Residential Property Prices in a Sub-Market of South Africa: Separating Real Growth from Attribute Growth

MICHAEL ELS AND DIETER VON FINTEL

Stellenbosch Economic Working Papers: 14/08

KEYWORDS: HEDONIC PRICING, HOUSING MARKET, GROWTH RATES
JEL: C21, C23, R31

MICHAEL ELS
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
DEPARTMENT OF ECONOMICS
DUKE UNIVERSITY
E-MAIL: MICHAEL.ELS@DUKE.EDU

DIETER VON FINTEL
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: DIETER2@SUN.AC.ZA

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

Residential Property Prices in a Sub-Market of South Africa: Separating Real Growth from Attribute Growth

MICHAEL ELS[1] AND DIETER VON FINTEL[2]

ABSTRACT

This paper analyses the South African residential housing market using hedonic price theory. It builds and tests pooled OLS, fixed effects OLS, pseudo-panel and quantile regression models. The main findings are in agreement with most modern related literature. This paper highlights how house price growth rates have been calculated incorrectly due to the changing aggregate house sold every year. It calculates more accurate growth rates for the property market, yielding surprisingly different growth patterns from those originally thought. It illustrates that much of the recent house price growth was caused by attribute inflation rather than pure price inflation. It also shows that most of the pure inflation occurred at the bottom end of the market while most of the attribute inflation occurred at the top end of the market. Furthermore, it shows that house price determinants change across the house price distribution

The data used was sourced from the Residential Property Price Ranger and covers 1930 house sales measured half yearly over three years; from 1 September 2004 to 31 August 2007. These sales were recorded in the towns of Stellenbosch, Somerset West, Strand and Gordon's Bay.

Keywords: : Hedonic pricing, Housing market, Growth rates

JEL codes: C21, C23, R31

[1] Department of Economics, Stellenbosch University and Department of Economics, Duke University, michael.els@duke.edu

[2] Department of Economics, Stellenbosch University, dieter2@sun.ac.za

1 Introduction

Purchasing property is often the biggest investment most people will make in their lives and it is one of the key investment areas for large financial institutions. It is therefore important to find an accurate means of assessing the value of such property. Since the market measures the value of a property only at the time of sale, the data available for price estimates is limited. It is important to create price models for the measurement of property values between transactions, as many parties hold an interest in the value of property during this timeframe.

A popular method used for this purpose is hedonic price modelling, whereby property is valued according to its characteristics. From this type of model it is then possible to estimate property prices based on attributes and to assign each of these an implicit price. Research into hedonic pricing, in the context of the property market, has surged over the last decade, and so has the development and refinement of hedonic theory in general. A major trend in current geographically linked hedonic research has been on how to deal with the spatial correlation between units of observation: this is a topic also confronted in this paper.

The main aim of this paper is to create a hedonic price model for the geographic area of Stellenbosch, Somerset West, Strand and Gordon's Bay, an area to the east of the Cape Town metropole in South Africa, for the time period from September 2004 to August 2007. It also discusses the main techniques used in building such a model, as well as each of their problems. This paper then goes on to establish the drivers of property price growth. Once controlling for attributes, it is evident that growth in this market is not as impressive as often reported. Much of property's value has therefore been derived from improvements in attributes rather than true growth in the market. These effects are furthermore investigated in the context of market segments to uncover how the property market has performed along these different sub-classifications.

The following section reviews the South African residential property market and provides the motivation for the study. The third section of this paper covers some of the foundations of hedonic theory and it describes the premises of such a model. The fourth section is a literature review briefly covering some influential studies in hedonic property pricing. It reviews many of the attributes which have been found to be important in previous hedonic studies, and also discusses the alternative statistical techniques which can be used for this type of study. The fifth section describes the data used, as well as its limitations. The sixth section discusses the modelling techniques applied. They include OLS, quantile regression, spatial statistics and pseudo-panels. The seventh section discusses the empirical findings of these models, and the eighth section concludes.

2 The South African Residential Property Market

In all countries with established property rights, residential property is a large part of many investment portfolios of companies and is naturally also important for individual consumers. It is thus imperative that proper analysis and dissection of such a market is carried out.

Two well-known property market analysts operate in South Africa: ABSA and Rode Property Valuations, although many others exist. Both of these firms have similar measurement methodologies. Rode's House Price Index use median sales prices of suburbs divided into price categories using data obtained from the Deeds Office (Rode 2005: 156). ABSA's House Price Index tracks the average sales price of properties categorized by house size using data from finance applications which they have received (Rode 2005: 156).

Both these research institutions' indices have shown massive growth. The sales price of residential property in South Africa since 2000 is in the region of 10% and more per year (nominally) with signs of a slowdown in late 2007 and 2008 (Rode 2005: 157 & ABSA 2008: 2). In a similar study on international house price trends, Shiller shows that since the mid to late 1990's there has been a sharp increase in house price growth rates in the USA, UK, the Netherlands and Norway in a similar magnitude through to 2006 (2007:41-46). This growth has altered since late 2007 and 2008, analysts are speculating about a market crash in residential property prices, as sudden declines in house prices is being experienced globally (ABSA 2008:2).

The techniques currently used are not sufficient to correctly analyse the growth trends in the market and treats all houses as homogenous units while they clearly are not. Therefore using ABSA's and Rode's methodologies is necessarily not the optimal empirical strategy to come to the given conclusions about property price growth in South Africa. This study challenges the current market methodology for valuing residential property in South Africa and the proclaimed growth rates in residential property. In particular, raw growth rates do not differentiate between true increases in value as opposed to attribute inflation. Hedonic models are able to isolate the effect of measurable attributes by controlling for their individual effects over time, allowing ‘pure’ property inflation to be measured – as is illustrated in section 7.2.1. This study uses microeconomic data, which allows us to control for these attributes and the heterogenous characteristics of individual properties. Most South African property studies use ABSA and Rode’s indices and analyse aggregated prices by region and sub-market over time, rather than focussing on individual sales. In these studies it is assumed that each property provides a similar, homogenous “housing service” (Burger & Janse van Rensburg, 2008: 292), rather than focussing on the utility derived from individual characteristics of houses. Burger and Janse van Rensburg (2008) use panel unit root tests to show that in the middle price segment, middle-sized and large houses each constitute respective “single markets” across metropolitan areas of South Africa, while smaller units form separate markets across regions. Gupta & Das (2008) find similar conclusions estimating VAR and Spatial Bayesian VAR forecasting models. This study focusses on one geographical sub-market of South Africa for which micro-level data was obtained. Given the conclusions of the above studies, there is a case to be made that the results presented below can cautiously be extrapolated to the rest of the South African property market.

3 Hedonic Theory

As previously mentioned, the basis of this study is founded in what is known as hedonic price theory. The expression ‘hedonic’ comes from the word ‘hedonism’ in Greek philosophy which pertains to or involves pleasurable or painful feelings or effects. A hedonic study is thus one which aims to measure something by virtue of its inherent pleasures and pains (Hidano 2002:1-2). A key reason for using the hedonic method in this study is that it is not limited to *homogenous goods*.

When analysing a standard competitive market for a *homogenous good* one can rely on supply and demand forces to reach equilibrium at the optimal price (Day 2001:23). When analysing the housing market, however, it is not as simple because one is then dealing with *differentiated goods*.

When considering this supply and demand in hedonic theory, landlords are equivalent to producers while households wishing to reside in these properties are to be considered the consumers. Each property j is then described by a vector $\mathbf{z}$ of quantifiable and inseparable attributes which determines its price[3]:

$$\boldsymbol{z}_j = (z_{j1}, z_{j2}, z_{j3}, \dots, z_{jk})$$

Thus when a household chooses a particular property j, they have in fact chosen a vector $\boldsymbol{z}_j$ of attributes to purchase (Day 2001: 23-4).

Freeman (1993:371) uses the analogy of consumers considering the housing market as a huge supermarket with many different goods, but the consumer can only choose from prefilled shopping carts. Consumers can thus only increase the quantity of one good in the shopping cart if they can find a cart where all the other goods are still provided in exactly the same quantity.

The price of one of Freeman's 'shopping carts' is determined by its vector of attributes:

$$P_j = P(\boldsymbol{z}_j) = P(z_{j1}, z_{j2}, z_{j3}, \dots, z_{jk})$$

where more positive attributes increase the price and more negative attributes decrease the price, *ceteris paribus*. This is derived from the standard microeconomic theory of implicit prices (Lancaster 1966). One can thus interpret the price of property j as a function of its attributes $\boldsymbol{z}_j$. This function is now defined as a hedonic price function, as it is determined by the good's various attributes.

As not all households (or consumers) are purchasers of property, the link between the sales market and the rental market is the long term equilibrium assumption that the purchase price equals the present value of future rents. Purchase price $(P) = \sum_{t=1}^{T} \frac{P(\boldsymbol{z})}{(1+d)^t}$; with time index t, over total lifetime T and discount rate d.

It is now possible, through regression analysis, to calculate implicit prices for each characteristic i of property j. This is represented as:

$$P_i(\boldsymbol{z}_j) = \frac{\partial P}{\partial \boldsymbol{z}_i}$$

[3] It is assumed that all the properties in one geographical area represent the products in that property market. The consumers are the households residing in that area and the producers are the property owners in that area (Day 2001:23).

This price, $P_i(\boldsymbol{z}_j)$, is considered an implicit price from Lancaster's (1966) theory as there is no direct market for the attributes $\boldsymbol{z}_j$. One could infer that this price represents the value added to a property for a unit increase of a given attribute. This represents the value to investors wanting to maximise returns, as they could weigh up the costs of expanding a property with an attribute, against the gains in sales price of having the additional attribute.

One is now able to view the hedonic function in the linear form (Hidano 2002:2):

$$\begin{aligned} &value\ of\ good \\ &\qquad = (value\ of\ attribute\ 1) \times (quantity\ of\ attribute\ 1) \\ &\qquad + (value\ of\ attribute\ 2) \times (quantity\ of\ attribute\ 2) + \ \ldots \\ &\qquad + (value\ of\ attribute\ n) \times (quantity\ of\ attribute\ n) \end{aligned}$$

To continue with Freeman's (1993:371) analogy: by adding a good to the 'shopping cart,' the price of the basket will rise. However, as is standard in economics, one might witness diminishing marginal returns when adding goods to the 'shopping cart.' That is, the jump in the price of a house with 3 bedrooms to a house with 4 bedrooms will not necessarily be the same as the jump in price from a 4 bedroom house to a 5 bedroom house, *ceteris paribus*. That is:

$$\frac{\partial^2 P}{\partial z_k^2} < 0$$

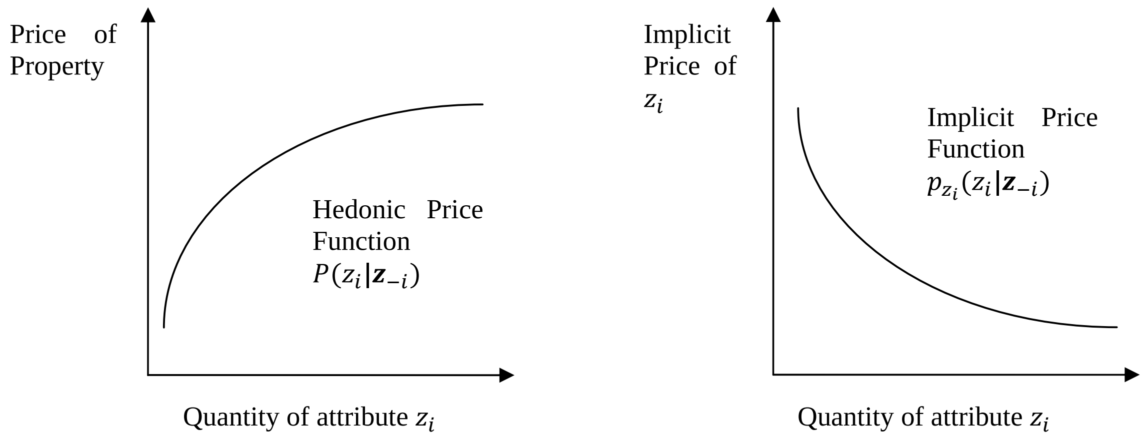
These diminishing marginal returns are shown in Figure 1 below by the dampening of the hedonic price function for a given attribute z_i and by a decreasing implicit price of attribute z_i.

The second implication obtained from extending Freeman's (1993:371) theory is that the price of one characteristic may depend on the quantity of other characteristics. This implies that the value of an additional bedroom is related to the number of bathrooms, kitchens and other attributes which the property already possesses. This is represented as:

$$\frac{\partial P}{\partial z_i} = p_{z_i}(z_i|\boldsymbol{z}_{-i})$$

where $p_{z_i}(z_i|\boldsymbol{z}_{-i})$ is the marginal price of attribute z_i given all other attributes $\boldsymbol{z}_{-i}$ (Day 2001:28). The empirical model below links well with this theory, as the coefficients of the model represent implicit prices – after controlling for the other characteristics that are important for determining the price of the property.

Figure 1 Hedonic Price Function and Implicit Price Function for attribute z_i



Source: Day (2001:28)

This theoretical analysis extends beyond the requirements of this paper. The core of the theory is that both suppliers of property and consumers of property will be in equilibrium along the hedonic price function in Figure 1 (Day 2001:28). Any price above the curve would not be paid by a consumer, while similarly, any price below the curve would not be offered by any supplier. The analysis naturally extends to all dimensions covering every attribute. The theoretical analysis is also based on solid micro-foundations with utility maximising consumers and profit maximising suppliers[4].

The hedonic approach is the technique of estimating these implicit prices. This approach is economically strong as it is based on the revealed preferences of producers and consumers in actual market conditions. Revealed preference techniques are strong in that they measure what consumers have actually done or paid rather than their stated preferences of what they would be willing to do or pay.

Hedonic theory does, however, also rely on the assumption of *perfect information*. For markets to function correctly, each consumer and producer needs to be fully informed as to

[4] This is not discussed further here, but a formal discussion is provided in Rosen (1974), Epple (1987) and Day (2001).

the true value and costs of each of the characteristics jointly as well as separately (Hidano 2002:3).

From an econometric perspective, it is worth briefly mentioning that should one of these characteristics be missing from the empirical model or should the consumers not be aware of it, then there will be omitted variable bias in the model (Greene 2003:148-9). This is particularly true if the omitted attribute is a strong complement (strongly positively correlated) or substitute (strongly negatively correlated) to an included attribute. By implication, misspecification will lead to biased estimates of implicit prices.

Hedonic theory further assumes *costless mobility*. Consumers must be able to freely move from one good to another should circumstances in the market change. This is necessary to ensure that true equilibrium prices are always attained (Hidano 2002:3). This is clearly not the case in practice and is likely to create noise in the estimation process However, the markets observed are large enough to diminish this problem.

For further reading on hedonic theory, Bunzel (2003) provides an overview of the most influential papers and theories. Most empirical research has been similar to this paper in that it estimates implicit prices from market data. However, researchers could model explicitly the utility functions of consumers if household data were also available.

The hedonic price method's first and greatest advantage is that it is based on actual market choices made by economic agents and hypothetical statements. Secondly, the property market is relatively large, so one might expect a reasonable amount of competition, large sample sizes and on average accurate valuations. Thirdly, property records tend to be particularly reliable and readily available, as real estate is big business in most countries which presents the motive to maintain accurate databases.

Its most prominent disadvantages are: firstly that neighbourhood characteristics cannot easily be isolated. There are many common characteristics endemic to a particular neighbourhood, but limited means to isolate these effects. Secondly, these neighbourhood characteristics can only be measured to the extent to which people are willing to pay for them in property transactions and to the extent to which they are even aware that any such characteristics exist. Thirdly, it assumes that the exact combination of property characteristics can be purchased. This may not be as problematic as it may at first seem, due to the size of the property market. It should not be too difficult to find a house approximating the desired bundle of characteristics. Fourthly, the method is relatively complicated and a certain level of statistical knowledge is required for implementation, as its interpretation and correct choice of functional form is not straightforward. Lastly, although the data is readily available, it still requires ‘cleaning’ and a large sample.

4 Hedonic Property Market Models - A Review

Hedonic theory first appeared when Griliches (1961) developed an econometric application for the valuation of automobiles. He termed the method hedonic price estimation because the method used the automobile characteristics to determine motor vehicle prices (as discussed earlier). This was the first groundbreaking research that paved the way for future hedonic price estimation.

Rosen (1974) was one of the first academics to adapt Griliches (1961) theory of hedonic pricing in order to analyse housing markets. The core of Rosen’s (1974) theoretical work is summarised in Section 2 above. Epple (1987) extended and refined the more technical aspects of Rosen’s work regarding hedonic theory, identification and estimation methods. It is Epple’s extensions that most researchers have applied and adapted when conducting empirical research.

Beyond the general extensions to hedonic theory, the most profound insight in the context of property markets has been that of spatial autocorrelation. Spatial autocorrelation exists when there is some form of cross-sectional dependence related to the geographical location of the observations (Anselin 2006: 902). Thus, the covariance between observations is determined by the geographic location of observations, which violates the classic OLS assumption of uncorrelated error terms.

Bourassa *et al.* (2007:146) discuss the link between spatial dependence and housing submarkets. This concept of housing submarkets implies some form of substitutability. Pairs of goods are likely to be substitutes when they share similar characteristics, and when the price of one goes up, the price of the other tends to increase also. Hence the spatial autocorrelation problem is very closely related to the concept of housing submarkets.

Bourassa *et al.* (2007:146) argue that this spatial autocorrelation of the error terms is more likely to occur within submarkets than across them[5]. They suggest controlling for submarkets

[5] Gupta & Das (2008) investigate the spatial dependency of housing markets in South Africa by estimating Spatial Bayesian VAR models across six metropolitan areas in South Africa. They conclude that spatial dependence is only important for large middle segment houses, while other sub-markets do not conclusively

with dummy variables or estimating separate regressions for each submarket to deal with the problem. This implies that one must have some predefined submarkets or a means to define such submarket. Geographical areas are typically used as predefined submarkets. Palm (1978:211) defines submarkets according to real estate definitions while Bourassa *et al.* (2003:12-13) have defined them according to buyers' value. By controlling for geographic areas in a model, one can drastically reduce the problem of spatial autocorrelation.

Approaching the spatial autocorrelation problem with formal spatial statistics is another way to address this problem. There are a number of methods employed to deal with the phenomenon. However, the spatial autoregressive process (SAR) is by far the most common and also the oldest method. It was first put forward by Whittle (1954). This method breaks up geographical areas into a regular rectangular lattice. Using the cartographic coordinates of the properties, each error component is the sum of its own true error and the errors of surrounding properties. This basic model has been studied in great detail and many forms exist. It is, however, computationally intensive and it requires the exact geographic coordinates of each property in the study.

These methods utilise spatial weighting matrices when dealing with the spatial autocorrelation problem. This does not involve weighting each area by itself with a dummy variable, but rather assigning each neighbouring area a certain weight. This is a controversial procedure, as there is no clearly defined weight that each neighbour should take. A variety of methods have been used, such as contiguity, whereby each area i gets the weight $1/n_i$ (where n_i is the number of neighbours of i), and distancing, whereby each observation gets a relative weight to every other observation based on geographical or economic distance in a set functional form (Anselin 2002:256-260).

Can (1992) makes a theoretical argument that clarifies the relationship between a spatial statistics approach versus a submarket approach. She defines two types of locational effects. The first is *adjacency effects*, which encapsulate all the characteristics associated with the immediate location of the property. This type would be modelled using spatial statistics. The second type is that of *neighbourhood effects*, which would encapsulate the fixed effects of the

display this phenomenon. The analysis remains silent on spatial dependency *across* sub-markets, though this is a less likely source of concern. Burger & van Rensburg (2008) offer a possible explanation for spatial dependency in the large housing sub-market: it is possible that these properties are not acquired by individuals, but by institutional investors who rent these properties out rather than acquiring them for own use. The heterogneity (in price and attributes) among these types of properties across regions, suggests that arbitrage will occur, particularly as these transactions do not necessarily require the investor to be mobile, as would be the case if the property was acquired for personal use.

characteristics of a particular neighbourhood that affect property demand for that area. These effects can represent anything that sets one neighbourhood apart from others, such as good views, high quality schools, low crime rates, quiet surroundings or even being more socially prestigious. Wang and Li (2004:69) have found that these neighbourhood effects are far more important than individual property characteristics when choosing a property to purchase. Such neighbourhood effects can be modelled for, as Bourassa *et al.* (2007) suggest, with dummy variables, and they directly represent the premium paid to live in a given area.

Due to data constraints in this study, no spatial statistics can be used[6]. This is, however, not as problematic as it seems, as Bourassa *et al.* (2007) have shown that by controlling for submarkets properly in the analysis, one can obtain more interpretable results using a simpler model. These authors show that the neighbourhood effects predominate and thus matter more than adjacency effects. They support this hypothesis by testing a submarket OLS model against four of the most popular spatial econometric models in use[7]. They use a large sample of approximately 5000 observations and remove 20% of them for an out of sample test after estimating the model. Their results indicate that a well designed OLS model can improve upon any of the spatial econometric models while being significantly simpler.

Bourassa *et al.* (2007) also ran separate hedonic models using OLS for each of the submarkets in order to analyse the consistency of the estimated coefficients. They find that the coefficients are indeed area dependent. This should, however, not be all that surprising, as many areas have houses with roughly the same characteristics and it is unlikely that the marginal effect of explanatory variables will always be linear in nature.

Zietz *et al.* (2007) had a similar idea when they constructed a hedonic model using quantile regression. The hypothesis under consideration was that a part of the variation in the implicit prices of the characteristics differs across the distribution of property prices. Their study indeed showed that consumers in different house price categories value the same characteristics differently.

Other authors have highlighted several issues of slightly lesser importance but which still do add value to the empirical analysis. One is the matter of school quality. Chiodo *et al.* (2005) review and extend upon the recent attempts of measuring the value added to residential

[6] Spatial statistics usually require geographical coordinates unique to each observation or at least some other identification lattice unique to the latitude and longitude of each observation. These were not available for the housing under investigation.

[7] Bourassa *et al.* (2007) compare their OLS output to SAR and CAR models as well as two weighted geostatistical models with different weighting functions.

property prices by ‘good’ schools. Many authors (particularly in the United States) have shown conclusively that areas with ‘good’ schools do indeed have higher house prices, but not as much as initially thought[8].

A second characteristic which Kaufman and Cloutier (2006) have shown to significantly influence residential property prices is that of environmental quality. These authors utilise hedonic regressions and include the variables ‘brownfields’ and ‘greenspaces’ to directly measure the value lost when land is polluted as well as value added for well preserved areas of land[9].

The crux of the previous two arguments, as well as of the neighbourhood effect of Can (1992), stems from Tiebout’s (1956) paper where he argues that consumers “vote with their feet”. Here this is applied in the residential property market where it is argued that people will move to neighbourhoods and pay the neighbourhood premium if they believe that all the other offsite characteristics are worth the premium. Any model which includes fixed effects for neighbourhoods will capture all such neighbourhood effects.

Another issue brought to light by Bourassa and Peng (1999) is that of society’s own aesthetic opinions, traditions and superstitions. They conducted a hedonic study in Auckland, New Zealand, in an area which is predominantly Chinese and where *feng shui* is regarded highly. According to such beliefs, certain numbers are considered to be unlucky while others are considered to be very lucky. These authors assumed that residents with significantly more expensive or cheaper homes did indeed pay a premium for ‘lucky’ numbered houses or were in fact subsidised for ‘unlucky’ numbered houses due to the correlation between market prices and house numbers.

A final matter complicating any hedonic study done over time is that of the relationship between real estate markets and stock markets. Okunev *et al.* (2000) have shown that in many circumstances there was strong unidirectional causality from the stock market to the real estate market in the United States of America from 1972 to 1998, which is also consistent with any structural breaks in the data. Such a link could greatly help to explain the

[8] Black (1999) has shown that due to the reverse causality between good areas and good schools there is in fact an endogeneity problem in such hedonic regressions and that the effect, although still statistically significant, is exaggerated.

[9] Brownfields are defined by the U.S. Environmental Protection Agency as “real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant.” (www.epa.gov/swerosps/bf/glossary.htm#brow). Greenspaces are generally accepted to mean any piece of land which is ‘healthy’ in terms of natural well-being.

movements of house prices over time, as could other financial indicators. It would, however, be controversial to include any such variables in hedonic studies as stock market movements and other financial indicators are hardly considered to constitute ‘property attributes’.

5 Data

The data used in this study has a unique set of attributes which does not lend itself to traditional time series and panel analysis. However, as described in the following section, it is both possible and necessary to account for both time and fixed effects.

The data is taken from house sales over three years, giving the data a sporadic time series element, as houses are purchased and sold at relatively random intervals. However, it also contains details about each property sold, giving it a cross sectional element. Since each house is sold only once, the observations cannot, however, be followed over time, so no panel or times series of any sort exists. The dataset was obtained from Residential Property Price Ranger (RPPR). RPPR collect data from confirmed sales from subscriber real estate agencies – which encompass almost all market sales in the areas concerned and thereby minimises possible sample selection bias along this dimension (though potential for this phenomenon may still be prominent along other dimensions, as discussed below).

Thus, the data’s generating process is of the form:

$$Y_{ijt} = \beta X_{ijt} + \mu_i + \lambda_t + \eta_{ij} + \varepsilon_{ijt} \quad i = 1, \dots, N;\ j = 1, \dots, K; t = 1, \dots T$$

where Y_{ijt} is the selling price of property j in area i which was sold in time period t. β represents the vector of coefficients of the effects that each variable in X_{ijt} has in determining property selling prices. μ_i represents the individual area effect of a neighbourhood and the aggregate house selling price in that area. λ_t represents the aggregate time trend in house prices, as experienced by the entire market. η_{ij} represents the individual house effects within various neighbourhoods. Lastly, ε_{ijt} represents the idiosyncratic error term. An important point to make is that it is impossible to model η_{ij} or some τ_j directly to represent any house level fixed effects, because each house is only observed once in the dataset (at the point of sale) – adding a fixed effect at the house level would exhaust degrees of freedom in hypothesis testing. It is, nevertheless, possible to use this mixed dataset to trace aggregate time trends, despite not following the same properties over time. Here the assumption is that a similar set of houses to represent each area are analysed in the different years, even though each cohort is represented by different units of observation across time[10].

The dataset required a moderate amount of 'cleaning' and consisted initially of 2054 observation. This 'cleaning' primarily entailed the removal of duplicate and incomplete observations. It also included the removal of golf estate properties as they are too few to form their own submarket and are clearly outliers in terms of price[11].

The final data covers 1930 house sales measured half yearly over three years; from 1 September 2004 to 31 August 2007. These sales were recorded in the towns of Stellenbosch, Somerset West, Strand and Gordon's Bay. These are all neighbouring areas that form a large property sub-market close to the Cape Town metropolitan area in South Africa. They also cover all income brackets from lower class to higher upper class.

These towns were sub divided into 32 areas of analysis (i). Areas were chosen firstly by geographic criteria (neighbours separated by main roads or other natural boundaries) and secondly by means of sale prices within each suburb. This is to ensure that each group of houses shares similar characteristics to take advantage of any group level fixed effects (μ_i) that may exist.

The full list of explanatory variables available in the database, area groupings and maps of areas are included in Section 1 of the Appendix.

It must also be understood that when comparing property attributes, quality is not taken into account. This implies that a house with three luxurious bedrooms will be considered equivalent to one with three rundown bedrooms. The "condition" variable attempts to control for part of this problem. However, it is not possible to control for subjectivity of different agents. This variable has been omitted from the analysis for this reason and due to a lack of statistical significance. The Spearman's rank correlation between the area dummies and the condition dummies show that there is also correlation between these variables which explains its lack of explanatory power.

Furthermore, because the data is only constructed from properties that are sold in the time period under consideration, there is a strong chance of sample selection bias (Heckman

[10] Failure of this assumption may result in sample selection bias, particularly if highly priced houses with better attributes are sold later in the sample period as a result of changing preferences, rather than a real change in the average house in each area.
[11] The Spier and the Erinvale Golf Estate were excluded.

1979:153-4). Munneke and Slade (2000) conducted an empirical study into the sample selection bias in this kind of property data in Phoenix in the United States. They constructed an investment based model to determine property prices and compare the property values in the rental market to the property values in the sales market. According to hedonic theory, both markets should value housing attributes equally. They find that there is indeed sample-selection bias, but that the difference between corrected and uncorrected property price indices is not statistically significant. However, Gatzlaff and Haurin (1998) conducted a similar study, but in Florida, USA and found that this difference was indeed statistically significant. It is important to be aware of this potential problem for this paper, although it is beyond the scope of this paper to investigate this problem, due to the absence of information on the representative sample of properties that are not sold in the sample period.

A significant virtue in this analysis is that there is minimal measurement error, as the characteristics of the houses are well documented and accurate. The only possible concern with regard to measurement error is that the end sales price of the houses may not be equivalent to that of a truly competitive market – a requirement mentioned in hedonic pricing theory. But unless the measurement error in sales price (being the dependent variable) is correlated with any of the explanatory variables, the measurement error will be absorbed into the idiosyncratic error term. This leaves the estimated coefficients unbiased and consistent (Wooldridge 2002: 71).

6 Methodology

As far as empirical methodology goes, there is little consensus in the literature as to what is acceptable and what is not. Four methods will be used here. The first method discussed is pooled ordinary least squares (OLS). The second method controls for area dummies in an OLS model, which mimics a fixed effects panel model. The third method involves a pseudo-panel which has been adapted from other areas of study in an attempt to find consistent estimates. The final method is quantile regression and is also somewhat new in this context. It must also be mentioned that one of the most common methods, that of modelling property distances explicitly in the error term (such as SAR), has been omitted. This is primarily due to data constraints (as full geographic coordinates are required) and secondly due to Bourassa *et al.'s* (2007) convincing illustration that controlling for area fixed effects is far simpler and provides at least as good an estimator.

6.1 ***Pooled OLS***

The best place to begin most empirical studies is with the standard pooled OLS model where one completely ignores the time dimension and any fixed area effects in the data. This model assumes the structure of the data to be:

$$Y_{ijt} = \beta X_{ijt} + \gamma_{ijt} \quad for \quad i = 1, \dots, N;\ j = 1, \dots, K; t = 1, \dots T$$

where $\gamma_{ijt} = \mu_i + \lambda_t + \eta_{ij} + \varepsilon_{ijt}$

and Y_{ijt} is the log of sales price (taking the log is standard in the literature, as it gives a higher R-squared and is a stabilising monotonic transformation).

Thus one has a much simpler model which can be estimated using OLS where all the standard assumptions should apply, the 'key' assumption being $E(\boldsymbol{x}'\gamma) = 0$ (Wooldridge 2002:52). If this assumption is violated then the estimates are biased and inconsistent. If there is any correlation between μ_i, η_{ij} or λ_t and X_{ijt}, then β will be biased and inconsistent (Wooldridge 2002: 256). One would in fact expect to find some trace of these correlations as house prices, as well as housing attributes, have been increasing rapidly over the sample period, implying that $E(\boldsymbol{x}'\lambda) \neq 0$ (Du Toit 2007:2). The literature review also suggests that there are definite area effects in determining house prices which will cause $E(\boldsymbol{x}'\mu) \neq 0$ due to correlations between house size and other attributes by area. The *a priori* expectation, therefore, is that this strategy will lead to biased and inconsistent estimates.

In estimating such a model, the standard OLS estimator is used:

$$\hat{\beta} = (X'X)^{-1}(X'y)$$

with variance[12]:

$$Var(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$$

6.2 ***Group level fixed effects***

This model moves one step closer to approximating the true data generating process by adding dummy variables for each area, as well as for a time trend. This, in effect, gives one the mathematical equivalent of a "within" fixed effects estimator of the model (Wooldridge 2002: 273), even though the data is in essence far from a balanced panel, as repeated sales for the same property are not common over the timeframe.

[12] Robust standard errors were used for all models tested.

It assumes the form:

$$Y_{ijt} = \beta X_{ijt} + \mu_i + \lambda_t + \omega_{ijt} \quad i = 1, \dots, N;\ j = 1, \dots, K; t = 1, \dots T$$

where $\omega_{ijt} = \eta_{ij} + \varepsilon_{ijt}$

Here ω_{ijt} is a new idiosyncratic error term, μ_i is again the area effect and λ_t is again the time trend. The latter are modelled by the dummy variables.

Thus, this model can incorporate significantly more information, which, theoretically, should matter a great deal in the estimation output. However, there remains the problem that if there is any correlation between η_{ij} and X_{ijt}, then β will be biased and inconsistent. It is difficult to give any definite *a priori* reasoning to explain whether this correlation exists or not and it is not possible to test for it, as the houses in the sample are not followed over time. The area effects have been assumed to be time invariant to absorb all unobserved and constant area effects. There is now allowance for correlation between μ_i and X_{ijt} as well as between λ_t and X_{ijt} without losing in consistency or gaining in bias (Wooldridge 2002: 268). However, should these effects be time variant, then their full effect will not be caught by the fixed effect coefficients and whatever is remaining will be absorbed by the idiosyncratic error term; hence a risk of endogeneity bias remains[13].

This model effectively represents the 'superior' one put forward by Bourassa *et al.* (2007), when compared to those defined with spatial statistics, and does so without the truncation bias of running separate regressions for each area (Heckman 1979). It incorporates Can's (1992) notion of neighbourhood effects rather than adjacency effects to control for spatial correlations. This model then gives a measure of area value (or premia) that people are willing to pay for. It thus gives empirical clout to Tiebout's (1956) notion of voting with ones feet and estimates the true value of each neighbourhood under the assumption of perfect competition in the property market of this region.

6.2.1 Testing for spatial autocorrelation

As mentioned previously, little can be done directly in terms of spatial statistics with this dataset. However, by also aggregating the residuals of respective models by area in each time period, and by assuming that the number of properties sold in each time period per area is similar, one can apply some basic tests for spatial autocorrelation.

[13] This is a considerable risk in the current dataset, as the units of observation representing each area in subsequent time periods differ.

This new data set consists of 191 observations[14]: the average property price and characteristics for each area. It is then possible to construct a spatial weighting matrix as each observation now has a uniquely defined geographic position with definable neighbours. With this data setup it is possible to conduct Moran's I test for spatial autocorrelation (Anselin 2006:932-3):

$$I = \frac{(e'We)/S_0}{(e'e)/n}$$

where e is typically the OLS residuals[15], W is the spatial weighting matrix and S_0 is a normalizing factor relating to W[16]. It thus tests whether one area's errors are related to any of its neighbour's errors.

The test relies on the asymptotic normal approximation of this statistic[17]. The null hypothesis of this test is that no spatial autocorrelation exists. It tests whether e is normally distributed and it tests whether this distribution is random. This test will show whether or not there was indeed any spatial autocorrelation and whether the group level fixed effects model was able to eliminate it.

6.3 ***Fixed effects and random effects pseudo panel***

Since each house is observed only once, no true panel data exists for this type of study. But one can construct a pseudo-panel, first proposed by Deaton (1985:109-110). Pseudo-panels are created by aggregating sets of cross-section observations according to a set of characteristics over time to give the desired panel structure. This technique adds a dynamic view to the data which the periodic cross-sections cannot.

One can achieve this by averaging the data over individual houses and then using neighbourhoods as units of observation of average house characteristics for each time period. This will then give a panel data structure of the form:

$$\overline{Y}_{it} = \beta\overline{X}_{it} + \delta_i + \lambda_t + \xi_{it} \quad i = 1, \dots, N;\ t = 1, \dots T$$

[14] The number of areas multiplied by the number of time periods, $32 \times 6 = 192$, less one area with no observations in one time period.
[15] In this case it represents the OLS residuals, aggregated by area
[16] The weighting matrix is structured by creating a contiguity lattice of neighbouring areas weighted by $1/n_i$, as discussed in section 4.
[17] This is somewhat worrying as 191 observations are relatively few to invoke asymptotic theory. Hence the test should be seen as only mildly informative and not decisive.

where $\delta_i = \mu_i + \overline{\eta}_i$ since η_{ij} has been averaged over j, $\overline{\eta}_i = \left(\frac{1}{j}\right)\sum_j \eta_{ij}$

It is now possible to fully model the fixed effects, so there is less reason to expect biased and inconsistent estimates of β, unless $\overline{X}_{it}$, λ_t or δ_i are correlated with the idiosyncratic error term ξ_{it} (Wooldridge 2002:268). The within transformation will remove any fixed effect that exists for a given area or time period (Wooldridge 2002:267).

A potential problem with using pseudo-panels is that the mean of a sampled variable is not necessarily representative of the population's true mean. This is particularly a worry with this type of study, where, as mentioned earlier, there is a sample selection bias in that one can only observe the prices of houses actually being sold. Also, if the mean values have large standard errors or are derived from few observations, then the accuracy of inferring them to be representative of the true sample means is questionable (Von Fintel 2007:8-9). This is likely to be a concern as many neighbourhoods have few observations once separated over the time dimension, although the price variations within neighbourhoods is somewhat less worrying. Furthermore, different houses are used to construct the means to represent the same neighbourhood in each period, making it more difficult to justify a panel structure if the law of large numbers cannot be invoked. If these problems do exist, then the estimates could nevertheless be biased and inconsistent.

The standard fixed effects estimator is applied to the group aggregated data:

$$\hat{\beta} = \left(\overline{X}'M'M\overline{X}\right)^{-1}\left(\overline{X}'M'M\overline{y}\right) = \left(\tilde{X}'\tilde{X}\right)^{-1}(\tilde{X}'\tilde{y})$$

where $M = I_K \otimes \left(I_T - \left(\frac{1}{T}\right) \underline{1}_T\, \underline{1}'_T\right)$, which is known as the demeaning matrix. I_T and I_K are identity matrices of sizes T and K respectively and $\underline{1}_T$ is a column vector of one's of length T. It is clear to see that the fixed effect estimator is the same as the pooled OLS estimator but with the time demeaned and averaged $\tilde{X}$ and $\tilde{y}$ instead of the usual X and y (Wooldridge 2002:267-9).

This fixed effects model is also contrasted with the random effects model. The two models are mathematically very similar. The random effects model is the matrix weighted average of the fixed effects and the between effects estimators (Hardin 2000). It uses the assumption that δ_i follows the normal distribution around zero and not some constant. The result is that the model assumes that the $\hat{\beta}$ coefficients have the same effects in the cross-section and time series dimensions (Gould 2001).

6.3.1 *Ecological fallacy*

A problem arises with grouped data, typically geographical data as in this case, when it is analysed in terms of aggregates. When inferences are made from group level data about individual level characteristics, a substantial risk arises that cross-level biases can occur. This problem is known as the ecological fallacy (Greenland 2002:389).

Greenland (2002:390-1) provides the following simple mathematical model to illustrate the problem:

If one were to consider a model at the individual level, where each individual belongs to a group, and both individual level and group level variables are include in the model as follows:

$$y_{ijt} = \alpha + x_{ijt}\beta + \bar{x}_{it}\gamma + \varepsilon_{ijt}$$

where x_{ijt} are the characteristics of individual *j* belonging to group i, $\bar{x}_{it}$ is the average of all characteristics of individuals in group i in time *t*. y_{jit} is the dependent variable of individual *j* in group i in time t and ε_{ijt} is the idiosyncratic error term. β is then known as the *individual effect* and γ is known as the *contextual effect;* similar to Can's (1992) neighbourhood and adjacency effects.

When the data is aggregated by group i over all individuals in each group, the following model arises:

$$\bar{y}_{it} = \alpha + \bar{x}_{it}(\beta + \gamma) + \bar{\varepsilon}_{it}$$

The new error term can now become heteroskedastic if the groups do not contain an equal number of members and the independence of errors assumption is violated. It is also clear to see that the coefficient of $\bar{x}_i$ is now a combination of individual and contextual effects. This will lead to misinterpretation of these coefficients unless either β or γ are equal to zero (Greenland 2002:390-1). These problems are likely to be of some concern in this analysis.

6.4 ***Quantile regression***

An enlightening extension on standard OLS regression is that of pooled quantile regression. Where OLS is constrained to explaining characteristics at the mean of the dependent variable, quantile regression aims to explain the dependent variable at any point in the dependent variable's distribution and not just at the mean (Koenker and Hallock 2001:143). For this paper's purposes, quantile regression allows one to see if housing characteristics are valued similarly at different points in the distribution of property prices. This method would be similar to segmenting the data into groups based on property price but avoids the truncation bias of running separate regressions for different house price brackets as it still utilises the entire set of data (Heckman 1979:153-4). This would control for some group level fixed effects as houses in similar price categories are found in similar geographical areas, and share other characteristics. However, it is not wise to control for group level fixed effects here using dummy variables due to the strong correlation between the house price quantiles and their geographic areas.

Another caveat of this method is that of the bracket creep, which occurs when ‘grouping’ by price, while the sample spans three years. This will also cause a bias, as the date dummies and the regressed quintiles will be correlated, due to the increasing price trend that is known to exist.

Whilst OLS regression minimises the sum of squared residuals, quantile regression minimises the weighted sum of absolute deviations:

$$\min_{\{b_p\}_{p=0}^{z}} \sum_{j} \left| y_j - \sum_{p=0}^{z} b_p x_{pj} \right| h_j \qquad for \quad p = 0, \dots, z; \;\; j = 0, \dots, K$$

where $y_j$represents the *j*th element of the dependent variable, b_p indicates the *p*th regression coefficient and x_{pj} is the *p*th regressor for observation *j*.

h_j is defined as:

$$h_j = 2q$$

when the residual for the observation is positive, or as:

$$h_j = 2 - 2q$$

when the residual is smaller than or equal to zero, and q is the quantile at which the equation is being estimated and is thus between zero and one $(0 < q < 1)$ (Koenker and Hallock 2001:145).

It should also be mentioned that quantile regressions do not use the usual OLS standard errors for the coefficient estimates, but bootstrapped standard errors (Gould 1992, 1997)[18]. These standard errors are significantly more robust to heteroskedasticity.

7 Estimation Results

The following section reviews the estimated results of the models discussed in the previous section. In all methods nonlinear effects for attributes are included even when not statistically significant in order to investigate the existence of diminishing marginal returns to attributes expected by hedonic theory.

7.1 Pooled OLS

Below, in table 1, is the regression output for the pooled OLS model with log of sales price as the dependent variable.

This semilog construction implies that all coefficients can be interpreted as percentage effects on the price [19]. All the coefficients' signs are as expected and more of each characteristic results in a more expensive property.

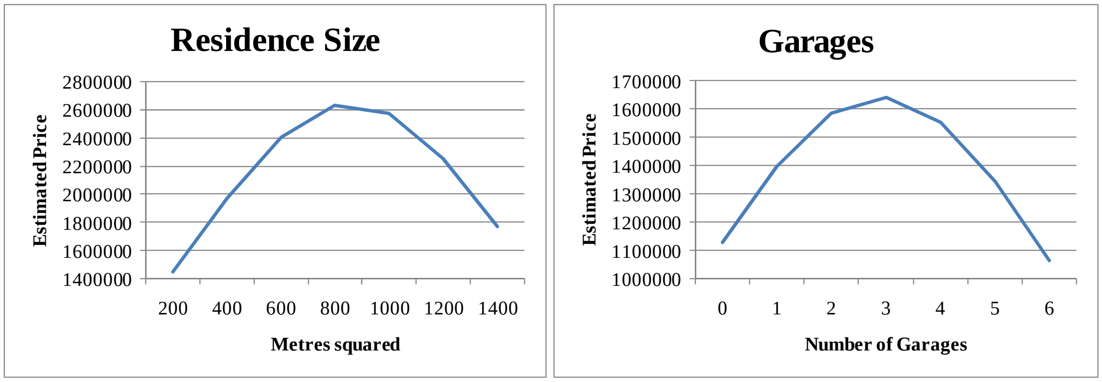
The exponent of the constant represents the base price for a property, which is approximately R297 542. Of the remaining statistically significant coefficients, a second storey increases the value by 17.6%, while the number of days on the market increases the property price by 0.02%. Prices increase with longer exposure to the market, as low offers are likely to be rejected initially and subsequent offers converge to the high asking price – sellers that are convinced of the value of their "attribute basket" are willing to wait longer to obtain a higher price[20]. .

The residence size and garage variables both have statistically significant non-linearities and are best interpreted graphically (by showing predicted property values from the model) as in

[18] The bootstrap standard errors were constructed using 1000 repetitions.
[19] The transformed coefficients in all the models are interpreted as percentage changes of sales price for a one unit change in the attribute ($percentage\ change = [\exp(\beta) - 1] \times 100$) as explained in Gujurati (2003: 320).
[20] Note that the potential for endogeneity bias exists in this case, as causality may also run from a higher list price to a longer selling period. However, no approriate instruments were available in the dataset to account for this possibility.

figure 2 below. These graphs show that not only do diminishing marginal returns exist, but negative marginal returns eventually occur[21].

Figure 2 Conditional Pooled OLS Non-Linearities



Having an additional reception area adds 4.2% to the value of a property while each additional study area adds 8.8% to the value of the property.

Each bathroom adds 19.7% to the price of the property, while a swimming pool adds 6.4% to value of the property and a parking bay adds 3.8%[22]. The statistically insignificant non-linearities indicate that the assumption of diminishing marginal returns does not hold.

The date of sale dummies are defined in Section 1 of the Appendix. They represent six month increments starting from September 2004 to August 2007. These models use the earliest date as the base category and as such, the coefficients represent the percentage growth in subsequent periods, after controlling for attributes.

In testing the OLS assumptions, the summary statistics of the residuals indicate that the errors are distributed with a mean of zero but with a much higher than normal kurtosis of 9.5 and skewness of -0.5. The Ramsey RESET test with the fitted dependent variable squared indicates model misspecification[23].

Since the standard OLS assumptions are violated, the coefficients are biased and inconsistent and one can conclude that the bias is probably upwards. Also, by violating the normality assumption, the hypothesis tests should be treated with caution.

[21] These negative marginal returns are somewhat questionable as they occur towards the extremes of the variable distributions and their robustness is uncertain.
[22] Throughout the paper only the percentage of the linear variable is given when the non-linearity is not statistically significant. This is not strictly accurate as the square of the variable is modelled, but the approximation is expected to be relatively accurate as a result of the non-significant non-linearities.
[23] The diagnostic tests are presented in Section 2 of the Appendix.

Table 1 Pooled OLS Hedonic Price Function

Variable	Coefficient	Percentage	p-value
CONSTANT	12.60331		0.000
DOUBLE STOREY	0.16213	17.601	0.000
TRIPLE STOREY	-0.06189	-6.001	0.554
DAYS	0.00019	0.019	0.027
RES SIZE	0.00238		0.000
RES SIZE SQUARED	0.00000		0.007
STAND SIZE	0.00018	0.018	0.074
STAND SIZE SQUARED	0.00000		0.824
AGE	-0.00125		0.631
AGE SQUARED	0.00004		0.375
BEDROOMS	-0.00266		0.958
BEDROOMS SQUARED	-0.00108		0.854
RECEPTION	0.04097	4.182	0.004
STUDY	0.08361	8.721	0.000
BATHROOMS	0.17958	19.672	0.000
BATHROOMS SQUARED	-0.00612		0.376
GARAGE	0.25822		0.000
GARAGE SQUARED	-0.04465		0.000
POOL	0.06246	6.445	0.003
DATE2	0.12580	13.406	0.000
DATE3	0.14728	15.868	0.000
DATE4	0.18534	20.362	0.000
DATE5	0.19406	21.417	0.000
DATE6	0.25228	28.695	0.000
DOMESTIC ACCOM.	-0.02858	-2.818	0.263
PARKING BAY	0.03776	3.849	0.002
PARKING BAY SQUARED	-0.00221		0.260
R-squared			0.692
Root MSE			0.339
Number of Observations			1930

Source: Own calculations from RPPR data

7.2 Group level fixed effects

The group level fixed effects model is this paper’s imitation of Bourassa *et al.‘s* (2007) method of dealing with neighbourhood effects. The results for this regression are presented below in Table 2 which is again contrasted against the pooled OLS results.

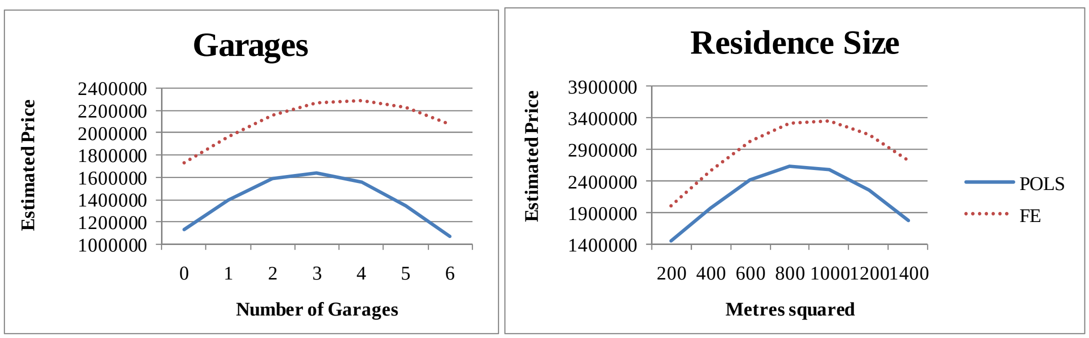
Upon closer inspection of Table 2 it is clear that all the statistically significant fixed effects coefficients are smaller than their pooled OLS counterparts in absolute terms. It can be concluded from this observation that much of the neighbourhood effects that were captured in the other characteristic variables have been controlled for and are captured in the dummy variables for area. The inclusion of the neighbourhood effects has now drastically reduced the bias in the other characteristic coefficients.

The signs of the statistically significant coefficients have not changed and thus the interpretation of the coefficients has also not changed since the previous section. The base price of houses in the fixed effects model is R603 372.

The premium for having a double storey house has dropped by 11 percentage points to 6.2 %. The coefficient for the number of days that the property spent on the market lost its statistical significance.

The effects of residence size and garages are illustrated in figure 3 below due to their statistically significant non-linearities. These effects are also contrasted against their pooled OLS equivalents. As in the pooled OLS model, the diminishing marginal returns also later become negative marginal returns.

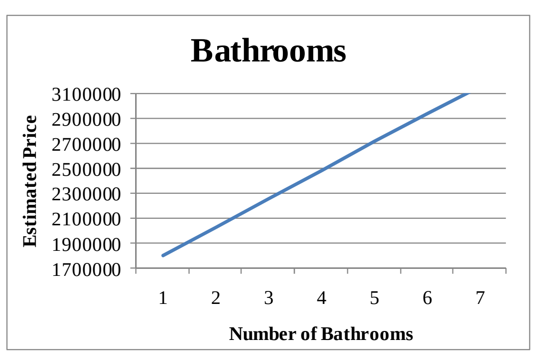
Figure 3 Fixed Effects Non-linearities



The effect of an additional reception rooms has dropped to a return of only 3.9% while the return on a study dropped to 2.9%. The age variable has now become statistically significant and indicates that for each additional year after a house is built, the property value decreases by 0.35%.

The bathroom attribute’s non-linearity has become statistically significant, thus its interpretation is illustrated in figure 4 below. However, it is clear that although the non-linearity is statistically significant, the magnitude of its effect is negligible.

Figure 4 Bathrooms Non-linearity



The swimming pool variable dropped to a return of 5.9% and the parking bay variable lost its statistical significance.

The date variables were also all dampened by the inclusion of area effects and are discussed in more detail in the next section regarding house price growth.

The area dummies indicate which areas are higher priced relative to the others. The base area is area 1 which is the most southern suburb of Stellenbosch. The highest premium paid for property was in area 7, the central area of Stellenbosch, with a 22.8% higher property price. The second dearest area is area 9, just east of area 7, with a premium of 15.6% on the base property price. At the other end of the spectrum, area 32, South End in Strand, has the largest negative effect by decreasing the base price by 62%, followed by area 31, just north west of area 32, with a negative effect of 53% to the base price.

From these results it is clear that neighbourhood effects clearly do matter; they are statistically significant and they remove the area effects correlated with other characteristic variables. By controlling for these neighbourhoods, the model has drastically reduced the bias in estimated coefficients and greatly improved the model. The R-squared also increases substantially from 0.69 to 0.82 upon adding the area dummies. It is thus likely that these neighbourhood effect variables have controlled for the spatial correlation as suggested by Bourassa *et al.* (2007).

Table 2 Group Level Fixed Effect Hedonic Price Function

Variable	Pooled OLS			Fixed Effects		
	Coefficient	Percentage	p-value	Coefficient	Percentage	p-value
CONSTANT	12.60331		0.000	13.31029		0.000
DOUBLE STOREY	0.16213	17.601	0.000	0.05982	6.164	0.001
TRIPLE STOREY	-0.06189	-6.001	0.554	-0.06522	-6.314	0.532
DAYS	0.00019	0.019	0.027	0.00006	0.006	0.443
RES SIZE	0.00238		0.000	0.00180		0.000
RES SIZE SQUARED	0.00000		0.007	0.00000		0.005
STAND SIZE	0.00018	0.018	0.074	0.00013	0.013	0.039
STAND SIZE SQUARED	0.00000		0.824	0.00000		0.791
AGE	-0.00125		0.631	-0.00350	-0.349	0.089
AGE SQUARED	0.00004		0.375	0.00005		0.183
BEDROOMS	-0.00266		0.958	0.02027		0.452
BEDROOMS SQUARED	-0.00108		0.854	-0.00242		0.310
RECEPTION	0.04097	4.182	0.004	0.03795	3.868	0.000
STUDY	0.08361	8.721	0.000	0.02838	2.879	0.048
BATHROOMS	0.17958	19.672	0.000	0.13111		0.000
BATHROOMS SQUARED	-0.00612		0.376	-0.00469		0.000
GARAGE	0.25822		0.000	0.14889		0.000
GARAGE SQUARED	-0.04465		0.000	-0.01972		0.000
POOL	0.06246	6.445	0.003	0.05727	5.894	0.000
DATE2	0.12580	13.406	0.000	0.12376	13.174	0.000
DATE3	0.14728	15.868	0.000	0.14461	15.559	0.000
DATE4	0.18534	20.362	0.000	0.17951	19.663	0.000
DATE5	0.19406	21.417	0.000	0.18354	20.146	0.000
DATE6	0.25228	28.695	0.000	0.23219	26.136	0.000
DOMESTIC ACCOM.	-0.02858	-2.818	0.263	-0.02712	-2.676	0.173
PARKING BAY	0.03776	3.849	0.002	0.01613	1.626	0.145
PARKING BAY SQUARED	-0.00221		0.260	-0.00119		0.621
AREA 2				-0.04312	-4.221	0.384
AREA 3				0.02724	2.762	0.631
AREA 4				-0.23421	-20.881	0.000
AREA 6				-0.66197	-48.416	0.000
AREA 7				0.20527	22.785	0.002
AREA 8				-0.06974	-6.736	0.139
AREA 9				0.14528	15.637	0.012
AREA 10				-0.35405	-29.816	0.000
AREA 11				-0.59467	-44.826	0.000
AREA 12				-0.53326	-41.331	0.000
AREA 13				-0.17013	-15.644	0.005
AREA 14				-0.31880	-27.298	0.000
AREA 16				-0.16089	-14.862	0.004
AREA 17				-0.35052	-29.568	0.000
AREA 18				-0.16065	-14.841	0.005
AREA 19				-0.11723	-11.062	0.029
AREA 20				0.02846	2.886	0.704
AREA 21				-0.15181	-14.085	0.045
AREA 22				-0.34255	-29.005	0.000
AREA 23				-0.32224	-27.547	0.000
AREA 24				-0.21624	-19.446	0.000
AREA 25				-0.01278	-1.270	0.809
AREA 26				-0.30973	-26.635	0.000
AREA 27				-0.36778	-30.773	0.000
AREA 28				-0.53214	-41.265	0.000
AREA 29				-0.69640	-50.163	0.000
AREA 30				-0.45807	-36.750	0.000
AREA 31				-0.75444	-52.972	0.000
AREA 32				-0.96692	-61.975	0.000
AREA 33				-0.64281	-47.419	0.000
AREA 34				-0.14394	-13.406	0.085
R-squared		0.692				0.815
Root MSE		0.339				0.265
Number of Observations		1930				1930

Source: Own calculations from RPPR data

7.2.1 *Understanding property price growth rates*

Upon closer inspection of the growth rates provided by the fixed effects estimates[24], one notices that they are dissimilar to those quoted by most property market analysts.

The standard method used to calculate the yearly growth in residential property prices is $Growth\ rate = [\frac{new\ average\ sales\ price}{old\ average\ sales\ price} - 1] \times 100$. This method clearly does not account for changes in property characteristics over time and thus introduces an indeterminate bias in the estimate. The fixed effects method, however, can isolate the change in price paid over time from the change in all other variables over time in the regression. The fixed effects method can thus drastically reduce the bias in calculating residential property price growth rates and give a more accurate estimate of the rate that actual residential property prices have been growing at. Table 3 below contrasts residential property price growth rates calculated using the fixed effects output, an OLS model controlling only for time periods[25] and the standard method using unadjusted average sales prices over time.

From Table 3 and Figure 5 it is clear to see that residential property price growth rates tend to be overstated if we do not control for changes in attributes (as is the case with the standard method and the uncontrolled OLS estimates). The fixed effects method enforces the *ceteris paribus* assumption by holding all other variables constant and thus calculates more accurate growth rates. In this case these growth rates are substantially lower.

[24] A house price index (after controlling for attributes and areas) was constructed from the estimates of time dummies in the fixed effects model. The omitted period, February 2005, is also used as the base year for the index. Growth rates presented above are year-on-year estimates based on the index.

[25] This index was constructed similarly to the fixed effects index, though only controls for time (and no other attributes or area effects) were included in the model. The results represent unconditional "least square means" in the ANOVA sense, and assist the transition from standard arithmetic growth rates to rates calculated from linear models. These OLS means are similar to standard growth rates, and very high compared to the fixed effects variants. The difference between these estimates and the fixed effects growth rates therefore suggests that controlling for attributes *is* important.

Figure 5

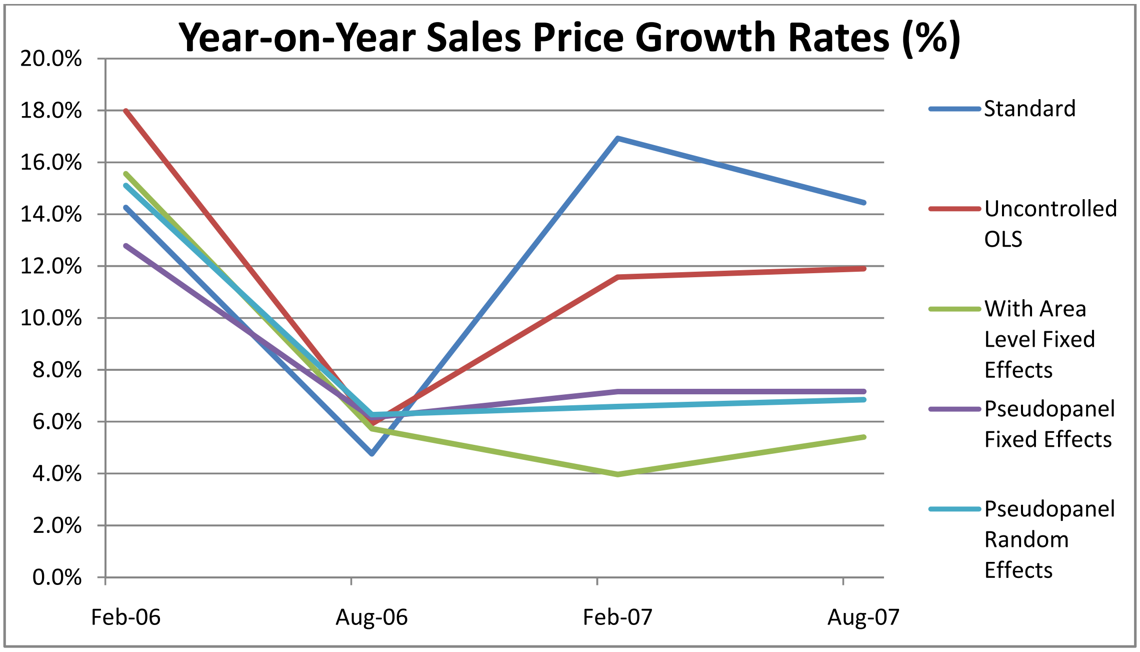


Table 3 House Price Growth Rates Using Different Methods

	Period of Sales					
	Feb-05	Aug-05	Feb-06	Aug-06	Feb-07	Aug-07
Standard :						
Index	100.0	114.1	114.3	119.6	133.6	136.8
Year on Year Growth			14.3%	4.8%	16.9%	14.4%
Uncontrolled OLS :						
Index	100.0	117.4	118.0	124.3	131.6	139.1
Year on Year Growth			18.0%	5.9%	11.6%	11.9%
With Area Level Fixed Effects :						
Index	100.0	113.2	115.6	119.7	120.1	126.1
Year on Year Growth			15.6%	5.7%	4.0%	5.4%
Pseudopanel Fixed Effects :						
Index	100.0	110.9	112.8	117.7	120.9	126.2
Year on Year Growth			12.8%	6.2%	7.2%	7.2%
Pseudopanel Random Effects :						
Index	100.0	113.0	115.1	120.1	122.7	128.4
Year on Year Growth			15.1%	6.3%	6.6%	6.8%

Source: Own calculations from RPPR data

This difference in growth rates between the standard and fixed effects methods must be caused by changes in the explanatory variables over time. A table providing a summary of the average explanatory variables over time is given below in Table 4. From this table it is clear to see that the average house that is sold has been changing significantly over the observed time period. It is thus incorrect to calculate residential property price growth rates based on average sales prices alone, as the average house is changing over time. This standard and incorrect calculation is the mathematical equivalent of comparing apples with oranges. By controlling for most other property characteristics, one obtains a truer reflection of the actual growth in property prices.

The bias caused by the explanatory variables can be seen by both the different growth rates in table 3 and the clear changes in the average house attributes over the years in table 4. This analysis illustrates that house price inflation can be separated into 'attribute inflation' and 'real inflation'. Attribute inflation is the increase in the house prices due to the increase in the average housing attributes over time. The real inflation is the actual growth in house prices over time with all else being constant.

These findings indicate that there were initially high house price growth rates in the market. Then, market participants reacted by expanding supply of houses on the property market which later outstripped the demand, or by investing more in attributes that provide high returns, caused the real growth rates to fall. The last two years of claimed 'high' growth in the residential property market (as calculated using the standard growth method) has mainly been driven by attribute inflation and not by real increases in the house price.

Table 4 Summary Statistics

	Period of Sales					
	Feb-05	Aug-05	Feb-06	Aug-06	Feb-07	Aug-07
SALES PRICE	1294135.74	1476850.01	1478646.9	1547188.79	1728887.42	1770647.64
DAYS	65.841	67.679	74.436	74.719	95.911	90.379
RES SIZE	210.161	218.807	204.229	215.802	220.483	225.416
RES SIZE SQUARED	63315.120	64290.640	53212.410	66070.510	65189.910	70624.690
STAND SIZE	727.260	775.339	893.805	732.925	810.966	765.747
CONDITION	4.041	4.104	3.951	3.958	3.966	4.130
CONDITION SQUARED	17.486	18.189	16.914	17.263	17.301	18.396
BEDROOMS	3.148	3.089	3.109	3.147	3.312	3.198
RECEPTION	1.819	1.904	1.925	1.982	2.065	2.020
STUDY	0.265	0.339	0.327	0.287	0.281	0.304
BATHROOMS	2.152	2.052	2.070	2.115	2.197	2.193
BATHROOMS SQUARED	6.575	5.149	4.884	5.065	6.354	5.475
DOMESTIC ACCOM.	0.151	0.182	0.173	0.213	0.209	0.174
GARAGE	1.403	1.438	1.453	1.482	1.490	1.491
GARAGE SQUARED	2.588	2.662	2.693	2.787	3.017	2.973
PARKING BAY	0.585	0.586	0.624	0.599	0.630	0.614
POOL	0.275	0.314	0.308	0.308	0.360	0.396
Number of Observations	465	280	266	334	292	293

Source: Own calculations from RPPR data

7.2.2 *Testing for spatial autocorrelation*

Moran's I test is applied to the area aggregated residuals from the pooled OLS regression and from the fixed effect OLS regression. This aggregation is not at risk of losing meaning, as the residuals per area grouping would still be consistently biased in one direction should spatial autocorrelation exist. Thus any aggregation would not lose the bias, but would approach the true area bias. Table 5 below shows the results of this test.

Table 5 Moran's I tests for spatial autocorrelation in the pooled OLS and fixed effects OLS residuals

	Pooled OLS residuals				**Fixed Effects residuals**			
	statistic	expected value	standard deviation	p-value	statistic	expected value	standard deviation	p-value
Moran coefficient I	0.511	-0.032			-0.085	-0.032		
normality			4.653	0.000			-0.449	0.653
randomisation			4.651	0.000			-0.448	0.654

Source: Own calculations from RPPR data

From this output it is clear that the original OLS model did suffer from spatial autocorrelation. The Moran I test shows that the normality and randomness assumptions of the error terms are violated. The areas even in aggregate form depended on the neighbouring area's error components.

By controlling for these area effects the model has by construction removed all spatial autocorrelation at an aggregate level. This is illustrated by the almost surely normally and randomly distributed residuals from the Moran I test.

Thus, by controlling for area effects, it is possible to remove all of the contextual effects, but with this data there is no way to test the extent of individual effects remaining in the residuals. These findings show again that the fixed effects model is the most viable model for this study.

7.3 Fixed effects and random effects pseudo panel

Table 6 below compares the estimates of the fixed effects (FE) and random effects (RE) pseudo-panel.

Table 6 Pooled OLS, Pseudo-panel Fixed Effects and Pseudo-panel Random Effects Hedonic Price Functions

Variable	Fixed Effects			Random Effects		
	Coefficient	Percentage	p-value	Coefficient	Percentage	p-value
CONSTANT	12.99130		0.000	12.56972		0.000
DOUBLE STOREY	-0.06909	-6.676	0.257	0.04719	4.832	0.505
TRIPLE STOREY	-0.68804	-49.744	0.027	-0.60400	-45.338	0.094
DAYS	0.00006	0.006	0.758	0.00004	0.004	0.867
RES SIZE	0.00189	0.189	0.002	0.00317	0.317	0.000
RES SIZE SQUARED	0.00000		0.153	0.00000		0.002
STAND SIZE	0.00010		0.070	0.00016		0.014
STAND SIZE SQUARED	0.00000		0.142	0.00000		0.023
AGE	-0.00450		0.142	-0.00552		0.118
AGE SQUARED	0.00009		0.063	0.00015		0.007
BEDROOMS	-0.03282		0.667	-0.10128		0.271
BEDROOMS SQUARED	-0.00013		0.986	0.01093		0.195
RECEPTION	0.09401	9.857	0.008	0.02480	2.511	0.546
STUDY	0.09478	9.942	0.097	0.17903	19.605	0.008
BATHROOMS	0.13179		0.019	0.23313	26.254	0.000
BATHROOMS SQUARED	-0.00448		0.142	-0.00720		0.051
GARAGE	0.25485	29.027	0.001	0.36836	44.536	0.000
GARAGE SQUARED	-0.04989	-4.866	0.013	-0.06998		0.003
POOL	0.07732	8.039	0.122	0.09055	9.478	0.136
DATE2	0.10352	10.906	0.000	0.12265	13.048	0.000
DATE3	0.12031	12.785	0.000	0.14069	15.107	0.000
DATE4	0.16326	17.734	0.000	0.18348	20.140	0.000
DATE5	0.18948	20.863	0.000	0.20449	22.690	0.000
DATE6	0.23249	26.173	0.000	0.24974	28.369	0.000
DOMESTIC ACCOM.	-0.10004	-9.520	0.106	-0.14433	-13.440	0.053
PARKING BAY	0.00704		0.824	0.06356		0.089
PARKING BAY SQUARED	-0.00072		0.798	-0.00901		0.006
R-squared* (overall)			0.78			0.89
R-squared* (within)			0.71			0.67
R-squared* (between)			0.87			0.93
Number of Time Periods			6			6
Number of Observations			191			191

**The Panel R-squareds are not directly comparable with the Pooled OLS R-squared*

Source: Own calculations from RPPR data

Both panel models appear to be poor fits of the data, as the R-squareds are high, but many of the coefficients are not statistically significant. It appears that by averaging the already compressed log of sales, much of the variation in the data has been lost.

Many of the statistically significant coefficients of the FE and RE models are similar to their pooled OLS counterparts. It is important to also note again that, as Greenland (2002:390) proves, it is not possible to distinguish between what part of the pseudo-panel coefficients would be individual effects and what part is contextual effects, as both are measured jointly. Most notably, however, growth rates (as presented in Table 3) essentially tell the same story as the disaggregated models. These coefficients do not suffer from the same problem as those of aggregated variables, as they are modelled neither at the individual or group level in any of the models presented, but rather for *all* houses in respective time periods.

The analysis of the pseudo panel is of little importance due to the few observations and weakness of technique in this scenario. It is shown as an alternative investigation in the data generating process which happened to be unsuccessful in determining implicit prices for inidividual-level attributes, but which underscores the approximate magnitudes of conditional growth rates.

7.4 Quantile regression

The output for the quantile regressions is given below in Table 7. Quantile regressions were run for the 25th, 50th and 75th percentiles.

As can be seen below, some of the variables from the OLS model lose their statistical significance in the quantile regression models. It is also interesting to see how the coefficients change across the quantiles, indicating that hedonic prices are sensitive across the property price distribution, as Zeitz *et al.* (2007) have indicated.

A second storey on a house seems to matter slightly more as one moves up the price distribution. The number of days for which the property is on the market positively influences house prices, as would be expected[26].. It is, however, only statistically significant for the middle and upper percentiles of the property price distribution and is stronger higher up the distribution. This is also to be expected, as people buying property at the top end of the distribution are likely to shop around more for the 'right' property basket and sellers are willing to market their properties' superior attributes for a longer period by rejecting poor offers. Conversely, it is not significant at the bottom end of the price distribution: people are willing to sell their properties more quickly since their real returns are greater[27], as opposed to the need to spend longer periods in getting the right price for attributes. Furthermore, the

[26] The same reasons and caveats as above apply in this context.
[27] This is illustrated *vis-à-vis* the quantile growth rates in table 8 below.

"attribute basket" in this segment is likely to vary less, so that less "shopping around" is done.

The effects of residence size and garage variables are represented in the figure 6 due to their statistically significant non-linearities. Again, one notices not only the diminishing marginal returns, but negative marginal returns as well.

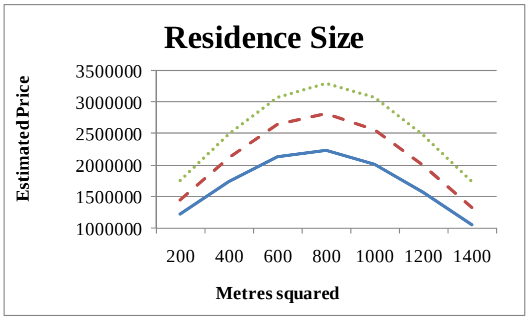
Table 7 Quantile Hedonic Price Functions

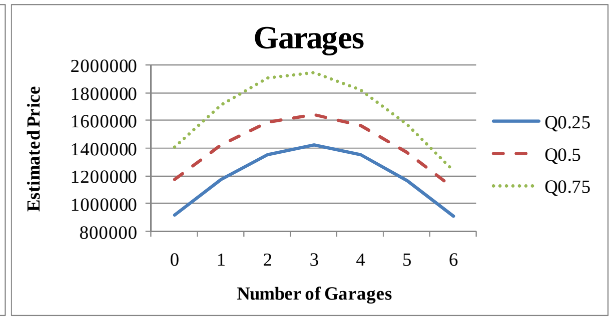
Variable	Q0.25			Q0.5			Q0.75		
	Coefficient	Percentage	p-value	Coefficient	Percentage	p-value	Coefficient	Percentage	p-value
CONSTANT	12.44437		0.000	12.55994		0.000	12.71553		0.000
DOUBLE STOREY	0.13154	14.059	0.000	0.13391	14.329	0.000	0.16164	17.544	0.000
TRIPLE STOREY	-0.06300	-6.106	0.646	-0.11601	-10.953	0.417	-0.08092	-7.773	0.626
DAYS	0.00018	0.018	0.148	0.00017	0.017	0.064	0.00027	0.027	0.042
RES SIZE	0.00288		0.000	0.00309		0.000	0.00288		0.000
RES SIZE SQUARED	0.00000		0.010	0.00000		0.000	0.00000		0.000
STAND SIZE	0.00018	0.018	0.007	0.00019	0.019	0.000	0.00022	0.022	0.000
STAND SIZE SQUARED	0.00000		0.802	0.00000		0.698	0.00000		0.690
AGE	-0.00252	-0.251	0.114	-0.00187		0.230	-0.00379		0.092
AGE SQUARED	0.00003		0.181	0.00004		0.195	0.00010		0.015
BEDROOMS	-0.05652		0.422	-0.00163		0.973	0.02461		0.651
BEDROOMS SQUARED	0.00167		0.864	-0.00081		0.894	-0.00374		0.524
RECEPTION	0.04318	4.413	0.003	0.02797	2.836	0.044	0.00464	0.465	0.749
STUDY	0.05356	5.502	0.015	0.05131	5.265	0.024	0.08545	8.921	0.002
BATHROOMS	0.18619	20.466	0.001	0.18854	20.749	0.001	0.19354	21.353	0.000
BATHROOMS SQUARED	-0.00566		0.587	-0.00624		0.570	-0.00681		0.427
GARAGE	0.29090		0.000	0.23497		0.000	0.23501		0.000
GARAGE SQUARED	-0.04851		0.000	-0.04061		0.000	-0.04276		0.000
POOL	0.04624	4.732	0.048	0.04796	4.913	0.013	0.05337	5.482	0.050
DATE2	0.13096	13.992	0.000	0.10394	10.953	0.000	0.09913	10.421	0.006
DATE3	0.18083	19.821	0.000	0.13830	14.832	0.000	0.10195	10.733	0.003
DATE4	0.21806	24.367	0.000	0.18325	20.111	0.000	0.15093	16.292	0.000
DATE5	0.17855	19.548	0.000	0.18061	19.795	0.000	0.23116	26.006	0.000
DATE6	0.28301	32.711	0.000	0.24462	27.714	0.000	0.25241	28.713	0.000
DOMESTIC ACCOM.	-0.02513	-2.481	0.335	-0.06202	-6.013	0.022	-0.06272	-6.079	0.054
PARKING BAY	0.02584		0.246	0.01835		0.231	0.04854	4.974	0.021
PARKING BAY SQUARED	-0.00018		0.980	-0.00145		0.741	-0.00258		0.548
R-squared			0.49			0.49			0.46
Number of Observations			1930			1930			1930

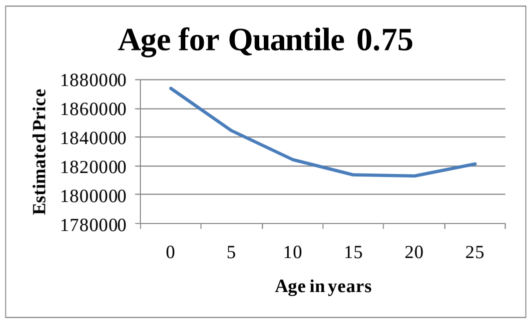
Quantile regression R-squareds are psuedo R-squareds calculated as 1 minus (sum of deviations about the estimated quantile / sum of deviations about the raw quantile)

Source: Own calculations from RPPR data

Figure 6 Pooled Quantile Regression Non-Linearities







Stand size matters marginally, and slightly more so at the upper end of the property price distribution.

Property age only has a noticeable effect at the upper end of the property price distribution and may be used by "high end" buyers as a potential indicator of quality. Quality is expected to be a more definitive factor in the high end market. These effects are illustrated in figure 6 above.

Bedrooms seem to be meaningless across the property price distribution. Reception rooms appear to matter more to the lower end of the property price distribution, probably because these rooms become standard in the upper end of the spectrum.

Bathrooms are of more value to higher priced properties than to lower priced properties and the nonlinear terms are statistically insignificant; this is probably due to higher priced property also having higher quality bathrooms.

Having a swimming pool seems to matter for the whole property price distribution, but more so for the upper percentiles. This is probably because it may be considered an expensive item to the lower end of the market while being a 'necessary' luxury item to the upper end.

Having accommodation for a domestic employee plays a statistically significant role in the middle and upper part of the property price distribution and decreases the price of the property. The number of parking bays on a property appears to only have a linear effect and only for the upper percentiles of the property price distribution.

Table 8 below shows the quantile regression corrected growth estimates for the period under investigation, also based on indices constructed from the date dummies. The index reveals that the lower end of the property market has experienced relatively higher growth rates over the entire period, though most recently the high end market has outpaced the lower end. This indicates a possible convergence in property price across sub-markets[28]. This convergence could partially be due to the increasing standards of living in South Africa[29], thereby causing an increased demand for low end housing. It could also indicate that some people are being priced out of the high end of the property market (as a result of initial growth in that market) into the lower end, thereby increasing the lower end prices more than the higher end prices. These results also indicate that more of the 'real inflation' in house prices have occurred at the bottom end of the market while the top end have experienced more 'attribute inflation'. These results must be considered with some caution as the quantile regression model was unable to control effectively for area effects.

Table 8 Quantile Growth Rates

	Period of Sales					
	Feb-05	Aug-05	Feb-06	Aug-06	Feb-07	Aug-07
Q0.25						
Index	100.0	114.0	119.8	124.4	119.5	132.7
Year on Year Growth			19.8%	9.1%	-0.2%	6.7%
Q0.5						
Index	100.0	111.0	114.8	120.1	119.8	127.7
Year on Year Growth			14.8%	8.3%	4.3%	6.3%
Q0.75						
Index	100.0	110.4	110.7	116.3	126.0	128.7
Year on Year Growth			10.7%	5.3%	13.8%	10.7%

Source: Own calculations from RPPR data

[28] Note that this is convergence *across* markets and not convergence within different market segments across geographic areas, as analysed in Burger & Janse Van Rensburg (2008) and Gupta & Das (2008)

[29] See van der Berg et al (2008) for an analysis of post-Apartheid poverty reduction, which is linked to social grants. But more importantly, the rise of the black middle class is likely to stimulate demand in the residential property market.

Due to the nature of the quantile regression technique, there are few diagnostic tests available. In Section 2 of the Appendix, output of a *link test* for the quantile regression model is presented. This test is similar to the Ramsey RESET test for model misspecification. The result of the statistically significant variables in the test indicates that the quantile regressions also suffer from model misspecification. This leads one to conclude that its coefficient estimates are also biased and inconsistent, but it is encouraging that they are similar to their OLS counterparts.

8 Conclusion

The aim of this paper was to construct a well defined hedonic price model for Stellenbosch, Somerset West, Strand and Gordon's Bay for the time period from September 2004 to August 2007. This aim was met as best possible given the data constraints.

As is apparent from the literature review, the largest problem was dealing effectively with the correlation of property sales values with other nearby property sales values: spatial autocorrelation. It was clear that the standard OLS model gave biased and inconsistent estimates and it was shown that the residuals of this regression were probably spatially correlated. This was effectively dealt with by adding neighbourhood fixed effects to the OLS regression as done by Bourassa *et al.* (2007). Spatial tests then indicated that the contextual spatial autocorrelation was removed by the fixed effects model.

This paper also highlights some shortcomings regarding assumptions made when calculating property price growth rates. Research reports treat average monthly growth as a homogenous measure of change, which is not true, as the average house changes every month. This leads to biased and inconsistent estimates of property price growth rates. An alternative method was presented that corrected the bias in these estimated growth rates by controlling for changes in property characteristics over time with a fixed effects model. The results showed that much of the more recent property price growth was caused by attribute inflation and not real inflation.

It was also shown, with quantile regression, that housing characteristics have different implicit prices at different points in the price distribution. This is what was expected according Freeman's (1993:371) explanation of hedonic theory and was also shown by Zietz *et al.* (2007). This model showed that relatively more of the inflation in property prices came from the lower end of the market. This suggests that real property prices are converging across markets, while there is a divergence in attribute levels and prices.

Furthermore, attribute inflation dominates at the high end of the market. This suggests that in this segment, price increases have been fuelled by either improved levels of attributes or by higher values of existing attributes. If the former is a true reflection, then it suggests that homeowners in this segment are pursuing a high price by improving attributes and not by lifting their 'pure' growth rates by engaging in these activities. This may be a case of demand satiation: where the initial property boom stimulated attribute investment, whose rewards declined as supply of such attributes became more widespread. This paper shows that attribute investment in the high end market may lead to misleading increases in property value, suggesting that 'pure' growth of the initial house is not dramatically lifted by such investments.

However, these attribute investments are not required at the bottom end of the market, as this is where more real growth has in fact taken place. This shows that the bottom end was either a high return market during this period, or that people in the high segments have been priced into lower segments, thus shifting demand patterns.

In summary, these findings indicate that buyers have responded strongly to higher past returns to property investments, but that investments in new attributes (supposedly to capitalise on these returns), have been unwarranted. This is indicative of a proverbial "bubble": real growth has been overvalued by buyers and sellers in that period. The effects are currently playing out in a more subdued South African housing market.

9 References

ABSA (2008) "ABSA Housing Review - Second Quarter2008" [online] available at http://www.finforum.co.za/absa/publications/property/Property%20perspective.pdf (9 June 2008)

Anselin, L. (2002) "Under the hood. Issues in the specification and interpretation of spatial regression models", *Agricultural Economics*, 27(3), pp 247-267

Anselin, L. (2006) "Spatial Econometrics", in T. C. Mills (ed.) and K. Patterson(ed.), *Palgrave Handbook of Econometrics*, New York: Palgrave Macmillon

Basu, S. and Thibodeau, T.G. (1998) "Analysis of Spatial Autocorrelation in House Prices", *Journal of Real Estate Finance and Economics*, 17(1), pp 61-85

Black, S. E. (1999) "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114(2), pp 577-599

Bourassa, S. C. and Peng, V. S. (1999) "Hedonic Prices and House Numbers: The Influence of Feng Shui", *International Real Estate Review*, 2(1), pp 79 - 93

Bourassa, S. C., Cantoni, E. and Hoesli, M. (2007) "Spatial Dependence, Housing Submarkets, and House Price Prediction", *Journal of Real Estate Finance and Economics*, 35(1), pp 142-160

Bourassa, S. C., Hoesli, M., and Peng, V. C. (2003) "Do housing submarkets really matter?", *Journal of Housing Economics*, 12(1), pp 12–28

Bunzel, H. (2003) *Hedonic Theory and Econometric Specification*, Aalborg University, [online] available at www.econ.au.dk/fag/2785/e03/econ-topic2.pdf (7 December 2007)

Burger, P. and Janse van Rensburg, L. (2008) "Metropolitan House Prices in South Africa: Do They Converge?" *The South African Journal of Economics.* 72(2): pp 291-297.

Can, A. (1992) "Specification and estimation of hedonic housing price models", *Regional Science and Urban Economics*, 22(3), pp 453–474

Chiodo, A. J., Hernadez-Murillo, R. and Owyang, M. T. (2005) "Nonlinear Hedonics and the Search for School District Quality", Federal Reserve Bank of St. Louis - Working

Paper Series, [online] available at http://research.stlouisfed.org/wp/2003/2003-039.pdf (7 December 2007)

Day, B. (2001) "The Theory of Hedonic Markets: Obtaining welfare measures for changes in environmental quality using hedonic market data" [online] available at http://www.cserge.ucl.ac.uk/Hedonics__Chapter_1_.pdf (11 December 2007)

Deaton, A. (1985) "Panel Data from Time Series of Cross-Sections", *Journal of Econometrics, 30(1)*, 109-126

Du Toit, J. (2007) "Residential Property Perspective – Third Quarter 2007", *Absa Group Economic Research*, [online] available at http://www.finforum.co.za/absa

Epple, D. (1987) "Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products", *Journal of Political Economy*, 95(1), pp 59-80

Freeman, A. M. (1993) *The measurement of environmental and resource values: Theory and methods*, Resources for the Future, Washington, D.C, [online] available at http://books.google.com/books?hl=en&lr=&id=eiLIJj74szsC&oi=fnd&pg=PR13&dq=%22The+measurement+of+environmental+and+resource+values:+Theory+and+methods%22+freeman&ots=taHzFTEAzN&sig=JbZDVTDj1mc8-PjRPF0xfdtQzLI#PPR13,M1

Gatzlaff, D. and Haurin, D. (1998) "Sample Selection and Biases in Local House Value Indices," *Journal of Urban Economics*, 43(1), pp 199-222.

Gould, W. (1992) "Quantile Regression with Bootstrapped Standard Errors", *Stata Technical Bulletin*, 9, pp 19-21

Gould, W. (1997) "Interquantile and Simultaneous-Quantile Regression", *Stata Technical Bulletin*, 38, pp 14-22

Gould, W. (1997) "Between estimators" [online] available at http://www.stata.com/support/faqs/stat/xt.html (9 January 2008)

Griliches, Z., (1961) "Hedonic Price Indices for Automobiles", in Z. Griliches (ed.), *Price Indices and Quality Change: Studies in New Methods of Measurement*, Cambridge: Harvard University Press, pp 55- 87

Greene, W. H. (2003) *Econometric Analysis, Fifth Edition*, New Jersey: Pearson Education

Greenland, S. (2002) “A review of multilevel theory for ecologic analysis”, *Statistics in Medicine*, 21(2), pp 389-395

Gujurati, D.N. (2003) *Basic Econometrics.* 4th ed. New York: McGraw-Hill

Gupta, R. and Das, S. (2008) Spatial Bayesian Methods of Forecasting House Prices in Six Metropolitan Areas of South Africa. *The South African Journal of Economics*, 72(2), pp 298-313.

Hardin, J. (1996) “Fixed-, between-, and random-effects and xtreg” [online] available at http://www.stata.com/support/faqs/stat/xtreg.html (9 January 2008)

Heckman, J. J. (1979) “Sample Selection Bias as a Specification Error”, *Econometrica*, 47(1), pp 153-61

Hidano, N. (2002) *The Economic Valuation of the Environment and Public Policy. A Hedonic Approach*, Cheltenham, UK: Edward Elgar

Kaufman, D. A. and Cloutier, N. R. (2006) “The Impact of Small Brownfields and Greenspaces on Residential Property Values” Journal of Real Estate Finance and Economics, 33(1), pp 19-30

Koenker, R. and Hallock, K. F. (2001) “Quantile Regression”, *Journal of Economic Perspectives*, 15(4), pp 143-56

Lancaster, K. J. (1966) “Change and innovation in the technology of consumption”, *American Economic Review*, 56(1), pp 14-23

Munneke, H. J. and Slade, B. A. (2000) “An Empirical Study of Sample-Selection Bias in Indices of Commercial Real Estate”, *Journal of Real Estate Finance and Economics*, 21(1), pp 45-64

Okunev, J., Wilson, P. and Zurbruegg, R. (2000) “The Causal Relationship Between Real Estate and Stock Markets”, *Journal of Real Estate Finance and Economics*, 21(3), pp 251-261

Palm, R. (1978) “Spatial segmentation of the urban housing market”, *Economic Geography*, 54(3), pp 210–221

Residential Property Price Ranger (2006), Data available [online] via request at http://www.rppr.co.za (12 September 2006*)*

Rode (2005) "Rode's Property Report 2005" available [online] at http://www.rode.co.za/publications/demos/RR_2005_4_-_15_-_House_Mark.pdf (9 June 2008)

Rosen, S. (1974) "Hedonic prices and implicit markets: Product differentiation in pure competition", *Journal of Political Economy*, 82(1), pp 34-55

SaMap (1999) software by *Chicago Map Corporation*, http://smap.ocx

Shiller, R. J. (2007) "Understanding Recent Trends in House Prices and Home Ownership", *Federal Reserve Bank of Kansas City 2007 Symposium on "Housing, Housing Finance, and Monetary Policy"*, available [online] at http://www.kc.frb.org/publicat/sympos/2007/PDF/2007.09.27.Shiller.pdf (24 June 2008)

Tiebout, C. M. "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 64(5), pp 416-424

Van der Berg, S., Louw, M. And Yu, D., (2008) "Post-Transition Poverty Trends Based on an Alternative Data Source", *The South African Journal of Economics*. 76(1), pp 58-76.

Von Fintel, D. (2007) "Rising Unemployment in South Africa: A Dynamic Birth Cohort Panel Analysis" Unpublished Masters Thesis, University of Stellenbosch

Wang, D.G. and Li, S.M. (2004) "Housing preferences in a transitional housing system: the case of Beijing, China", *Environment and Planning* 36, pp 69-87

Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: The MIT Press

Zietz, J., Zietz, E. N. and Sirmans, G. S. (2007) "Determinants of House Prices: A Quantile Regression Approach", *Journal of Political Economy*, Forthcoming as of December 2007

10 Appendix

10.1 <u>Section 1 – The Data</u>

10.1.1 Variables

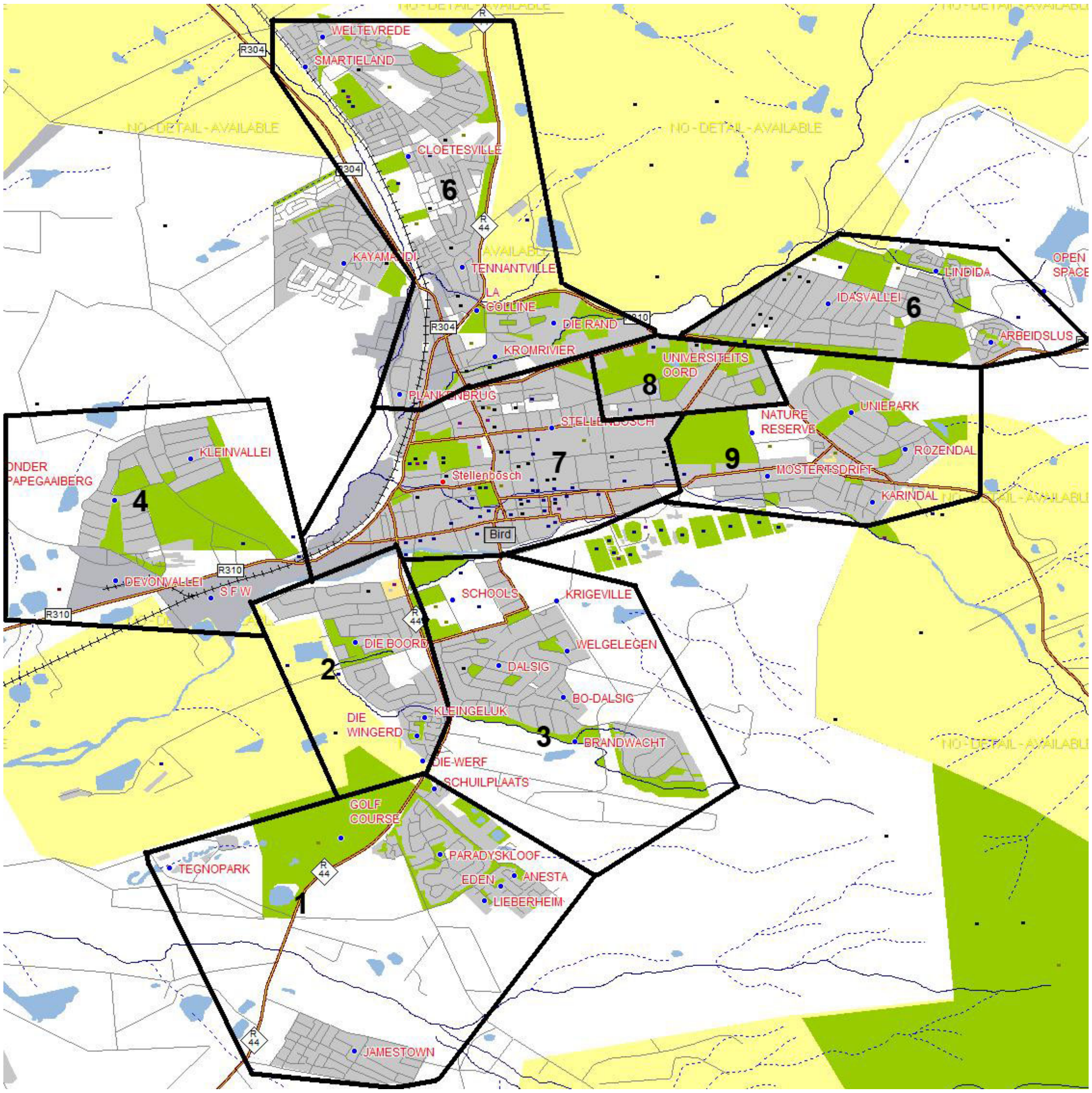
Variable	Description	Unit/Input	Min	Max	Median
Type	Type of residential structure:				
	Single storey house	1			
	Double storey house	2			
	Triple storey house	3			
SaleP	Selling Price	In Rands	45 600	8 750 000	1 300 000
ListP	Listing Price	In Rands	55 000	8 960 000	1 395 000
Date	Sales Date:				
	September 2004 - February 2005	1			
	March 2005 - August 2005	2			
	September 2005 - February 2006	3			
	March 2005 - August 2006	4			
	September 2006 - February 2007	5			
	March 2007 - August 2007	6			
Days	Number of days on the Market	As an integer	0	1 095	51
Res_Size	House size in metres squared	In metres^2	0	1 727	200
Stand_Size	Plot size in metres squared	In metres^2	0	43 196	600
Age	Age of house	In Years	0	190	7
Condition	Condition of House (subjective impression):				
	Excellent	1			
	New	2			
	Good	3			
	Renovated	4			
	Fair	5			
	Poor	6			
Bedrooms	Number of bedrooms	As an integer	1	15	3
Reception	Number of reception rooms	As an integer	0	8	2
Study	Number of study rooms	As an integer	0	2	0
Bathrooms	Number of bath rooms	As an integer	0	25	2
DomAcc	Number of domestic servant quarters	As an integer	0	2	0
Garage	Number of Garages	As an integer	0	6	2
Parking bay	Number of parking bays	As an integer	0	17	0
Pool	Dummy variable for pool	1 for pool			
No.13	Dummy for house with address no. 13	1 for no. 13			
Area	Dummy variable for area:	**Area Number**			
	Stellenbosch - Jamestown	1			
	Stellenbosch - Paradyskloof	1			
	Stellenbosch - Anesta	1			
	Stellenbosch - Eden	1			
	Stellenbosch - Mont Blanc	1			

	Stellenbosch - Le Montier	1			
	Stellenbosch - Tegnopark (no obs)	1			
	Stellenbosch - La Pastorale	1			
	Stellenbosch - Schuilplaats	1			
	Stellenbosch - Die Boord	2			
	Stellenbosch - Kleingeluk	2			
	Stellenbosch - Fairways (no obs)	2			
	Stellenbosch - Die Wingerd (no obs)	2			
	Stellenbosch - Die Werf (no obs)	2			
	Stellenbosch - Harrington Place (no obs)	2			
	Stellenbosch - Brandwacht	3			
	Stellenbosch - Dalsig & Bo Dalsig	3			
	Stellenbosch - Krigeville	3			
	Stellenbosch - Welgelegen (no obs)	3			
	Stellenbosch - Onderpapegaaiberg	4			
	Stellenbosch - Stellenoord	4			
	Stellenbosch - Devon Park	4			
	Stellenbosch - Devon Valley (no obs)	4			
	Stellenbosch - Spier (outlier)	5			
	Stellenbosch - Cloetesville	6			
	Stellenbosch - Green Oaks (no obs)	6			
	Stellenbosch - La Colline	6			
	Stellenbosch - Idas Valley	6			
	Stellenbosch - Central	7			
	Stellenbosch - Die Weides	7			
	Stellenbosch - Simonswyk	8			
	Stellenbosch - Aanhouwen	9			
	Stellenbosch - Karindal	9			
	Stellenbosch - Mostersdrif	9			
	Stellenbosch - Jonkerspark	9			
	Stellenbosch - Rozendal	9			
	Stellenbosch - Uniepark	9			
	Stellenbosch - Welgevonden	10			
	Stellenbosch - De Wijnlanden	10			
	Stellenbosch - Klein Welgevonden	10			
	Somerset West - Audas	11			
	Somerset West - Bridgewater	11			
	Somerset West - Helderzicht	11			
	Somerset West - Somerset Ridge	12			
	Somerset West - Westridge	12			
	Somerset West - Boskloof	13			
	Somerset West - Dennegeur	13			
	Somerset West - Schapenberg	13			
	Somerset West - Somerset Heights	13			
	Somerset West - Rome Glen	13			
	Somerset West - Bizweni	14			
	Somerset West - Bayview Heights	14			
	Somerset West - Erinvale Golf Estate	15			
	Somerset West - Land & Zeezicht	16			

	Somerset West - Morningside	16			
	Somerset West - Natures Valley	16			
	Somerset West - Central	17			
	Somerset West - Stuarts Hill / Jacques Hill	17			
	Somerset West - Golden Acre	18			
	Somerset West - Golden Hill	18			
	Somerset West - World's View	18			
	Somerset West - Helderbrand	18			
	Somerset West - Parel Vallei	19			
	Somerset West - Spanish Farm	20			
	Somerset West - Fairview Heights	20			
	Somerset West - La Sandra	20			
	Somerset West - Mont Clair	21			
	Somerset West - Pinegrove	22			
	Somerset West - Roundhay	22			
	Somerset West - Martinville	22			
	Somerset West - Die Wingerd	23			
	Somerset West - The Links	23			
	Somerset West - Briza	23			
	Somerset West - Monte Serena	24			
	Somerset West - Helena Heights	24			
	Somerset West - Bakkershoogte	25			
	Somerset West - La Concorde	25			
	Somerset West - Belair	25			
	Somerset West - Helderberg Estate	25			
	Somerset West - Berghowe	26			
	Somerset West - Illiare	26			
	Somerset West - Steynsrus	26			
	Somerset West - Emeralds View	27			
	Somerset West - Heldervue	27			
	Somerset West - Van Der Stel	28			
	Somerset West - Heritage Park	28			
	Strand - Central East	29			
	Strand - Central West	30			
	Strand - South End	31			
	Gordons Bay - North	32			
	Gordons Bay - Central	33			
	Gordons Bay - South	34			

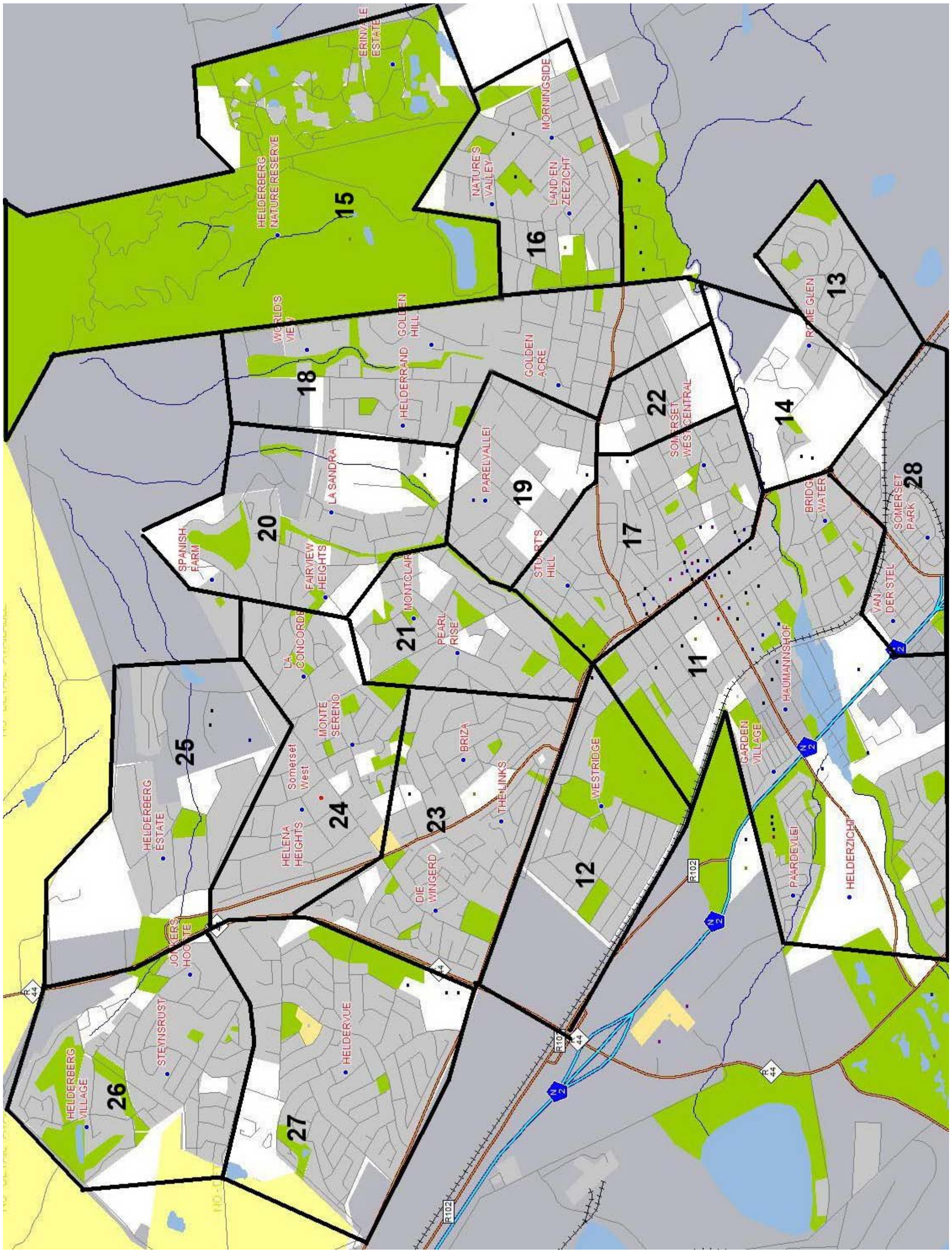
10.1.2 Maps of Areas[30]

Stellenbosch:



[30] Source of maps: SaMap

Somerset West:



Gordon's Bay:


11
30
29
31
32
33
34
Strand
STRAND CENTRAL
SOMERSET PARK
OPEN SPACE
NOMZANO
SOUTH END
HELDERBERG PARK
RUSTHOF
SERCOR PARK
BROADLANDS
GUSTROUW
CASABLANCA
GORDON STRAND ESTATE
SOUTHFORK
Gordon's Bay
ONVERWACHT
ANCHORAGE PARK
GORDONS BAY CENTRAL
TOWN

10.2 Diagnostics

10.2.1 Pooled OLS

Ramsey RESET

F(3,1899)	40.100
Prob > F	0.000

Residuals Tests

Mean	0.000
Skewness	-0.490
Kurtosis	9.565
Prob normal	0.000

10.2.2 Group level fixed effects

Ramsey RESET

F(3,1868)	36.050
Prob > F	0.000

Residuals Tests

Mean	0.000
Skewness	-1.014
Kurtosis	19.022
Probability of being normally distributed	0.000

10.2.3 Quantile Regression

Link test for simultaneous quantile regression

	coefficient	t-value	p-value
lsalep_hat	4.421	6.730	0.000
lsalep_hat squared	-0.123	-5.230	0.000
constant	-23.594	-5.150	0.000

Dependent variable is lsalep