RESIDENTIAL LIVING STRUCTURE AS A BASIS FOR THE SPATIAL DELINEATION OF RESIDENTIAL SUBMARKETS

TONY LOCKWOOD and NEIL COFFEE
University of Adelaide

ABSTRACT
Whilst it is generally accepted that residential submarkets exist, this is not the case for either a definition or delineation of residential submarkets. There is a need to incorporate a full range of property attributes (including structural, environmenta, and socio-economic) describing all dwellings in the study area in order to provide an understanding of the whole residential structure in which submarkets reside. This paper proposes the derivation of a residential living structure (RLS) to achieve this. Principal Component Analysis (PCA) is used to derive the underlying RLS from a wide range of variables including age, size, building construction, house condition, amenity, accessibility to various services and the CBD, together with a range of socio-economic variables. Altogether 65 variables were collected for each of the approximate 440,000 residential properties in the study area. The results presented in this paper reveal 15 underlying components of the RLS in the metropolitan area of Adelaide, South Australia explaining 74% of the variation in the original data. This study extends the existing work in this area by including a wide range of property attributes for all properties in the study area and by establishing the resulting principal components as a set of comprehensive and independent surrogate property characteristics as a basis for further submarket analysis.

Keywords: Residential living structure, submarkets

INTRODUCTION

The residential real estate market is unlike many other commodity markets in terms of its complexity and dimensions. The complexity derives from its heterogeneous nature, as each property is unique, immobile, modifiable and expensive to both buy and sell (Galster, 1996). The residential 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 market place (Evans, 1995). Also, the residential market is unlike many other markets in terms of its dimensions, as it has both a price quantity equilibrium and a spatial equilibrium dimension (Thrall, 2002) and may be viewed as a series of interconnected submarkets defined by both structural and spatial elements (Grigsby, 1963; Watkins, 2001) . There is a need to understand the underlying dimensions of residential submarkets, particularly in the areas of housing policy formulation, the urban planning process (Meen, 2001) and in the implementation of mass appraisal systems (McCluskey et al., 2002).

The literature identifies two important issues in identifying submarket dimensions. First, the need to understand the underlying residential structure of the study area in which the submarkets exist. Second, if optimal submarket segmentation is to be achieved, it is necessary to express this structure in the hedonic modelling constructed to detect submarket differences rather than relying on the individual property characteristics. From this, a definition of a spatial submarket as being a significant change across space in the market's interpretation of the underlying residential structure has been adopted.

Deriving spatial submarket boundaries from the data and not relying on some form of *a priori* definition has a better opportunity to find an optimum spatial boundary, as it is not constrained by an artificial spatial boundary drawn without regard to any submarket criteria. Studies adopting this approach have addressed these issues to varying degrees. For example, they have promoted the need to derive measures of the residential structure that incorporate the complex interaction between the individual property characteristics rather than just the characteristics and have used principal component analysis (PCA) to achieve this. Some have included an indicator of market value, so that the resulting residential structure is based on economic criteria and hence a basis for submarket segmentation. However, these studies have used only sample data from the study area, either dictated by properties sold (Watkins, 1999; Bourassa et al., 2002) or through survey samples (Bourassa et al., 1999) in order to construct the underlying residential structure. These studies then cluster the resulting components and test for submarkets.

As a contribution to the delineation of spatial submarket boundaries, this paper proposes the derivation of a residential living structure (RLS) from property attribute data using PCA, building upon existing research and extending it in the following ways:

1. extends the derivation of the underlying RLS to include all the properties in the study area rather than a sample. This eliminates any bias that may be present in the sample and more accurately reflects the existing underlying structure.
2. extends the data used in formulating the RLS to include relevant data items from all land related data sources (socio economic, environmental, and structural dwelling characteristics, namely size; condition; number of rooms and construction type).
3. deliberately excludes a market value attribute, leaving the resulting RLS describing only the property together with its externalities and therefore the resulting RLS cannot be considered to represent submarket segments (as in other studies (Bourassa et al., 2002; Bourassa et al., 1999; Watkins, 2001)), but rather only dimensions of the underlying structure.
4. deliberately excludes any attribute that describes property location from the RLS derivation process. It is the subsequent mapping of the analysed data that displays the significance of location.

5. provides 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 space will be used to indicate potential spatial submarket boundaries.

The project to develop a methodology to derive the spatial component of submarket boundaries has been divided into two stages. Stage 1 is to derive the RLS of the whole study area in which the submarkets reside. Stage 2 will relate that structure to the market place at a given point in time to indicate the spatial component of submarket boundaries. The objective of this paper is to present the results of stage 1 of this project, namely the need for and construction of the RLS.

LITERATURE REVIEW

There is general support in the literature for the theoretical existence of residential real estate submarkets (Adair et al., 1996; Goodman and Thibodeau, 2003), although there is very little consensus as to what the model of a submarket looks like (McCluskey et al., 2002; Watkins, 2001). The literature contains many definitions of submarkets at the local urban level that may be defined spatially or structurally or indeed both (MacLennan and Tu, 1996). A comprehensive summary of the various studies covering the different definition types is given by Watkins (2001) who outlines suggested reasons for the failure to develop submarket models to include a lack of submarket definition and identification procedures.

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 market place is of fundamental importance to the wellbeing of our society. Hence, the importance of understanding the underlying residential living structure and its relationship to the market place (Rothenberg et al., 1991).

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). It is important for the architects of housing policy to have an understanding of the market structure to improve their decision making process (MacLennan and Tu, 1996; Meen and Meen, 2003). Galster (1996) reinforces the view that the formulation of 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), which suggests that market indicators should be used to determine when planning intervention may be appropriate in the development process. A better understanding may be achieved by identifying submarkets based on the underlying residential living structure.

Can and Megbolugbe (1997) 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, measure housing demand, establish price trends and formulate and design housing and mortgage policies, all of which relies on an understanding of the structure of residential submarkets.

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 as at a specific date for the entire jurisdiction, often many hundreds of thousands of properties (Eckert, 1990). 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 identification, or delineation of the various submarkets that may be present in the jurisdiction concerned. This enables more accurate and reliable statistical models to be developed for each particular submarket rather than one model for the whole jurisdiction (Figueroa, 1999; Fletcher et al., 2000; Gallimore et al., 1996; McCluskey et al., 2002; McCluskey et al., 1999; 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.

Current situation

There is a body of literature in which the spatial component of submarkets are thought to be important, but are assumed to conform with existing administrative boundaries such as suburbs, postcodes districts, local government jurisdictions, electoral boundaries and other forms of *a priori* definitions proposed by real estate agents and other land professionals (Adair et al., 1996; Bourassa et al., 2002; Jones et al., 2004). In studies based on *a priori* definitions, separate submarket status is based on the now accepted test 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 use an approach where the data is used to determine the spatial submarket boundary (Bourassa et al., 2002; Maclennan and Tu, 1996; Watkins, 2001). These studies approach the identification of submarkets by firstly recognising that *a priori* spatial segmentation does not necessarily give an optimal solution and secondly, that submarkets are determined by more than individual property characteristics and perhaps should focus on a more complete picture of the residential structure of the study area. It is critical when determining submarket boundaries that the complete urban structure is used and not just individual dwelling characteristics. Watkins (2001) refers to microeconomic theory, which states that:

> *"… 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." (p.2241)*

Watkins stresses two important points; first, both product groups (representing the segmented supply characteristics) and consumer groups (representing the segmented demand characteristics) need to be reflected in the model and second, such a model has to be market based as submarkets are clearly a market entity. This second point is vital and supported by Pryce (2004) who reports that the use of non-economic criteria to delineate an economic entity "contains an essential flaw" (p2-10). 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 market place so the structures can be recognised as economic entities.

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 PCA to extract a reduced set of independent components from structural and local government data to describe the underlying structure of the data and then used cluster analysis to appropriately group these components into submarkets. Adopting a different approach Costello (2004) used property transaction data in Perth to establish market segmentation. Costello (2004) started with *a priori* postcode districts and examined price variation across these to build a price-location model of market segregation. An interesting spatial aspect emerges relating to the contiguity displayed by his price-location model namely, they are not necessarily contiguous, a finding supported by Rothenberg et al. (1991). Although not specifically investigating 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 a number of socio-economic variables into a number of independent components that were related to house prices. Reed (2001) 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.

These studies identify a number of fundamental issues that need addressing to achieve submarket delineation:

1. understand the underlying structure and dimensions of the market place in which submarkets reside (Meen, 2001; Rothenberg et al., 1991; Watkins, 2001).
2. recognise the overemphasis in existing studies on the individual dwelling characteristics to determine submarket boundaries (Bourassa et al., 2002).
3. recognise the importance of the relationship of the market place to the underlying residential structure (Pryce, 2004).

Contribution of the RLS

RLS is not a new concept, but rather builds upon the earlier work of Bourassa et al. (2002), and Maclennan and Tu (1996) by using PCA to reduce a large number of relevant data, collected for every property in the study area, to a smaller number of components that represent the underlying residential structure of the whole study area. The contribution of RLS lies in the fact that:

1. 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 giving a truer understanding of the total structure as at the study date. This contributes to issue 1 identified above.
2. by calculating component scores for each property in the study area and adopting them as surrogate property characteristics the overemphasis on the individual dwelling characteristics is removed. This contributes to issue 2 identified above.
3. it provides a comprehensive and independent set of surrogate property characteristics that can be used in stage 2 of the project. The housing market transaction price attribute is deliberately omitted in the construction of the RLS. The market transaction price will be used as the dependent variable and the surrogate property characteristics from the RLS as the independent variables in a geographically weighted regression model to indicate submarket boundaries. This is a point of difference with the current literature in that any group of properties that transact representing the study date can be related to these 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. This contributes to issue 3 identified above.

METHODOLOGY

The methodology adopted in this study to calculate the RLS was developed from previous work by Bourassa et al. (2002), Bourassa et al. (1999), Maclennan and Tu (1996) and Watkins (2001). This study further develops the use of PCA by applying it to all properties in the study area and incorporating as many available variables that are appropriate in describing the property to derive a reduced number of components from the original data set and calculate component scores for each component for each property within the study area.

Study area

The study area contains all the urban census collection districts (CCDs) as defined in the Australian Standard Geographic Classification (ASGC) 2001 within the metropolitan area of Adelaide, South Australia. There were 2059 CCDs with a population of approximately 1.1 million people out of a total population of approximately 1.5 million for the state of

South Australia (ABS, 2001) . The study area contains approximately 440,000 residential dwellings.

Analysis techniques

PCA is used in this research to reduce a large number of observed variables to a smaller number of principal components. This is achieved by summarising the correlation patterns amongst the original variables into groups or components. For more details of the PCA procedure, together with worked example, see Kline (1994). The number of adopted components are chosen based on their respective eigenvalue and the scree plot interpretation. Eigenvalues greater than 1 indicate components that contribute more than single variables. A scree plot displays the eigenvalue (y-axis) and the associated components in order of extraction (on the x-axis) and indicates the relative importance of each component (larger eigenvalues having more importance). The point of inflexion may assist in indicating an appropriate cut off in the number of components to be selected. These two criteria should be considered together (Field, 2005; Hair et al., 1992). Component scores are assigned to each property in the study area by multiplying the component 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:

$$Y_{i} = b_1X_1 + b_2X_2 + _{\ldots} + b_nX_n + \text{error.}$$

where: Y_i is the i^{th} (1 to 15) component
b is the component loading
X is the original variable

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 and 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) are summarised in italics as follows:

1. *the solution does not have a criteria 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 using geographic information system (GIS).
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."*
   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. Field (2005) argues that orthogonal rotation should never be used for any data involving humans. 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 adopted as it provided a more sensible interpretation of the components and was done with minimal correlation so as not to invalidate their use in future research.

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.

The variables included in the construction of RLS were selected on the basis of their contribution to the market value of property. Structural variables were identified from previous work in the study area (Lockwood, 2003; Rossini and Kershaw, 2005). Environmental, accessibility and socio-economic attributes were identified based on a Canadian study (Kestens et al., 2002). A recent study in Brisbane, Australia highlighted Aesthetic, Amenity, and Social Interaction as important factors in the decision to choose a residential location (Chhetri et al., in press). The link between social constructs (demography and income) and property values was established in a Brisbane study by Reed (2001). Based on these studies, an initial set of variables was identified for use in this study. They were further refined for PCA, based on various criteria as suggested by Hair et al., 1992) and are summarized as follows:

- Significance of variable loadings on any particular component.
- Importance of the variable to the research objectives.
- Communality index.

This resulted in 65 variables being used in this research (see Table 1).

Table 1: Variables used in the study

Variable name	Variable type
Single Dwelling	Boolean (0,1)
Multiple Dwelling	Boolean (0,1)
Home Unit	Boolean (0,1)
Rural Living (non primary production)	Boolean (0,1)
Dwelling construction – brick	Boolean (0,1)
Dwelling construction – stone	Boolean (0,1)
Dwelling construction – rendered	Boolean (0,1)
Dwelling area	Continuous standardised variable
Dwelling condition	Continuous standardised variable
Dwelling added value	Continuous standardised variable
Dwelling age	Continuous standardised variable
Dwelling number of main rooms	Continuous standardised variable
Land area	Continuous standardised variable
Amenity score (NDVI) index	Continuous standardised variable
Road distance of dwelling from:	
- GP surgery	Continuous standardised variable
- primary school	Continuous standardised variable
- secondary school	Continuous standardised variable
- major shops	Continuous standardised variable
- urban shops	Continuous standardised variable
House hold mortgage repayment	
- Low	Continuous standardised variable
- Below average	Continuous standardised variable
- Average	Continuous standardised variable
- Above average	Continuous standardised variable
- High	Continuous standardised variable
House hold rental repayment	
- Low	Continuous standardised variable
- Below average	Continuous standardised variable
- Average	Continuous standardised variable
- Above average	Continuous standardised variable
- High	Continuous standardised variable
House hold size	
- Small	Continuous standardised variable
- Average	Continuous standardised variable
- Large	Continuous standardised variable
House hold tenure	
- Owned	Continuous standardised variable
- Mortgaged	Continuous standardised variable
- Rental	Continuous standardised variable
House hold income	
- Low	Continuous standardised variable
- Below average	Continuous standardised variable
- Average	Continuous standardised variable
- Above average	Continuous standardised variable
- High	Continuous standardised variable
House hold – number of cars	
- 0 cars	Continuous standardised variable
- 1 car	Continuous standardised variable
- 2 cars	Continuous standardised variable
- 3 cars	Continuous standardised variable

Length of same occupancy	
- 1 year	Continuous standardised variable
- 5 years	Continuous standardised variable
Individual place of birth	
- NW Europe	Continuous standardised variable
- SE Europe	Continuous standardised variable
- SE Asia	Continuous standardised variable
- NE Asia	Continuous standardised variable
- South Central Asia	Continuous standardised variable
- Australia	Continuous standardised variable
Individual status	
- Married	Continuous standardised variable
- Sole Parent	Continuous standardised variable
- Lone	Continuous standardised variable
- Dependant Children	Continuous standardised variable
Individual age	
- 0 to 20 years	Continuous standardised variable
- 21 to 34 years	Continuous standardised variable
- 35 to 54 years	Continuous standardised variable
- 55 to 65 years	Continuous standardised variable
.-.Greater than 65 years	Continuous standardised variable
Languages spoken at home	
- English only	Continuous standardised variable
- English and another	Continuous standardised variable
Employment status	
- Not in labour force	Continuous standardised variable
- Unemployed	Continuous standardised variable
TOTAL NUMBER OF VARIABLES	65

Data

The data describing the structural attributes of dwellings are from the South Australian Valuer General's database and include the age, style, size, condition, and building materials. Current site and capital values of the property calculated for taxation purposes are 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.

The residential properties in the study area are grouped into four dwelling types, single dwellings, multiple dwellings, home units, and non-primary producing rural dwellings. In addition, data describing the accessibility (in terms of road distance) of each property to various services such as education, health and shopping derived by the University of Adelaide was used as accessibility attributes (GISCA, 2002) .

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 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.

The socio-economic data is sourced from the Australian Bureau of Statistics (ABS), Census of Population and Housing 2001. Census data include measures of age, income, family structure, employment and tenure. The smallest spatial unit at which these data are available is the census collector district (CCD), which include approximately 200-300 households. The variables used to describe ‘household mortgage repayments’, ‘household rental repayments’, ‘household size’, ‘household tenure’ and ‘household income’ as ‘low’ to ‘high’ are calculated based on the distribution about the mean value for the whole study area. For each CCD, the number of households or individuals that fit each category is expressed as a percentage of the total for that CCD and then standardised.

As the focus of the study is at the individual property level, it was important to collect variables at that level. However, due to confidentiality restrictions, socio-economic data are only available at the CCD level and therefore all properties within the CCD are attributed with the average CCD attributes. It may be argued that an average of the CCD is a better measure than the actual value for the household. For example, if the property is sold, the individual property value may change and the buyer may well have considered the surrounding attributes to be of more relevance than those of the seller. However, it must be acknowledged that the extent to which this misrepresents the households within the CCD may weaken the quality of the result. The spatial reference for the data is based on the property cadastre which is a digital representation of the Valuer General’s designated property boundaries.

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. This is to eliminate, as far as possible, any differences due to time.

RESULTS

The PCA analysis in this study derived 15 principal components from the original 65 variables that described various aspects of all the residential dwellings within the study area. These components are shown in Table 2, with a description and the original variables that were found to correlate with that component. The resulting components explain approximately 74% of the variation in the original variables and may be thought of as representing the dimensions of the total existing residential dwelling stock within the study-area.

Table 2: Principal components from 65 original variables

Principal Component	Description	Main Contributing Variables (correlation between variable and component) Oblique rotation (delta = 0)	Variance Explained by Component	Cumulative variance explained
1	***Families*** (Large - average income)	- No. of cars 2 (.68); 3 (.5) - Married (.63) Dependent children (.35) - Household Income Abve Avg (.632); Avg (.4); High (.31) - Mortgaged dwelling (.55) - Household size Large (.51); Average (.48) - Age structure 0-20yrs (.33); gt 65yrs (-.35) - Distance from major shops (.26)	17.4%	17.4%
2	***Families*** (Disadvantaged)	- Household mtge - Below avg (.83); low (.5); Rent below avg (.54) - Household income – Below avg (.75); low (.47) - Sole parent (.57) - Unemployment rate (.51); Not in labour force (.36) - No. of cars 1 (.43) - Dwelling area (-.39)	11.9%	29.3%
3	***Families*** (Well-off, older & established)	- Same as 1 yr ago (.82); 5 yrs ago (.8); - Owned dwelling (.67) - Married (.48); Dependent children (.65); Age 0-20yrs (.25); 35-54 (.29); 55-65 yrs (.26) - No. of cars 3 (.6); 2(.36) - Amenity (.51) - Household size Large (.46)	6.8%	36.1%
4	***Ethnicity*** (Australian/ European born)	- English only spoken at home (.96) - Place of birth Australian (.69); NW European (.58) - Mortgaged dwelling (.4) - Amenity (.31)	6.3%	42.4%
5	***Families***	- Multiple dwelling (.85)	4.4%	46.8%

	(Poor – multiple dwellings)	- Unemployment rate (.53); Not in labour force (.3) - Sole parent (.45) - Household mortge Low (.39) - Rental dwelling (.35) - Household income (.27)		
6	**Distance from services**	- Distance from GP surgery (.81); Secondary sch (.81); major shop (.71); primary sch (.67); urban shops (.61) - No. of cars 2 (.34) - Mortgaged dwelling (.34) - Household size (.29)	3.9%	50.7%
7	***Dwelling*** (redevelopment potential)	- Dwelling condition (-.67) - Dwelling added value (-.65) - Dwelling age (.44) - Household size (.35) - No. of cars 1 (.32) - Age structure 0-20 yrs (-.27); gt 65 yrs (.21)	3.3%	54%
8	***Dwelling*** Costs (low leverage)	- Household mortgage average (.76) - Household rent average (.6) - Household income above average (.55) - No. of cars 3 (29); 2 (.34) - Married (.38) - Unemployment (-.57); not in labour force (-.4) - Australian born (.25)	3.1%	57.1%
9	***Dwelling Type*** (Rural Living)	- Dwelling type – Rural living (.85) - Site area (.87) - No.of cars 3 (.3); 2 (.29) - Amenity (.27)	2.8%	59.9%
10	***Dwelling Type*** (single dwelling &	- Single dwelling (.85) - Household size Large (.58); Average (.48) - Dwelling area (.52) - No. of cars 2 (.51); 3 (.45)	2.6%	62.5%

	well-off marrieds)	- Married (.45) - Dependent children (.38) - Age 0-20 yrs (.49) - Household income Abv. Avg (.35)		
11	***Dwelling Type*** (brick or stone)	- Rendered (-.93) - Brick (.74) - No. of rooms (.27)	2.5%	65.0%
12	***Families*** (Young – not so well off)	- Age structure 0-20 (.58); 55-65 yrs (-.8); gt 65yrs (-.26) - Mortgaged dwelling (.37); Owned dwelling (-.5) - Household size Large (.37); Small (-.33) - Australian born (.33) - Sole parent (.24)	2.5%	67.5%
13	***Dwelling type*** (brick or rendered)	- Stone (-.89) - Brick (.67) - No. of rooms (.53)	2.2%	69.7%
14	***Families*** (Large, young well – off)	- Age structure 35-54 yrs (.77); 0-20 yrs (.45); gt 65 yrs (-.88) - No. of cars 2 (.59); 3 (.54) - Household income Abv Avg (.6); High (.54) - Mortgaged dwelling (.59) - Household size Large (.58); Average (.56)	2.1%	71.8%
15	***Ethnicity*** (Asian born)	- Scent Asian born (.80) - NE Asian born (.71) - Mortgage - high (.22); above average (.23) - Amenity (.27) - Lone person (.24)	2.1%	73.9%

The description of each principal component (or dimension) is subjective and based on an interpretation of a number of variables which group along the same axis in data space. The interpretation of components is important in understanding the 'sense' of the results and is not a trivial exercise. As stated above, this is why oblique rotation was used to assist in the interpretation. To gain a more comprehensive understanding of the components and the complexity of the residential structure they represent, a closer examination of the contributing variables will be required.

An identified issue in the literature is the overemphasis on individual dwelling characteristics in hedonic housing models constructed to differentiate submarket structures (Bourassa et al., 2002; Watkins, 2001), rather than capturing the underlying residential structure (Rothenberg et al., 1991). The results of PCA (see Table 2) display the complex relationship and interdependency between the 65 individual variables. A good example of this complex interaction of the structural dwelling characteristics and the exogenous dwelling characteristics, which together form a more complete picture of the dwelling structure, is given in the six components broadly described as FAMILIES (components 1,2,3,5,12, and 14 in Table 2). The description FAMILIES is based on age structure, individual characteristics and household size, with the differences between the six components described by other variables including the structural dwelling characteristics (dwelling area in component 2, and multiple dwelling type in component 5). Even when the structural dwelling variables are grouped as components in their own right (components 7,8,9,10,11 and 13 in table 2), they still comprise combinations of many different structural dwelling variables, as well as other variables. An advantage in using DISTANCE FROM SERVICES as a surrogate property characteristic is that it can be considered an enriched form of the original data, to the extent that it includes the relationship between distance as well as the effect of the number of cars, mortgaged status and household size as one characteristic, instead of modelling them as separate characteristics. Again, the results show the interdependence of the individual variables in forming the residential living structure.

The importance of these results is that the individual dwelling characteristics are significantly correlated with many other variables that affect the residential market and form only part of the complex structure in which submarkets exist. Therefore, the use of these components as surrogate property characteristics captures the complexity of the structure overcoming the identified issues and providing a sound basis for further research.

One method of examining these results is to map the resulting components across the study area using geographic information systems (GIS). As each property has a derived component score for each component, these can be classified and mapped showing where properties with like component scores are geographically clustered. If these maps do not make intuitive 'sense' (Tabachnick and Fidell, 1989), then the results are in doubt. The mapping of the PCA components reinforced the appropriateness of the results by displaying an intuitively 'sensible', and expected, spatial distribution for each of the 15 components. Due to publication constraints, the 15 maps cannot be provided in this article, but are available for viewing at www.gisca.adelaide.edu.au/rls.

CONCLUSION

The objective of this paper was to provide a method for the derivation of spatial submarket boundaries that relied on the analysis of individual property data without relying on any form of *a priori* boundary definitions. In achieving this objective, the literature identified two issues. Firstly, not rely on the individual dwelling characteristics, and second, relate the underlying residential structure to the market place. RLS contributes to the resolution of both of these issues by quantifying the underlying residential structure in terms of surrogate property characteristics constructed from attribute data collected at the individual property level and not from an *a priori* unit. This provides the basis for further research in which these surrogate property characteristics can be used as independent variables in a geographically weighted regression model designed to detect changes in submarket boundaries.

RLS calculated for this study comprised 15 components representing the underlying dimensions of the residential structure of the study area, with each component containing a complex mix of property attribute data including socio-economic, amenity, accessibility and structural attribute data. The results make ‘sense’ (Tabachnick and Fidell, 1989) and indicate that the decision making process within the market place is based on much more than just the individual dwelling characteristics, but rather is built upon a complex relationship of variables that are related both in terms of data and geographic space.

New data are becoming available for social science research as land information management techniques improve and computing hardware and software are more able to store and manipulate larger quantities of data in a cost effective and efficient manner. RLS methodology allows the researcher to take advantage of this through the addition of appropriate data to the RLS when available. For example, ‘view from a property’ may be significant when calculating property value. In addition, other property characteristics may be important such as noise pollution, proximity to undesirable land uses and internal dwelling variables such as quality of fixtures, state of modernisation of kitchen and bathrooms, all of which may significantly impact on the marketability of property. Addition of such data may improve the quality of the RLS components and any further research incorporating these derived surrogate property characteristics.

The RLS methodology proposed in this paper provides the ability to improve as more of these data becomes available. It is important to include all relevant data to capture the true dimensions of the residential living structure. In an ideal world, all data would be available for analysis and would reflect reality, but this cannot be the case and modelling can only be built upon imperfect data (both quantity and quality). However, the RLS proposed in this paper provides a sound basis for the further research into the spatial delineation of residential submarkets.

REFERENCES

ABS (2001), Australian Bureau of Statistics, Census of Population and Housing, 2001 Census Community Profile Series. South Australia.

Adair, A., Berry, J. and McGreal, W. S. (1996), Hedonic Modelling, housing submarkets and residential valuation, Journal of Property Research, 13, 67-83.

Barker, K. (2004), Review of Housing Supply Delivering Stability: Securing our Future Housing Needs (Final Report - Recommendations), HM Treasury, London.

Bourassa, S., Hameling, F., Hoesli, M. and Mcgregor, B. (1999), Defining Housing Submarkets, Journal of Housing Economics, 8, 160-183.

Bourassa, S. C., Hoesli, M. and Peng, V. S. (2002), Do Housing Sub-Markets Really matter?, http://cura.osu.edu/research/roundtables/data/Bourassapaper.PDF, 05/09/2006.

Can, A. and Megbolugbe, I. (1997), Spatial Dependence and House Price Index Construction, Journal of Real Estate Finance and Economics, 14, 203-222.

Chhetri, P., Stimson, R. and Western, J. (in press), Modelling the Factors of Neighbourhood Attractiveness Reflected in Residential Location Decision Choices, Policies and Applications, Studies in Regional Science.

Costello, G. (2004) Price Discovery and Information Diffusion in the Perth Housing Market 1988 - 2000, PhD Thesis, In Business School, University of Western Australia, Perth.

Eckert, J. (Ed.) (1990) Property Appraisal and Assessment Administration, International Association of Assessing Officers.

Evans, A. (1995), The Property Market: Ninety Per Cent Efficient?, Urban Studies, 32, 5-29.

Field, A. (2005) Discovering Statistics Using SPSS, SAGE Publications, London.

Figueroa, R. A. (1999), Modelling the Value of Location in Regina Using GIS and Spatial Autocorrelation Statistics, Assessment Journal, 6, 29-38.

Fletcher, M., Gallimore, P. and Mangan, J. (2000), The modelling of housing submarkets, Journal of Property Investment & Finance, 18, 473-487.

Gallimore, P., Fletcher, M. and Carter, M. (1996), Modelling the Influence of Location on Value, Journal of Property Valuation and Investment, 14, 6-19.

Galster, G. (1996), William Grigsby and the Analysis of Housing Sub-markets and Filtering, Urban Studies, 33, 1797-1805.

GISCA (2002), An Index of Metropolitan Accessibility for Adelaide, South Australia, The University of Adelaide,

Goodman, A. and Thibodeau, T. (2003), Housing market segmentation and hedonic prediction accuracy, Journal of Housing Economics, 12, 181-201.

Grigsby, W. (1963) Housing Markets and Public Policy, University of Pennsylvania Press.

Hair, J., Anderson, R., Tatham, R. and Black, W. (1992) Multivariate Data Analysis with readings, Maxwell Macmillan.

Jones, C., Leishman, C. and Watkins, C. (2004), Intra-Urban Migration and Housing Submarkets: Theory and Evidence, Housing Studies, 19, 269-283.

Kestens, Y., Theriault, M. and Des Rosiers, F. (2002) House Prices and Environmental Quality: Combining GIS, Image Analysis and Hedonic Modelling, Sixth Annual Integrating GIS/CAMA 2002 conference, Reno, Nevada.

Kline, P. (1994) An Easy Guide to Factor Analysis, ROUTLEDGE.

Lockwood, A. (2003) Analytical Derivation of Residential Real Estate Sub-Market Group Boundaries., Masters Dissertation in Spatial Information Science, In Discipline of Geography and Environmental Studies, The University of Adelaide, Adelaide.

Maclennan, D. and Tu, Y. (1996), Economic Perspectives on the structure of Local Housing Systems, Housing Studies, 11, 387-405.

McCluskey, W. J., Deddis, W. G. and Lamont, I. G. (1999) The Application of Surface Generated Interpolation Models for the Prediction of Residential Property Values, The Cutting Edge, RICS,

McCluskey, W. J., Deddis, W. G. and Lamont, I. G. (2002) The Application of Spatially Derived Location Factors within a GIS environment, PRRES conference, Christchurch, New Zealand.

Meen, D. and Meen, G. (2003), Social Behaviour as a Basis for Modelling the Urban Housing Market: A Review, Urban Studies, 40, 917-935.

Meen, G. (2001) Modelling Spatial Housing Markets. Theory, Analysis and Policy, Kluwer Academic Publishers.

Pryce, G. (2004), The Micro and Macro Effects of the Location of New Housing Supply, Office of the Deputy Prime Minister, London.

Reed, R. (2001), The Significance of Social Influences and Established Housing Values, Australian Property Journal, May 2001, 524-527.

Rossini, P. and Kershaw, P. (2005) Using a Hybrid Automated Valuation Model to Estimate Capital and Site Values, Eleventh Annual Pacific-Rim Real Estate Society Conference, Melbourne, Australia.

Rothenberg, J., Galster, G., Butler, R. and Pitkin, N. (1991) The Maze of Urban Housing Markets, University of Chicago Press.

Schnare, A. and Struyk, R. (1976), Segmentation in Urban Housing Markets, Journal of Urban Economics, 3, 146-166.

Tabachnick, B. and Fidell, L. (1989) Using Multivariate Statistics, Harper & Row, New York.

Thrall, G. I. (2002) Business Geography and New Real Estate Market Analysis, Oxford.

Watkins, C. (1999), Property Valuation and the Structure of Urban Housing Markets, Journal of Property Investment & Finance, 17, 157-175.

Watkins, C. (2001), The Definition and Identification of Housing Submarkets, Environmental and Planning A, 33, 2235-2253.