes are levied. An important feature in the mass appraisal process is the ability to identify the various submarket groups that may be present in the jurisdiction. This enables more accurate and reliable statistical models to be developed for each particular submarket rather than one model for the whole jurisdiction (Adair et al. 1996b). A mass appraisal system requires account to be taken of location (Figueroa 1999, p.1; Fletcher et al. 2000, p.473; Gallimore et al. 1996, p.8; McCluskey et al. 2002, p.1). There are a number of ways of achieving this but one fundamental approach is to calibrate separate predictive models for separate geographic areas (Goodman & Thibodeau 2007, p.209). It is envisaged that the geospatial submarkets will fulfil this role. The question is to determine the number of geospatial submarkets that would be appropriate for this exercise. There are practical considerations that involve the geographic size necessary to generate enough sales with which to base a model calibration, as well as having a satisfactory homogeneity of price within any given geospatial submarket. As part of this methodology, the Valuer-General of South Australia, who is responsible for the mass appraisal function across the study area, has indicated a strong interest in examining the results of this methodology as to date the important task of spatial submarket delineation is carried out on an ad hoc basis. Given the importance of the real
estate market to the overall economy a clear understanding of the nature of submarkets is a necessary first step to analysis of the market as a whole.

### 2.4 Definition of a submarket

#### 2.4.1 General comment

The imperfect nature of the real estate market as a whole appears to be the driving force for segmentation into smaller submarkets in which a less imperfect set of market conditions exist. The literature contains many definitions of residential real estate submarkets, as discussed below.

There is consensus within the literature supporting the theoretical existence of submarkets (Adair et al. 1996b, p.69; Goodman & Thibodeau 2003, p.182). However there is very little agreement as to what the model of a submarket group looks like (McCluskey et al. 2002, p.2; Watkins 2001, p.2235). This lack of coherent and consistent submarket definition may be responsible for the submarket not being used as a fundamental building block in housing market research, as it is encouraged to be (Barker 2004; Bates 2006). Undoubtedly, the many uses to which the concept of a submarket can be applied may influence the way it is defined and derived. Watkins (2001, p.2235) made five observations as to why submarkets have not been widely used in housing market analysis:

First, there is no single, coherent definition of a housing submarket … Second, even when researchers agree on a definition based on structural or locational features, there is little consensus as to how submarkets should be identified in practice … Third, the urban area under examination varies from study to study … Fourth, the time period from which market data are drawn varies across studies, and, as such, so do market conditions. Fifth, the means of statistical testing for submarket existence also differs between studies.
Watkins goes on to show that fundamental differences in the various studies range from considering submarkets to be broadly segmented on structural and or spatial identifiers down to more detailed, and perhaps area specific identifiers. Different study areas represent different markets and thus different consumer behaviour patterns, making a unified definition of a submarket at the specific identifier level very difficult.

2.4.2 Substitutability

Galster (1996) and Megbolugbe et al. (1996) recognize Grigsby as providing the seminal foundation upon which a sound concept of a submarket can be based. Grigsby’s theory of substitutability is the basic concept behind submarket definition.

Substitutability, as discussed by Grigsby (1963), describes the market’s willingness, when purchasing housing, to substitute one house, and its bundle of attributes, with another in order to optimise housing needs within a given price constraint. For example, while a four-bedroom brick house next to a primary school may be the preferred housing choice, a three-bedroom rendered dwelling may have to be substituted for proximity to the primary school to be achieved. Conversely, if the number of bedrooms is more important, proximity to the primary school may have to be compromised. The various combinations of attributes involved in this process can be numerous and complex making an obvious solution extremely unlikely.

Where this ability to substitute becomes so weak that it has no recognizable affect on the price paid then the dwellings may be considered to belong to different submarkets. Therefore, the concept of a submarket is based on substitutability for a given price.

However, Grigsby (1963, ch.3) went on to describe a more detailed model based on this principal of substitution. Submarkets are interrelated through the concept of a ‘chain of
substitution’. For example, although home buyers in market ‘A’ may not directly affect market ‘C’ they may affect market ‘B’ which in turn may affect market ‘C’ thus establishing this ‘chain’ affect of substitution. The further the chain goes the weaker the price change generated becomes until it is said to be ‘defused’. Each movement creates the potential for another movement and may allow the level of housing satisfaction to rise or fall. This filtering process is well recognized in the literature (Galster 1996, p.1800) and is seen as a fundamental mechanism of the housing market.

Grigsby (1963, p.34) also clearly acknowledges that the substitutability factor may be a continuum and the boundaries between the various submarkets may be ‘fuzzy’:

… in the real world there are no clean cut off points between two submarkets, the chain of substitution being a continuum with sharp breaks or gaps being the exception rather than the rule.

He goes on to note that because the strength of the substitutability is based on a composite strength of the attributes it is therefore possible to have two totally different houses as substitutes for one another. From a property valuation perspective this may have an important implication for the way in which comparable properties are chosen, in the sense that they do not necessarily have to appear similar in terms of property characteristics and location alone but may be chosen in terms of their composite strength as an indication of their substitutability.

Two important points come out of the concept of substitutability when deriving a methodology for the spatial delineation of submarkets. First, the closeness of attributes described by Grigsby refers to closeness in data space and the challenge for this thesis is to represent this closeness or otherwise in geographic space. Second, the attributes that are bundled together that describe a property to the consumer include both structural and
spatial property attributes. This raises the question as to whether submarkets should be
defined in spatial or aspatial terms or both.

2.4.3 Spatial or aspatial submarket identifiers
A good summary of reported studies defining submarkets in spatial and/or aspatial terms is
provided by Watkins (2001, p.2235) who concludes:

… the dimensions of the housing submarkets are determined by both spatial and
structural factors simultaneously.

The literature also suggests that structural factors may be nested within the spatial structure
(Adair et al. 1996b, p.78; Maclennan & Tu 1996, p.395; Tu, 1997 p.340). In particular,
dwelling type nested within broad a priori spatial submarkets suggests that submarkets
based on the theory of substitutability should not necessarily have to be geographically
contiguous. Indeed, as Rothenberg et al. (1991, p.63) point out, the theory of
substitutability extends across all characteristics making it theoretically feasible to have
closely substitutable properties that may not be contiguous. Rather than negate, this
enhances the importance of geography as Bates (2006, p.7) concludes:

Because the dominance of locational factors is defining housing quality, it makes
sense to conceive of submarkets as spatial realities, not abstracted market spaces.

This complexity highlights the difference between housing as a commodity and other
traded commodities in that market equilibrium must exist not only between price and
quantity but also spatially. Thrall (2002, p.38) discusses spatial equilibrium in terms of
the balance of forces attracting consumers closer to nodal regions offering goods and
services, against those pushing them away, namely differing land prices. It is important to
recognize that properties that are close substitutes need not necessarily be geographically
close. However, if they are close substitutes, but not necessarily geographically contiguous, the market may perceive them as belonging to a different submarket.

The concept implicit in the discussion of spatial and structural attributes is that structural attributes are not location dependent. This may not be the case as argued by Orford (1999, p.50):

For example, although the age of the property, the size of the building parcels and the patterns of tenure are all structural attributes, they will tend to vary systematically across urban space, reflecting historic growth patterns.

Supporting this, Orford (1999, p.50) continues, by pointing out that studies have revealed spatial autocorrelation exhibited by structural characteristics.

It is intuitively sensible to think that, in general terms, both structural and spatial components should be part of a submarket definition and the degree to which this might be the case in a particular study area at a particular time would depend on the market. It is the structural and spatial attributes together that describe the real estate geography that constitute the bundle of attributes traded in the marketplace. As Evans (1995, pp.6–7) explains:

The buyer is purchasing a property which is a bundle of characteristics. So, in the case of a house, the purchaser buys a location relative, say, to shops and workplaces, ‘fertility’ in the sense of the quality of the environment, and also a house where attributes of the house – central heating, number of bathrooms, size and number of rooms – cannot be detached and sold separately.

The question now becomes one of adopting a suitable submarket identifier that satisfies the needs of the submarket delineation.
2.4.4 Specific submarket identifiers

At the more detailed level there are many examples of market specific submarket identifiers reported in the literature. Race and religious identifiers are important in some areas and to segment the market using these criteria would be a logical step. For example, Palm (1978, p.210) segmented the market on the key components of:

... real estate board jurisdictions, the racial-ethnic compositions of neighborhoods, and the average house price of neighborhoods.

The study concluded that the real estate board jurisdiction was statistically better than the others in terms of accounting for variance in price change and generally submarkets should be used in the analysis of housing. Similarly, Adair et al. (2000) defined submarkets using the identifiers of religion and locality in Belfast while Goodman and Thibodeau (1998) used the quality of public education as a segmentation criteria. One of the most important property attributes is ‘price’ and Costello (2004, ch.7) examines market segregation along the ‘price-location structure’ of a spatially disaggregated market based on a priori spatial boundaries, in his case post codes. He found that the economic submarkets defined in this way have a spatial component although again they do not have to be contiguous.

Other definitions of submarkets include looking at migration of people as indicating potential areas of housing choice. This is an important influence in the make up of the urban social space and intra-urban mobility should be studied as part of the housing submarket structure (Stimson 1982, p.143). Although Grigsby (1963, p.76) talks about the movement of families as having important influences on the supply and demand of housing, more recently Jones et al. (2004, p.273) returned to household migration as an important ‘first principles’ indicator of the existence of submarkets. The basis for using
migration patterns as an indicator of submarkets is challenged, in certain circumstances, by Pryce (2004b, p.2–9) when he argues:

… households that move (other than for job or family reasons) may actually be more likely to be switching to a different (hopefully preferable) submarket. After all, why else move? So unless we know in great detail the reasons for each and every observed move, population movements in themselves are unlikely to tell us a great deal about the location of submarkets.

Others take a more generalized approach. For example, Thrall (2002, p.20) describes submarkets in terms of urban form and urban morphology and defines the boundaries:

… with the objective of minimizing the variation in the descriptive characteristics and phenomena that characterize the submarket.

Although this is a somewhat arbitrary way of defining boundaries it acknowledges the need for some type of homogeneity within a submarket and an understanding of the interrelationship between submarkets.

Homogeneity can also be expressed in terms of price and as submarkets are a market entity this appears an appropriate submarket identifier. Rothenberg et al. (1991, p.204) defines a submarket as:

… a collection of dwellings in a metropolitan area whose myriad of attributes, though different, are evaluated as a whole by demanders and suppliers as closely equivalent. The metric collapsing of these attributes of the housing package into a single dimension, quality, is the basis for arraying the housing stock in an urban area in a spectrum of submarkets, from highest- to lowest-quality levels.

The identifier that quantifies the ‘metric collapsing’ is the hedonic index. It is derived from regressing market price on ‘constituent components of the package’ (Rothenberg et al.
1991, p.360). This provides the ability to not only identify but to quantify and compare submarkets so identified. This may be an important and useful aspect of submarket identification making this submarket identifier an interesting one to build on. Goodman and Thibodeau (2007, p.213) discusses the various identifiers used in previous studies and concludes by also proposing a submarket delineation process based on price.

### 2.4.5 Diagrammatic submarket concepts

The literature contains many concepts of a submarket spatial definition, but two general concepts are of particular interest.

The first presents a hierarchical submarket group concept as suggested by Maclennan and Tu (1996, p.395) and also Tu (1997, p.340) and shown in figure 2.1.

#### Figure 2.1 Local housing submarket structure

![Diagram](image)

**NOTE:** This figure is included on page 39 of the print copy of the thesis held in the University of Adelaide Library.

Note: N1 and N2 are the neighbourhood sub-markets. SS1…SS22 are the sectoral sub-markets. D111…D222 are individual dwellings.


It portrays submarkets constructed on a hierarchical basis. It addresses the question as to the submarket structure. It shows the neighbourhood submarkets as spatial and have nested within them the sectoral dwelling ‘product group’ submarkets defined by their structural dwelling attributes (Tu 1997, p.339).
When discussing these ‘product groups’ Watkins (2001, p.2237) writes:

Each product group will be composed of relatively homogenous dwellings which represent reasonably close substitutes to the demanders of housing. It is the way in which segmented demand is matched to the differentiated housing stock which is likely to give rise to housing submarkets, and to cause differential prices to be paid for given attributes in different market segments.

This is important to residential submarket theory as it draws the link between homogeneity and substitutability. Watkins explains (2001, p.2237) that homogeneity means that dissimilar dwellings are linked ‘by the degree of indifference of consumers’ which is the essence of substitutability as it is the marketplace that determines the ‘indifference of consumers’ and therefore the submarket delineation. The inclusion of the marketplace in submarket definitions is of fundamental importance as submarkets are an economic entity and to define them using non economic criteria ‘contains an essential flaw’ (Pryce 2004b, p.2–10).

Another conceptual submarket model is illustrated in figure 2.2. It shows that within each submarket there is a different relationship between the bundle of property attributes. These different relationships are brought about by different submarkets and may be modelled separately, which is a useful concept as it allows the market as a whole to be modelled in its component parts (submarkets) rather than relying on one model to encompass the whole market.
Figure 2.2  Separate models for each submarket

The figure 2.2 concept model has particular appeal in the field of mass appraisal and is basically the model adopted in South Australia in the late 1980s onwards with the submarket boundaries being delineated from the professional valuer's sense of where submarket boundaries should be (Lockwood & Reynolds 1992). However, rather than relying on a subjective boundary delineation of the valuer, there is a need to be able to confidently and consistently delineate more optimal spatial submarket boundaries in which simple, understandable and transparent models can be formulated and used to predict property values for taxation purposes (Deddis et al. 2002, p.5).

From looking at these studies it is clear that there is no agreement as to which identifiers are appropriate for submarket definition, and therefore no conclusive definition of a submarket. This is not surprising, as every study area has a different residential living structure composition and therefore different market behaviour patterns. As Bourassa et al. 

Source: Abstract interpretation of functions in an heterogeneous property market described in a mathematical conceptual space, (Lewis et al. 2001, p.108)
(2003, p.12) points out an appropriate definition of a submarket depends upon the use to which it is put. This requires a clear definition as to the user requirements and the application for which the submarket delineation is to be used.

### 2.4.6 Summary of spatial submarket definitions

In general terms Tu (2003, pp.40–1) summarises submarket structure approaches into three groups. Firstly, the ‘topographically-based housing submarket’ based on the clustering of housing product groups and testing for significant differences in housing attribute price. Secondly, ‘a quality-based’ housing submarket suggested by Rothenberg et al. (1991, p.204) and thirdly, the hybrid model suggested by Tu (2003, p.41) which suggests a topographically-based submarket inside of which are different sectorally based submarkets based on dwelling type. Regardless of the definition of submarkets there is a recognition of the importance of the spatial component of submarkets. The important question of delineating the spatial boundaries remains. From the wide variety of definitions summarized by Watkins, (2001, p.2238) there are broadly three ways of currently determining the spatial dimension and structure of submarket groups with a fourth, highly theoretical approach, an interesting one to contemplate, but not included in Watkins' summary.

The first group is based on real estate practitioners who define the spatial boundaries based upon their experience in an *a priori* way, either in terms of drawing boundaries on a map or defining the submarket group in terms of property characteristics (Adair et al. 1996b; Jones et al. 2002; Watkins 1999). Other forms of *a priori* segregation may use the spatial boundaries of existing administrative units such as the postcode, suburb, local government area or census area unit. Submarket status was then established using a commonly employed test developed by Schnare and Struyk, (1976, p.151). This test is based on the
'law of one price' and is made up of three stages. Firstly, a hedonic price model is calibrated in each of the proposed submarkets and compared for significant difference with other similarly calibrated submarket models using a chow test which uses an $F$-statistic to test for parameter equality. The hedonic model uses various property characteristics as independent variables and the dwelling price as the dependent variable. Rothenberg et al. (1991, p.64) suggests caution is needed when using this approach for two reasons. Firstly, any differences probably only have a small practical significance and secondly, the differences may not be statistically meaningful unless the submarkets being tested are relevant to the marketplace.

Only if the submarkets defined $a$ priori are in fact the relevant ones (in the usual economic sense of substitutability in the marketplace) can any conclusions be drawn about the cross-sectional stability of the hedonic relationship from these studies.

Submarkets are defined using an $a$ priori approach and are not necessarily relevant and tested with ‘the law of one price’ are likely to be questionable.

A search for an optimal and relevant solution is the goal of the second group who take the view that the data itself should determine the boundaries (Bourassa et al. 1999; Dale-Johnson 1982; Goodman & Thibodeau 1998; Kauko 2000; Maclennan & Tu 1996). Maclennan and Tu, (1996, p.395) argue that it is necessary to make no assumptions about the structure of submarkets but to ‘establish it empirically’ based on knowledge of the underlying structure. An Australian study examined submarkets derived from a variety of sample data from Melbourne and Sydney including socio-economic as well as property characteristic data, and using principal component analysis (PCA) suggest that the clustering of the various principal components of these data display a submarket structure (Bourassa et al. 1999).
In establishing a method to quantify the underlying structure, Maclennan and Tu, (1996) also use PCA to derive the dimensions of the property characteristics, including accessibility, to form the ‘product group’. Tests for a significant difference in hedonic attribute values between these ‘product groups’, with the additional test of persistence over time, establish the presence of a submarket structure. Dale-Johnson (1982) used physical property characteristics, various environmental and socio economic data collected for properties that had sold (representing only a sample of the total population of dwellings in the study area). Using PCA, five factors were extracted from the original 13 variables. These factors represent the underlying structure of the data. An important characteristic of this second group of research studies is the realization that it is necessary to include a wide variety of property descriptors for submarket segregation so that the underlying structure can be identified. This gives the consistency of results required by the many housing policy and market analysis applications seeking an understanding of the market structure. To acquire this understanding is not a trivial exercise as the factors that influence the formation of market segments are complex and diverse (Dale-Johnson 1982, p.331).

While preserving a quantitative approach, the third group of researchers argue that discontinuity between submarkets generated by both of the above approaches may not represent reality, and the use of response surface analysis whereby continuous surfaces depict a change in market status may be a more appropriate approach (Deddis et al. 2002; Figueroa 1999; Gallimore et al. 1996; Lamont et al. 2002; McCluskey et al. 2002). McCluskey et al. (2002, p.1) suggest their research shows that the real world represents a more fuzzy submarket group boundary than the choroplethic submarket delineations of other studies discussed in the first group, in which case a ‘fuzzy’ surface may be better represented through response surface analysis. Grigsby (1963, p.34) described the interdependence of submarkets as a continuum of substitutability from strong to weak, also
implying that the boundaries between them may be ‘fuzzy’. Lamont et al. (2002) takes this approach one step further, arguing that the locational effect of a data point may be determined through variography. This is a technique based on Tobler’s first law of geography (Tobler 1970) that says over a given range of influence, events closer together are more likely to be the same than those further apart. Calculating this range of influence is the art and science of variography. Property market transaction analysis can be carried out within this determined range of spatial influence on property characteristic differences alone as the effect of location can be accounted for by the distance weights shown by the variogram. The link between variography and submarkets is unclear except to suggest this may be another way of determining spatial areas of market interest.

The fourth method of understanding submarket groups may be through the broader understanding of the structure of urban development and the process that is driving the development. Torrens (2000) argues that in order to understand the complexity of urban development new approaches are necessary. He discusses complexity studies and their potential to simulate human behaviour in such systems as the residential location market. Batty (1996, p.304) talks about a ‘new physics of urban development’ and ‘fractal cities’. Sui (1998, p.651) also suggests the need for new models for ‘informational cities’ to be incorporated within the GIS environment giving the ability to:

… espouse new computational models … [to] enable us to think above and beyond the technical issues that have occupied us during the last ten years.

More recently Torrens (2007, p.202) discusses a methodology that simulates residential mobility to form artificial property submarkets.
The application of these theories is still in the development stages and their contribution to the understanding of the urban form remains, as yet, largely unknown but may well provide the new tools necessary to unravel the complexities of the real estate marketplace.

2.5 Issues identified in the literature

The underlying and consistent premise upon which submarket definition is built is that of substitutability (Bourassa et al. 2003, p.14; Grigsby 1963, ch.3). The issues that emerge from the literature involve the recognition that submarkets are best defined simultaneously in terms of both spatial and structural identifiers (Watkins 2001, p.2250) and should be derived from the data itself rather than on the basis of some a priori definition. The data should reflect the underlying residential real estate structure of the study area and not rely on selected property characteristics alone. However, it is insufficient to identify the underlying structure and geography only, as submarket definition requires expression in terms of the marketplace in order to give it an economic entity status (Pryce 2004b, p.2–10).

2.5.1 Opportunity for contribution

There is an opportunity to more fully quantify the underlying residential real estate living structure through building on the work of Bourassa et al. (1999), Bourassa et al. (2003) and Maclennan and Tu, (1996), by including all properties in a study area rather than only a sample. This would give a complete picture of the underlying residential structure in which submarkets reside. Also an opportunity in that each of these properties can be described by a wider range of property descriptors than has currently been undertaken so as to more fully describe the real estate geography being traded. This study has included data from socio economic, environmental, accessibility and structural property characteristics
obtained at the individual property level thus avoiding any assumption of *a priori* boundary definition. This structure can then be related to the marketplace by monitoring changes in the hedonic regression coefficients for each component of the underlying residential structure across geographic space as indicators of submarket change.

### 2.6 Submarket definition adopted for this study

The objective of this study is to delineate the spatial boundaries of residential real estate submarket structures. From the literature a nested structure exhibiting a continuous surface appears to be an appropriate concept model. There appears to be no reason to suppose that the submarket group drivers necessarily abruptly change at an artificial administrative boundary in a discontinuous choroplethic fashion. Indeed, even at natural geographic boundaries such as rivers, or at urban infrastructural barriers such as main roads, change may be rapid but should be continuous. In fact interdependency between submarkets is suggested as an attribute of a submarket (Galster 1996, pp.1803–4). Any definition of a submarket should be based on the concept of substitutability for two reasons.

Firstly, it is a broad definition that approaches the problem, not from a particular set of property attributes that may or may not be relevant to the market at a particular point in time, but rather from the perspective of the behaviour of the players in the marketplace. As Jones et al. (2004, p.273) explains:

> … there must be a clear identification of the differential linkages between different classes of property on the supply side and different groups of consumers on the demand side.

This requires taking into account the relevant property attributes on the supply side and relevant consumer attributes which may influence that behaviour on the demand side.
Together, these data make up the underlying residential real estate geography which is transacted in the marketplace. It is therefore necessary to quantify the dimensions of this structure in order to understand how the marketplace prices them. It is this price that provides the basis for delineating their spatial boundaries.

Secondly, because substitutability is based on market behaviour any resulting submarket boundaries will have to reflect the marketplace. The literature suggests that a market related housing structure is what housing policy and market analysis should be based on (Can and Megbolugbe, 1997, p.270) and more particularly the complexity of that structure (Meen & Meen 2003). Any market analysis is essentially about human behaviour and the way people manoeuvre themselves to maximize their own outcome. Given the unique nature of real property, the ability of a household unit to make a choice about satisfying the need for shelter intuitively revolves around the ability to substitute one property for another within their economic constraints.

The literature has come to recognize property as bundles of attributes that cannot be unravelled (Baker 2002; Evans 1995; Lancaster 1966). Each property has a unique bundle of attributes consisting of both locational and structural components (Evans 1995). These are bundles of attributes that taken together describe the underlying residential real estate structural dimensions of the study area rather than considering individual attributes in an unbundled state. The attributes attached to the property itself and contributing to its value are well summarized by Wyatt (1996) as shown diagrammatically in figure 2.3.
Some of the attributes are common to a wide range of properties such as the general economic climate, planning constraints and the general law. Other factors apply to a smaller group of properties, perhaps at an urban level such as accessibility, neighbourhood and environmental factors, while other structural attributes are about the individual property itself, for example the age, structure and size of the physical improvements.

The locational attributes often serve as a ‘proxy for numerous unobserved variables’ (Pavlov 2000, p.249) and are many and varied and as Orford (1999, p.63) points out are potentially limitless in number. Wyatt (1996) suggests in figure 2.3 ‘location’ may be described under the broad headings of planning constraints, environment, neighbourhood and access. The effect of planning constraints on property value may be reflected, in part, by examining the added value of the structure to the price of the property. Planning laws may change from time to time allowing higher density living or they may further restrict the use of the land. Planning constraints are a valuable tool available to Government to control the development of an area. For example, larger inner suburban land parcels restricted to single dwelling usage may, as part of an urban regeneration programme, have
that restriction lifted allowing for higher density living. In this case the added value of the existing dwelling may fall, becoming economically obsolete (Whipple 1995, p.10). Access portrayed as a subset of the ‘location’ factors in figure 2.3 is the proximity of a particular property to various services and can also be considered as property attributes that influence price (Kestens et al. 2004, p.539). The environmental subset of property attributes (figure 2.3) may be complex and difficult to measure as noted by Kestens et al. (2004, p.541), but nevertheless be an important part of the price structure of property. The neighbourhood subset of property attributes may include socio-economic indicators of the surrounding area that attach to the property itself in terms of its marketability. For example, a purchaser may be attracted to a particular property because the average income or age of the surrounding area is of a particular make up or the ethnicity of the surrounding area is of a particular composition. These average surrounding socio-economic indicators can be considered attributes of a particular property in terms of its marketability (Jackson et al. 2007; Reed 2001). The market value may be visualised as the end product of the bargaining process between buyer and seller and represents the total value of all the individual structural and locational attributes taken together.

This study adopts the general submarket concept as being of a nested hierarchical structure (Tu 1997, p.340) summarized by Tu (2003, p.40) as the hybrid model. Basically this involves a spatial submarket in which sectoral submarkets may exist. It is the spatial component of this model for which this study seeks to delineate the spatial boundaries and is termed the geospatial residential real estate submarket for the purposes of this study and is defined as follows:

A geographic area in which the prices of the individual components of the underlying residential living structure have a predefined degree of homogeneity in their price pattern.
The definition of the spatial component is considered a complex continuous surface of the market’s price of the underlying residential living structure across geographic space. Specifically, the definition interprets the hedonic parameter estimates as indicators of the price pattern of the underlying structure. As the regression model moves across geographic space the change in the price pattern indicates change in the geospatial residential real estate submarket. The spatial delineation of the geospatial submarket requires a break to be nominated along this complex continuous surface. This is achieved by the user defining the degree of homogeneity of the price patterns required within the resulting geospatial submarket. The definition by the user of the submarket is in keeping with Bourassa et al. (2003, p.12) when stating:

We maintain that the appropriate definition of submarkets depends on the use to which they will be put.

For example, the hierarchical concept that there may be residential submarkets nested within the geospatial submarket representing the various types of different residential dwellings based on various dwelling characteristics becomes important when building predictive valuation models within the geospatial submarkets. This particular use would demand a particular size of geospatial submarket and would therefore require specific professional user input to define this.

2.6.1 The use of the term ‘submarket’

Submarket delineation has been created on an ad hoc basis using a variety of identification criteria and therefore can not be sure that the resulting sub-markets are identified in an optimal manner (Tu 2003, p.51). The lack of consistent definition is also acknowledged by Watkins (2001, p.2235) who suggests various reasons for it being so as previously discussed.
However, once identification has been made there tends to be a more consistent approach to confirming that the identified submarket is indeed a submarket. Importantly, in acknowledging the ad hoc identification process Watkins (2001) goes on to discuss the test for submarket existence, as being conceptually that described by Tu (1997, p.339) namely:

A housing sub-market is defined to exist if there are persistent, significant disparities in the hedonic housing prices of the non-spatial, key dwelling components between this dwelling product group and others.

As Watkins (2001, p.2241) explains, this test was introduced by Schnare and Struyk (1976) and used in many other submarket studies that have been well summarised by Watkins (2001, p.2242). In this summary (discussed in more detail earlier in this chapter) the definition of the submarket (spatial or structural or both) and the ad hoc submarket identifiers are given together with the statistical test used to establish their submarket status. For example, postcodes, racial composition of districts, structural property characteristics as well as nested *a priori* spatial groups within which are nested property types (for example, detached, and terraced dwelling types) are used. However defined, a statistical test for significant difference in the hedonic relationship generally described by Tu (1997) is applied to give submarket status.

Again this is supported by Goodman and Thibodeau (2003, p.184) who also observe that identifying submarket boundaries has typically been carried out on an ad hoc basis with researchers often imposing submarket boundaries (school districts, racial divisions, and housing types) rather than deriving them. The status of submarket is deemed appropriate if:

Hedonic regressions are estimated separately for the individual submarkets and F tests determine whether the resulting reduction in sum of squared residuals is significant. If
the reduction is significant, then the posited submarkets are assumed to be appropriate, conditional on the particular specification of the submarkets.

This is an important aspect of the housing submarket literature and poses the question as to whether or not a spatial entity, not subjected to this form of testing, can be accorded the status of ‘submarket’. The submarket definition adopted in this study does not formally satisfy this test.

However, the submarket identifier is price; and more particularly the homogeneity of the individual component prices of the underlying residential structure even though they are not tested for significant difference. Pryce and Evans, (2007, p.10) comment that these tests are generally used to test differences between *a priori* (arbitrary) property groupings and the tests:

… are not sufficiently precise to allow one to apply a fixed threshold for what constitutes a break. Depending on how one sets up the regression equation, one can make all break points exceed the standard statistical threshold of 5% significance, or none at all.

As will be seen in chapter 8, the resulting numbers of geospatial submarkets may be large, and although to subject each one to this test may be possible, it would be impractical to carry out and an appropriate level of statistical significance would have to be determined, perhaps depending on the application for which the geospatial submarkets were derived. Further research involving the level of similarity (homogeneity) within the geospatial submarket cluster of parameter estimates may also satisfy this requirement and will be investigated as part of further research.
Not all submarket identification processes involve the satisfaction of this particular test. Three criteria are suggested by Tu (2003, p.51) in identifying urban housing sub-markets, being:

1. Simplicity – a solution with few sub-markets is superior to a solution with many sub-markets.

2. Similarity – the housing attributes associated with a dwelling should be as similar as possible within a sub-market. In other words, the dwellings in a sub-market should have a high degree of homogeneity (or substitutability), while the dwellings in different sub-markets should be subjected to a higher degree of heterogeneity.

3. Compactness – dwellings located in contiguous areas should be more likely to be grouped into one sub-market than dwellings in more distant areas.

The adopted definition of a geospatial submarket complies with these criteria and may be summarised, as regards Tu’s headings above, as follows:

Simplicity – the number of submarkets in a geospatial submarket solution is tailor made by the user for a particular application. This would ensure that there are not a predefined number of submarkets covering all situations.

Similarity – the housing attributes associated with the dwelling in this study are the surrogate property attributes derived so as to represent the underlying residential living structure. By definition and design of the geospatial submarket these represent a high degree of homogeneity (or substitutability) within the geospatial submarket while these attributes represent a high degree of heterogeneity between the submarkets.

Compactness – dwellings located in contiguous areas are more likely to be grouped into the one submarket than those in more distant areas due to the nature of the geographically weighted hedonic regression model which uses a similar group of sales data in similar areas with different weightings. More distant areas use differently weighted sales data and more likely to group differently. However, in geospatial
submarket groups with relatively low internal degrees of homogeneity there may be discontinuous spatial groupings. If this occurs and is designated unacceptable by the user then the non contiguous spatial submarkets can be defined and treated as separate spatial submarkets.

While this study recognises that some researchers may find this definition technically unacceptable it is not the intention to redefine the submarket term. It is however considered the most suitable and clearest term to describe the aims and objectives of this study. For the sake of clarity the term *geospatial submarket* has been adopted.

### 2.7 Conclusion

The literature reveals consensus as to the existence of submarkets but an ad hoc approach to their identification. However, there are emerging themes that present themselves as a basis for a more consistent identification process and which have been adopted in this study. Namely that submarket:

- Structure can have both spatial and structural dimensions that exist simultaneously
- Structure should be derived empirically and not based on *a priori* structures
- Structure should be based the underlying residential structure of the study area and the marketplace
- Spatial structure is a continuous complex surface based on the concept of substitutability constrained by price.

The submarket literature offers several opportunities to contribute to the understanding of the spatial delineation of residential submarket boundaries. These include:

- Identifying the underlying residential living structure (RLS) of the study area as a basis for understanding its submarket structure
- Deriving the RLS from a broad range of variables (both structural and spatial) that together describe as much of the real estate geography as possible
Relating the RLS to the marketplace to reveal the underlying submarket structure.

This thesis accepts these opportunities to contribute by defining a geospatial submarket boundary as a change in the market’s value of the underlying RLS across geographical space.

Although each study area would understandably reveal a different residential submarket structure the definition and derivation process would be generic and therefore consistent.
CHAPTER 3 THE NATURE OF THE STUDY AREA

3.1 Introduction

The aim of this chapter is to provide an understanding of Adelaide, South Australia (the study area) in terms of:

♦ Its physical extent and location, including the broad residential real estate structure and the market operating within that structure.

♦ The creation of the data sets used to describe the study area together with the associated limitations on their use and the selection criteria employed.

Adelaide is the capital city of the state of South Australia. As one of eight Australian states and territories, South Australia’s comparative population and geographic area is shown in table 3.1.

Table 3.1 Populations of Australian states and their respective capital cities, 2001.

<table>
<thead>
<tr>
<th>State of Australia</th>
<th>Capital City</th>
<th>Population 2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>New South Wales</td>
<td>Sydney</td>
<td>4,908,749</td>
</tr>
<tr>
<td>Victoria</td>
<td>Melbourne</td>
<td>4,907,081</td>
</tr>
<tr>
<td>Queensland</td>
<td>Brisbane</td>
<td>4,197,400</td>
</tr>
<tr>
<td>South Australia</td>
<td>Adelaide</td>
<td>1,261,000</td>
</tr>
<tr>
<td>Western Australia</td>
<td>Perth</td>
<td>1,701,500</td>
</tr>
<tr>
<td>Tasmania</td>
<td>Hobart</td>
<td>1,066,000</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>Darwin</td>
<td>1,037,000</td>
</tr>
</tbody>
</table>

NOTE: This table is included on page 57 of the print copy of the thesis held in the University of Adelaide Library.

As can be seen from the table, South Australia has the second lowest population of the various states and territories although its geographic size is relatively large.

Figure 3.1  Location context of study area
This chapter details the geographic extent and controlling influences on the physical boundaries of the study area and the data available to describe the residential real estate geography and residential real estate markets contained therein. This study seeks to delineate the spatial submarket boundaries as at a particular point in time. The important question as to the temporal dimensions of the submarket structure is outside the scope of this study and is left to future research.

The date of the study is chosen to be August 2001. This was the date of the last quinquennial population census for which data were available and all data sets used in this study have been taken to be as close as possible to that date. As discussed in the previous chapter, this study proposes not to rely on existing spatial units for geospatial submarket delineation but rather to rely on the analysis of land-related attribute data collected at the individual property level to reveal their whereabouts. As in many jurisdictions there are a number of land-related data custodians important to undertaking research of this nature. It is not only the quality of the individual data sets but also the land information management systems that allow such data to be integrated that is of equal importance in establishing an appropriate data set for the integrated nature of social science research. The wide variety of data collected, the reasons for their choice, and the caveats attached to their use are discussed in this chapter.
3.2 Geographical dimensions

Although the study area can be broadly defined as the Adelaide Statistical Division (ASD), it is more particularly delineated as being those urban centres within the Adelaide Statistical Division. An urban centre is defined as follows (Australian Bureau of Statistics 2002b, p.30):

Each urban centre with a population of 20,000 or more is to consist of a cluster of contiguous urban CDs and other urban areas.

A census collection district (CD) in urban areas contains, on average, about 220 dwellings (Australian Bureau of Statistics 2002b, p.7). An urban CD is defined by the Australian Bureau of Statistics (2002b, pp.30-31).

As at the date of this study (August 2001), there were 2078 urban CDs within the 12 Adelaide urban centres containing a population of 1.05 million people out of a total population of 1.5 million for the state of South Australia (Australian Bureau of Statistics 2002b).

The study area runs approximately 100 km north to south and approximately 30 km east to west. There is a non-contiguous component of the study area represented by a group of suburbs in the Adelaide Hills. They are approximately 20 km from the CBD and connected by part of the main national highway system making travelling time of the order of half an hour at peak times of the day. The study area contains 343 suburbs that are contained in 19 local government areas. The geographic extent of the study area and its locational context within Australia is shown in figure 3.1.

An important factor in controlling the geographical extent of the study area is the powers given to the State Government under The Development Act 1993 as amended and the
associated regulations. This legislation provides for the orderly approval of new land divisions by both state and local government. It refers to an urban growth boundary in the planning strategy (Government of South Australia 2006) designed to ensure the spread of land development is able to be supported by the level of local infrastructure.

3.3 The residential structure

When Europeans first settled South Australia in the mid-19th century as a free colony, Adelaide was laid out by Colonel William Light and land sold to free settlers in South Australia under what was termed the ‘Wakefield System’ of land allocation. A detailed account of the workings of this system and the progress of settlement patterns of the state in general is given by Williams (1974). Unlike other Australian states, a feature of this system was that it controlled occupation of the land by requiring a survey before it was offered for sale to selective immigrants to ensure the efficient working of the land (Sedunary 1987, p15). Initially settlers came from the United Kingdom, but later post-war immigrants from Europe and, more recently, Asia and Africa have influenced the population makeup of Adelaide.

By looking at the spatial distribution of the dwelling stock by age (figure 3.2) it can be seen that residential development started around the CBD and has spread out in a general concentric pattern to the geographic extremities of the study area. Areas of earlier housing development also occurred along the coastline as well as through the Adelaide Hills where early residential areas developed as summer retreats from the often oppressive summer heat experienced on the plains.
Figure 3.2  Dwelling age distribution

Spatial distribution of Dwelling Age and Regional Activity Centres

Legend
- Study Area
- Regional Activity Centres

Dwelling Age
- Years old
  - 1 - 5
  - 6 - 15
  - 16 - 25
  - 26 - 35
  - 36 - 45
  - 46 - 55
  - 56 - 65
  - 66 - 75
  - 76 - 85
  - Greater than 85
The residential mosaic of the study area has changed over the years. Up to the 1980s urban growth roughly followed a concentric ring pattern. In the dwelling age spatial distribution (figure 3.2) the older dwellings closer to the CBD and along the early established route between Port Adelaide and the CBD are distinguishable in a north westerly direction from the CBD. Beginning in the 1990s, inner suburbs began to undergo urban consolidation programmes with the backing of government policy.

The residential redevelopment of the CBD also commenced during this period which Forster (2004, p.112) suggests was partly in response to the fall off in office accommodation demand. This activity was well underway at the date adopted for this study.

In broad terms, a key socio economic indicator of household income (figure 3.3) shows the eastern suburbs and older established areas in the Adelaide Hills and to some extent along portions of the coast as being economically better off than the economically poorer areas in the north and south. The poorer areas include suburbs to the north described by Baum et al. (2005, ch.3, p.22) as ‘old-economy extremely disadvantaged localities’ along with suburbs in the south. These locations are predominately driven economically by long-established car manufacturing industries that are increasingly threatened by current globalisation trends.

However, relatively new opportunities for Adelaide in the form of the naval shipbuilding and allied electronic defence industries may provide a source of replacement employment opportunities in the northern areas where they are located.
Chapter 3 – The Nature of the Study Area

Figure 3.3 Income distribution

Spatial distribution of Household Income

Legend

Study_Area
Census 2001 data
Household income
Low

CBD
AIRPORT
Adelaide Hills suburbs

Legend

Study_Area
Census 2001 data
Household income
Low
High

0 3.75 7.5 15 Kilometers
The population’s changing spatial patterns are an important aspect of the residential living structure and associated real estate markets. Historically, population growth appeared highest on the urban fringe of most Australian cities while the growth pattern shown by the 2001 census appears to have changed as reported by Hugo (2003, p.49):

Certainly areas of population growth are found on the expanding urban fringe but there is also growth in several inner suburbs and in a scattering of the older inner and middle suburbs especially along main transport routes and coastal areas.

A total of 443,619 residential properties are used in this study. Dwellings are defined as a self-contained residence and at the date of the study (August 2001), they were made up of four dwelling types (Office of the Valuer-General April 2003) represented as follows:

1. Single dwellings. Defined as detached dwellings they constitute the largest group in this study at 76% of the total.

2. Home units. Multiple dwellings in one building or separate buildings in the one parcel of land defined as a dwelling capable of separate ownership but with shared facilities (access ways and/or grounds). Separate legal ownership is created under the *Strata Titles Act 1988* and *The Community Titles Act 1996*. This group constitutes 17% of the total in this study.

3. Multiple dwelling. Multiple dwellings in one building each with separate title but not under the home unit category described above. They are usually the older semi-detached or row (terraced house) and constitute 6.5% of the total in this study.

4. Rural living properties. These are residential dwellings within a rural setting, but not in the business of primary production. These constitute 0.2% of the total.

Each of these four groups represents dwellings that are capable of separate sale on the open market as well as separate occupancy, and as such represent the ‘owner-occupied housing sub-market structure’ referred to by Tu (1997, p.337) as being more specifically the group of residential properties within which further submarkets exist and the submarket structure adopted in this study. Dwellings not capable of separate sale are considered in a separate
submarket and do not form part of this study. For this reason the dwelling type designated ‘flats’ were omitted from this study. Dwellings in groupings of two or more dwellings, but legally capable of separate ownership, are included and described as either home units (including townhouses), or multiple dwellings.

The State Government recognises that the metropolitan area of Adelaide has experienced a long period of low-density urban sprawl both to the north and the south. In 2002 formal controls were imposed to limit this sprawl and the controls are endorsed in the current planning strategy (Government of South Australia 2006). As part of this strategy a number of regional activity centres are nominated around which mixed development, including community facilities, offices and residential development, can take place (Government of South Australia 2006). The location of the regional activity centres can be seen in figure 3.2. The planning strategy signals the government’s intention not only to focus residential development within the existing metropolitan boundaries but also to concentrate development around a number of regional centres through mixed land uses and higher density living.

3.4 The residential market

Real estate markets can be discussed both at the global and at the local level. Although this study is concerned with local variation (within the study area), it is useful also to look at the real estate market from a metropolitan perspective so that it can be compared with other Australian capital cities.

In making this comparison, the Adelaide real estate market can be seen to languish behind most other Australian capital cities. As expected, Sydney and Melbourne are not only Australia’s largest capital cities, they also have the highest median prices and rates of
growth in median price as measured by the September quarter median price movement shown in figure 3.4.

At the time of the study the Adelaide market had the lowest median price with the exception of Hobart, capital city of Tasmania, Australia’s smallest state. Historically, during the late 1980s, Perth in Western Australia, and Brisbane in Queensland were lower than Adelaide, but at the date of the study this was starting to change. This change has become more pronounced in current times caused mainly by migration to Queensland and the commodities boom in Western Australia. The parity with interstate markets can be influenced by the relative investment and employment opportunities.

**Figure 3.4 Median quarterly price movements**

![Figure 3.4 Median quarterly price movements]

Source: Charted figure from raw data supplied by Real Estate Institute of Australia 2007.

However, it is the local market variation that is of interest in this study. At the date of this study the South Australian economy was dominated by manufacturing industry, representing 15% of gross state product, compared with 11% for Australia as a whole.
Chapter 3 – The Nature of the Study Area

(Gelber 2005, p.1). The automotive industry occupies two significant manufacturing plants, one in the north and one in the south, employing the largest number (22%) of full-time employed males in 2002 compared with other industries (Spoehr & Barrett 2005, p.147). A significant portion of the housing market in the northern and southern areas is therefore influenced and, to some degree determined by, the car manufacturing industry.

In general terms the pattern of spatial variation in the local market can be seen in figure 3.5. This displays an interpolation of the sale prices used in this study (stage 2) and provides a simple overview of market price variation across the study area.

For example, those suburbs immediately north, south and east of the CBD including the eastern suburbs generally, the coastal properties with increasingly sought after sea views and the Adelaide Hills suburbs have higher selling prices indicating the more desirable Adelaide locations.

These higher value areas are generally in line with the high household income patterns shown earlier in figure 3.3. The western suburbs between the CBD and the coast and the suburbs in the north and south generally have lower selling prices. This may be expected as these areas generally are the more socially disadvantaged parts of the study area. They have the higher rates of unemployment, unskilled and semi-skilled workers, and single parent families (Glover, Hetzel et al. 2006, pp.85, 87, 99). In addition, the suburbs between the CBD and the coast are influenced by Adelaide’s major airport which has become busier with increasing tourism and the ability to handle larger international flights. This has had a negative effect on local property values (Rossini et al. 2002) and caused an increase in local traffic as airport land is increasingly being developed as a commercial retail centre.
Figure 3.5   Interpolated sale price (August 2001)
3.5 Creating the data set

As discussed in the previous chapter, the objective of this study was to delineate residential submarket boundaries without making any *a priori* assumptions as to where the spatial boundaries might be. Therefore the data were collected at the individual property level for each of the 443 619 properties in the study area. The objective of the stage 1 data analysis was to derive the dimensions of the underlying residential living structure (RLS) and this stage required the majority of the data used in this study. Stage 2 of the study used data derived in stage 1 together with the sale transaction price of properties that had sold within a timeframe that represented the market value at the study date of August 2001. All data used in this study were secondary data or derived from the secondary data that were readily available both in terms of time and cost. Housing data has traditionally been divided into those data describing the structure of the dwellings and those describing its location (Orford 2000, p.106). This section describes the data environment, the selection process and, where appropriate, the derivation processes used to formulate the structural and locational data attributes used in this study.

3.5.1 The land information management environment

The management of land-related data in an efficient and effective manner is a challenging task in many jurisdictions around the world due to the necessary coordination of many different data custodians in order to produce integrated data sets. The problem of providing land-related information in a timely and effective manner was first recognised in South Australia in May 1974 with the establishment of a full-time study team to investigate the establishment of a land information system (LIS) to support various government functions (Sedunary 1987, p.112). The first fruits of this initiative were realised through the early 1980s with the implementation of a computerised database referred to as the Land
Ownership and Tenure System (LOTS). By 1984 LOTS included integrated information from the land title registration and land valuation function including an integrated sales history system providing on-line access to ownership, valuation and property transactions data. Part of the early vision of the state’s land information system was the creation of a spatial component to the LIS. This took the form of a digital cadastral database (DCDB) that was a digital representation of current cadastral land parcels of the state and became available to the public by mid-1986 (Sedunary 1987, p.132). Although designed as part of the state’s land information system linking data from LOTS (especially the valuation file) and the DCDB involved a ‘many to many’ relationship between a valuation parcel and a DCDB parcel, making the integration of the data sets a non-trivial exercise. To overcome this problem the property cadastre was developed and became operational in 2000. This was created as an overlay on the DCDB and represented the spatial units deemed by the Valuer-General to form a valuation assessment under the provisions of the Valuation of Land Act 1971 as amended. This represents the same spatial unit against which the Valuer-General collects the property characteristic data including the site and capital values used as a basis for the property taxation function. Together, these data provide a rich source of land-related information, not only for government but also for research purposes. A comprehensive history of the development of the South Australian Land Information System is given by Sedunary (1987).

However, the current South Australian land-related data environment often holds the data in integratable, but not necessarily integrated, form. Generally caution needs to be exercised in the integration of the data, not necessarily from a technical point of view but more importantly from a data suitability perspective. The advent of the property cadastre removes a significant integration problem. However, in the case of this study, information as to the accuracy of the valuation data on an individual attribute basis was not always easy
to establish. For example, the quality of a prima facie useful property characteristic (presence of an ensuite bathroom) was not included after investigations revealed anecdotal evidence of unreliable quality. Ultimately the onus is on the user to establish the suitability of the data for a particular application. This is true when obtaining data from any data source and must be recognised as a caveat on the accuracy of the results.

3.5.2 Data identification and selection

The focus of this study is on real estate, its geography, and the behaviour of the property markets in relation to that geography. In stage 1 the data selection criteria are to identify a bundle of property attributes that describe the real estate geography traded in the residential real estate marketplace. Historically, hedonic modelling did not include location attributes limiting themselves to mainly structural attributes and were referred to by Orford (1999, p.48) as ‘location insensitive’ models. In part this may have been due to data availability which in current, more mature, land information management environments, is changing allowing access to a wider range of locational data including description of the neighbourhood amenity, accessibility, and socio economic indicators all of which have been shown to influence price allowing a more complete description of the real estate geography (Jackson et al. 2007; Kestens et al. 2004; Kestens et al. 2006; Reed 2001). This requires collecting data from a number of different sources and was done on the basis of their contribution to the description of the real estate geography. The increase in the wealth of available data has promoted a parsimonious selection process that is both required and adopted in this study. This necessitates that attributes adopted have to show cause for selection in terms of their ability to contribute to the answering of the research question. Attributes that cannot satisfy this are omitted as their presence could mask the end result. The data selection process is now discussed in more detail for both stage 1 and stage 2 of this study.
3.5.3 Stage 1

The research question to be answered in stage 1 focuses on the identification of the dimensions of the underlying residential real estate structure (RLS). It is the underlying real estate geography that constitutes the residential living structure, incorporating both the locational as well as the structural attributes that people purchase in a dwelling. The attributes that contribute to the ‘housing quality’ and ‘housing prices’ are explained by Tu (2003, p.38):

Spatial immobility implies that characteristics associated with a location are inherent in the bundle of attributes of a dwelling found at that location. These locational attributes include socio-economic status of the neighbourhood, its physical conditions, as well as wider notions of accessibility to any desired destinations, in terms of jobs, relatives and friends, private goods or public facilities. Combined, all of these attributes contribute to differences in housing quality and housing prices across locations.

Each of these categories are explored for appropriate attributes that may be considered as significant in this regard and compared to other studies reported in the literature that have similarly attempted to identify such attributes. An interesting question when considering the selection criteria of data in studies of this type is one investigated by Chhetri et al. (2006, pp.393–4) when discussing the influence of amenity on housing choice. It is the question as to whether or not such influence is based on an objective or a perceived understanding of the data. This may be of fundamental importance if the correct outcome is to be derived. Although that study was inconclusive as to this question (Chhetri et al. 2006, p.409), it certainly provides an interesting step forward in the understanding of the ‘geographic reality’ of a study area and is noted as a topic for further research in the development of this methodology.
Accessibility attributes

The accessibility of a property to various services is an essential part of choosing a location in which to live (Kestens et al. 2004, p.540). Accessibility factors included in this study are road network distances of the property from various services such as:

- Health: in particular a general medical practitioner’s surgery as this is often the first and most frequently accessed health provision.
- Education: especially primary and secondary schools as students frequently travel to school by walking or bicycling and closer proximity to the home is deemed to be an advantage.
- The CBD: as this often provides employment, entertainment and shopping facilities and properties further away are considered less favourably as transport costs to commute are higher.
- Regional shopping: being close to regional shopping can often partly negate the need to be close to the CBD.

There are various ways of describing accessibility including Euclidean distance, actual road distance and ‘car time’ to destination. Kestens et al. (2006, p.90) shows that ‘car time’ from major activity centres in Quebec City is negatively linked to property value. Where ‘car time’ data are available it is probably more relevant than other distance measures. There has been much work done at The University of Adelaide (GISCA 2002) in calculating the road network distance from the centroid of each property polygon to the centroid of the property polygon with a specific land use providing such broad services as education, health, shopping, financial services, public transport and the CBD. Although this data set does not include travel time data it has been used to represent accessibility of each property in the study area to the general practitioner’s surgery, primary school, secondary school, major shopping centres, local urban shops and the CBD as listed above.
Amenity attributes

The amenity of a neighbourhood is an important contributor to the property value (Chhetri et al. 2006, p.393). The definition of amenity may vary between study areas and it could become a complex task to quantify all aspects of what is a subjective measure. This study was constrained by time to derive and model a measure of amenity and also by the cost of purchasing the high resolution urban remotely sensed data often necessary to do so. Therefore, the amenity attribute was derived from the freely available 25 metre resolution satellite imagery as described later. The selection criteria for an amenity indicator can also be complex depending on how generally the term is defined. In a broad definition Orford (1999, p.1) refers to ‘locational externalities’ as describing the built environment that he argues is becoming increasingly important to understand and model as the urban structure changes through regeneration. The whole question of selection criteria for housing attributes that affect choice must be seen in the light of market behaviour and hence their relative importance as perceived by the players in the marketplace. Many attributes describing neighbourhood amenity can be selected as summarised by Kestens et al. (2004, p.541). However, the selection criteria are taken one step further by Chhetri et al. (2006, pp.393–4) by comparing subjective assessments with objective data. Perhaps this is in keeping with Kestens et al. (2004, p.541) who observed that it is what people can see from and around their property that influences the price. Trees are easily seen and have been shown to positively influence price (Morales et al. 1976). Therefore, trees (greenness) has been used as a proxy in this study for amenity, especially as they can be seen from and around their property enhancing the importance of the perception of the amenity as being important. This study has been restricted to the use of 25 metre resolution satellite imagery through the prohibitive cost of higher resolution remotely sensed data. Also, the time to research more exacting amenity data attributes is outside the scope of this study. This was the reason the
normalised density of vegetation index (NDVI) data obtainable from the available satellite imagery has been used to establish a ‘greenness’ index, particularly representing a proxy for the tree cover in the study area. An index was derived combining both the NDVI score for the individual property and its surrounds and the derivation for this study is discussed in more detail later in this chapter.

Socio economic attributes

Another important group of locational data important to housing price are socio economic attributes. They are selected on the basis of their contribution to the description of the underlying real estate geography. For example, attributes describing the average household, ethnicity, employment status, age and education structures of the census collection district (CD) in which the individual property is located are attributes of that property, forming part of the decision-making process in the marketplace. A number of studies have investigated the importance to housing choice of socio economic indicators. The effect of income on rental value and positive relationship between the per cent of highly educated households in a neighbourhood and price paid for housing was established in a Canadian study (Quebec) (Kestens et al. 2006). Of more relevance to this study, Reed (2001) investigated the various social constructs of suburban Brisbane between 1976 and 1996 to determine the relationship between social constructs and residential housing values and found (p.525) that suburbs with higher socio economic scores showed a strong positive correlation with house price. The socio economic scores comprised income, education and occupation type indicators. It is another question as to whether it is the price that is a cause or an effect of such attributes.

Another Australian study, not specifically concerned with the study area, addressed the question of the effect of socio economic indicators on the residential real estate market in
Darwin (Jackson et al. 2007). This study used data collected in the Australian Bureau of Statistics 2001 Census. Items included the ‘age’ of the individuals, dwelling type, individual income, ethnicity, occupation and length of stay in the same dwelling. The authors concluded that the social constructs of these indicators do play a role in price determination. The six social constructs, formulated using principal component analysis, that were found important were discussed by Jackson et al. (2007, p.6):

1. ‘Family’ (composed of age structure, dwelling tenure and individual status). High positive correlation with price for coupled families with children.

2. ‘Mobility’ (composed of living in the same address five years ago, tenure and age structure). High correlation with price for people living in the same address for five years, older people who owned their own home.

3. ‘Socio economic status’ (composed of income, bachelor degree, occupation). High positive correlation with high income.

4. ‘Low to middle skilled’ (composed of income, occupation). High positive correlation of price with trade occupations.

5. ‘Owned outright’ (composed of one-parent families, people who speak English only). High positive correlation of price with one-parent families.

6. Ethnicity (composed of place of birth and people who speak other languages). High positive correlation with price for people born overseas and speak other languages.

Socio economic data adopted in this study reflect the social constructs found important to the choice and price paid for housing in previous studies. The ‘household size’ (small, average and large), ‘tenure’ (owned, mortgaged and rental) and ‘income’ (low, below average, average, above average and high) describe the overall household while the ‘age structure’, ‘individual status’ (married, sole parent, lone, and dependent children), ‘length of occupancy’ (one year, five years), ‘tertiary qualification’ and ‘employment status’ give a profile of the household structure while ‘individual place of birth’ and ‘languages spoken at home’ highlight the ethnicity dimension of the residential structure. The age structure
categories were chosen to represent children at home (0–20 years), young adults (21–34 years), parental age (35–54 years), older, pre-retirement ‘baby boomer age (55–65 years) and grand parent and/or retiree age (greater than 65 years). These are an important group of data that contribute to property price and therefore form an important component in the derivation of the underlying residential real estate living structure.

*Structural attributes*

Finally, structural attributes of the dwelling affect the value of the property and therefore properly belong as part of the attribute bundle that describes the real estate geography (Orford 1999, p.48). Prominent among these attributes are the size, age, construction, condition and type of dwelling. Structural variables have been used in hedonic modelling over a long period of time (Grether & Mieszkowski 1974; Lancaster 1966; Rosen 1974). More recent studies examining structural attributes carried out in the Adelaide study area show that dwelling size, condition, number of rooms, property type and land area of the property each contribute to property value (Lockwood 2003; Rossini & Kershaw 2005). Dwelling size was selected as a proxy for the number of rooms as it was considered to be more reliable and consistent within this data set. An additional variable describing the added value of the dwelling structure to the property was included and the derivation is discussed later in this chapter. Although the use of this as a dwelling attribute has not been found in other studies there is recognition in the planning strategies for high density living as part of the urban development programmes. This attribute is used as a proxy for those properties that may be economically suitable for redevelopment to a higher density land use and therefore considered an appropriate descriptor of the residential living structure for this study area. The structural attributes adopted for this study were the dwelling types expressed as ‘dummy’ variables (single dwelling, multiple dwelling, home unit and rural living), the major construction types also expressed as ‘dummy’ variables (brick, stone and
rendered) and the dwelling area, condition, added value and the land parcel size all expressed as continuous variables from, or derived from, the Valuer-General’s data sets.

3.5.4 Stage 2

Sale transaction price

In stage 2 of the study the RLS is related to the marketplace using geographically weighted regression techniques to interpret the submarket structure so as to give the resulting spatial boundaries a market entity status. The market is characterised by the property transaction prices for the time concerned. The transaction prices were obtained from the South Australian Valuer-General for the time span July 2001 to September 2001 which was deemed to represent the market at the date of the study namely August 2001. These data consisted of records that represent the sales transactions as registered with the South Australian Lands Titles Office and that have been investigated by the Valuer-General and found to represent market value. This process eliminates errors and outliers from the data set. Sales transaction data analysed in this manner are not normally available but were supplied by the Valuer-General in this case due to the interest in the potential use of the results. This resulted in a dataset containing 7155 property transaction records.

Surrogate property characteristics from stage 1

The objective in stage 1 of this study was to quantify the underlying dimension of the residential living structure of the entire study area. This was achieved by describing this structure in terms of the property attributes that were thought to contribute to the choice process of selecting a property in which to live. These attributes (50 in number collected on all residential properties in the study area) were reduced using principal components analysis to 10 factors that are deemed to represent the dimensions of the underlying structure. Factor scores were calculated for each of the 10 factors for each of the 443 619
residential properties in the study area. These 10 factors were then used as surrogate property characteristics in stage 2 of the study and added to the data set containing the analysed sale prices.

3.6 Data sources

3.6.1 Australian Bureau of Statistics


3.6.2 The University of Adelaide

GISCA – The National Centre for Social Applications of GIS (road network distances from each property centroid to various services), The University of Adelaide. A caveat is in the fact that the road distances take no account of time taken to travel the road or of ‘one-way’ systems and in reality may in some cases not be the shortest distance.

3.6.3 The State Government of South Australia

Department of Environment and Heritage (remotely sensed vegetation data and the property cadastre).

Department for Administrative and Information Services, Land Services Group, Office of the Valuer-General, South Australia, (structural dwelling attributes and property transaction data).

3.7 Preparation and derivation of certain data attributes

Some data attributes used in this study were derived from the secondary data collected. Each of the derivation processes of these data are described as follows.
3.7.1 Amenity attribute

Amenity was based on a vegetation greenness index, where an area with abundant vegetation scored higher than an area with sparse vegetation. This index is constructed by using a normalised density of vegetation index (NDVI) derived from a 25-metre resolution satellite image provided by the South Australian Department of Environment and Heritage. An amenity value was calculated for each property so as to reflect the amenity of the property itself together with its more immediate surrounds. The rationale for this is that when a property is transacted a buyer would take notice of the general amenity in which the property is situated to a greater extent than the actual amenity of the property itself. The property itself is under the control of the property owner and thus able to be changed whereas the surrounding amenity is not. Therefore, the construction of the amenity score was undertaken as follows:

- The 25 metre grid was converted to a 5 metre grid with each of the 5 metre grid cells assigned the 25 metre grid cell value. This allowed the smaller property polygons that were contained in or only occupied a portion of the 25 metre grid to be summarised.

- The neighbourhood amenity factor was taken as an average of 15 cells by 30 cells about the centre of each property. This is an arbitrarily chosen area, done to reflect the importance of the immediate surrounds visible from the property.

- The property amenity factor was taken as the average of a 5 metre grid over property itself.

The amenity score was calculated by taking twice the neighbourhood value plus the property value. While the factor of two for the neighbourhood was arbitrary, it acknowledges a higher weighting for the neighbourhood than for the property itself as discussed above. It is acknowledged that this is a rudimentary estimate of amenity but was the only measure the time and cost constraints allowed. Further research will be undertaken.
to improve the quality of this important attribute in future developments of this methodology.

### 3.7.2 Socio-economic attributes

The socio-economic data are sourced from the Australian Bureau of Statistics (ABS), quinquennial Census of Population and Housing 2001. Census data include measures of age, income, family structure, employment and tenure. The smallest spatial unit at which this data is available is the census collection district (CD) level, which includes on average 220 households. The categories for the variables describing ‘household size’, ‘household tenure’ and ‘household income’ as ‘low’ through ‘high’ are based on the distribution about the mean value for the whole study area as given in the 2001 census data (Australian Bureau of Statistics 2002b). For each CD the number of households or individuals that fit each category is expressed as a percentage of the total for that CD. These scores are then standardised. As all other data are collected at the individual property level every property within a CD was assigned the above described socio-economic characteristics of that CD. It is argued in this study that an average of the CD is a better attribute than the actual attribute for the household as upon sale the actual attribute may change and the incoming occupant may well have considered the surrounding attributes to be of more relevance than the actual attributes of the outgoing occupant.

### 3.7.3 Added value structural attribute

Site and capital values of the property calculated for taxation purposes are also obtained and used to determine the added value of the structural component of the property as another attribute. This was calculated as a ratio of the capital value to the site value with a value of 1 indicating no added value of the structure as both capital and site values are equal, with progressively higher values indicating an increase in the added value of the
structural component. This is so because, in South Australia, the site and capital values are market values calculated independently from one another and therefore their difference is an approximate added value of the structure as perceived by the market place. This variable provides a reflection of the functional and/or economic obsolescence of the dwelling structure. As Whipple, (1995, p.10) observes:

… if functional or economic obsolescence overtakes improvements their value is all but destroyed even though they may be in first class condition.

This may occur for a number of reasons. A prime example in the residential context is a permitted land use change allowing a higher and better land use. Using ‘added value’ as a dwelling attribute it is possible to see this planning and market phenomenon reflected in the underlying structure of the real estate geography of the study area by observing its correlation (if any) with other variables in the study.

### 3.8 Data adopted for the study

The data from various sources are, as far as possible, taken as at a particular point in time. The study date is August 2001, as that is the date of the last available census data and the other data sets have been taken to be as close as possible to that date.

Spatial reference for the data is based on the property cadastre that is a digital representation of the Valuer General’s designated property boundaries. The data used in this study, and described in this chapter are summarised in table 3.2.
### Table 3.2 Variables used in the study for Stage 1

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Variable type</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single dwelling</td>
<td>Boolean (0,1)</td>
<td>Valuer General (VG)</td>
</tr>
<tr>
<td>Multiple dwelling</td>
<td>Boolean (0,1)</td>
<td>VG</td>
</tr>
<tr>
<td>Home unit</td>
<td>Boolean (0,1)</td>
<td>VG</td>
</tr>
<tr>
<td>Rural living (non-primary production)</td>
<td>Boolean (0,1)</td>
<td>VG</td>
</tr>
<tr>
<td>Dwelling construction – brick</td>
<td>Boolean (0,1)</td>
<td>VG</td>
</tr>
<tr>
<td>Dwelling construction – stone</td>
<td>Boolean (0,1)</td>
<td>VG</td>
</tr>
<tr>
<td>Dwelling construction – rendered</td>
<td>Continuous standardised variable</td>
<td>VG</td>
</tr>
<tr>
<td>Dwelling area</td>
<td>Continuous standardised variable</td>
<td>VG</td>
</tr>
<tr>
<td>Dwelling condition</td>
<td>Continuous standardised variable</td>
<td>VG</td>
</tr>
<tr>
<td>Dwelling added value</td>
<td>Continuous standardised variable</td>
<td>VG</td>
</tr>
<tr>
<td>Land (site) area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road distance of dwelling from:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✦ GP surgery</td>
<td>Continuous standardised variable</td>
<td>The University of Adelaide</td>
</tr>
<tr>
<td>✦ primary school</td>
<td>Continuous standardised variable</td>
<td>The University of Adelaide</td>
</tr>
<tr>
<td>✦ secondary school</td>
<td>Continuous standardised variable</td>
<td>The University of Adelaide</td>
</tr>
<tr>
<td>✦ major shops</td>
<td>Continuous standardised variable</td>
<td>The University of Adelaide</td>
</tr>
<tr>
<td>✦ urban shops</td>
<td>Continuous standardised variable</td>
<td>The University of Adelaide</td>
</tr>
<tr>
<td>✦ CBD</td>
<td>Continuous standardised variable</td>
<td>The University of Adelaide</td>
</tr>
<tr>
<td>Amenity value</td>
<td>Continuous standardised variable</td>
<td>Department of Environment and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heritage</td>
</tr>
<tr>
<td>Household size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✦ Small</td>
<td>Continuous standardised variable</td>
<td>Australian Bureau of Statistics (ABS)</td>
</tr>
<tr>
<td>✦ Average</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Large</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>Household tenure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✦ Owned</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Mortgaged</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Rental</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>Household income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✦ Low</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Below average</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Average</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Above average</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ High</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>Length of same occupancy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✦ 1 year</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ 5 years</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>Individual place of birth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✦ NW Europe</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ SE Europe</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ SE Asia</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ NE Asia</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Australia</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>Individual status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✦ Married</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Sole parent</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Lone</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Dependent children</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>Individual age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✦ 0 to 20 years</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ 21 to 34 years</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ 35 to 54 years</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ 55 to 65 years</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
<tr>
<td>✦ Greater than 65 years</td>
<td>Continuous standardised variable</td>
<td>ABS</td>
</tr>
</tbody>
</table>
3.8.1 Previous data selection processes

The selection of the 50 variables described above represents the results of a continuing development process. Previous work in this study adopted 65 variables as detailed in Lockwood and Coffee, (2006, p.359). Principal component analysis of these data revealed 15 factors explaining 74% of the variation in the data. Analysis in stage 2 of the study found that 15 factors were too many for ease of interpretation and a more parsimonious model would be more practical. The 65 variables were examined and data such as ‘Household mortgage repayment’ (five variables) and ‘Household rental repayments’ (five variables) were considered to be accounted for by the ‘Household tenure’ and ‘Household income’ variables in so far as it was particularly the tenure and the income influence rather than the repayment details that were considered important for this study and were therefore omitted. Similarly, the four variables describing the number of cars as a measure of wealth were again adequately accounted for by the ‘household income’ and were omitted. As discussed earlier the ‘Dwelling number of rooms’ variable was omitted as the ‘Dwelling area’ was deemed a proxy for this attribute and for this data set was considered a more reliable and consistent variable. The variable ‘Dwelling age’ was omitted due to the consistently low communality score (consistently less than 0.2 in several models with various data attribute combinations) and did not significantly load on any factor (Hair et al. 1992, p.241). Although thought to be an important variable in explaining the composition of the real estate geography it was omitted from this particular data set with investigations.
to be undertaken in future research to establish a more comprehensive understanding as to the reason. An extra variable was added namely that of the distance of each property to the CBD. It was omitted from the original data set based on the ‘multi-nuclear’ nature of the study area. However, given the increasingly residential nature of the CBD, this was considered inappropriate and was reinstated to the variable list. The above described deletions and the one addition reduced the number of variables in the stage 1 analysis from 65 to 50 and reduced the number of factors from 15 to 10, achieving the desired objective of a more parsimonious model. It is considered part of the ongoing work with land-related data custodians to investigate the identification and collection of appropriate variables to further improve the parsimonious nature of the model.

3.9 Conclusion

Testing the efficacy of the methodology is an important aspect of this study. The creation of an integrated data set containing appropriate property attributes is central to this. The study makes no assumptions as to where the spatial boundaries might be but instead relies on the individual property data to reveal their whereabouts. As with the creation of many social science data sets this involves the integration of data from a number of different data custodians and the ease with which this can be done is influenced by the land information management environment of the study area concerned. Also, in common with many social science projects, suitable attribute data often has to be derived from existing data sets. This may result in a compromise with reality, but representing reality in a modelling exercise such as this can never be fully achieved, and the extent to which this cannot be achieved, must be offered as a caveat in the interpretation of the results.
It is a study in its own right to derive attributes that adequately describe the influence on property choice of such subjective concepts as amenity and socio economic status. As indicated by Chhetri et al. (2006) it may be the perceived rather than the objective measure that is the important aspect to quantify. This offers a necessary challenge for future research.

The data attributes adopted in this study represent a wide range of property characteristics that contribute to the process of choosing a property to buy and thus the price paid. The attribute selection process was focused on answering the research questions of firstly describing the underlying residential real estate structure of the study area, and secondly understanding its relationship with the marketplace operating within it. The formulation of the RLS is an important part of the total process and is a non-trivial exercise. The identification and collection of appropriate data is an ongoing exercise aimed at achieving a suitably parsimonious model as a basis for the delineation of the geospatial submarket boundaries.
CHAPTER 4 METHODOLOGY OVERVIEW

4.1 Introduction

The aim of this chapter is to examine existing methodologies adopted for determining spatial submarket boundaries. This will enable opportunities to be identified that can be used to build upon existing work in order to make a contribution to this field of research.

The analytical techniques used in this study are not new. Principal component analysis (PCA) (stage 1), cluster analysis and hedonic modelling (stage 2) are well supported in the submarket literature (Bourassa et al. 1999, p.161; Bourassa et al. 2003, p.167; Kestens et al. 2004, p.543; Maclennan & Tu 1996, p.398; Watkins 1999, p.165). In the case of geographically weighted regression (GWR) (Fotheringham et al. 2002), although the technique is not new, it has not currently been specifically reported as an analytical tool in the submarket literature.

The contribution this study makes is the way in which these techniques are used, in particular using PCA to construct independent variables for use in the hedonic GWR model to detect the spatial variation in the model’s parameter estimates as indicators of submarket boundaries.

This study investigates the spatial dimensions of submarkets at a particular point in time. The temporal aspects of submarkets are of importance but are outside the scope of this study. A date of study has been adopted as August 2001 being the date at which the last Australian quinquennial population census was undertaken and from which the socio
economic data for the study area was taken. Other data have been collected as close to this
date as possible as outlined in chapter 3. The aim of the methodology is to delineate spatial
submarket boundaries without reference to existing spatial structures.

The methodology adopted in this study involves two steps:

1. Deriving and mapping the dimensions of the residential living structure (RLS) of the
   study area. It is proposed to do this by deriving the principal components of a wide
   variety of attributes for all the properties in the study area collected at the individual
   property level. Using principal component analysis (PCA) as a quantitative
   expression of the closeness of attributes in n-dimensional data space, GIS is used to
give expression of the derived components in two-dimensional geographic space as
   a means of testing the efficacy of the PCA results.

2. Relating the derived RLS to the marketplace by using a hedonic regression
   algorithm. In this study a geographically weighted regression (GWR) model is
   adopted with the sale price of a property as the dependent variable and the principal
   components of the RLS (derived in stage 1) as the independent variables. The study
   is based on the premise that real estate markets are dynamic. Location matters,
   geography matters, and to detect where one local market effect begins and ends is
   the essence for establishing the spatial submarket boundaries. Markets are about
   price and this methodology uses price as the submarket identifier. In particular, the
   price identifiers used are the price patterns of the hedonic parameter estimates from
   the GWR model calculated at each regression point as it moves across the
   geographic space of the study area. Such a model takes the general form as shown in
equation 4.1.

**Equation 4.1**

\[
\text{Price} = b_0(u,v) + b_1(F_1)(u,v) + \ldots + b_n(F_n)(u,v) + \text{error}(u,v)
\]

where: \((u,v)\) are the location of the sale

\(F_1 \ldots F_n\) are the principal components of the supply and demand attributes taken together from stage 1.

Differences in the patterns of the GWR parameter estimates \((b_0 \text{ to } b_n\) in equation 4.1) moving across the study area are taken as a signal of change in the market’s assessment of
the RLS. It is the thesis of this study that this change may signal a change in the spatial
component of the submarket boundary. Cluster analysis is used to group like patterns of
the parameter estimates \( b_0 \ldots b_n \) in equation 4.1) so as to maximise internal cluster
homogeneity and maximise between cluster heterogeneity. The resulting clusters are
considered to be geospatial submarkets.

This chapter discusses the analytical techniques used in the real estate literature and
matches those and other spatial techniques with the underlying market theory in order to
derive a methodology that will delineate the geospatial submarket boundaries in a
transparent manner.

4.2 Analytical techniques

4.2.1 A context

As discussed in chapter 2 there are many and varied definitions of housing submarkets.
This may be due in part to the number of different applications to which knowledge of
submarket structure may be put. Agreement on a definition has not been forthcoming and
Watkins (2001, p.2235) suggests that this may be one of the many reasons why there has
been so little development of submarket models. Although this study investigates the
derivation of spatial submarkets for general use in the micro housing policy area, it is
particularly focused on its use in the mass appraisal industry to derive a fair and equitable
market price of property that is used as the basis for the property taxation system in many
jurisdictions around the world (Eckert 1990, p.6). Hence, a submarket definition in terms
of substitutability in which human marketplace behaviour is recognised as a fundamental
element of establishing housing price is particularly appropriate.
Having adopted a definition of a housing submarket (chapter 2), the task now becomes one of identifying an appropriate set of analytical tools that are capable of translating the definition into the identification of the spatial delineation of submarket boundaries.

As data becomes better managed, more freely available, and computer hardware and software become more powerful and cheaper, there is a growing opportunity for practitioners to look more analytically at the structure of local housing markets. This may be beginning to happen, and as Jones et al. (2004, p.270) point out ‘it is only relatively recently that the existence of submarkets has been deemed to be of analytical importance’. Perhaps this is because both data management and affordable high-powered desktop computing allow it to be. This view is expanded to include both methodological and operational impediments to the ‘spatial’ treatments of real estate markets as discussed by Anselin (1998). The methodological impediment broadly involves the recognition of the importance of the ‘two-dimensional nature of spatial interaction (or spatial auto-correlation)’ (p.113), and the operational impediment is about the ‘lack of dissemination of the methods of spatial econometrics’ (p.114) to real estate practitioners and their inclusion within GIS software. These questions, although important, do not address a major impediment to social science research generally, namely the translation of analysis in n-dimensional space to its 3-dimensional representation in geographic space. This is a complex question, beyond the scope of this study and should be viewed as a matter for further research.

The task of the analyst is to match the theory and definition of a submarket with the theory and application of the analytical techniques. As long as the definition of the submarket being derived is clearly established, users can benefit from consistently derived submarket boundaries that should lead to a better understanding of the local housing market structure.
This chapter reviews the most commonly used techniques as described in the literature and attempts to choose the most appropriate for the geospatial submarket definition adopted for this study and described in chapters 1 and 2. Where appropriate techniques do not currently exist within the literature then the search will need to be extended.

This study adopts a methodology that has been divided into two stages. The first is to develop an understanding of the residential structure (RLS) of the whole study area in which the submarkets reside and then the second stage is to relate that structure to the marketplace at a given point in time so as to indicate the spatial component of submarket boundaries. This chapter discusses the overall methodology and its relationship with the current literature. Chapter 5 details the methodology for stage 1 and chapter 6 for stage 2.

4.3 The submarket literature

The submarket literature is concerned with understanding the structure of a market entity (Pryce 2004b; Watkins 2001). Regardless of how submarkets may be delineated, the evidence they exist is based on a now accepted standard test that according to Jones (2002, p.554) was first set out by Schnare and Struyk (1976). Hedonic techniques are the basis for this test. Founded on ‘the law of one price’, which in general terms states that in an efficient market a commodity must have a single price or else an arbitrage opportunity would exist returning the market to the one price. In the real estate market Pryce (2004a, p.2–11) interprets this to mean that in an hedonic model ‘within any one market, each attribute will have the same market value’ whereas Schnare and Struyk (1976, p.163) conclude that it is up to the researcher to decide whether to apply the law to the attribute price or the overall price of the property. The method involves comparing the difference between hedonic pricing models and, if significant, then they are deemed to be different
submarkets. The test appears intuitively sensible but Rothenberg et al. (1991, p.64) place a caveat on the use of the test to produce meaningful results where the *a priori* submarkets have been determined by methods not based on substitutability. They claim that Schnare and Struyk (1976) showed variations in coefficients only have a small difference on the ability of the equation to either explain or predict what is happening and it was unlikely that such variations are economically or statistically meaningful. The geospatial submarkets derived in this study have not been subjected to this test for submarket status as discussed in chapter 3.

The use of analytical techniques in the local residential housing market literature is large and has been well reviewed (Kauko 2000; Watkins 2001, p.2242). For the purpose of this study the various types of analytical techniques may be regarded as falling into three broad groups as summarised in table 4.1.

<table>
<thead>
<tr>
<th>Author</th>
<th>Objective</th>
<th>Analytical technique</th>
<th>Location &amp; data used</th>
<th>Sample size</th>
<th>Comments grouping (as per text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adair et al. 1996a</td>
<td>Homogeneity in <em>a priori</em> structural &amp; spatial submarkets</td>
<td>Hedonic modelling – stepwise regression</td>
<td>Belfast. Property, census &amp; environmental</td>
<td>Approx. 1000 sales</td>
<td>Link submarkets (structural &amp; spatial) to valuation process GROUP 1</td>
</tr>
<tr>
<td>Goodman &amp; Thibodeau 2003</td>
<td>Test predictive accuracy of hierarchical submarkets</td>
<td>Hierarchical techniques</td>
<td>Dallas County</td>
<td>28,939 sales</td>
<td>Similar results with lower variance GROUP1</td>
</tr>
<tr>
<td>Jones et al. 2004</td>
<td>Examine intra migration patterns</td>
<td>Migration patterns</td>
<td>Glasgow</td>
<td>Approx 16,000 open market transactions</td>
<td>Housing market insights through migration patterns. Used predefined submarkets. GROUP 1</td>
</tr>
<tr>
<td>Rothenberg et al. 1991</td>
<td>Ranking by quality index</td>
<td>Hedonic modelling</td>
<td></td>
<td></td>
<td>Quality spectrum. (hierarchical degree of substitutability) GROUP 1</td>
</tr>
</tbody>
</table>
## Chapter 4 – Methodology Overview

<table>
<thead>
<tr>
<th>Author</th>
<th>Objective</th>
<th>Analytical technique</th>
<th>Location &amp; data used</th>
<th>Sample size</th>
<th>Comments grouping (as per text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watkins 2001</td>
<td>Comparison of spatial, structural &amp; nested SMGs</td>
<td>Hedonic regression</td>
<td>Glasgow</td>
<td></td>
<td>A priori spatial &amp; structural submarkets.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chow test</td>
<td></td>
<td></td>
<td>GROUP 1</td>
</tr>
<tr>
<td>Clapp 2003</td>
<td>Account for location in valuation models.</td>
<td>Local Regression Model (LRM)</td>
<td>Six towns in Massachusetts</td>
<td>5713 sales</td>
<td>Produced a continuous location value surface.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OLS – holding prop chars constant</td>
<td>Prop Char data</td>
<td></td>
<td>Argued no need for submarkets. GROUP 2</td>
</tr>
<tr>
<td>Deddis et al.</td>
<td>Locational value distributions as a CAMA base.</td>
<td>Kriging interpolators</td>
<td>Londonderry.</td>
<td></td>
<td>CAMA/GIS discussion GROUP 2</td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figueroa 1999</td>
<td>Measure the value of location in different regions for mass appraisal.</td>
<td>Location ‘blind’ MRA residual model</td>
<td>Regina.</td>
<td>7413 sales</td>
<td>Use of derived location factor in the predictive model.</td>
</tr>
<tr>
<td>Gallimore et al.</td>
<td>Derive a location value surface from location factors</td>
<td>Location ‘blind’ MRA residual model</td>
<td>Stafford</td>
<td>218 sales</td>
<td>Improves statistical models. GROUP 2</td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kauko et al.</td>
<td>Derive housing market segmentation</td>
<td>Neural networking</td>
<td>Helsinki metropolitan area</td>
<td>1993 dwelling transactions</td>
<td>Limited success GROUP 2</td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td>4 structural and 6 locational variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orford 1999</td>
<td>Accounting for location externalities</td>
<td>Multilevel modelling</td>
<td>Cardiff Location (not SMGs)</td>
<td></td>
<td>Geographic and economic approach GROUP 2</td>
</tr>
<tr>
<td>Thrall 2002</td>
<td>Delineation of housing submarkets</td>
<td>Geodemographic techniques</td>
<td></td>
<td></td>
<td>GROUP 2</td>
</tr>
</tbody>
</table>

### GROUP 2. location included in modelling process

<table>
<thead>
<tr>
<th>Author</th>
<th>Objective</th>
<th>Analytical technique</th>
<th>Location &amp; data used</th>
<th>Sample size</th>
<th>Comments grouping (as per text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clapp 2003</td>
<td>Account for location in valuation models.</td>
<td>Local Regression Model (LRM)</td>
<td>Six towns in Massachusetts</td>
<td>5713 sales</td>
<td>Produced a continuous location value surface.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OLS – holding prop chars constant</td>
<td>Prop Char data</td>
<td></td>
<td>Argued no need for submarkets. GROUP 2</td>
</tr>
<tr>
<td>Figueroa 1999</td>
<td>Measure the value of location in different regions for mass appraisal.</td>
<td>Location ‘blind’ MRA residual model</td>
<td>Regina.</td>
<td>7413 sales</td>
<td>Use of derived location factor in the predictive model.</td>
</tr>
<tr>
<td>Gallimore et al.</td>
<td>Derive a location value surface from location factors</td>
<td>Location ‘blind’ MRA residual model</td>
<td>Stafford</td>
<td>218 sales</td>
<td>Improves statistical models. GROUP 2</td>
</tr>
<tr>
<td>Kauko et al.</td>
<td>Derive housing market segmentation</td>
<td>Neural networking</td>
<td>Helsinki metropolitan area</td>
<td>1993 dwelling transactions</td>
<td>Limited success GROUP 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 structural and 6 locational variables</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### GROUP 3 spatial delineation determined by the data

<table>
<thead>
<tr>
<th>Author</th>
<th>Objective</th>
<th>Analytical technique</th>
<th>Location &amp; data used</th>
<th>Sample size</th>
<th>Comments grouping (as per text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bourassa et al.</td>
<td>Do submarkets produce better out-of-sample predictions? (i.e. mass appraisals)</td>
<td>PCA, K-means Cluster analysis.</td>
<td>Auckland NZ Property characteristics Census Euclidean dist from CBD.</td>
<td>8421 transactions 34 sales groups (valuer’s opinion of homogenous areas)</td>
<td>Spatial submarkets do matter. GROUP 3</td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bourassa et al.</td>
<td>A priori (constrained) – vs data driven (unconstrained)</td>
<td>PCA, K-means Cluster analysis.</td>
<td>Household survey data in Melbourne &amp; Sydney.</td>
<td>4661 cases (both Sydney &amp; Melbourne)</td>
<td>Only 1 unconstrained submarkets showed significantly better results than a priori definition GROUP 3</td>
</tr>
<tr>
<td>1999</td>
<td>submarket definitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hedonic modelling to test differences.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

94
The first group sought to confirm that certain *a priori* spatial and or structural boundaries were better than none at all in terms of establishing more homogeneous housing markets. This was not an attempt to find the optimum solution but rather to confirm that smaller submarkets did exist and were each more homogenous than the global market. Various

<table>
<thead>
<tr>
<th>Author</th>
<th>Objective</th>
<th>Analytical technique</th>
<th>Location &amp; data used</th>
<th>Sample size</th>
<th>Comments grouping (as per text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bourassa et al. 2007</td>
<td>Compare methods for controlling spatial dependency</td>
<td>Modelling error structure</td>
<td>Auckland NZ</td>
<td>4880</td>
<td>Valuer defined geographical submarkets preferred GROUP 2 &amp; 3</td>
</tr>
<tr>
<td>Cano-Guervos et al. 2003</td>
<td>Dividing a city into homogenous districts</td>
<td>PCA Theory of regionalised variables.</td>
<td>City of Granada Housing characteristics</td>
<td>298 previously occupied flats.</td>
<td>Results better than <em>a priori</em> admin boundaries. GROUP 3</td>
</tr>
<tr>
<td>Dale-Johnson 1982</td>
<td>Define market segmentation</td>
<td>PCA</td>
<td>Santa Clara County</td>
<td>3021</td>
<td>Complex task. Diverse data required. GROUP 3</td>
</tr>
<tr>
<td>Goodman &amp; Thibodeau 1998</td>
<td>Define submarkets using hierarchical techniques</td>
<td>Hierarchical techniques</td>
<td>Dallas County</td>
<td>28 939 sales</td>
<td>submarket defined as constant price per unit of housing quality Hierarchical techniques are useful frameworks for deriving submarkets GROUP 3</td>
</tr>
<tr>
<td>Maclennan &amp; Tu 1996</td>
<td>Empirical evidence RE: the existence of submarkets.</td>
<td>PCA Structural attributes forming product groups</td>
<td>Glasgow Structural spatial</td>
<td>Used <em>a priori</em> spatial units. GROUP 3</td>
<td></td>
</tr>
<tr>
<td>Meen &amp; Meen 2003</td>
<td>Empirical models of urban housing markets</td>
<td>Cellular Automata</td>
<td></td>
<td>New approach GROUP 3</td>
<td></td>
</tr>
<tr>
<td>Pryce 2004a</td>
<td>Delineate submarket boundaries</td>
<td>Hedonic structural break scanning</td>
<td></td>
<td>GROUP 3</td>
<td></td>
</tr>
<tr>
<td>Watkins 1999</td>
<td>Structural submarkets</td>
<td>PCA</td>
<td>Glasgow Supply-side attributes</td>
<td>Individual CAMA models for each structural submarket. GROUP 3</td>
<td></td>
</tr>
</tbody>
</table>
submarkets are selected *a priori*, to be either structural and or spatial and then tested using hedonic housing price differences to see if the results were significantly improved in the smaller submarkets (Adair et al. 1996b, p.69). The advantage of this method is that submarkets can be quickly identified in a cost-effective manner and if market difference is significant they may be used to generate, in the case of the mass appraisal function more reliable predictive value models than if submarkets were not used (Adair et al. 1996b, p.69; Goodman & Thibodeau 2007, p.209). There is also concern in the literature that if hedonic pricing models do not recognise the existence of submarkets (however delineated) they may be subject to aggregation bias (Watkins 1999, p.157).

However, the question always remaining unanswered is that the chosen spatial and structural groups may not be appropriate for the given market conditions present at the time the study was conducted and are not optimal since their formation was not related to the marketplace. In Adair et al. (1996b, p.73) the spatial groupings chosen were planning districts that appeared meaningful, but to be meaningful for planning may not necessarily coincide with being meaningful for the prevailing property market conditions. This will always be a problem with groupings not specifically designed to meet a particular submarket definition or market conditions. This is supported by a study in Auckland in which Bourassa et al. (2007, p.158), and Bourassa et al. (2003, p.27) found that spatial submarkets nominated by valuers performed better than delineations that were not, suggesting the importance of basing the spatial delineations on the market place.

The second group of analytical techniques recognises the importance of location in residential real estate value and attempts to include it as part of the housing hedonic price modelling process, thus removing the need to be aware of submarket boundaries (Bourassa et al. 2007; Clapp 2003; Figueroa 1999; Fik et al. 2003; Gallimore et al. 1996; McCluskey
Chapter 4 – Methodology Overview

et al. 2002; Pryce & Evans 2007; Tu et al. 2007). The attraction of these analytical techniques is that they recognise every property as having a unique location factor and therefore accounting for location on an individual property basis and removing the need to assume a priori boundaries. The effect due to location is a continuous geographic surface. This group incorporates location into their modelling by variously determining a unique location factor for each property. Clapp (2003, p.303) relies on local regression models incorporating a local polynomial regression to calculate a location factor as a function of the latitude and longitude. Fik et al. (2003, p.626) uses higher order polynomial functions incorporating both the ‘x’ and ‘y’ coordinates and location ‘dummy’ variables to provide a unique ‘location-value signature’ for each property, while Pryce and Evans (2007, p.22), build on this to incorporate the temporal dimension of the market place and establish continuous ‘inflation surfaces’ as indicators of submarket behaviour across geographical space. Earlier attempts to generate value location surfaces constructed global hedonic regression models incorporating only structural property attributes (leaving the model deliberately ‘blind’ to location) and treated the residuals as being indicative of the effect of ‘location’ and model error (Figueroa 1999; Gallimore et al. 1996; Lockwood 2003). Interpolation of the residuals is then considered to be the effect on value due to location. The implicit advantage of the approach of this second group is that firstly, it does not make a priori assumptions as to where the effects of location on property value may be and secondly, the models only rely on readily available data thus removing the time and cost of measuring the individual components of externalities deemed to comprise ‘location’. This can be of significant practical importance in jurisdictions where such data are unavailable and indications of the influence of location are necessary for inclusion in property valuation predictive models to support a property taxation system. Interestingly, Bourassa
et al. (2007, p.158) who compared the modelling of the error structure to other spatial market delineations, concluded that, for the data in their study:

the valuer-defined geographical submarkets are more useful in a mass appraisal context than the more fluid concept of submarkets implied by formal modelling of the spatial dependence of residuals.

However, the disadvantage of the general approach taken by the second group may be that the resulting location value surfaces can not be explained in terms of the underlying real estate geography that includes various location attributes such as amenity, accessibility, socio-economic and environmental indicators. In addition, the surface interpolators used to generate the value gradients may not suitably detect the changes that may be exist in reality. For example, as Clapp (2003, p.304) points out, there appears a potential conflict between using polynomials of sufficiently high degree to capture the flexible value surface and collinearity and loss of degrees of freedom in the higher order polynomial models. The disadvantage of the ‘blind’ residual model approach (Figueroa 1999; Gallimore et al. 1996; Lockwood 2003) is the uncertainty as to the degree of model error that may be present in the residual used as a proxy for location.

The third group focuses on the delineation of submarket boundaries from the data alone, also without the need to assume a priori spatial boundaries. These data include product group attributes and consumer group attributes sometimes used separately or together. The most common analytical technique appears to be the use of principal component analysis (PCA) to group attributes that represent the dimensions of marketplace. These components can then be represented spatially either using cluster analysis (Bourassa et al. 1999; Bourassa et al. 2003) or using geostatistical interpolation (Cano-Guervos et al. 2003) to give more homogenous districts. The principal components of the structural attributes are
termed product groups by (Maclennan & Tu 1996, p.396) and tested as being significantly different using the accepted test laid down by Schnare and Struyk (1976) and thereby redefined as submarkets. A drawback with the principal component approach is that it is not necessarily related to the current marketplace and therefore cannot be viewed as an optimal submarket delineation which is an economic entity. A new approach is taken by Pryce (2004b) who, in effect, uses the accepted test for submarket differentiation itself as a means of delineating the submarket boundaries in the first place. Based on the ‘law of one price’ a technique Pryce (2004b, p.2–9) calls ‘hedonic structural break scanning’ is basically a hedonic pricing model that moves across the study area in a grid pattern calculating the marginal price of the attributes and comparing for significant statistical change using a chow testing procedure, although, at the time of writing, is still in the process of development.

Hedonic modelling plays a fundamental role in each of these broad groups of submarket determination. However, the significant difference lies in how the technique is used to accommodate the spatial component of submarkets. In group 1 hedonic modelling is used to show that certain a priori groups are better than the global models while making no attempt to optimise the result. In group 2 the spatial component is included as an independent variable in the hedonic modelling process that provides recognition of the location of the individual properties. Group 3 uses hedonic modelling in much the same way as group 1 (Bourassa et al. 1999; Bourassa et al. 2003), namely to show significant differences between submarkets derived by some other means. In the case of group 1 it is a priori submarkets and in group 3 submarkets derived from the underlying data structures and grouped in terms of their homogeneity by cluster analysis.
Across all groups the most predominately used analytical technique is the hedonic model that has been the foundation of economic real estate research for some time (Watkins 1999, p.159). It has been used to identify the various bundles of property attributes and their contribution to market price (Lancaster 1966). Also, the hedonic model is used to establish differences in submarket groupings, *a priori* or otherwise (Schnare & Struyk 1976). PCA has been used to understand the underlying residential structure of the marketplace and cluster analysis used to formulate similar groups of homogeneity amongst property variables (Bourassa et al. 2003; Maclennan & Tu 1996; Watkins 2001).

### 4.4 Choice of analytical techniques

Of the three groups of techniques summarised in table 4.1, it is both the second and the third that make the least number of assumptions about the location of the spatial submarket boundaries. The objective is to delineate the geospatial submarket boundaries based on the underlying residential living structure of the study area and in doing so recognise the complex and continuous nature of the geospatial submarket surfaces. Therefore the methodology requires appropriate analytical techniques to both define the underlying structure and the appropriate breaks along the continuous surfaces to form the geospatial submarket boundaries. The overall methodology is to approach the problem in two distinct stages, adopting the appropriate analytical techniques in each. These are summarised diagrammatically in figure 4.1 and discussed in more detail as follows.
Figure 4.1  Geospatial submarket delineation process

**STAGE 1**

- **Study area**
  440,000 residential dwellings
  50 variables for each dwelling

- **PCA**
  on 50 variables for 444,000 properties

- **RLS**
  10 principal components
  10 factor scores for each of 440,000 properties giving
  10 surrogate prop chars

**STAGE 2**

- **Sales (7143)**
  July–Sept 2001

- **GWR**
  Step 1 & 2 (Chp. 6)

- **CLUSTER MODELLING**
  Step 3 (Chp. 6)

- **GEOSPATIAL SUBMARKETS**
4.4.1 Stage 1: Principal component analysis

The objective of stage 1 is to describe, as fully and parsimoniously as possible, the real estate geography of the entire study area. A dwelling lies in, and is part of, the real estate geography of the study area and it is this geography, as much as (if not more than) the dwelling itself, the consumers purchase when satisfying their housing requirements (Evans 1995, p.6). The identified need to base submarket identification on the underlying structure of the potential marketplace (Maclennan & Tu 1996, p.389) has historically limited itself to using only samples of a study area described by a limited number of property attributes (Bourassa et al. 1999). This study builds upon this work by creating a RLS that quantifies the spatial and structural data (that are relevant and important in the description of the real estate market) into its underlying dimensions. It achieves this by collecting a wide range of property attributes (described in chapter 3) at the individual property level for each property in the study area. This is described more fully in chapter 5.

The analytical technique used to derive the RLS is principal component analysis (PCA). This technique is a multivariate technique used to understand the underlying dimensions of data by reducing them to a smaller number of independent factors. The use of PCA in this manner is well supported in the submarket literature (Bourassa et al. 1999; Bourassa et al. 2003; Maclennan & Tu 1996; Watkins 2001). It is an important feature of this methodology that the derivation of the RLS deliberately excludes market value of the property from the original data and therefore cannot be considered a market entity. It is simply a description of the structure of the original data. It may be viewed as a residential real estate geography construct that allows the real estate geography to be expressed in terms of its dimensions (principal components) and in a form that can be used in the further
analysis (stage 2) in which these dimensions are expressed in terms of the marketplace, thus deriving the economic entity of the real estate submarket.

The use of PCA has been used by geographers for a long time in the broader field of understanding urban residential structures. Referred to generally as factorial ecology, a large and important body of literature evolved during the 1970s and the 1980s describing the essence of the Australian urban residential structure. The three broad social constructs identified by Shevky and Bell (1955), may be presented as socio economic status, familism and ethnicity, and were generally confirmed in several studies of Australian cities summarised by Burnley (1980, p.199) for Melbourne and Sydney. However, two separate studies of Adelaide in the same summary from the 1966 and 1971 census revealed six constructs. A subsequent study investigating the factorial ecology of the metropolitan area of Adelaide from the 1971 census reveals four constructs ‘Familism/growth’, ‘Socio economic status’, ‘Household structure/urbanisation’, and ‘Ethnic structure’ (Stimson & Cleland 1975, p.189). This study was able to ‘conclusively state’ (p.193) that the characteristics of the residential areas of Adelaide should be understood in terms of at least four dimensions. Stimson poses the question whether or not these four dimensions are just adding to the understanding of the three broad constructs or indeed overriding them. Although carried out some time ago, these studies are amongst the more recent and provide an important insight into the residential living constructs of the study area and provide a basis upon which interpretation of more recent results can be made. They focus on social constructs whereas this study is specifically orientated towards a construct of the residential real estate geography. However, several socio economic variables are common and results should bear similarities. Collecting and identifying appropriate data is important to such constructs and although factorial ecology provides a powerful method of reducing a larger number of housing and socio economic characteristics into a more
manageable number of underlying dimensions, there are a number of limitations explained by Badcock (1984, p.13). He perceives these as falling into three categories. Firstly, the provision of service amenities is generally not included although this may be partially alleviated in this study by including a distance to various services. Secondly, the presence of ‘social networks, group affiliations’ which as Badcock (1984) points out are hard to measure as is the final category including mechanisms for social control and the distribution of power and authority. Not only are they important descriptors of the social system, they also influence the desirability to live in certain locations and hence may be included in a measure of a residential real estate living structure. The question as to how to derive these measures may well be undertaken as part of further research.

4.4.2 Stage 2: Geographically weighted regression and cluster analysis

The dimensions of the RLS derived in stage 1 are used as a set of independent surrogate property characteristics that describe the underlying real estate geography of the study area. It is this total geography that consumers purchase rather than just a dwelling structure. The objective of this stage is to show that the relationship of the RLS to the marketplace describes an economic entity that can be referred to as the geospatial submarket defined for this study. The analytical technique used to establish this relationship is the hedonic regression model using the property transaction price as the dependent variable and the surrogate property characteristics identified in stage 1 as the independent variables. The hedonic model structure is expanded to incorporate the behaviour of the model across geographical space. This is achieved using the geographically weighted regression (GWR) technique that in essence is an hedonic model designed to incorporate the local variability (where it exists) of the independent variables across geographic space. The use of the single hedonic index describing housing quality (Rothenberg et al. 1991, p.360) in a priori spatial units, is substituted by the hierarchical
clustering of patterns of the individual parameter estimates moving across geographic space (price) as submarket identifiers. As Rothenberg et al. (1991, p.62) explain:

The estimated marginal prices from the hedonic regression are used in the behavioral model to provide price information for the estimation of demand and supply, and to provide a basis for the segmentation of the housing market into its component submarkets.

Rothenberg et al. (1991, pp.62, 67) goes on to point out, that this is essentially the procedure suggested by Rosen (1974). The clustering of the similar implicit prices of the surrogate property characteristics indicates a higher degree of substitution between properties within the cluster. This is an important argument that links the analytical techniques used in submarket theory and is explained by Meen (2001, p.31) when he argues:

Areas, which exhibit similar implicit prices of housing characteristics, may be aggregated into sub-markets and used as a basis for analysis, since similarity of price indicates a high degree of substitution between properties within the area.

This study uses the price patterns of the group of individual parameter estimates of the GWR model at the individual regression points as a basis for monitoring changes across geographical space to denote changes in a submarket without assuming any pre existing form of spatial unit. It is the similarity of these price patterns that are the geospatial submarket identifier for the geospatial submarket boundaries.

Each of the parameter estimates represents a continually changing surface across geographic space. The complex interaction of these many continuous surfaces is very difficult to represent in two-dimensional geographic space. To do so would invoke the use of manifold theory that is beyond the scope of this study. Submarkets are a continuous
surface and the question of where along that complex continuous surface is appropriate to define a submarket boundary depends upon the degree of homogeneity (substitutability) required to be present within the submarket. In theory it can be zero (or nearly zero) degree of homogeneity as would be the case if there was only one nominated submarket for the entire study area or perfect homogeneity in the case of nominating a separate submarket for each property within the study area with each property being its own unique market. The answer lies somewhere between these two extremes and may depend on the application for which the submarket are to be used (Bourassa et al., 2003, p.12). In the case of mass appraisal, for example, a submarket size geographically large enough to include the sales evidence required to support the valuation function would be necessary. A practical method for representing the spatial breaks along the submarket’s complex continuous surface would be to cluster the parameter estimate’s price patterns into a nominated number of clusters. This would represent a predefined degree of homogeneity between the regression points (in terms of the parameter estimates) within each of the nominated clusters. An appropriate clustering algorithm is one that maximised internal homogeneity (substitutability) and maximised between group heterogeneity thus preserving the submarket theory and definition that requires maximum similarity (substitutability) of the pattern of the parameter estimates amongst the regression points within submarkets and maximum dissimilarity between submarkets. The details of the cluster algorithm used are given in chapter 6. The representations of the predetermined clusters in two-dimensional geographic space are defined to be geospatial submarkets. This method has not been reported in the literature before and coupled with the use of the principal components of the underlying residential living structure as independent variables again provides also a new approach to the delineation of spatial submarket boundaries.
4.5 Computer software

The methodology employed to analysis the data involved several computer software packages. The software used and the interaction between the packages is summarised in figure 4.2.

The geographical information system (GIS) software provided the spatial environment in which the data in the study were managed. Data sets described in chapter 3 were assigned as attributes of the individual properties with each property given spatial reference through the property cadastre provided by the South Australian Valuer-General. From the GIS environment relevant data was imported into the statistical package (SPSS) for both principal component analysis (PCA) and cluster analysis (CA) after which the results were imported back into the GIS environment for further analysis and the plotting of results. A similar concept was adopted for the geographically weighted regression (GWR) analysis. Input files (described in chapter 3) were required to be in ‘.csv’ format for the GWR software and were created in the spreadsheet environment. The GWR output files were in ‘.e00’ ESRI format and were able to be directly imported into the GIS environment. The
migration of data between the GIS and general statistical software was achieved through the ‘.dbf’ file format structure.

4.6 Verification of the results

This methodology is dependent on the appropriateness of PCA and cluster analysis techniques for the delineation of the geospatial submarket boundaries. An often cited criticism of such techniques is the lack of any objective criteria against which the results can be tested. Tabachnick and Fidell (1989, p.598) suggest that an appropriate test is to look at the ‘sense’ of the result. A good PCA ‘makes sense’ whereas a bad PCA does not. The question then becomes one of testing the ‘sense’ of the result in an appropriate manner. This study has adopted the approach of comparing the results with those of other studies that have investigated similar geographic patterns that, in various ways, describe the underlying residential living structure or residential mosaic. In this regard a body of research that is important to the understanding of the housing submarket structure discusses the urban residential mosaic from a social geography perspective.

The residential location behaviour characteristics are considered by Stimson (1982, p.145) who describes residential preferences in terms of perceived desirability of selected attributes for the Adelaide area. Many of these attributes may have been beneficial to adopt if the scope of the study allowed for their collection. These included indication as to a seaside location and ‘elevated hills location’ that could be translated into the ‘view’ attribute considered important to property value within the study area. Time constraints did not allow for this data attribute to be collected. From an economic perspective, Orford (1999, ch.3) expands the number of locational externalities even further when valuing the built environment through the mechanism of the urban housing market by including
proximity to various other land uses. Randolph (2004) describes the ‘new patterns’ of residential structures that result from pressures that are driving change in Australian cities. They provide a valuable insight into the nature and dimensions of the emerging patterns of these residential structures and will be beneficial in testing the ‘sense’ of the outcomes of stage 1. The study by Baum et al. (2005, ch.3) examines the social mosaic of the Australian capital cities by defining suburbs of ‘advantage and disadvantage’. Also the RLS derived in this study should have some similarity with the underlying social constructs of a city identified by Shevky and Bell (1955) and built upon in terms of the study areas by Stimson and Cleland (1975) to include a broader basis for Adelaide’s typology. A picture of the more general residential mosaic is painted well by Forster (2004, ch.4–5) in which the changing nature of Australian city residential structures are traced and described in terms of their demographic composition and locational behavioural patterns. All of these studies provide a useful framework in which to understand and interpret and test the veracity of the resulting RLS in stage 1 in this study.

4.7 Conclusion

The importance of recognising the appropriate analytical techniques to derive submarkets defined for the study based on sound submarket theory has been achieved by extending work already done in two significant ways. Firstly, the data used involve the whole population and a wide variety of appropriate attribute data collected at the individual property level. Secondly, by reducing these data into their principal components a manageable and statistically meaningful number of underlying residential real estate market dimensions is derived. These can be used as surrogate property characteristics in a hedonic geographically weighted regression model. The changing marginal price of the
parameter estimates defines the complex continuous geospatial submarket surface. The submarket identifier is price.

Adopting a methodology that approaches the task in two stages allows the study to focus in turn on the two important components of submarket delineation. Firstly, to base such delineation on the underlying residential structure of the study and to quantify the dimensions of that structure so as they can be used in the second component as independent surrogate property characteristic variables in a hedonic relationship to derive submarket entities. This satisfies the identified need that submarkets should be an economic unit and based on the underlying residential living structure of the study area. The two-staged approach allows transparency in the methodology. Each of the derived geospatial submarkets can be analysed in terms of the original data that contributed to the formation of the principal components making it easier for the professional user to understand and interpret the outcomes. There is no assumption in any part of the two-stage methodology as to how the results may represent themselves spatially. Any resulting spatial pattern is solely the data talking.

Chapter 5 details the methodology for stage 1 and chapter 6 for stage 2. The proposed methodology approaches spatial submarket delineation from a geospatial perspective without any reference to existing spatial boundaries. It recognises that geospatial submarkets are continuous surfaces in n-dimensional data space that are continually changing in a complex manner defined by the changing patterns of the underlying residential living structure across geographical space. The number of geospatial submarkets depends upon the degree of homogeneity of the price patterns required within the geospatial boundaries which in turn depends upon the user application. The two-stage
methodology is transparent so as to be as understandable as possible to the end user and through them the general public.
CHAPTER 5 METHODOLOGY: STAGE 1

DERIVATION OF THE RESIDENTIAL LIVING STRUCTURE (RLS)

5.1 Introduction

As outlined in previous chapters the overall methodology adopts a two-stage approach. This chapter details the methodology for stage 1, namely the derivation of the residential living structure (RLS) for the study area. The literature identifies a number of issues that, if addressed, would contribute to the delineation of submarket boundaries. They can be summarised as the need to:

- Understand the underlying structure and dimensions of the marketplace in which submarkets reside (Maclennan & Tu 1996, p.389; Meen 2001, ch.5).
- Recognise the overemphasis in existing studies on the individual dwelling characteristics to determine submarket boundaries (Bourassa et al. 2003, p.14; Orford 1999, ch.3).
- Recognise the importance of the submarket identity as being about the market place (Pryce 2004b), thus adopting price as the submarket identifier.

In recognising the importance of relating the submarket derivation to the underlying structure of the study area, it is proposed to derive as much of a complete picture as existing data will allow. This is done by building upon existing work and extending it in the following ways:

- Extending the derivation of the underlying RLS to include all the properties in the study area rather than a sample (Bourassa et al. 1999, p.166). This eliminates any
bias that may be present in the sample and more accurately reflects the existing underlying structure. The resulting surrogate property characteristics that are derived and used in stage 2 will therefore reflect the dimensions of the whole structure.

- Extending the data used in formulating the RLS to include relevant data items from all land-related data sources (socioeconomic, environmental, and construction dwelling characteristics) as described in chapter 3.

- Deliberately excluding a market value attribute, leaving the resulting RLS describing only the property together with its externalities. Therefore the resulting RLS cannot be considered to represent submarket segments as in other studies (Bourassa et al. 1999; Bourassa et al. 2003), but rather only dimensions of the underlying structure.

- Providing a basis for further research where the market value is introduced as the dependent variable in a geographically weighted regression model with the resulting components from the RLS used as a more comprehensive and independent set of surrogate property characteristics to investigate the spatial boundaries of submarkets. Change of significance in the model over geographic space will be used to indicate potential spatial submarket boundaries in stage 2.

The methodology adopted was developed from previous work (Bourassa et al. 1999; Bourassa et al. 2003; Maclennan & Tu 1996; Watkins 2001) in which PCA is used to quantify the underlying structures of the data. The terms factor analysis and principal component analysis are often used interchangeably. Hair et al. (1992, pp.224–5) describes factor analysis as ‘a generic name given to a class of multivariate statistical methods whose primary purpose is data reduction and summarization’. Two of these methods are common factor analysis and component analysis. Principally the difference lies in the type of variance used in their determination. Common factor analysis uses only the ‘common variance’ defined as ‘that variance in a variable that is shared with all other variables in the analysis’, whereas component analysis uses the total variance which includes common variance plus specific variance associated with only a specific variable and error variance associated with the unreliability in the data collection (Hair et al. 1992, p.230). This study
adopts the principal component analysis (PCA) as it is conceptually less complex and in reality the solutions ‘differ little’ in situations with data sets being greater than 30 variables (Field 2005, p.631). Importantly, the use of PCA follows the previous work in the field as discussed above.

This study builds on the use of PCA by applying it to all properties in the study area and incorporating as many available variables as are appropriate and are available into the analysis. By including the total population, and not just a sample, the resulting constructs can be clearly thought of as those representing the whole study area without relying on the extrapolation from a sample that may not be valid (Field 2005, p.629). This may be important because this study then takes the further step of calculating the factor scores for each factor for each property in the study area and using these as surrogate property characteristics in hedonic modelling in stage 2 which is across the total population.

5.2 Analytical steps

PCA is used in this research to reduce a large number of observed variables to a smaller number of principal components or factors. This is achieved by following a methodology set out by Hair et al. (1992, p.227) in which seven broad steps are clearly identified culminating in a set of factor scores that summarise the underlying residential living structure of the study area.

These procedural steps are summarised in figure 5.1. It is proposed to follow these seven steps as a sequence of analysis in the methodology for stage 1. Each step of the analysis is described below.
Step 1: The research problem

The research problem in stage 1 of this study is to quantify the underlying nature of the residential living structure (real estate geography) of the study area. Each property has a large number of attributes associated with it including structural, socio-economic, environmental, and proximity attributes, all of which play a role in determining the type of real estate market in which a particular property may participate.

Therefore, the variables included in the construction of RLS were selected on the basis of their contribution to the market value of property as discussed in chapter 3. They were further examined for inclusion in the factor solution by examining their loadings on the
resulting factors in terms of their importance to the research question and their communality index. If a variable does not significantly load on any factor then it should be examined and perhaps rejected if it does not contribute to the research question or has a low communality score (Hair et al. 1992, p.241) as was discussed in chapter 3 regarding the variable describing the dwelling age.

This resulted in 50 variables being used in this research. These variables were taken as standardised continuous variables where possible. Dummy variables (coded 0 or 1) were formed where their nature was not of a continuous form. Examples of this include dwelling type and construction, as discussed in chapter 3. The important question of the sample size adequacy does not arise in this study as it is the total population that is being used.

**Step 2: The correlation matrix**

The nature of this research is to investigate underlying data dimensions in a large set of variables and therefore it is the correlation between the variables that is of interest. This is achieved using the ‘R’ factor analysis as described by Hair et al. (1992, p.228).

**Step 3: The factor model**

Component analysis is adopted as the resulting factors are to be used in later analysis and the minimum number of factors is required to account for the maximum portion of the variance in the original data.

**Step 4: Extraction method**

To obtain a simpler and more meaningful solution, rotation of the factors is essential (Hair et al. 1992, p.234; Kline 1994, p.174). Rotating factors changes the factor loadings and therefore the meaning, but explains the same amount of variance in the original variables (Kline 1994, p.61). Apart from the varimax orthogonal rotation, which by definition
preserves the independence of the resulting components, there are the oblique rotations that allow varying degrees of correlation between the components, often leading to a more ‘sensible’ interpretation. The disadvantage of this rotational technique is that independence between the components is compromised which, if they are to be used as variables in further analysis, may be of concern. The advantage is that the correlation allowed may improve the interpretability of the components as in real life such correlation does exist and therefore some argue that orthogonal rotation should never be used for any data involving human behaviour (Field 2005, p.637). If the resulting components are to be used as independent variables in further regression analysis a compromise between allowing some correlation, giving a more realistic component interpretation, and a loss in absolute component independence, may have to be reached. In this study oblique rotation is investigated to determine if it provides a more sensible interpretation of the components with minimal correlation so as not to invalidate their use in future research. Tabachnick and Fidell (1989, p.637) suggest the better way of deciding between the two rotation methodologies is to attempt both and investigate the correlation between the factors in the case of the oblique rotation. The methodology adopted in this study will be to use both types of rotation to check that no underlying correlated patterns in the original data are missed.

Step 5: Number of factors

A combination of methods is used to determine the optimum number of factors to be selected as there appears to be no definite rule of selection, as noted by Watson (1998, p.1362):

Admittedly there is no approved method of selecting the number of factors to be rotated from principal component analysis.
The number of adopted components may be chosen based on their respective Eigen value and the scree plot interpretation. Eigen values greater than 1 indicate components that contribute more than single variables. A scree plot displays the Eigen value (y-axis) and the associated components in order of extraction (on the x-axis) and indicates the relative importance of each component (larger Eigen values having more importance). The point of inflexion may assist in indicating an appropriate cut-off in the number of components to be selected. It is suggested these two criteria should be considered together (Field 2005; Hair et al. 1992). An aspect of appropriateness in the choice of original variables for the number of factors extracted is demonstrated by the degree of communality shown by the original variables. Hair et al. (1992, p.224) define communality as being ‘The amount of variance an original variable shares with all other variables included in the analysis’. Therefore the closer to 1.0 the communality value is the better the resulting factors are at explaining the original data (Field 2005, p.634).

Step 6: Factor interpretation

Once a factor structure has been derived, the interpretation of factors in terms of their contributing variables is a subjective exercise that requires an understanding of the underlying data and research objectives. However, the literature offers some basic principles as guidelines, although the question as to what constitutes a significant contribution is less clear. The factor loadings are a measure of the relationship between the variable and the derived factor. The ‘rule of thumb’ is that all variables that load on a particular factor with a value of 0.3 or more are investigated for a common theme (Tabachnick & Fidell 1989, p.639). This is expanded on by Hair et al. (1992, pp.239–40)
who introduce the effect of sample size and various significance levels and summarise the interpretation in three basic guidelines:

(1) the larger the sample size, the smaller the loading to be considered significant; (2) the larger the number of variables being analysed, the smaller the loadings to be considered significant; (3) the larger the number of factors, the larger the size of the loading on later factors to be considered significant for interpretation.

As this study is large, factor loading below 0.30 may be considered significant, at least in the first few factors, although this may be higher depending on the number of factors being extracted. However, only one loading on any factor for each variable should be recognised (Hair et al. 1992, p.249) with at least three variables loading on each factor and the cleanest factor structure perhaps being the most appropriate (Costello & Osborne 2005, p.3).

In the case of oblique rotation, factor interpretation should be derived from both the structure matrix and the pattern matrix output. Because the oblique rotation technique allows correlation between the factors the correlations between the variables and the factors is given in the ‘structure matrix’ whereas the loading in the ‘pattern matrix’ represent ‘a measure of the unique relationship between the factor and the variable’ (Tabachnick & Fidell 1989, p.639).

From the above discussion the basic ‘rule of thumb’ of 0.30 was adjusted for the size of the study and the number of factors, and a cut-off factor loading of the original variable and the factor of 0.40 was adopted for this study and only one loading included on any factor for each variable, where possible, in the interpretation of the factors.
Step 7: Factor scores

There are many procedures for determining factor scores (Tabachnick & Fidell 1989, p.641). This study adopts the regression approach used in the statistical package SPSS and described by Field (2005, pp.626–7). Factor scores are assigned to each property in the study area by multiplying the factor score coefficient for each variable by the value of that variable for the property concerned and summing the results. This gives a value for each component for each property in the study area based on the general linear form shown in equation 5.1.

Equation 5.1

\[ Y_i = b_1X_1 + b_2X_2 + \ldots + b_nX_n + \text{error} \]

where:
- \( Y_i \) is the \( i \)th (1 to 10) component
- \( b_j \) is the factor score coefficient
- \( X_i \) is the original variable

The factor score coefficient is determined by dividing the factor loadings by the correlation coefficients. This gives a purer measure of the relationship between the variable and the factors than using the factor loadings alone.

Each of the 10 factor scores are then weighted by the amount of variance (after rotation) explained by each factor. This gives recognition as to their relative importance in explaining the variation amongst the original variables and hence their contribution to the RLS construct.

The weighted scores are then added to the spatial database within the GIS environment and each factor score for each property is plotted in one half standard deviation groups from the mean and represented on a colour ramp plot.

5.2.1 Limitations
Although the literature supports the use of PCA as an appropriate technique for exploring the dimensions of data generally (Hair et al. 1992) and residential dwelling data in particular (Bourassa et al. 1999; Maclennan & Tu 1996; Watkins 1999), PCA is not a precise technique and has documented limitations. Three main problems associated with this analytical technique, as discussed by Tabachnick and Fidell (1989, p.598) are summarised in italics as follows:

1 *The solution does not have a criterion against which it can be tested.* ‘A good PCA or FA (factor analysis) “makes sense”; a bad one does not’.

In this regard one method of testing the ‘sense’ of the results is by mapping the output component scores, which in this study is calculated and plotted on an individual property basis within a geographic information system (GIS) environment and comparing its’ ‘sense’ with similar studies.

2 *After extraction of the factors [components] there are an infinite number of rotations available to assist in the factor interpretation ‘all accounting for the same variation in the original data, but with factors defined slightly differently’.*

The methodology presented in this study will include two broad forms of rotation and compared, as discussed earlier, for the most appropriate to use.

3 *PCA can be used to cover poorly conceived research due to its capacity to produce an ordered set of results from poorly collected and presented data.*

This places an important emphasis on the alignment of the research objectives with the data collection and preparation phase of the research. Often PCA involves data collected by various sampling techniques. Questions as to the appropriateness or otherwise of these techniques does not arise in this case as the whole population is used in the analysis.

A summary of the criteria used to assess the results in the derivation of the RLS using PCA may be as follows:

- Both orthogonal and oblique rotations should be explored using social science data as there may be hidden structures due to the correlated nature of the data.
The criteria upon which the selection of the number of factors to be extracted should include the investigation of the following:

- scree plot
- Eigen value greater than 1
- communalities of the underlying data
- sense of factor interpretation.

Factor loadings considered significant of 0.4 and above have been adopted in this study together with only one loading on any factor for each variable.

5.3 Conclusion

The objective in stage 1 is to contribute to the field of submarket boundary delineation by basing it on a better understanding of the underlying residential structure of the study area. This structure relies on a wide variety of data collected to represent as many facets of property as appropriate. The methodology used to achieve this objective is PCA with the derived linearly independent factors representing the underlying residential living structure of the whole study area. These factors can be regarded as surrogate property characteristics and used in further analysis to test for submarket boundaries as set out in stage 2.

The use of PCA is not an exact science. Judgements have to be made by the researcher throughout the analysis ranging from the appropriate data to be used, to the appropriate rotation and interpretation methods to be employed. This will always be a caveat on the final results. However, subject to this, if the analyst is faithful to the research objectives and the results ‘make sense’ then the methodology can provide valuable insights and significantly contribute to the research objectives.

The increased availability of relevant data describing the underlying real estate geography may, paradoxically, present a problem. Hedonic modelling incorporating too many variables becomes cumbersome and subject to multicollinearity. The question as to how
best to accurately portray a description of a complex construct such as the real estate geography can therefore perhaps be best answered using PCA as described in this methodology as providing a manageable number of statistically independent surrogate characteristics representing a large number of variables describing the complex underlying residential real estate geography of the study area.
CHAPTER 6 METHODOLOGY: STAGE 2
DERIVATION OF GEOSPATIAL RESIDENTIAL REAL ESTATE
SUBMARKET BOUNDARIES

6.1 Introduction

The objective in this stage of the methodology is to relate the residential living structure (RLS) derived in stage 1 to the market place. For this study, a geospatial submarket was defined in chapter 2 as being the change in the market’s value of the underlying dimensions of the RLS across geographic space. The recognition of submarkets as an economic entity is achieved by constructing a hedonic model that relates the underlying residential structure to the market place. This provides the market’s interpretation of the RLS at a particular point in time and allows the resulting submarket delineation to be given the status of an economic entity which, in this study, is defined as a geospatial residential submarket.

The basis of the methodology in this stage is to use the surrogate property characteristics from stage 1 as independent variables in a geographically weighted regression (GWR) model with the sale price of the property as the dependent variable. The observed change in the price pattern of the independent variable parameter estimates across geographic space indicates change in market’s influence, and hence, by the definition adopted in this study, indicates a different geospatial submarket.
The analytical technique adopted to observe this change is geographically weighted regression (GWR) (Fotheringham et al. 2002). GWR makes no assumptions as to spatial variation of the independent variables and is based on the ability of a series of local regression models to capture the change in the price patterns of those variables across space. This change is a complex continuous surface comprising the interwoven individual continuous surfaces of each of the individual independent variables. A 300 metre regular grid is established creating 9075 regression points evenly distributed across the study area. At each point a hedonic parameter estimate is calculated for each of the surrogate property characteristics including an intercept value. Together, these constitute a price pattern that is the market’s interpretation of the underlying RLS at each regression point. The essence of this thesis is that the clustering of the similarity of the price patterns constitutes geospatial submarkets. The methodology therefore is to cluster these parameter estimates based on their similarity of patterns into a predetermined number of clusters which then can be considered the geospatial submarkets of the study area. The cluster technique used is Ward’s hierarchical algorithm. The number of clusters requested can be varied depending on the use the resulting submarket groups are to be put. Each of these techniques is further discussed in detail.

6.2 Data sets

6.2.1 Property transaction price

The transaction price is the fundamental data set used in this stage of the analysis and was described in chapter 3. It contains the sale price of those properties, across the whole study area, which sold between July and September 2001 and are considered to represent the market at the date of the study, namely August 2001. The factor scores for each factor, for each of these properties, derived in stage 1, were added and this data set and was then
converted to a ‘.csv’ file format and used as input to a GWR model. The factor scores were weighted by the amount of variance each factor explained in the formation of the RLS in stage 1. These 10 weighted factor scores are standardised and called surrogate property characteristics and are used as the independent variables in the GWR modelling. Two GWR models are constructed, one performing local regressions at the data points (sale property locations generating 7143 regression points, one for each sale location), and a second performing the local regressions over a regular 300 metre grid generating 9075 regression points. Information from the first GWR model is needed to calibrate the second GWR model. Each of these models generated its own output and associated GIS data sets in the ESRI ‘.e00’ format summarised as follows.

The first data set, where the local regression points are the same as the data points, have the following output fields:

- values of the estimates of the parameters at each regression point
- values of the estimates of the standard errors of the parameters at each regression point
- pseudo-t values
- observed dependent variable value
- predicted dependent variable value
- unstandardised residual
- HAT statistic
- standardised residual
- cook’s distance statistic
- pseudo-R² values.

The second data set, where the local regression points are the regular 300 metre grid across the whole study area, have the same first three fields as above. The other fields are not
relevant as there is no local data point upon which to make the necessary calculations. In addition there is a unique identifier field added to this data set.

6.2.2 Regular 300 metre grid
A regular 300 metre grid is constructed, within the GIS environment, across the study area together with their respective ‘X’ and ‘Y’ coordinate values and associated unique identifier of each grid point and is written to a ‘.csv’ file format for use in the second GWR model. There are 9075 such points in this data set, representing the 300 meter regular grid.

6.3 Analysis techniques
There are two analysis techniques employed to delineate the spatial submarket boundaries in this stage of the methodology. They are geographically weighted regression and cluster analysis. Each is discussed in turn.

6.3.1 Geographically weighted regression (GWR)
The use of GWR is a relatively new technique in the housing market literature. However, it has intuitive appeal as it specifically addresses the question of spatial non-stationarity among the independent variables in the regression model. The essence of geographically weighted regression methodology (Fotheringham et al. 2002) is that it does not assume spatial stationarity of the independent variables and allows local variation to be captured where it exists. This is exactly the reason for using it in this study; indeed it is the central part of the analysis as it is the difference in the local variation across space which forms the basis to the definition of the geospatial submarket boundaries in this study allowing their delineation without reference to any existing spatial boundary. At each regression point across space the data are geographically weighted to give relevance to their position in geographic space.
The weighting can be flexible in terms of its functional form. One functional form would be a Gaussian Function shown in figure 6.1, but a bi-square weight function is also used.

### Equation 6.1. Weighting function in the Gaussian kernel

\[ w_{ij} = \exp(-1/2(d_{ij}/b)^2) \]

where: \( d_{ij} \) is the distance of the data point from the regression point and \( b \) is the bandwidth

Source: (Fotheringham et al. 2002, p.45).

### Figure 6.1 A spatial kernel

In situations where the distribution of the data points varies across space the bandwidth can be allowed to adapt from one geographical location to the next, depending on the density of data points, so as to include an optimum number of cases. When the GWR model moves across a regular grid where the regression points are not the data points the optimum bandwidth which gives the optimum number of nearest neighbours used in the adaptive kernel is calculated by minimising the Akaike Information Criteria (AIC) statistic. This statistic is derived using the GWR software using the form described in equation 6.2.
Chapter 6 – Methodology: Stage 2 – Delineation of Geospatial Submarket Boundaries

**Equation 6.2**

\[
\text{AIC} = 2n \log_e(\hat{\sigma}^2) + \log_e(2\pi) + n \left( \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right)
\]

where:
- \(n\) is the number of observations and
- \(\hat{\sigma}\) is the estimated standard deviation of the error term
- \(\text{tr}(S)\) is the trace of the HAT matrix of the GWR.

Source: (Fotheringham et al. 2002, p.61).

This optimisation process not only takes into account the goodness of fit of any given model but also penalises models for having a greater number of parameters so as to favour models that are parsimonious and also fit the data well.

This study adopts the “adaptive” kernel methodology using a bi-square weighting function.

The general GWR model may be represented by the following form set out in equation 6.3.

**Equation 6.3**

\[
\text{Sale Price} = \beta_0 (u,v) + \beta_1 (F_1)(u,v) + \ldots + \beta_n (F_n)(u,v) + \text{error}(u,v)
\]

where:
- \((u,v)\) are the location coordinates of the sale
- \(\beta_0 \ldots \beta_n\) are the parameter estimates
- \(F_1 \ldots F_n\) are the surrogate property characteristics (principal components from stage 1).

Equation 6.3 is recalibrated at each point on a regular 300 metre grid set up over the study area yielding 9075 sets of regression coefficients (\(\beta_1 \ldots \beta_n\) in equation 6.3). These, together with the intercept \(\beta_0\) for each of the regression points, are the variables used in the clustering algorithm (in standardised form) and together constitute the submarket identifiers in this study.

The GWR methodology followed in this study performs the regressions over a regular grid rather than at the data points as the resulting parameter estimates are estimates of a continuous surface that can either be approximated by an interpolated grid or a recalibration of the GWR model in a uniform manner on a regular grid over the entire study area. It is considered more realistic to rely on the recalibration than on an
interpolated surface in this study but is a topic that requires further research. A regular 300 metre grid was considered appropriate for this study as the accuracy of the data may not warrant a higher density grid.

6.3.2 Cluster analysis (CA)

Cluster analysis is then used to group together the submarket identifiers into appropriate homogeneous geospatial submarkets. There are many different types of clustering algorithms and the important question of choosing the most appropriate for the task must be based on an understanding of their differences. The submarket theory suggests that areas of similarity are areas of homogeneity in which consumers are indifferent to choice based on price. Therefore, the within cluster homogeneity needs to be maximised while the between cluster heterogeneity needs to be maximised. In this case the price parameter estimates show the market’s value of each RLS dimension and therefore the closer the price pattern amongst the regression points within a cluster the more homogeneous the market’s interpretation of the RLS. The most appropriate clustering algorithm for achieving this was considered to be Ward’s hierarchical method because it minimises the within group variance and maximises the between group variance closely simulating the submarket theory.

Figure 6.2 shows the various type of distance measure used in cluster algorithms. The appropriateness can be seen from the concept diagrams in figure 6.2 with Ward’s method being the one that minimises the within cluster variance as well as maximising the between cluster variance whereas the other concentrate on between cluster distance only. Ward’s method is also supported in the submarket literature as being appropriate (Bourassa et al. 1999, p.170).
The important aspect of Ward’s approach is that it minimises the within cluster sum of squares over all clusters formed at any particular stage in the agglomeration process. And the distance between two clusters is the sum of squares summed, in this case, over all the 11 parameter estimates. It is therefore important to take the standardised score of each parameter estimate so as to eliminate ‘spurious effects’ resulting from unequal variances of the parameter estimates (Hair et al. 1992, p.271).

A general form, together within the derivation of the algorithm is given by Anderberg, (1973, p.143) and summarised in equation 6.4.

**Equation 6.4**

\[
\Delta E_{pq} = \frac{m_p m_q}{m_p + m_q} \sum_{i=1}^{i=n} (x_{ip} - \bar{m}_p)^2
\]

where:

- \(\Delta E_{pq}\) is the minimum increase in the squared Euclidean distance between the centroids of the merged clusters
- \(m_p\) and \(m_q\) are the next two clusters to be merged
- \(x_i\) is the score on the \(i\)th of \(n\) variables
The algorithm clusters the like patterns of the regression coefficients as shown conceptually in figure 6.3. For each resulting cluster the factors with strongest market
influence on the formation of the cluster can be seen in both the magnitude and direction of the standardised mean value of each of the parameter estimates. The degree of homogeneity within the cluster can be measured using a methodology developed and described later in this chapter. These are, by definition, distinctive for each cluster and may be described as the geospatial submarket signature patterns that can be examined to provide an understanding of the structure of each cluster or geospatial submarket group. To further add meaning and transparency to the structure of the cluster, each of the independent variables that most influence the cluster formation can be expressed in terms of the original 50 variables collected on each property in stage 1. The analytical steps adopted in this study are individually described in the next section.

**Figure 6.3 Cluster signatures**

<table>
<thead>
<tr>
<th>PE</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>F2</td>
</tr>
</tbody>
</table>

**Note:** Cluster 1 & 2 lines (cluster signatures) represent examples of groups of regression points with similar price patterns. PE SCORE is the mean parameter estimate score for each principal component (including the intercept - see equation 6.3) of the group of similar regression points.

### 6.3.3 Sequence of analytical steps

Three broad steps have been derived in this study as a methodology for the delineation of the geospatial submarket clusters. Each is discussed in turn and summarised in diagrammatic form in figure 6.4.
Step 1 Preliminary analysis

The model is calibrated using GWR 3.0 software (Fotheringham et al. 2002). The GWR software used in this analysis calibrates a regression model at each data point or alternatively over a nominated regular grid. The methodology adopted in this study is running the GWR (using an adaptive kernel) across a regular 300 metre grid covering the whole of the study area. However, before this is done, preliminary analysis needs to be undertaken in order to specify the GWR model for step 2. This needs to be undertaken by
calibrating the regression model at each data point as the required statistical output needs
known regression point dependent variable data. They are discussed as follows:

1. The stationarity aspects of the GWR model needs to be understood. If the model is not
influenced by location, the surrogate property characteristics are not spatially
dependent and therefore will not show geospatial submarket behaviour. The
methodology used in this study adopts three measures of difference between the ‘local’
and ‘global’ models. Firstly, a visual comparison between the standardised residuals of
the ‘local’ and the ‘global’ models. Secondly, a comparison of the AIC statistic in the
‘local’ and ‘global’ model and finally the Monte Carlo test for non-stationarity,
provided as part of the GWR software, output is used. In essence the Monte Carlo test
establishes whether the “local” model better describes the relationships in the data than
does the “global” model. The “p-value” expresses the significance as to whether the
variation in the parameter estimates are smaller in the “local” model than in the
“global” model and hence shows if the GWR model is stationary or not.

2. Any outliers that may exist need to be identified and excluded from the data set if
appropriate. The failure to predict statistic (FP) and standardised residuals greater than
+3 or less than -3 standard deviations from the mean and whose influence statistics
(both HAT and COOK’s D) were above their respective critical values discussed later
in this chapter, are identified and assessed as to the reason they are outliers.

3. The identification of ‘Hot Spot’ locations that may exist, indicating geographic areas
for further investigation, is important as this allows the opportunity to include
information about specific locations that may have been missed when deriving the
residential living structure (RLS) in stage 1. The methodology for determining the ‘Hot
Spot locations’ adopted in this study is described as follows:

- Using the same statistics as in point 2 above and having removed outliers due
to bad data, these statistics were again used assuming outliers were this time
due to missing information. The aim was therefore to identify the locations of
an arbitrary concentration of these outliers.

- A raster surface was determined for each of the statistics in point 2 above and
a kernel density function weighted by the outlier statistic being investigated
using a broad search area calculated for each of the residual statistics. The
kernel density function weights nearby events more heavily than distant ones
emphasising the importance of location (O’Sullivan & Unwin 2003, p.87). This is achieved in the GIS environment. The resulting density measures were standardised and added together with the highest density locations (arbitrarily chosen to be greater than three standard deviations) plotted as a single ‘Hot Spot’ indicator. The locations can then be investigated for possible reasons as to why they should present as locations of unusual influence. Because this procedure is experimental, only extreme locations were identified as being areas of concern and further investigation. The criterion used was therefore only to identify those locations where there was visually more than one contiguous point. The ability to identify potential areas for further investigation is important. Further research is needed to enhance this aspect of the methodology, and it is proposed to investigate the development of this aspect with the end users of the methodology.

4. For the sake of a potentially more parsimonious model, the parameter estimates that do not significantly influence the global model may be tested for their influence on the “local” model.

5. The optimum bandwidth for the GWR model for the given data set can only be derived by running the model with the regression points the same as the data points so as to apply the algorithm given in equation 6.2 that requires the trace of the HAT matrix which can only be calculated at a regression point with a known dependent variable value (where regression points are equal to data points).

These above investigations can only be achieved by first running the model using the data points as regression points so that residuals may be determined. The output containing these results is imported into the GIS environment where the residuals can be analysed. From these results the following three types of residuals are examined (including two influence statistics).

1. **Failure to predict statistic (FP).** This is calculated as the ratio of the actual dependent variable value to the predicted dependent variable value and is calculated as follows for
each of the local regression points and can be interpreted as a measure of the local model’s failure to predict (FP) as shown in equation 6.5.

**Equation 6.5**

\[
FP = \frac{\text{actual selling PRICE}}{\text{predicted selling PRICE}}
\]

A value of 1.0 indicates a perfect prediction, whereas a value greater than 1.0 would indicate an under prediction and conversely value less than 1.0 would indicate an over prediction of the selling price. Plotting where these variations occur may help indicate some local phenomena that have not been accounted for.

2 *Influence statistics.* These two statistics are produced as a standard part of GWR v3.0 software output.

- **HAT statistic** (h_{ij}). This is the leading diagonal in the HAT matrix S and is defined in the following manner with ‘y’ the dependent variable

  \[
  \hat{y} = Sy
  \]

  as described by Fotheringham et al. (2002, p.55) and can be interpreted as the contribution of the independent variables to the overall model performance.

- **COOK’S Distance statistic** (D_i) measures the influence of the combined effect of the dependent and independent values to the overall model performance.

It is considered a common criterion for the h_{ij} statistic as exhibiting too much influence of the independent variables to the overall picture as being when:

h_{ij} > 2(k+1)/n,

and from the D_i statistic values when:

D_i > 4/(n-k-1)

where:

k is the effective number of parameters given by the sum of the HAT matrix diagonal given in the GWR output file (Washington University 2007).
3 *Standardised residual*. A caveat is applied to the interpretation of this statistic as a local statistic as there is a different standard error at each regression point, so residuals larger than -3 or +3 are considered to be unusual (Fotheringham et al. 2002, p.215). From the results of these investigations the GWR model can be specified for step 2.

**Step 2 Calibration of the GWR model over a regular 300 metre grid**

Having obtained the GWR results from the initial run using regression points as data points the GWR model is rerun using regression points to be a regular 300 metre grid with an adaptive kernel and a bandwidth derived in step 1. The results are a set of parameter estimates, one for each independent variable. The GWR generates output in an ESRI format (.e00 files), that can be incorporated as attributes in the GIS environment.

This provides an opportunity to investigate the sense of the results by individually interpolating the parameter estimates across the study area. A standard inverse distance weighted (IDW) algorithm is used. These interpolations can be interpreted as the changing influence each independent variable (surrogate property characteristic) has on price across the study area. Each of the parameter estimates are individually plotted to visualise the variation in influence of the various factors as individual components.

**Step 3 Derivation of geospatial submarket boundaries**

The objective of this step is to represent the complex interaction of the surrogate property characteristic surfaces in two-dimensional geographical space as geospatial submarket boundaries. Cluster analysis is used to group the like marginal hedonic values (parameter estimates) for each surrogate property characteristic across all cases giving similar pattern groupings of the values within each resulting cluster. The agglomeration schedule for the Ward’s hierarchical clustering algorithm describes each stage of the clustering process. The algorithm starts with as many clusters as there are points. It progressively groups...
together either previous groups or single points that give a minimum increase in the overall squared Euclidean distance between the cluster centroids, eventually ending up with only the one cluster. At each merger the measure of the clustering coefficient is given as the squared Euclidean distance between the two cases or clusters being merged (Hair et al. 1992, p.281). The difference in this value from one stage to the next is examined to give an indication as to when abnormally high increases take place and optimum clusters may be reached for the particular data set. This is sometimes referred to as ‘error variability measures’ by Hair et al. (1992, p.270) and is used as part of the methodology for investigating the appropriate numbers of clusters to adopt.

Although there are guidelines as to the number of clusters that should be formed, no definitive selection procedures exist. Amongst these guidelines is to examine some ‘intuitive conceptual or theoretical relationship’ as being able to suggest an appropriate number of clusters (Hair et al. 1992, p.279). As there is more than one use for these clusters it may suggest that more than one set of cluster solutions should be adopted. This is in keeping with the theory of submarkets as being a continuous surface to be broken into the two major uses identified in chapter 2, namely the planning function and the mass appraisal function. Each function demands different levels of homogeneity within the clusters and each will be consulted to ‘intuitively’ and ‘theoretically’ impose the number of clusters to be formed. This is supported by Bourassa et al. (2003, p.12) when they assert that the appropriate submarket definition depends on the use to which it is put. However, to assist the user in making the appropriate choice of cluster numbers, it is an important part of this methodology to provide an understanding of the major influences of price in the various cluster solutions that may be chosen, which is to provide an understanding of the geospatial submarket structure.
6.4 Understanding the structure of geospatial submarkets

After selecting the appropriate number of clusters, and after each of the 9075 regression points has been assigned a cluster number from the Ward’s hierarchical agglomeration clustering process, these results are spatially joined to a regular 300 metre poly grid structure within the GIS environment so that for each of the 9075 points there is a 300 metre polygon each with the point cluster number (derived using Ward’s cluster algorithm in SPSS) attribute assigned. These polygons are then dissolved on their cluster number and geographically plotted. When dissolving, non contiguous polygons with the same cluster number are allowed. In effect, this permits non-contiguous geospatial submarkets to be formed that may be considered legitimate for some applications (Costello 2004 p.359; Rothenberg et al. 1991, pp.63–64) but not for others (Pryce & Evans 2007, p.12). The selection of the appropriate number of geospatial submarkets for a particular application is an important part of the methodology and should be made by the user based on specific professional requirements. It is therefore important to give the user as much understanding of the structure of particular cluster (geospatial submarket) solutions as possible in order to make an informed decision. This methodology assists the user by plotting the average standardised parameter estimates of each of the independent variables in the hedonic GWR model for all the regression points that were included in the particular cluster. This indicates which of the variables were of influence in determining price for the particular cluster. The standard deviations from the mean are also displayed for each of the independent variables. This provides the basis for indicating the homogeneity between the regression points included in the cluster. Because the independent variables are orthogonal factors derived in stage 1, a homogeneity measure (HM) that measures the variability of the regression points within a cluster may be interpreted from a calculation of the standard error as shown in equation 6.6.
Chapter 6 – Methodology: Stage 2 – Delineation of Geospatial Submarket Boundaries

Equation 6.6

$$HM_{(n,s)} = \sqrt{\frac{\sum_i \text{var}(\beta_i)}{i}}$$

HM is the homogeneity measure in a particular cluster (n) in a particular predetermined cluster solution (s), where i is the number of regression coefficients at each regression point (which in this particular study is equal to 11, being made up of the intercept and 10 independent variables) and

$\beta_i = \text{particular average regression coefficient for a given cluster.}$

Note: $\Sigma \text{var}(\beta_i)$ is the total (between grid reference) cluster variance (because $\beta_i$ are based on principal components)

The lower the resulting value of HM, the more homogenous a particular cluster is in terms of the variability of the regression points within a cluster. As the coefficients at the regression points represent price, it is a homogeneity of price of the underlying real estate geography that defines the geospatial submarket. It is not a single price but the price pattern of the individual components that forms the definition.

The plots together with the HM are considered the ‘signature pattern’ of the cluster or geospatial submarket and describe its nature in terms of the underlying dimensions of the RLS which are influencing the market shown by the degree of similarity amongst the parameter estimates in this particular geospatial submarket. The question as to what HM value constitutes an acceptable degree of homogeneity will depend on the application and will have to be ‘learnt’ through trial and error by the user. The signature pattern gives the user a ‘sense’ of what is driving the market and how this changes across geographic space.

The methodology adopted in this study is to involve two principal users in assessing the results of this study for their particular applications. They were described in chapter 2 as the property taxation and the planning policy functions. Each of these two major functions were consulted as to the appropriateness of the methodology, and the resulting geospatial submarkets delineations, to their respective functions. Their feedback forms and important part of the results and, although preliminary at this point in time, are outlined is in chapter 8.
6.5 Conclusion

The essence of the thesis that has been captured in the methodology has two important elements. Firstly, the methodology described recognises that geospatial submarkets are a continuous surface generated by the complex interaction of variables in n-dimensional data space. The resulting two-dimensional cluster boundaries are not a choroplethic solution as the two-dimensional boundaries may suggest. The use of such boundaries merely gives a practical way to view a boundary of a particular minimum degree of homogeneity in two-dimensional geographic space. The submarket surface is not static across a cluster surface but continually changing down into smaller and smaller groups until eventually each property is its own market. This is impractical for most applications and therefore a compromise, or a degree of tolerance of a certain amount of heterogeneity, has to be accommodated and that degree will depend on the application. The methodology outlined in this chapter adheres to essential submarket theory as adopted in this study and provides the flexibility to provide cluster solutions for different applications. The interpretation of the results is guided by the ‘sense’ of the result to the professional user and this is enhanced by the transparency the methodology provides through the generation of the ‘cluster signatures’, the associated coefficients of variation within the cluster, and the ability to trace the influence on price of the underlying real estate geography to the original 50 variables collected for each of the properties in the study area. Secondly, the geospatial submarket identifier is price. In particular it is the similarity in the price pattern of the parameter estimates calibrated at 9075 regular grid points across the study area. The degree of similarity is the homogeneity of the price patterns amongst the regression points within the cluster group. This cluster group is defined as the geospatial submarket. An experimental ‘Hot Spot’ analysis component has been included to enhance the quality and efficacy of the results. The success of the methodology is measured, at least in part, by the
confidence the end users have in the results and the ‘Hot Spot’ analysis should contribute to this. Chapters 7 and 8 detail and discuss the results of the application of the two-stage methodology outlined in chapters 4, 5 and 6.
CHAPTER 7  RESULTS OF STAGE 1: DERIVATION OF THE RLS

7.1 Introduction

The aim of this chapter is to present the resulting residential living structure (RLS) derived from the seven-step methodology outlined in chapter 5. This was achieved by expressing the dimensions of the RLS as the principal components of the original 50 variables (described in chapter 3) collected for each of the approximate 440,000 properties in the study area. An often-stated criticism of the use of PCA is the lack of ability to test the veracity of the results (Tabachnick & Fidell 1989, p.598). The results presented in this chapter attempt to address this criticism by mapping the results and examining the ‘sense’ of the spatial patterns of the resulting components. This is achieved by comparing, where appropriate, the resulting spatial patterns with other studies that have examined various socio economic and demographic trends to have come out of analysis of census data for the study area, for example, (Australian Bureau of Statistics 2002a). Although the RLS has been deliberately designed to describe the real estate geography of the study area some of the resulting dimensions should coincide with aspects of these other studies due to common data attributes. Interpretation of their spatial patterns and comparisons can enhance the understanding of both while supporting the legitimacy of the various spatial patterns of the RLS dimensions. Notable amongst these studies are in the area of urban residential and social geography (Baum et al. 2005; Forster 2004; Randolph 2004; Stimson 1982), and the earlier factorial ecology studies leading to the development of a typology
for the Adelaide area (Burnley 1980, Stimson & Cleland 1975). Using the assessment criteria detailed in chapter 5 the results of the analysis for both the selection and the interpretation of the principal components are presented and discussed. This chapter presents the results of the PCA by first discussing the choice of the number of factors to include in the model, then the appropriate rotation of those factors and finally presenting and discussing the interpretation of the factors loading matrix and the spatial distribution of the resulting RLS construct. The objective of stage 1 is not to present a factorial ecology of the study area. Presenting the spatial patterns of the underlying dimensions is therefore not a key aim of the study, but offered purely to assist in making sense of the results.

7.2 Selection of the number of factors

Selecting the number of factors to be extracted is not an exact science. As discussed in chapter 5, the literature suggests a number of different methods that may be employed to assist with the final decision, either together or taken individually. These include the ‘scree plot’ interpretation, the number of factors with Eigen values greater than 1, the interpretation of the ‘degree of communality’ (Field 2005; Hair et al. 1992), and running a number of different factor solutions to test the sensibility of the various models (Costello & Osborne 2005). In this study the selection was based on a consideration of all the above. The scree plot results, the number of factors that have an Eigen value above 1, and an acceptable level of explained variance in the underlying variables and the degree of communality among the variables were all considered. A number of different factor solutions were then run to test the sensibility of the various solutions. The above indicators are discussed and are presented in this section as the basis for choosing the appropriate number of factors for this study.
7.2.1 The scree plot

The scree plot result is shown in figure 7.1 and because there is no clearly observable inflexion point a definitive indication as to the number of factors to include in the model is not possible by this method alone (Field 2005, p.633). Visual examination would indicate that an appropriate number may lie somewhere between 8 and 15 factors but as the Eigen value is less than 1 for any factors above 12 the optimum solution should lie, on these criteria, in the range of 8 to 12 factors.

![Scree plot output](image)

7.2.2 Sensibility of factors

A number of PCAs were undertaken requesting various numbers of factors to be extracted ranging from 8 to 12 in order to investigate the possibility of a clearer, more interpretable result. This is a method recommended for checking the optimum number of factors to extract (Costello & Osborne 2005, p.3).
The 12-factor model was first examined as this represented the maximum number of factors that contributed more than a single variable on its own (Eigen values greater than 1). However, the interpretation of factors 11 and 12 in this model were not sensible. Not enough variables significantly loaded on factors 11 and 12 to give a sensible interpretation. From these criteria the 10-factor model appears an appropriate solution as it afforded a sensible interpretation of the factor loadings and the variance explained by these factors of the original 50 variables was acceptable at approximately 70%.

### 7.2.3 Communality

Communality describes ‘the amount of variance an original variable shares with all other variables included in the analysis’ (Hair et al. 1992, p.224). This is important in assessing the PCA results as the closer the communality values are to 1 after the rotation ‘the better does the variable serve as an indicator of the associated factors’ (Bartholomew et al. 2002, p.153).

The communalities after the extraction of the 10 factors were found to have a mean value of 0.71 and a standard deviation of 0.17. This represents a high and consistent result demonstrating the appropriateness of the variables for the factor model. This result supports the conclusion that the selected variables contribute well to the derived dimensions of the RLS. A comparison of the communality for the 8, 9 and 10 factor models is given in table 7.1

<table>
<thead>
<tr>
<th>Factors extracted</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>% variation explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.27</td>
<td>0.95</td>
<td>0.65</td>
<td>0.20</td>
<td>65.52</td>
</tr>
<tr>
<td>9</td>
<td>0.29</td>
<td>0.95</td>
<td>0.68</td>
<td>0.18</td>
<td>68.35</td>
</tr>
<tr>
<td>10</td>
<td>0.32</td>
<td>0.95</td>
<td>0.71</td>
<td>0.17</td>
<td>70.85</td>
</tr>
</tbody>
</table>
Based on the ‘scree plot’, the ‘sensibility’ of the results and the ‘communality’ results of various factor solutions the 10-factor model was deemed appropriate for the data set. In comparison to earlier factorial ecological studies across Adelaide a number of factors between 4 and 10 explaining up to 80% of the underlying variables was considered desirable (Stimson 1978, p.30). Although this 10-factor solution is at the upper end of this range, it is a significantly more parsimonious solution than the 15-factor solution obtained earlier in the study as described in chapter 3 (Lockwood & Coffee 2006). This highlights the importance of the data selection phase of the methodology and makes it a prime focus for further research. The question as to the most appropriate form of rotation to employ is the final step in the determination of an appropriate factor model.

7.3 The choice of rotation – orthogonal or oblique

There were two types of rotation investigated and the adopted methodology investigates both types. The first is the ‘orthogonal’ rotation that is well supported in the real estate literature (Bourassa et al. 1999, p.170; Bourassa et al. 2003, p.19) as a technique for quantifying the underlying real estate data structure. An additional advantage in this study is that orthogonal rotation generates a series of components that are linearly independent and therefore suitable for use in further analysis. The second is oblique rotation in which the factors are allowed to correlate. This rotation is suggested in the broader social science literature as being appropriate for data involving ‘humans’ (Field 2005, p.637) as in real life these types of data tend to be correlated.

Therefore, both the ‘orthogonal’ and ‘oblique’ rotations were undertaken with the ‘oblique’ rotation allowing only minimal correlation between the factors. The amount of correlation allowed in the oblique rotation was kept low by setting the value of delta to zero which
ensures high correlation between the factors is not allowed (Field 2005, p.637). As Tabachnick (1989, p.630) point out the fact that an ‘oblique’ rotation is requested does not mean that the factors will automatically be correlated, often they are not and this indicates that no correlated patterns in the underlying data have been missed by not allowing the factors to be correlated.

Table 7.2 Component correlation matrix using the oblique rotation – 10 factor solution

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
<td>-0.11</td>
<td>0.058</td>
<td>0.230</td>
<td>0.252</td>
<td>-0.043</td>
<td>-0.188</td>
<td>0.098</td>
<td>0.027</td>
<td>-0.074</td>
</tr>
<tr>
<td>2</td>
<td>-0.11</td>
<td>1.000</td>
<td>-0.073</td>
<td>-0.007</td>
<td>-0.028</td>
<td>0.006</td>
<td>0.211</td>
<td>-0.113</td>
<td>0.013</td>
<td>-0.166</td>
</tr>
<tr>
<td>3</td>
<td>0.058</td>
<td>-0.073</td>
<td>1.000</td>
<td>0.052</td>
<td>0.034</td>
<td>-0.187</td>
<td>-0.080</td>
<td>0.105</td>
<td>0.086</td>
<td>-0.092</td>
</tr>
<tr>
<td>4</td>
<td>0.230</td>
<td>-0.007</td>
<td>0.052</td>
<td>1.000</td>
<td>0.213</td>
<td>0.014</td>
<td>-0.090</td>
<td>0.082</td>
<td>0.008</td>
<td>-0.019</td>
</tr>
<tr>
<td>5</td>
<td>0.252</td>
<td>-0.028</td>
<td>0.034</td>
<td>0.213</td>
<td>1.000</td>
<td>0.011</td>
<td>-0.097</td>
<td>0.147</td>
<td>0.117</td>
<td>-0.026</td>
</tr>
<tr>
<td>6</td>
<td>-0.043</td>
<td>0.006</td>
<td>-0.187</td>
<td>0.014</td>
<td>0.011</td>
<td>1.000</td>
<td>0.027</td>
<td>-0.110</td>
<td>0.002</td>
<td>-0.072</td>
</tr>
<tr>
<td>7</td>
<td>-0.188</td>
<td>0.211</td>
<td>-0.080</td>
<td>-0.090</td>
<td>-0.097</td>
<td>0.027</td>
<td>1.000</td>
<td>-0.087</td>
<td>-0.015</td>
<td>-0.018</td>
</tr>
<tr>
<td>8</td>
<td>0.098</td>
<td>-0.113</td>
<td>0.105</td>
<td>0.082</td>
<td>0.147</td>
<td>-0.110</td>
<td>-0.087</td>
<td>1.000</td>
<td>0.027</td>
<td>0.015</td>
</tr>
<tr>
<td>9</td>
<td>0.027</td>
<td>0.013</td>
<td>0.086</td>
<td>0.008</td>
<td>0.117</td>
<td>0.002</td>
<td>-0.015</td>
<td>0.027</td>
<td>1.000</td>
<td>-0.046</td>
</tr>
<tr>
<td>10</td>
<td>-0.074</td>
<td>-0.166</td>
<td>-0.092</td>
<td>-0.019</td>
<td>-0.026</td>
<td>-0.072</td>
<td>-0.018</td>
<td>0.015</td>
<td>-0.046</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Extraction method: Principal component analysis. Rotation method: Oblimin with Kaiser normalization.

There was no appreciable difference in ease of interpretation and the resulting component correlation matrix (figure 7.2) showed minimal correlation between the factors and therefore the orthogonal rotation solution was chosen. However, without testing the ‘oblique’ solution it was not possible to see if the orthogonal solution was hiding any other structures. The decision to adopt the ‘orthogonal’ rotation also supports a prime objective of stage 1 namely to use the resulting factors as independent variables in the stage 2 of the analysis and ‘orthogonal’ rotation ensures their independence and suitability for this. On the basis of this discussion the 10-factor solution with orthogonal rotation was adopted as the appropriate model. The amount of variance explained in the original data for each factor and the Eigen value associated with that can be seen in appendix 7.1.
7.4 The interpretation of the factor loading matrix

Interpretation of the factor loading matrix is again not an exact science. However, the literature offers guidelines that assist (Aaker et al. 1995, ch.20; Costello & Osborne 2005; Hair et al. 1992, ch.6; Tabachnick & Fidell 1989, ch.12). Following such guidelines the results of the PCA on the original 50 variables with orthogonal rotation are given in table 7.3 together with the description label of each component subjectively derived from those original variables that most significantly load on that factor. The PCA results in 10 factors explaining approximately 70% of the variation in the original 50 variables. Because such descriptions are subjective they must be understood in a broad rather than in a specific sense as the interaction of the variables comprising the factors is complex and to a degree incomplete as not all relevant variables needed to produce a perfect model could ever be included in the analysis.

The rotated factor matrix is shown in appendix 7.2 showing the orthogonally rotated component matrix and the individual factor loadings. The factor loadings are an indication of the importance of a variable to the factor. As discussed in chapter 5 factor loadings of 0.4 and above were used to choose the variables that most significantly determine the composition of the factor.

These 10 factors represent the residential real estate geography constructs based on the 50 original variables chosen to represent the structural and spatial attributes that together contribute to the choice of a dwelling in which to live. Each of the derived factors are discussed in terms of their constructs and then their spatial patterns are examined to test the ‘sense’ of the results.
### Table 7.3 Description of the 10 factors comprising the underlying dimensions of the RLS

<table>
<thead>
<tr>
<th>Factor Description</th>
<th>Main contributing variables (correlation between variable and factor)</th>
<th>Variance explained by factor</th>
<th>Cumulative variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Families (young, large – average income)</td>
<td>Household size; large (.88); average (.75); small (-.93)</td>
<td>15.9%</td>
<td>15.9%</td>
</tr>
<tr>
<td></td>
<td>+ Dependents (.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Lone person household (-.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Tenure; mortgaged (.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Household income; above-average (.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Age structure; 0 to 20 yrs (.82); 35 to 54 yrs (.53); &gt; 65 yrs (-.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Families (disadvantaged)</td>
<td>Household income; below average (.76); average (.41); high (-.89)</td>
<td>10%</td>
<td>25.9%</td>
</tr>
<tr>
<td></td>
<td>+ Not in labour force (.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Education; tertiary (-.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Distance; CBD (.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Amenity (-.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Dwelling area (-.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Ethnicity (Australian born)</td>
<td>English only spoken (.94); English &amp; another language (-.96)</td>
<td>9.2%</td>
<td>35.1%</td>
</tr>
<tr>
<td></td>
<td>+ Place of birth – Australia (.67); SE European (.79); SE Asia (-.66); NE Asia (-.42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Families (older &amp; well established)</td>
<td>Same dwelling 5 yrs ago (.81); 1 yr ago (.78)</td>
<td>8.2%</td>
<td>43.3%</td>
</tr>
<tr>
<td></td>
<td>+ Owned dwelling (.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Married (.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Age structure; 21 to 34 yrs (-.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Dwelling tenure (low income, rental)</td>
<td>Tenure; rental (.62);</td>
<td>7.2%</td>
<td>50.5%</td>
</tr>
<tr>
<td></td>
<td>+ Multiple dwelling (.58);</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Household income; low (.53);</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Unemployment (.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Sole parent (.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Dwelling proximity (poor accessibility)</td>
<td>Distance from; GP surgery (.81); secondary school (.79); major shops (.68); urban shops (.59); primary school (.66)</td>
<td>6.1%</td>
<td>56.6%</td>
</tr>
<tr>
<td>7 Dwelling type (redevelopment potential)</td>
<td>Dwelling type; single (.72); home unit (-.73)</td>
<td>4.6%</td>
<td>61.2%</td>
</tr>
<tr>
<td></td>
<td>+ Dwelling added value (-.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Dwelling condition (-.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Dwelling wall construction; stone (.41);</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Dwelling type (brick)</td>
<td>Dwelling wall construction; brick (.82); rendered (-.81)</td>
<td>3.3%</td>
<td>64.5%</td>
</tr>
<tr>
<td>9 Dwelling type (rural living)</td>
<td>Dwelling type; rural living (.89)</td>
<td>3.2%</td>
<td>67.7%</td>
</tr>
<tr>
<td></td>
<td>+ Site area (.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Ethnicity (older European born)</td>
<td>NW European born (.56);</td>
<td>3.2%</td>
<td>70.9%</td>
</tr>
<tr>
<td></td>
<td>+ Age 55 to 65 yrs (.44)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7.4.1 Factor descriptions

Factor 1 is constructed from the household size, income and age structure variables. The household age structure represented by a positive correlation with the young (aged 0 to 20 years), suggesting dependent children (confirmed by the positive correlation with the dependent children variable), and the age structure of 35 to 54 years suggesting the parents in the family. Interestingly the age structure may also suggest that this type of family structure does not have the older generation living with them (the negative correlation with the age group greater than 65 years). This construct has an income that is above average but perhaps not large due to the fact that the property is often mortgaged with an age profile reflecting a young family still in the process of wealth accumulation.

In contrast, factor 2 is most significantly correlated with lower income, non-working, not tertiary qualified construct generally in a low amenity area and living further away from the CBD. It is a construct that has a negative correlation with dwelling area indicating the household tends to be accommodated in the smaller dwellings. These correlations strongly suggest a ‘disadvantaged’ family/household structure distinguishing it from factor 1.

Factor 3 introduces the dimension of ‘ethnicity’ expressed in terms of languages spoken and place of birth. The positive correlation with ‘Australian born’ supported by negative correlation with place of birth being south-eastern Europe or Asia and a high correlation with only English being spoken and a high negative correlation with more than one language spoken suggests that this is a definite dimension in the underlying residential real estate geography of the study area explaining 9.2% of the variation.

Factor 4 describes a different type of family structure namely that of an older, well-established household. The ‘well-established’ component of the descriptor is indicated by the positive correlation with both the variables describing the length of time spent in the
same dwelling namely one and five years. The age is indicated by the negative correlation with a relatively younger age group (21 to 34 years) and the positive correlation with the ‘owned dwelling’ variable suggesting the household has been able to repay any mortgage that might have been used in the original purchase of the property and therefore be economically better off, at least in terms of asset accumulation.

Factor 5 is the first of five dimensions that describe the nature and construct of the dwelling. This factor describes a dimension based on the rental ‘tenure’ and more particularly, low-income rental. There is a positive correlation with both the ‘rental tenure’ and ‘low household income’ and the positive correlation with ‘unemployment’ and ‘sole parent’ family structure would perhaps suggest a disadvantaged household. By comparison with factor 2 (disadvantaged families), factor 5 places more of an emphasis on rental accommodation than the family structure. Factor 2 is a more general description of the disadvantaged family structure. Both these factors have a socio economic bias.

Factor 6 is a less complex dimension expressing its construct in terms of the dwelling proximity to various services. It is positively correlated with the distance from a general practitioner’s clinic, both primary and secondary schools and the major and urban shopping facilities. This results in higher scores for those dwellings further away from the services. Interestingly the distance to the CBD does not form part of this dimension. Perhaps this reflects the multi-nuclei nature of the study area causing the proximity to various services to be more focused on the more local areas rather than on the CBD itself. This was pointed out as a feature of the study area by Stimson (1978, p.34) when the ‘concentrations of activities’ had a ‘multiple nuclei’ focus around the CBD, Glenelg (adjacent to the airport) and Port Adelaide. This has since been further compounded by
including additional ‘nuclei’ supported by state government policy as discussed in chapter 3 and spatially displayed in figure 3.2.

Factor 7 is a real estate geography construct describing the redevelopment potential of the dwelling. It describes dwellings that are negatively correlated with the ‘added value’ of the property therefore suggesting the dwelling tends not to add value to the property as a whole, potentially identifying them as redevelopment potential. It is negatively correlated with both ‘condition’ and being a ‘home unit’ dwelling type. This is consistent as the ‘home unit’ dwelling type generally tends to be of a younger age and therefore tends to be in better condition and of a higher added value to the property. The positive correlation with the ‘stone’ condition and ‘single’ dwelling type generally indicates an older type of property that would be expected to form part of this construct.

Factor 8 is a construct describing the structural characteristics of the dwelling. This factor may have some overlap with factor 7 as it describes a brick or a stone dwelling. The positive correlation with ‘brick’ construction variable and a negative correlation with ‘rendered’ construction suggest that it may be of ‘stone’ construction. This appears to be the a weak dimension in terms of their being only two variables that significantly load on this factor and thus incorporating a restricted variety of variables in its description. However, this construct does distinguish these dwellings from other construction types that may have importance in the market place in terms of describing the local effect on ‘price’ in stage 2.

Factor 9 describes the ‘rural living’ dwelling type although it is also significantly positively correlated with the larger site areas indicating that it would not be restricted to the ‘rural living’ dwelling type alone and include dwellings on larger blocks that may be away from the rural living environment.
Factor 10 returns to the ‘ethnicity’ construct this time positively correlated with the ‘north-western European’ born. In addition it is an ethnicity construct also positively correlated with the 55 to 65 year old age group.

It is interesting to observe that these results describe the residential real estate geography construct of the study area and share some similar basic constructs of socio economic status, familism and ethnicity as shown in the early social area analysis (Shevky & Bell 1955; Stimson & Cleland 1975). This is understandable as the marketplace does recognise social constructs (Jackson et al. 2007; Reed 2001). The two geographies should have some overlap and this can be seen in these results as discussed in the following section when examining the ‘sense’ of their spatial distributions.

### 7.4.2 Spatial distribution of factors

The 10 factors have three broad themes describing the underlying residential living structure. They are ‘families’, ‘dwellings’ and ‘ethnicity’. The most important in terms of describing the underlying structure is the ‘family’ theme that is composed of three factors comprising approximately 34% of the variance in the underlying variables. The next important is the ‘dwelling’ theme comprising five factors and explaining approximately 24% and the ‘ethnicity’ comprising two factors and explaining approximately 12%.

The sense of the derived factors can be appreciated better by representing the resulting factors spatially. A score was derived for each factor for each property and added as 10 new attributes for each of the approximate 440 000 properties within the spatial database. These individual scores were plotted for each individual property over the study area enabling the sense of the PCA results to be tested against *a priori* knowledge and other work investigating various social and economic characteristics of the study area.
In particular, the Social Atlas compiled by the Australian Bureau of Statistics (ABS) out of the 2001 quinquennial population census (Australian Bureau of Statistics 2002a) was used for such comparison. If such comparisons do not make sense then the results cannot be considered meaningful. Although these comparisons cannot be exact as the factors and spatial representation of the ABS data are not meant to represent the same thing they can be mutually beneficial in terms of visually examining common spatial patterns of both. The factors are complex interrelationships of many variables the main theme of which is summarised in the subjective factor description whereas ABS data are often only the spatial plot of one data item. In addition, the spatial patterns of social advantage and disadvantage as shown in various social geography studies are used as a general comparison (Baum et al. 2005; Randolph 2004). Again, the discussions of the changing urban structure as described by Forster (2004) are a useful touchstone in understanding the sense of the output.

The general social analysis literature describing the urban construct in terms socio economic status, family status and ethnic status (Murdie 1969; Shevky & Bell 1955; Stimson 1978) should also be recognisable in the real estate geography constructs displayed in these results. Where appropriate, the social analysis constructs are discussed as follows within each of the three identified broad themes identified in these results and compared with their spatial patterns.

*Family structure*

The strongest theme is that of ‘families’. Factors 1, 2 and 4 reflect different elements of the family residential living structure that together account for approximately 34% of the variance in the original variables. Factor 1 describes young families with large to average
income, who are likely to have mortgages and who do not have the grandparent generation living with them.

From the spatial distribution of the calculated factor scores for each property, figure 7.2 shows those properties scoring highly for factor 1 are located in the more recently developed housing parts of the study area. This is consistent with the factor description and may further contribute to that description by suggesting that they live in the newer type of project home developments. This is supported by a visual comparison with the spatial pattern (Australian Bureau of Statistics 2002a) of ‘Dwellings being purchased’ (p.36) and ‘Couples with dependent children’ (p.23). An important feature of this group is the age structure of the family household.

They are generally young and would not have had the opportunity to accumulate wealth suggesting their tendency to have above-average household income may be derived from two incomes. This would support a mortgage but probably only in the outer areas where property is more affordable. This dimension of the RLS may be Adelaide’s version of what Randolph (2004, p.489) refers to as the ‘new urban rim of affluence, aspiration and security’, although the factor 1 group appears to be younger and be living more in the outer, newer suburbs rather than the more inner suburbs described by Baum et al. (2005, ch.3–14). Factor 2 describes the more disadvantaged (lower socio economic) type of family structure living in smaller dwellings further away from the CBD with a higher level of unemployment, lower household income, without tertiary qualification and in an area with a lower amenity score. Within the social analysis typology this may be termed a socio economic status construct. This is consistent with the spatial pattern as shown in figure 7.3.
Figure 7.2  RLS construct – Factor 1

FACTOR 1 after varimax rotation

Families
(Young with average to large household income)

VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 1
- House Hold size: large (.88); average (.75); small (.53)
- Dependent children (.52)
- Lone person House Hold (-.83)
- Tenure: mortgaged (.77);
- House Hold Income: above average (.65)
- Age structure: 0 to 20 yrs (.62);
  35 to 54 yrs (.53); gt 65 yrs (-.76)

Variation explained
in original variables by this Factor is 15.9%

Legend
RESIDENTIAL LIVING STRUCTURE
Factor 1 - 1/2 std. dev breaks

= < -2.75
> -2.75 - -2.25
> -2.25 - -1.75
> -1.75 - -1.25
> -1.25 - -0.75
> -0.75 - -0.25
> -0.25 - 0.25
> 0.25 - 0.75
> 0.75 - 1.25
> 1.25 - 1.75
> 1.75 - 2.25
> 2.25 - 2.75
> 2.75

Study Area
Suburb boundaries
It scores highest in the far northern and southern parts of the study area, the south-western suburbs between the coast and the CBD and scores lowest in the eastern suburbs. These patterns are consistent with expectations and supported by the ‘one-parent families with dependent children’ visual patterns from the Social Atlas (Australian Bureau of Statistics 2002a, p.22). An interesting observation may be seen by comparing the spatial patterns of factors 1 and 2. There are similarities in terms of some general locations of both factors one describing a family structure while the other with more of an emphasis on the socio economic status. Both factors score highly in the northern and southern suburbs with some newer land divisions in the northern area (high score on factor 1) exist in close proximity to the more depressed socio economic areas that score highly on factor 2. These local areas of disadvantage may well be those parts of a more global area identified by Baum et al. (2005 ch.3–21) as being ‘old economy extremely disadvantaged localities’.

These are consistent with the suburbs laid out in the 1950s and 1960s to house workers employed at the northern car manufacturing plant and associated industries, while factor 1 locations are the newer land developments adjacent and to the east of the factor 2 areas.
Figure 7.3 RLS construct – Factor 2

FACTOR 2 after varimax rotation

Families
(Disadvantaged)

VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 2
- House Hold Income; below average (.76); average (.41); high (-.89)
- Not in labour force (.46)
- Education; tertiary (-.56)
- Distance; CBD (.45)
- Amenity (-.41)

Variation explained in original variables by this Factor is 10%

Legend

RESIDENTIAL LIVING STRUCTURE
Factor 2 - 1/2 std. dev breaks

-2.75
-2.25
-1.75
-1.25
-0.75
-0.25
0.25
0.75
1.25
1.75
2.25
2.75

Study Area
Suburb boundaries
Factor 4 describes the well-off older family structure and distinguishes itself from factor 1 in that they tend to display contrasting spatial patterns. The ‘older and well-off’ (factor 4) tend to occupy the older and longer established areas closer to the CBD and includes the eastern suburbs and western coastal and near coastal suburbs, as well as the Adelaide Hills suburbs as shown in figure 7.4. This understandably correlates highly with ownership of the family home rather than it being mortgaged and suggests, along with the older age structure, a higher accumulation of asset wealth in terms of ownership of the family home. The age structure of the household visually correlates with the ‘People aged 55 to 64 years’ (Australian Bureau of Statistics 2002a, p.10), which suggests that the age structure of this dimension of the RLS may well represent a portion of the ‘baby boomer’ generation. Conversely, again relying only on visual correlation of the spatial pattern may suggest that the Adelaide age cohort 55 to 64 years may have lived in their dwelling for at least five years and be married within the higher scoring spatial patterns shown in figure 7.4. This construct may represent, in part, what Forster (2004, p.117) refers to as the ‘middle suburb … characterised by relatively low rates of household turnover and therefore significant ageing in place’.
Figure 7.4  RLS construct – Factor 4

**FACTOR 4 after varimax rotation**

Families
(older & well established)

VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 4
- Same dwelling 5 yrs ago (.81); 1 yr ago (.78)
- Owner dwell (.73)
- Married (.44)
- Age structure: 21-34 yrs (.77)

Variation explained in original variables by this Factor is 8.2%

**Legend**

RESIDENTIAL LIVING STRUCTURE
Factor 4 - 1/2 std. dev breaks

以色表示
- <= -2.75
- >-2.75 - -2.25
- >-2.25 - -1.75
- >-1.75 - -1.25
- >-1.25 - -0.75
- >-0.75 - -0.25
- >-0.25 - 0.25
- >0.25 - 0.75
- >0.75 - 1.25
- >1.25 - 1.75
- >1.75 - 2.25
- >2.25 - 2.75
- >2.75

Study Area
Suburb boundaries

0 3.75 7.5 15 Kilometers
Dwellings

The broad theme of ‘dwelling’ covers many aspects of the term and is represented in this study by five factors. Each of these factors represents an independent dimension of the RLS involving a range of aspects of the dwelling itself accounting for a significant proportion of the variance in the underlying variables (24%). These dimensions describe a range of dwelling features from dwelling tenure to dwelling construction and aspects of dwelling location.

The first, factor 5 includes socio economic indicators as part of the construct of this dimension of the RLS as shown through the positive correlation with low income, sole parent household. From the perspective of this study this construct includes the ‘multiple dwelling’ type and the rental tenure of the dwelling. The spatial pattern of factor 5 (figure 7.5) shows these households live in property located in the inner north-western suburbs that broadly conforms with the Social Atlas (Australian Bureau of Statistics 2002a) map ‘Rented dwellings – government owned’ (p.37). This RLS dimension may contribute to the understanding of this distribution by describing the government-owned rental accommodation in terms of the type of disadvantaged households that may occupy this as being sole unemployed parents.
Figure 7.5 RLS construct – Factor 5

FACTOR 5 after varimax rotation

Dwelling Tenure
(low income, rental)

VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 5
- Tenure: rental (.62);
- Multiple dwelling (.58);
- House Hold income: low (.53);
- Unemployment (.38);
- Sole parent (.57)

Variation explained
in original variables by this Factor is 7.2%

Legend

RESIDENTIAL LIVING STRUCTURE
Factor 5 - 1/2 std. dev breaks

- <= -2.75
- >-2.75 - 2.25
- >-2.25 - -1.25
- >-1.75 - -1.25
- >-1.25 - -0.75
- >-0.75 - -0.25
- >0.25 - 0.25
- >0.25 - 0.75
- >0.75 - 1.25
- >1.25 - 1.75
- >1.75 - 2.25
- >2.25

Study Area
Suburb boundaries
There is a prima facie correlation between this dimension and the dimension described by factor 2 (disadvantaged families). However, as can be seen from their respective spatial patterns (figures 7.2 and 7.5), the disadvantaged families occupy more areas than just those described by only the low rental dwelling tenure in factor 5.

Factor 6 and 9 describe the location of the dwelling as distinguishing dimensions. Factor 6 describes the distance from various services whereas factor 9 describes the ‘rural living’ lifestyle. The spatial patterns generated by these factors support the description (see figures 7.6 and 7.7).

The first, factor 6, concentrates on the proximity of the dwelling to the whole range of goods and services included in this study. Interestingly, the CBD was not the most significant in this particular group, but rather it is the local or regional proximity that more significantly contributing to this factor. Figure 7.6 shows the spatial distribution of this factor score on an individual property basis and, as expected, those that score higher are those more remote properties. The influence of the provision of services at the regional level as opposed to that of the CBD can be seen in the lower scores of the northern and southern extremes of the study area and may suggest the influence of the ‘multi-nuclei’ structure of the study area as discussed earlier.
Figure 7.6  RLS construct – Factor 6

**FACTOR 6 after varimax rotation**

Dwelling Proximity  
(poor accessibility)

**VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 6**
- Distance from: GP surgery (.81); Secondary sch (.79);
  major shp (.68); urban shp (.50); primary sch (.66)

Variation explained in original variables by this Factor is 6.1%

**Legend**

**RESIDENTIAL LIVING STRUCTURE**
Factor 6 - 1/2 std. dev breaks

- Blue: <= -2.25
- Green: >-2.25 -1.75
- Green: >-1.75 -1.25
- Green: >-1.25 -0.75
- Green: >-0.75 -0.25
- Green: >0.25 - 0.25
- Orange: >0.25 - 0.75
- Orange: >0.75 - 1.25
- Orange: >1.25 - 1.75
- Orange: >1.75 - 2.25
- Orange: >2.25

Study Area
Suburb boundaries
Figure 7.7  RLS construct – Factor 9

**FACTOR 9 after varimax rotation**

Dwelling Type  
Rural living

**VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO**  
**FACTOR 9**  
- dwelling type: rural living (.89)  
- Site area (.86)

Variation explained  
in original variables by this Factor is 3.2%
Factor 9 defines itself by both property type and the general associated with larger land size. Together these describe the rural living lifestyle. Therefore properties scoring highly in terms of this dimension may also include properties on large parcels of land that are not the ‘rural living’ land use. These may be found in the older suburbs and associated with the single residential living dwelling. The spatial pattern of factor 9 indicates that this may be occurring. The highest scores are in geographic areas where both the ‘rural living’ land use and the larger parcels of land occur. High scores occur in some of the older suburbs closer to the CBD developed when land parcels were larger than today.

Factor 7 describes a RLS dimension that centers on the market potential of the property. Like all the RLS dimensions, the redevelopment potential of a property is not made up of one single attribute but a combination of attributes that in this case point to a potential higher and better use of the property. Randolph (2004, p.488) identifies this as one of the key changes facing Australian cities and describes locations with these attributes as the ‘stressed’ middle suburbs which are undergoing rapid change driven mainly by the market process. Examples within the study area may well be those areas in the middle and inner suburbs that Forster (2004, p.110) describes as having reached the ‘empty nest’ stage of the life cycle. The spatial patterns in figure 7.8 tend to support this view. The variables that correlate most significantly with factor 7 indicate that single dwellings in poorer condition of perhaps stone construction (maybe indicating, in the case of Adelaide, an older dwelling) and have a small added value may well be that type of property suitable for redevelopment. The spatial pattern as shown in figure 7.8 shows the higher scores in the inner to middle suburbs and some part of the Adelaide hills suburbs. This is consistent with where the older dwellings are more likely to be found and land use regulations more likely to have changed to allow for higher density living resulting in higher land prices and lower
added value of existing structures. This is a different but important dimension in the RLS in terms of explaining the residential real estate construct of the study area.

Factor 8 (figure 7.9) combines two attributes of the dwelling wall construction to form a RLS dimension indicating that wall construction is probably ‘brick’ but not rendered. This is the weakest dimension in terms of its interpretability although its spatial pattern supports the likely existence of brick or stone dwelling construction and its importance in conjunction with other variables in explaining the price structure may be more relevant than this plot may suggest.
Figure 7.8 RLS construct – Factor 7

**FACTOR 7 after varimax rotation**

Dwelling Type  
(Redevelopment potential)

VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 7:
- dwelling type; single (.72); Home Unit (.73)
- dwelling added value (.63)
- dwelling condition (.46)
- dwelling wall construction; stone (.41);

Variation explained in original variables by this Factor is 4.6%

Legend

**RESIDENTIAL LIVING STRUCTURE**

Factor 7 - 1/2 std. dev breaks

<table>
<thead>
<tr>
<th>Breaks</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>=&lt; -2.75</td>
<td>Blue</td>
</tr>
<tr>
<td>&gt;-2.75 - -2.25</td>
<td>Green</td>
</tr>
<tr>
<td>&gt;-2.25 - -1.75</td>
<td>Blue</td>
</tr>
<tr>
<td>&gt;-1.75 - -1.25</td>
<td>Green</td>
</tr>
<tr>
<td>&gt;-1.25 - -0.75</td>
<td>Blue</td>
</tr>
<tr>
<td>&gt;-0.75 - -0.25</td>
<td>Green</td>
</tr>
<tr>
<td>&gt;-0.25 - 0.25</td>
<td>Blue</td>
</tr>
<tr>
<td>&gt;0.25 - 0.75</td>
<td>Green</td>
</tr>
<tr>
<td>&gt;0.75 - 1.25</td>
<td>Blue</td>
</tr>
<tr>
<td>&gt;1.25 - 1.75</td>
<td>Green</td>
</tr>
<tr>
<td>&gt;1.75 - 2.25</td>
<td>Blue</td>
</tr>
<tr>
<td>&gt;2.25 - 2.75</td>
<td>Green</td>
</tr>
<tr>
<td>&gt;2.75</td>
<td>Purple</td>
</tr>
</tbody>
</table>

Study Area

Suburb boundaries

0  3.75  7.5  15 Kilometers
Figure 7.9  RLS construct – Factor 8

FACTOR 8 after varimax rotation

Dwelling Type
Brick

VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 8
- dwelling wall construction; brick (.82); rendered (>.81)

Variation explained in original variables by this Factor is 3.3%

Legend
RESIDENTIAL LIVING STRUCTURE
Factor 8 - 1/2 std. dev breaks

- <= -2.75
- >-2.75 - -2.25
- >-2.25 - -1.75
- >-1.75 - -1.25
- >-1.25 - -.75
- >-0.75 - -0.25
- >-0.25 - 0.25
- >0.25 - 0.75
- >0.75 - 1.25
- >1.25 - 1.75
- >1.75 - 2.25
- >2.25 - 2.75
- > 2.75

Study Area
Suburb boundaries

0 3.75 7.5 15 Kilometers
Ethnicity

Ethnicity is associated with two of the dimensions that make up the RLS of the study area. It is represented by factors 3 and 10 that distinguish between the Australian/European born and the older European born members of the population.

Interpretation may be made by comparing the spatial pattern of both the high and low scores of factor 3 (figure 7.10) with the ‘People not fluent in English’ (Australian Bureau of Statistics 2002a, p.15). The visual correlation between the two suggests that low scoring on factor 3 (low in terms of Australian/European born) corresponds with high percentages of people born in south-east Asia or Greece or Italy. By contrast the low factor 3 score in the eastern suburbs may correspond with the higher income Malaysian population.

Factor 10 (figure 7.11) describes a RLS dimension consisting of older (55 to 64 years) north-western European (not Australian) born population living in larger dwellings. Comparison with the Social Atlas (Australian Bureau of Statistics 2002a) shows this group (factor 10) does not all belong to the same income group.

For example, some outer southern suburbs and northern suburbs correspond with unemployed people 45 years and older (p.29) whereas the eastern edge of the study area (high factor 3 scores) correspond with the high income households (p.26).
Figure 7.10  RLS construct – Factor 3

**FACTOR 3 after varimax rotation**

Ethnicity
(Australian Born)

**VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 3**
- English only spoken (94); English & another (<96)
- Place of birth - Australia (97); SE European (<79); S.E.Asia (~66); N.E.Asia (~42)

Variation explained in original variables by this Factor is 9.2%

**Legend**

**RESIDENTIAL LIVING STRUCTURE**
Factor 3 - 1/2 std. dev breaks
- <=-2.75
- >-2.75 - -2.25
- >-2.25 - -1.75
- >-1.75 - -1.25
- >-1.25 - -0.75
- >-0.75 - -0.25
- >-0.25 - 0.25
- >0.25 - 0.75
- >0.75 - 1.25
- >1.25 - 1.75
- >1.75 - 2.25
- >2.25

Study Area
Suburb boundaries
Figure 7.11  RLS construct – Factor 10

FACTOR 10 after varimax rotation

Ethnicity
(older European born)

VARIABLES MOST SIGNIFICANTLY CONTRIBUTING TO FACTOR 10
- NIV European born (.56);
- Age 55-65 yrs. (.44)
- Dwelling area (.36)

Variation explained in original variables by this Factor is 3.2%

Legend
RESIDENTIAL LIVING STRUCTURE
Factor 10 - 1/2 std. dev breaks

<table>
<thead>
<tr>
<th>Range</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;= -2.75</td>
<td></td>
</tr>
<tr>
<td>&gt; -2.75 -2.25</td>
<td></td>
</tr>
<tr>
<td>&gt; -2.25 -1.75</td>
<td></td>
</tr>
<tr>
<td>&gt; -1.75 -1.25</td>
<td></td>
</tr>
<tr>
<td>&gt; -1.25 -0.75</td>
<td></td>
</tr>
<tr>
<td>&gt; -0.75 -0.25</td>
<td></td>
</tr>
<tr>
<td>&gt; 0.25 - 0.25</td>
<td></td>
</tr>
<tr>
<td>&gt; 0.25 - 0.75</td>
<td></td>
</tr>
<tr>
<td>&gt; 0.75 - 1.25</td>
<td></td>
</tr>
<tr>
<td>&gt; 1.25 - 1.75</td>
<td></td>
</tr>
<tr>
<td>&gt; 1.75 - 2.25</td>
<td></td>
</tr>
<tr>
<td>&gt; 2.25</td>
<td></td>
</tr>
</tbody>
</table>
7.5 Conclusion

The objective in this stage of the methodology was to quantify the dimensions of the RLS. The derived principal components adequately represent the large number of original property attributes (taken from a variety of sources) in a manageable number of statistically independent variables that can be used as surrogate property characteristics in the next stage of the methodology.

In this particular study the broad groupings of the 10 dimensions of the RLS are ‘families’, ‘dwellings’ and ‘ethnicity’. This thesis is based on the premise that it is within these 10 structural dimensions of the RLS that the real estate market participants make their housing choice decisions. This imposes a large obligation on the researcher to ensure that the variables selected properly represent the underlying residential real estate geography as fully and as accurately as possible. The extent to which this does not happen must be viewed as a caveat on the veracity of the final solution. The data collected represented the best that were available within the constraints of the study. The analysed results showed factor 8 to be weak in terms of the number of variables that significantly loaded on it. Future work will focus on improving the range and quality of data used to describe the real estate geography. This methodology allows for the inclusion of a large amount of data and if the data are appropriate should result in a more parsimonious factor solution. Another caveat on the methodology must be the subjective nature of PCA making the results difficult to verify.

However, the sense of the results, in terms of the formulated factor descriptions, and the resulting spatial patterns of these descriptions give confidence that the analysis makes sense as a good PCA should (Tabachnick & Fidell 1989, p.598). This sense is further supported and enhanced by being able to link the results of different studies through
similarity in spatial patterns. This enables a more complete picture to evolve of the RLS dimensions. These comparisons can be hard to make in anything but general terms as often they have been analysed on broad spatial units such as local government areas whereas the RLS has been derived on attribute data collected at the property level. However, subject to this caveat, these comparisons are important as property market structures are inexorably linked to the underlying real estate geography that these studies often describe. Conversely, the nature of the resulting market structures may also provide an insight into the composition of the social geographer’s construct of advantaged and disadvantaged areas.

The advantage of a good PCA for this study is that it provides a method for including a large number of relevant variables describing a complex structure by a smaller set of independent surrogate characteristics capable of representing the original variables in a practical and manageable manner in ongoing analysis. In this study a set of 10 surrogate property characteristics representing 50 original variables describing the real estate geography of the study area were derived and these can be used as a basis, in stage 2 of the analysis, as genuinely independent variables that sensibly represent the underlying residential living structure of the study area in which the submarkets reside.

The description and understanding of the dimensions of the RLS have particular significance in this research when (as will be shown in stage 2) they are classified as contributing to a particular submarket group. Subject to the caveats discussed, this will help the practitioner understand not only where submarket differences occur but also the characteristics of that difference. Chapter 8 presents the results of stage 2 of this study in which these dimensions are related to the marketplace through a hedonic geographically weighted regression model. This shows local the influence of price on each of these dimensions which ultimately lead to geospatial submarket delineation.