AN INVESTIGATION INTO THE HEDONIC PRICE ANALYSIS OF THE STRUCTURAL CHARACTERISTICS OF RESIDENTIAL PROPERTY IN THE WEST RAND

ROBERT SCOTT DODDS

A research report submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, in fulfillment of the requirements of the degree of Master of Science in Building.

Johannesburg 2010

ABSTRACT

A vast amount of literature on hedonic price modelling has been formulated on overseas property markets. Very little currently exists in South Africa and this poses a risk for sellers and estate agents of a residential property when listing it on the open market, as this could result in an extended list period, reducing the original asking price. This paper seeks to examine Gauteng's West Rand residential property market and formulate a multi-variate regression model to best predict property prices, determined by a property's structural characteristic. The research tracks residential sales from 1996 to 2009, a thirteen-year sample period from which a composite property index, to account for inflation and real house price growth, has been formalised. Correlation and regression analysis was used to interpret the data at the relevant significance level. In order to account for locational attributes present in property values, the data set was divided into locational quadrants and run as dummy variables. A further regression was run on a screened data set to create an ordinary least squares equation that could be used to show the relationship between property values and structural characteristics. The results indicated a good fit with an R^2 of 69.5%. This regression was then applied practically to predict property prices for houses that have transacted in the West Rand property market, and plotted along a value/price graph using the 45-degree true value frontier line. The relevant results were then interpreted, and recommendations given.

DECLARATION

I declare that this research report is my own unaided work. It has been submitted for the Degree of Master of Science in Building in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

____

(Signature of Candidate)

________ Day of __________________________ (year) ________

Hedonic price analysis of the structural characteristics of residential property in the West Rand

TABLE OF CONTENTS

ABSTRACT .... **i**

TABLE OF CONTENTS .... **iii**

LIST OF FIGURES .... **v**

LIST OF TABLES .... **vi**

1. INTRODUCTION **1**

1.1. Background to the study 2

1.2. Research Question 4

1.3. Statement of the problem 4

1.4. Statement of thesis 5

1.5. Sub problems 5

1.6. Significance of the study 6

1.7. Scope and delimination of work 7

1.8. Assumptions 7

1.9. Structure of the report 8

2. LITERATURE REVIEW **9**

2.1. Introduction 9

2.2. History, theory and construction of hedonic price modelling 11

2.3. Approach of hedonic price model 19

2.3.1. Advantages and disadvantages 20

2.3.2. Present day hedonic price modelling 22

2.4. Conclusion 26

3. RESEARCH METHODOLOGY 28

3.1. Types of research 28

3.1.1. Structured research 28

3.1.2. Unstructured research 29

3.2. Sample frame 31

3.3. Correlation analysis 33

3.4. Regression analysis 34

3.5 Regression application 38

4. RESEARCH DATA, EMPERICAL RESULTS AND ANALYSIS 39

4.1. Data capture 39

4.2. Data adjustment 39

4.3. Descriptive statistics 44

4.3.1. Quantitative variables 44

4.3.2. Qualitative variables 45

4.4. Correlation analysis 46

4.5. Regression analysis 49

4.5.1. Structural attributes on the full data set........ 49

4.5.2. Structural variables and area groupings........ 62

4.5.3. Structural attributes for selected pocket of suburbs........ 65

4.6. Application of the model for the entire West Rand region........ 69

5. CONCLUSION........ 72

6. RECOMMENDATIONS........ 74

7. REFERENCES........ 75

LIST OF FIGURES

Figure 4.1 Average house price growth for the West Rand........42

Figure 4.2 Heteroscedasticity for data set Table 4.8........54

Figure 4.3 Histogram of residual error for data set Table 4.8........55

Figure 4.4 Normal Probability Plot of residual error for data set Table 4.8........56

Figure 4.5 Heteroscedasticity for data set Table 4.8........59

Figure 4.6 Histogram of residual error for data set Table 4.8........60

Figure 4.7 Normal Probability Plot of residual error for data set Table 4.8........61

LIST OF TABLES

Table 4.1 Summary of West Rand house growth data..41

Table 4.2 Summary statistics of regression variables 44

Table 4.3 Number of observations for each region comprising the West Rand 45

Table 4.4 Number of observations for each qualitative factor 45

Table 4.5 Correlation matrix of structural variables for West Rand 46

Table 4.6 Correlation matrix of structural variables for pocket suburbs 48

Table 4.7 Regression output: included observations = 1 938...50

Table 4.8 Regression output: included observations = 1 548...52

Table 4.9 Regression output: Log of Adjusted SP included observations = 1 548……………........58

Table 4.10 Regression output: included observations = 1 548...63

Table 4.11 Regression output: Included observations = 1 548 64

Table 4.12 Regression output: Included observations = 233…………………………………………..66

1. INTRODUCTION

South Africa has experienced a massive growth rate in prices of housing over the past several years. Understanding these price changes is important for several reasons. Residential housing serves as a big source of an individual's wealth and any changes in the underlying value of a house dramatically affects one's consumer spending and savings decisions. This, in turn, has an overall affect on economic activity. It is therefore important to have an accurate measure of aggregate housing prices (Rappaport, 2007).

The sale of a residential house involves strategic interaction between a seller and a set of potential buyers. When a house is listed on the market, the seller advertises a certain listing price (usually through a property agent) and awaits a potential offer from a buyer. The listing price will influence the arrival of offers, which will determine the date on which the sale is concluded. As the time the property is on the market increases, the arrival rate of potential buyers decreases, and the probability of the listing price changing increases.

If the original asking price is high, the sale price is usually relatively high, but also experiences a longer time on the market (Ortalo-Magne, 2002). Taking the above into account, the pricing of a house sale is important.

This research analyses the structural determinants of residential property in Gauteng's West Rand, and develops a hedonic pricing model with statistical significance that could reasonably explain the open market value of a residential house, given a set of structural variables.

1.1. Background to the study

According to Hill and Melser (2008), who conducted research in Sydney Australia, using housing data for 14 regions over six years, existing house price indices usually measure the average or median price of houses sold during a given period of time. The problem with this method is that the mix of houses sold could change dramatically over time. For a price index to be useful, it must compare the sales of equivalent houses from one period to the next.

When interviewed on 20 June 2009, Mrs. N. Vergotine-Dube[1] confirmed that, in South Africa, several property house price indices are available. One of the most popular indices used amongst property valuers and bankers is that of the South African Property Transfer Guide (SAPTG). This is an online property sales registry that records all transactions registered in the South African Deeds Office. This information is useful as it illustrates sales volume and house price trends over a given period.

Other indices used in the business world to monitor house price growth, according to Mrs. N. Vergotine-Dube during an interview held on 20 June 2009, include the African Bank South Africa (ABSA), First National Bank (FNB), and Standard Bank of South Africa (SBSA), property indices. These indices use either the mode, median or mean to calculate the average increase in property prices, and would include repeat sales of an equivalent nature.

Previous models tried to overcome this problem by only dealing with repeat sales

[1] N. Vergotine – Dube is the manager of Investec Private Bank's Property Risk Team

occurring during the specific time set. This method was seen as an improvement on the average price approach but was still littered with problems (Hill and Melser, 2008).

(Hill and Melser, 2008) listed the following problems with the repeat sales method:

- It cannot account for spatial price indices as the same house could not sell twice during a given period in a different place;
- the maximum use of available data is not utilized as houses that do not sell twice during a period are omitted;
- data captured earlier on needs to be updated as new data is forever being created;
- the possibility of a property's condition being the same when it is sold for the second time during a particular period, several years since the initial sale, is difficult to guarantee as the model does not account for depreciation; and
- repeat sales data may tend to differ from single sale data as they may not follow the same sale path (e.g. flats could take longer to sell than houses) and different sales classes' (flats verses house) prices may tend to grow at different rates, resulting in the index becoming biased.

According to Hill and Melser (2008) a hedonic approach to house price construction has the potential to resolve the issues highlighted above and significantly improve the quality of house price valuation. The essence behind such an approach is to regress the price of the house in terms of its characteristics. This method has emerged as the general one used within the property industry for measuring the housing market.

According to Keng (1997), who conducted studies on the Malaysian property market from 1988 to 1997 using correlation and regression analyses to create a housing price index, no two houses are alike due to their heterogeneity, and house prices differ according to a wide variety of attributes such as locational and physical attributes.

Keng (1997) mentions that in order to indicate the price variations in the individual attributes from one house to the next, the price measure must be segregated. Multiple regression analysis (MRA) enables the estimation of changes in average price from one time period to another on a standardised basis, and hence forms the hedonic function.

1.2. Research Question

The research has asked: What statistically significant structural characteristics of a residential house will determine the open market value? The research examined the sales recorded by estate agents in relation to each property's structural characteristic, and found a solution to this question.

1.3. Statement of the problem

The problem statement to be addressed in this research study may be stated as:

Do the structural characteristics of a residential house determine its open market value, and if so, are they statistically significant?

Further, the research will also probe into the potential effects location will have on the selling price of a house.

The reason the above problem statement is being researched is for the investigation into an alternative method of property valuation that can replace the need for a physical inspection by a property valuer. This is because before a bank will bond and grant a loan on a property, it must assess the property's market value.

Under normal banking circumstances, a bank will acquire the expertise of a professional property valuer to conduct a physical property inspection of the property in question and write a report on its perceived market value.

Besides the time required for a property valuer to inspect and write up the report, which wastes valuable time for the bank and purchaser, a cost is attached to the creation of the property valuation report.

The West Rand is one of the more established residential areas in Johannesburg where property price ranges are not extreme, and property characteristics are fairly homogenous. The West Rand is one of Johannesburg's largest residential areas and would therefore comprise a big portion of any banks loan book. Bearing this in mind, it is seen as prudent that a hedonic model is created for banks to utilise when valuing property in the West Rand, to cut down on costs, and reduce turnaround times when allocating bonds.

1.4. Statement of thesis

The structural characteristics of a residential property will influence its open market value, and a basic multi-variate linear regression model can be used to estimate a property's value within certain bounds, given a set of statistically significant structural variables.

1.5. Sub problems

Sub problems were created in order to break the main research question down into constituent parts, for which workable solutions could be formulated. The following sub-problems were researched:

- Does the stand size of a property affect the open market value?
- Does the dwelling size of a property affect the open market value?
- Does the number of bedrooms affect the open market value?
- Does the number of bathrooms affect the open market value?

- Does the number of recreational rooms affect the open market value?
- Does the number of study rooms affect the open market value?
- Does a garage affect the open market value?
- Does the wall type of the property affect the open market value?
- Does the location of the property affect the open market value?

1.6. Significance of the study

When interviewed on 20 June 2009, Mrs. N. Vergotine-Dube confirmed that there is awareness amongst professionals within the property environment, that automated valuation reports generated for a residential property, rely on insignificant sales data. This poses a risk for sellers and estate agents of a residential property when listing it on the open market as this could result in an extended list period, reducing the original asking price.

Mrs. N. Vergotine-Dube further commented that risks are also extended to banks during a bond grant as the underlying security of the asset, calculated by the automated valuation report, is not always representative of the property's true open market value, and hence, the bank is under-or-over secured on the property. The spin-offs from this are that the client's loan to value, exposure and facility are calculated incorrectly, creating unnecessary risk for both parties.

This study investigates which structural characteristics of a residential property have a significant impact on the open market value, and develops a theoretical marginal value for each of the variables.

Once this is created, the model can be applied to properties currently on the market and help banks to hedge their risk further by identifying over or under capitalised market property transactions.

1.7. Scope and delimination of work

The property industry consists of many different types of transactions, mainly including residential, commercial, retail and industrial. Furthermore, these types can be broken up into different zoning categories which stipulate the certified rights attached to the property. This study focuses mainly on single residential property sales that have 'residential one' zoning. The study focuses initially on residential sales that have taken place in Gauteng's West Rand.

1.8. Assumptions

In a telephonic discussion held with Christo Wiid (16 March 2009)[2], Wiid reported that the following assumptions are well researched.

- That Property24.com's database is accurate.
- That Property24.com's reporting is in line with the rest of the property industry.
- That the database only contains data for single residential property zoned residential one.

According to Wiid, only data relating to actual sales of single residential properties within the West Rand are recorded in the excel database. Property estate agents responsible for the sale capture the information recorded. Wiid's final comments were that all empty fields within the spreadsheet can be treated as zeros (or are non-existent characteristics of the property in question).

---

[2] C. Wiid is the general manager of Property 24.com

1.9. Structure of the report

Following the above introduction, a review of past and present literature relating to hedonic price modelling is discussed.

The literature review will summarise some of the first multi-variate experiments carried out on the structural attributes and their affects on property prices, followed by the latest hedonic price analyses of the effects that schools, pollution and greenbelts have on property values.

After the theoretical background has been concluded, an in-depth outline of the methodology to be utilised for the purposes of this thesis is explained, which include correlation and regression analysis. The regression and correlation analysis employed for the purposes of testing the data included four separate statistical tests.

The final regression, which resulted in the highest regression (R^2), was then used to formulate the value/price chart to test the data against the 45-degree true value frontier. Upper and lower bounds were included to account for the standard error present in the model.

Finally, concluding remarks regarding the findings are summarised, and recommendations to improve further hedonic price analysis are given.

2. LITERATURE REVIEW

2.1. Introduction

After analysing studies carried out by Ball (1973), Bover and Velilla (2002) and Hill and Melser (2008) the last forty years has seen extensive research into property valuation methods using statistical models.

The initial studies employed techniques such as multiple regression, R^2 for best fit equation, ordinary least squares (OLS), and certain log equations used to estimate house prices. Since then, general views regarding residential property values have been formulated relating to location, micro-economy (demand and supply), general macro-economic climate, accessibility and environmental issues.

To date, research within a South African property context is sparse. The journals and sources researched in order to find data on hedonic models and regressions analysis's, proved fruitless. The main search engines included the following: Sabinet, Emerald, EBSCOhost Web, J Stor and Google Scholar.

Research of existing literature and the study of a hedonic price analysis of the structural characteristics of residential property in Gauteng, is considered value add for the model's potential use, amongst individuals and companies within the property industry. The following pages delve into previous research, and the lessons learnt for hedonic price modelling.

The only information on hedonic data sourced within a South African context was a study conducted by Van Rensburg *et al* (2004). The objective of the study was to provide a mathematical tool (regression model) for individuals to calculate the inherent value of red wine when making a purchase.

One of the issues highlighted in the paper is that the price of something is what you pay, and the value is what you get.

The paper based the regression on the premise that an informed consumer would 1) measure, and 2) maximise the value obtained for each unit of money spent. This was graphically depicted on a graph in their conclusion using a value frontier for red wines farmed in the Western Cape to illustrate wines that were either over priced, or under priced (Van Rensburg *et al*, 2004).

Van Rensburg *et al* (2004) used price as the dependent variable and regressed it against the independent variables wine type, quality, area of origin, percentage of red wine made by estate, type of wine-firm structure and size of farm. The independent variables or explanatory variables included in the regression were only included if they were perceived to add value to the consumer.

The term "hedonics" is derived from the Greek word hedonikos, which simply means pleasure. In an economic sense, it refers to the utility or satisfaction one derives through the consumption of goods and services (Leong, 2002).

According to Kathleen *et al* (2006), it is presumed that humans locate properties that have desirable characteristics (this would include land services), and go on to purchase the property with the most desirable set of structural characteristics. Further comments indicate that the variation in purchase prices and housing characteristics allows for a person's estimation of the price placed on individual characteristics.

According to Kathleen *et al* (2006), hedonic models are based on Lancaster's 1966 theory of consumer demand, where the price of a house is a function of a multitude of attributes. The application of a hedonic price model would therefore estimate the value placed on the different attributes present in a house, and then calculate the property's overall value.

According to Kathleen *et al* (2006) the property selected by an individual is a result of the satisfaction they would derive from it. Satisfaction would relate to the use a person could get out of the property.

In economics, satisfaction is known as 'utilisation' (Wikipedia, 2009). Therefore, as the number of bedrooms and bathrooms within a property increases, so would the utilisation an individual could gain from the property.

Furthermore, it is common property knowledge that as the number of bedrooms and bathrooms within a property increases, so does the price. Therefore, in order for an individual to gain maximum utility from a property, they would have to balance the satisfaction they expect to gain from it alongside the money they expect to pay.

2.2. History, theory and construction of hedonic price modelling

From the majority of the readings researched, Griliches is mentioned as being the pioneer in hedonic pricing techniques and research. His 1971 work is referenced most frequently. Gressel *et al* (1984) refers to Griliches being responsible for the initial formulation of the hedonic pricing model.

Goodman (1978) mentions that hedonic price approaches, as noted by Griliches, are based on the premise that a large number of models of a particular heterogeneous commodity can explained by a smaller number of attributes. The equation described is illustrated as follows:

$P = f(C)$ where,

P = is the selling price of an individual house

C = set of components which add to the selling price

The equation above sets out a very standard MRA. Mark (1988) highlights that an assumption of the above equation is that none of the independent variables are perfectly correlated with the dependant variable.

In a study carried out by Garrod and Willis (1992) on the environmental economic impact of woodland in Britain, they broke down the formation of a hedonic model into six specific stages. They mentioned that the hedonic model is used to estimate the implicit prices of a set of characteristics which differentiate between similar products in a particular product class.

Therefore, in order to apply this model to the property market, a set of structural and locational characteristics which define a house, and possibly influence its selling price, must be identified.

Dummy variables, or free variables as described by Garrod and Willis (1992), are variables that are known to affect property values, but are of no special interest to the study.

In the study carried out by Garrod and Willis (1992), only a few dummy variables were included in the regression model (such as socio-economic and locational factors). It was highlighted that the omission of such variables in the regression analysis did not affect the results sufficiently enough to warrant their inclusion.

Brown (2004) wrote a paper on hedonic regression models for the Bureau of Labour Statistics in the United States of America to help them meet their current and future needs in terms of the ever-evolving consumer price index. When deciding on the representatives of the data, Brown (2004) mentions that it is best to split the data into two issues.

The first issue is to decide on whether the data is representative of the market. It is mentioned however that if the representatives selected are not exactly the same but fairly similar, the cross data can be used, as the effects on the regression are not significant.

This study focuses on free standing residential houses as opposed to sectional title units, where although the underlying asset is similar in nature (i.e. residential dwellings), the various characteristics attached to each type of house, is expected to impact its value differently (i.e. size of sectional title unit is expected to have a massive impact on value proportionally compared to the size of house which is expected to have a smaller impact on value).

Brown (2004) mentions that the second issue in relation to the variables would be with where the data has been collected. Data collected from different sources could potentially affect the type of variables and prices. It is therefore deemed beneficial to collect the data from one source as its reliability and accuracy is improved.

One of the negatives attached to this approach is that it allows bias to creep into the data, as certain sales/types of houses could be excluded or prices deflated/inflated to suite the sale. This results in skewed sales data for the selected area, and diminishes the quality of the data.

In the study carried out by Brown (2004), data was collected from specific stores, indicating outlet name, business classification code, size and region category of the city in which the quote was collected.

The information collected for the study carried out by Brown (2004) was then converted into variables that control the effects different types of business practices and geographic locations may have on the product mix and type. This robust technique helped minimise the variation in parameter estimates for the price-determining characteristics in the regression model.

The data collected for this study came from one area, one source, and indicated universal sale prices depending on the structural attributes concerning the property. This is thought to have minimised the variation in the price-determining values of the various structural attributes.

Another reservation Brown (2004) had about the data collected was with regards to the quality of characteristic data used. It was stressed that well-defined data, leads to a reliable model, and this can only be achieved if considerable time is spent preparing and cleaning the data used for modelling.

The last concern Brown (2004) had relating to the data used, was its age. This study concerns sales that have taken place between 1996 and 2009, nearly a ten year period, and the market has definitely changed during this time frame.

This would be highlighted in the price paid for a 3 bedroom 1 bathroom house 10 years ago, compared to a 3 bedroom 1 bathroom unit in today's terms.

Roughly 10 years ago, the number of bathrooms present in a house would not have been a massive factor in determining the value of a house. This is compared to current market trends in which the number of bathrooms in a house is expected to match bedrooms accordingly, and if this situation is out of kilter, the house value is expected to drop. This would therefore effect the coefficients of the variables, as they would look substantially different if created from current data as opposed to dated data.

Brown (2004) mentions that a hedonic model is created from the variables found in the data collected. Once the variables are created, a functional form for the regression is selected based on a priori of assumptions about price-influencing characteristics. The functional form utilised most often is the semilogarithmic form. This form usually fits the data well and the coefficient estimates calculated, are interpreted as being the proportion of a good's price that is directly comparable to the respective characteristic of that good.

This semilogarithmic form is indicated by the equation below:

$$\ln P_i = b_0 + b_1X_1 + b_1X_2 + \ldots.. + b_nX_m + e_i$$

where $\ln P_i$ is the natural logarithim of the price of each good, b_0 is the value of the base good, b_1 is the coefficient of the characteristic variable X_1, and e_i is the residual error.

In a paper written by Meese et.al (1997), they discuss the construction of hedonic price indices and regressions models. They described two methods to control for variation in the types of homes sold over time. They are the hedonic regression and repeat sales approach.

The construction of a housing price index in a hedonic regression approach follows a two step process in which one needs to estimate a regression of house sales price on a set of house attributes, and a constant term for each period. The examples given for house attributes included the size under roof of a house, and the number of bathrooms, very similar to the independent variables selected for the regressions run in this paper.

Intercepts chosen above would account for any trend in housing prices over the selected sample period, whilst the hedonic attributes would control the types of homes sold during any given time period.

Secondly, estimates of the implicit attribute of prices need to be used to construct a housing price index. This is very similar to housing price index formulated later on in the paper to estimate current house prices.

Meese *et al* (1997) describes the following disadvantages of the hedonic price analysis approach:

1. Ignorance of the functional form of the relation, and of the appropriate set of house characteristics to include in the analysis, resulting in inconsistent estimates of the implicit prices of the characteristics. Meese *et al* (1997) commented that other researchers have overcome this problem by using either flexible or ideal parameters of dependent variables.

2. Consistent estimates of implicit hedonic prices will rely on a big assumption that all omitted variables are uncorrelated to those included in the analysis.

The second method analysed in this paper was the repeat sales method. Meese *et al* (1997) mentioned that researchers can control hedonic characteristics by looking only at the properties that have sold more than once during the sample period, without any change in a house's characteristics between sales.

When using this method, any sales that are identified as being repeat sales are run as dummy variables using the OLS regression. This creates a logarithmic price change on the defined dummy variable resulting in a consistent estimate of average house price changes for the sample period.

Two factors that must be taken into account when using the repeat sales method is that the subsample of homes sold twice is representative of all homes sold during that period. The second is that the implicit attribute prices are constant over time so that the attribute prices cancel in the construction of the housing price index.

Meese et.al (1997) describes the following disadvantages of the repeat sales approach:

1. The regression is based on a smaller data set when compared to the hedonic regression approach.

2. The sensitivity associated with any repeat sales data from a single period compared to repeat sales on average from a particular period (this would cause the index to jump up by the average sales for that unusual period)

During the early 1970's, research into the determinants of residential property prices got underway. Wilkinson (1971) used MRA that incorporated the general structural attributes of a residential house, locational factors that included the proximity to the central business district (CBD), population and schools density in the area. The study illustrated that locational factors explained 45% of the house price variance in the model.

Wilkinson (1971) described a set of assumptions that need to be considered during the creation of a hedonic model. They included the nature of the housing market and the behavior of consumers, as well as the specification and internal structure of the utility function. Furthermore, it was considered that the housing market contains numerous imperfections, where choice is constrained by the operation of organisations, and the flow of information to purchasers is usually incomplete.

In this study, it was found that the squares of the independent data, represented by the portion of the unit variance of each variate, and the sum of the squares of these numbers became known as the "communality", and gave rise to the R^2 in the MRA.

Ball (1973) conducted studies on three research papers, analysing the determinants of relative house prices.

The first paper calculated the OLS regression equation for a sample of houses within the London Metropolitan Region, with house price as the dependent variable, and housing attributes as the explanatory variables.

House attributes were divided into two categories: location and house type. Location variables for this category were divided into accessibility, environment and greenbelts. House type variables were split into floor area, age of house, garage and central heating. The results indicated that all variables from the house type showed significance at the 5% level.

The second study conducted by Ball (1973) researched the house price distance relationship, and the third looked at the housing demand situation. The study concluded that there was a high degree of fit for R^2 in all the studies, even though there were many different explanatory variables. Further comments related to the results in that they were suspiciously high, and a reason for this could have been either the variables have been accurately verified, or that the researchers selected samples that held certain variables constant.

Another problem highlighted by Ball (1973) was with the house price variable used, as it comprised average prices and not individual prices. The latter is preferred as averaging could easily obscure important variations.

Hedonic price methods use information on the changes in product characteristics to break down price variations into those attributable to changes in characteristics, and those that take place for given characteristics. Hedonic price methods also rely on considerable data collection, as information is not only related to product prices, but also on their related characteristics (Bover and Velilla, 2002).

Gressel *et al* (1984) explains that hedonic pricing models, as with most applications of economic theory, do not provide a complete quantitative characteristic of real land markets. Gressel *et al* (1984) further suggests that the researcher is expected to specify the hedonic relationship, select the appropriate variables and choose the proper functional form.

2.3. Approach of hedonic price model

Lancaster's 1966 consumer theory and Rosen's 1974 model set the platform for hedonic price modelling. The two approaches aimed to impute prices of attributes based on the relationships between the observed prices of differentiated products, and the number of attributes associated with these products (Leong, 2002).

Lancaster's model assumed a linear relationship between the price of goods and the characteristics of those goods. In the model, it was presumed that goods are members of a group and that some or all of the goods in that group are consumed in combinations, subject to the consumer's budget (Leong, 2002).

The above thoughts are very similar to those of Kathleen *et al* (2006) who mentioned that in order for an individual to gain maximum utility from a property, they would have to balance the satisfaction they would expect to gain from it alongside the money they expect to pay.

Rosen's model had two distinct stages. The initial stage served as an estimate of the marginal price for the attribute of interests by regressing the price of a commodity or good on its attributes. The first stage creates a measure of the price, but does not reveal the inverse demand function. The second stage estimates the inverse demand curve or marginal willingness to pay function (Leong, 2002).

In line with the above, the identification of the inverse demand function poses some problems as it depends on the assumptions made about the supply side of the implicit market for the attribute.

This means that if the supply side of an attribute is fixed, the marginal price of an attribute becomes exogenous in the estimation of the inverse demand function (Leong, 2002).

This was however seen to not be a problem, as the hedonic estimation problem is caused by the endogeneity of both prices and quantities of attributes in the context of a non-linear budget constraint. Hence, there is no necessity to model the supply side of the market (Leong, 2002).

2.3.1. Advantages and disadvantages

The following advantages and disadvantages are present in a hedonic property valuation method, as highlighted by Kathleen *et al* (2006):

Advantages:

- The data used in a hedonic model is based on actual sales.
- The change in zoning or use of a property can be incorporated into the model at any time.

Disadvantages:

- Any change in the zoning or use of the property will most likely affect the property's value. This will make statistical analysis of the values associated with the zoning and use difficult.

- Any change in the zoning or use of the property could create new market values that have not been experienced or reached in the market before. This will make it difficult to ascertain a current sales value that pertains to properties for which the analysis is needed.
- The impact of zoning and change of use is not limited to surrounding properties.
- The most desirable properties of the sample area will sell first, creating a sample selection bias. Therefore, only properties that have already traded within the sample area will be used during the regression analysis, leaving out all unsold properties.

Mark (1988) describes the following advantages and disadvantages of a MRA:

Advantages:

- Once the data collection is complete, the model can be used frequently allowing an infinite amount of assessment.
- The above allows revaluation of property at low cost.
- A close relationship between market values and the interpreted value of a property are formed. This is also made possible during booming or declining markets. Further, the relationship at hand, MRA can be more accurate than a physical valuation.

Disadvantages:

- The initial set up costs of a MRA model in terms of the data collection and computer software can be expensive. Continuous training for staff operating the program also poses potential costs an problems.

- Substantial errors exist in MRA, even though it is possible to to calculate the size of the errors.

Another advantage highlighted by Leong (2002) is that a hedonic approach only needs to have certain information such as the property price, the differing housing attributes, and the functional relationship between the two.

From this position, only the coefficients of the estimated hedonic regression are needed to indicate the structure, and no information about the individual characteristics or personal particulars of the house, buyers or the suppliers are required.

2.3.2. Present day hedonic price modelling

Today's hedonic price modelling is not limited to statistical research, but applied practically. Banks are using hedonic models in the form of computer aided generated valuations for property valuations before passing bonds, and estate agents are using them to determine the listing price of a property.

In a study carried out by Abelson *et al* (2005), they explained the change in real house prices in Australia from 1970-2003. Within this study, they made it clear that depreciation and maintenance costs do not need to be included in the model.

After analysing past literature, their reasoning for not including it was that maintenance costs did not vary much year from year, and that depreciation is subsumed in the expected real house price.

Further analyses were provided and the reasoning given was that, in the short run, the prices of new houses are determined by the value of the existing housing stock. The costs of a new house can only affect the price of existing housing if new house supply significantly affects the size of the housing stock.

In other circumstances, for example, changes in the cost of new houses (be it taxes on developers or increases in construction costs), reduces the value of land for new housing, and does not affect the price of new houses.

Another assumption made, given that the sample range of data is over a long time period, was that in the long run, real house prices adapt to economic fundamentals and establish equilibrium. The importance of such an assumption was due to the many boom and bust observations experienced in the market. The existence of co-integration between some of the variables studied by Abelson *et al* (2005), implied that they move together through time, tracing a long-run path from which they are disturbed by temporary shocks, but in which they continually readjust.

Fletcher *et al* (2000) examined heteroscedasticity in hedonic house price models during their 1994 study of property sales in Stoke-on-Trent England. Their study included 1 286 observations which, according to them, allowed extensive modelling of the housing variables.

According to Fletcher *et al* (2000) heteroscedasticity is where the variances of the disturbance term of the model are unequal. An example would be where the variance of the disturbance term differed between types of properties (detached, multi-story, terraced) or if the variance differed with the age of the property.

A simple example of heteroscedasticity would include the analysis of income to expenditure on meals. If an individual's income increases, so would the choice of food he has to purchase. A poor person's choice of food would stay relatively constant and most likely be spent on fast food. A wealthy person may occasionally spend their money on fast food but also on very expensive meals. Therefore, individual's with a higher income, display more variability in their food selection, and when there is a large difference amongst the sizes of the observations, heteroscedasticity exists (Wikipedia 2009).

There are numerous tests for heteroscedasticity as noted on Wikipedia (2009), but the most noticeable are Park, White, Cook-Weisberg, Spearman Rank Correlation Coefficient and the Harrison McCabe tests. Current day statistical packages include the tests for heteroscedasticity, and the effects of these on the multiple regressions run will be tested.

Two potential processes highlighted by Wikipedia (2009) to fix heteroscedasticity include viewing logging the data or applying different sets of variables. When viewing the logged data, information that is seen to be growing exponentially usually have increasing variability over time. However, the variability in percentages may be miniscule and have less of an effect on the regression. Using a different set of independent variables can also help reduce the heteroscedasticity by minimising the potential for vast variance (i.e. including number of garages as opposed to size of erf).

Flectcher *et al* (2000) explain that the variances calculated by standard OLS procedures are biased and this implies that the standard tests (t and F statistics) are unreliable. Usually, the variance is underestimated and this leads to larger than normal t – statistics.

Fik *et al* (2003) delved into absolute location and studied the inclusion of {x,y} coordinates within hedonic price modelling. They expected that the inclusion of such coordinates, known as dummy variables, would have a significant impact on a model's ability to explain house price variation.

To measure the effect the locational dummies would have on the model, they concluded several studies. Firstly, they created an “aspatial model” which would serve as the benchmark for all subsequent models, and then ran various regressions that included lot size, age and dwelling size as quantitative variables.

They found that the use of absolute location in hedonic price modelling is important when the researcher does not have adequate prior information on the boundaries of neighbourhoods within a city. Furthermore, they explained that discrete definitions of neighbourhoods could possibly create significant value discontinuities. For example, a property backing onto a green belt or park may fetch a much higher price than a property several metres away, having unobstructed views of overhead power lines. These positive and negative externalities are averaged out when locational dummy variables are used.

In the conclusion given by Fik *et al* (2003), each property within an area can be thought of as having a unique location value signature in relation to the sum total of all externalities, which affects a given property/location.

Further findings suggested that the degree to which any one or more externalities affects real estate values is highly variable over space, and any hedonic models that do not directly incorporate absolute location will most likely fall short of explaining the true impact that location has on market price.

Boyle and Kiel (2001) looked into the impact that environmental externalities had on house prices and what potential consumers were willing to pay for them. The externalities studied include the likes of air quality, water quality and distance to toxic and non-toxic locations.

The coefficients of air quality proved to be statistically insignificant, but their coefficients were sensitive to other variables. Given that the results indicated correlation amongst some of the air quality variables, a hint of multi-collinearity existed.

Water was found to be statistically significant with the correct coefficient signs and, in conclusion, environmental externalities were found to affect property prices.

In the study carried out by Keng (1997), only six predictors were used. The reason for this was parsimony. A parsimonious model includes the least number of explanatory variables that permits a good representation of the dependent variable (price), and is less likely to be affected by the problem of multi-collinearity. Further, multi-collinearity exists in almost all multiple regression models and finding two completely uncorrelated variables is rare.

Hedonic analysis can be viewed as being preemptive, as multi-variate regression analysis includes data that is thought to account for property values. The research papers mentioned herein all include 'ordinary least squares multiple regression' and 'correlation matrices' models in their methodology.

The methodology employed in this research therefore also makes use of these tried and trusted techniques; given the accessibility to modern technology, these can be performed relatively easily on most laptops and computers, utilising statistical software.

2.4. Conclusion

The research papers reviewed from 1968 to 2002 have mostly used cross-sectional data over a period of one year that is in contrast to this study, incorporating sales over a thirteen-year period. The biggest problems highlighted by most of the authors are the poor quality of data used, spatial correlation, a lack of selling prices and multi-collinearity ((Abelson *et al* (2005), Boyle and Kiel (2001), Fik *et al* (2003), Hill and Melser (2008), Keng (1997), Li and Brown (1980)).

Another point of interest regarding the literature collected was that the majority of the regressions created included a multivariable approach that comprised three variable types. They were namely ‘location’, ‘house-related’ and ‘environmental’. Not all of studies analysed made use of the three variables, as some opted to use just two.

This study makes use of house related variables only, as multivariable data within South Africa is simply not available, or very limited. A complete model will make use of all variables, but where possible, a certain degree of effort to avoid bias has been placed on locational attributes.

Specific notice should be placed on the simple valuation technique employed to kick out extreme outliers, where the selling rate/m^2 does not align with the sample data’s average.

There are many possible reasons for the existence of such extreme outliers but that does not form part of this research. The literature covered, surprisingly, does not include this technique and it is presumed to be unique within a South African context.

Purchasers commonly have to make quick decisions vis-à-vis whether they want to buy, and they are likely to vary the importance they attached to individual attributes and their evaluation thereof. The important assumption from this analysis is that purchasers will try to be rational.

Taking the above into account, this study seeks to identify the important structural characteristics that contribute to the selling price of a house in the West Rand.

3. RESEARCH METHODOLOGY

Previous residential sales from Gauteng's West Rand, captured in the database by Property24.com, were used as the foundation of the study. Sub questions for the study from available fields in Property24.com's database were formulated. A hedonic pricing model was then used to analyse the data.

3.1. Types of research

There are two types of data, namely quantitative and qualitative, that can potentially be researched and then analysed. These two types of data can be approached in either a structured or unstructured manner respectively.

Kumar (2005) says that the choice of a structured or unstructured approach should depend on the aim of the inquiry (exploration, confirmation or quantification) and the use of the findings (policy formulation or process understanding).

3.1.1. Structured research

Structured research makes use of the quantitative approach, whereby everything that constitutes the research process, from objectives, design, sample and the questions intended for the respondents, is predetermined. Therefore it is more appropriate to use this method of research when exploring the subject matter's nature (Kumar, 2005).

Quantitative data examines hypotheses that are composed of variables that are usually analysed individually or as a whole. The results of the analysis of the hypotheses are expressed numerically, and usually through the means of statistics (Creswell, 2003).

According to Yin (2003) quantitative research is considered to be data-driven, outcome-orientated and scientific.

The numerical data that is going to be sorted and analysed is housed within a data spreadsheet, namely Microsoft Excel. The results of the analysis of the hypothesis are therefore expressed numerically with statistical inferences, as stated earlier by Creswell (2003). The hedonic price analysis of the structural characteristics of residential property in the West Rand takes the form of quantitative research, and the variables used are described further on.

3.1.2. Unstructured research

According to Kumar (2005) unstructured research makes use of the qualitative approach. The qualitative approach allows flexibility in all aspects of structured research. It is more appropriate to use this approach when determining the extent of a problem, issue or phenomenon.

Creswell (2003) suggests that qualitative research is fundamentally interpretive, which would result in the researcher making interpretations of the data. This would mean making a description of an individual or setting, analysing the data for themes and/or categories, and finally making an interpretation or drawing conclusions about its meaning.

Kumar (2005) says that a study is classified as qualitative if the purpose of the study is primarily to describe the above where the information is gathered through the use of variables measured on nominal or ordinal scales.

Kumar (2005) further adds that a qualitative approach can be used if an analysis is to be done to establish the variation in a situation, phenomenon or problem without quantifying it.

According to Yin (2003) qualitative data cannot be readily converted to numerical values. Data of this nature can be represented by categorical data, by perceptual and attitudinal dimensions (e.g. colour perception) and by real-life events.

Neuman (2003) describes a qualitative researcher as someone who builds theory by making comparisons after observing an event, and as someone who immediately ponders questions and looks for similarities and differences in that event. Table 3.1, extracted from Neuman (2003), shows the differences between quantitative research and qualitative research:

Table 3.1 Quantitative versus qualitative research

QUANTITATIVE RESEARCH	QUALITATIVE RESEARCH
Test hypothesis that the researcher begins with.	Capture and discover meaning once the researcher becomes immersed in the data.
Concepts are in the form of distinct variables.	Concepts are in the form of themes, motifs, generalisations and taxonomies.
Measures are systematically created before data collection and are standardised.	Measures are created in an ad hoc manner and are often specific to the individual setting or researcher.
Data is in the form of numbers from precise measurement.	Data is in the form of words and images from documents, observations and transcripts.
Theory is largely causal and is deductive.	Theory can be causal or non-causal and is often inductive.
Procedures are standard and replication is assumed.	Research procedures are particular and replication is very rare.
Analysis proceeds by using statistics, tables or charts and discussing how what they show relates to hypothesis.	Analysis proceeds by extracting themes or generalisations from evidence and organising data to present a coherent, consistent picture.

Neuman (2003)

3.2. Sample frame

Numerous estate agencies operate in Gauteng's West Rand. The database constructed by Property24.com includes all sales registered by estate agents with Property24.com for the entire West Rand region, dating from 01 January 1996 to 31 January 2008.

All sales registered in the database with the required fields completed, are incorporated into the study.

The following candidate variables have been identified as possible characteristics of open market value for residential properties in the West Rand:

Quantitative variables

- Stand size (m^2)
- Dwelling size (m^2)
- Number of bedrooms
- Number of bathrooms
- Number of recreational rooms
- Number of study rooms
- Number of garages
- Number of flatlets
- Number of dining rooms

Qualitative variables

- Swimming pool (Yes/No)
- Type of house construction, i.e. brick, plaster or both
- Suburbs (Allens Neck, Constantia Kloof, Wilropark, Florida Park, Strubensvallei and Krugersdorp)

The characteristics that determine the open market value of property in the West Rand with regard to qualitative and quantitative variables are highlighted by means of a hedonic price model.

3.3. Correlation analysis

Hanke and Reitsch (1994) indicate that in MRA, the first step is to identify the dependent and predictor variables to be included in the prediction model. A random sample is then taken, and all the variables are recorded for each sampled item. The third step is to identify the relationships between the dependent and predictor variables, and also among the predictor variables. It is mentioned that this analysis can be done using a computer program that produces a correlation matrix for the variables.

Therefore, when examining the regression model created, it is important to measure the relationship between the structural variables (independent) and the dependent variable (house price) through the use of a correlation matrix. This tool will help analyse the direction and strength of the linear relationship using an upper and lower limit of –1 to 1. If the coefficient of correlation equals 1, then a positive linear relationship exists, and if it equals –1, a negative relationship exists. If a relationship is established, multi-variate regression can be applied for further analysis.

This approach was also employed in the study conducted by Keng (1997) to better recognise the direction and the bond of the co-movements of the dependent and independent variables. A further point highlighted by Keng (1997) was that one needs to be aware of the correlation between two variables as the relationship between them may not be casual, and that the r-value just gives the direction and strength of the co-movement of the two variables.

According to Hanke and Reitsch (1994), if any two-predictor variables in a MRA are too highly correlated, they will interfere with each other by explaining the same variance in the dependent variable. This is known as multi-collinearity suggesting that the predictor variables are not independent.

3.4. Regression analysis

When the regression analysis on the data is run, only structurally significant characteristics at the 5% level is used to formulate the model. All insignificant characteristics are excluded to prevent weakening the models explanatory power in determining the open market value of a house.

The OLS multi-variate regression approach has been selected to determine the statistically significant characteristics of the above mentioned candidate variables on the selling price of a single residential house in the West Rand.

The regression therefore determines the expected selling price based on the significant candidate variables identified. Keller and Warrack (2003) summarises this method as a means of producing a straight line drawn through the points so that the sum of the squared deviations between the points and the line is minimised.

The statistically significant variables were identified using the P-stat. According to Keller and Warrack (2003), the P-stat of a regression is the probability of observing a test statistic at least as extreme as the one computed, given that the null hypothesis is true. In other words, it therefore measures the amount of statistical evidence that supports the alternative hypothesis.

Table 3.2, as mentioned by Keller and Warrack (2003), can be used to describe the P-stat:

Table 3.2 P-Stat level of significance

P-STAT FINDINGS	P-STAT OUTCOME
If P-stat is less than 1%:	Overwhelming evidence to infer that the alternative hypothesis is true, i.e. highly significant.
If P-stat lies between 1% and 5%:	Strong evidence to infer that the alternative hypothesis is true, i.e. result is significant.
If P-stat lies between 5% and 10%:	Weak evidence to infer that the alternative hypothesis is true, i.e. result is not statistically significant.
If the P-stat is greater than 10%:	No evidence to infer that the alternative hypothesis is true.

Keller and Warrack (2003)

When interviewed on 14 October 2009, Mr O. Adetunji[3], confirmed that the P-stat shows randomness, and when the P-stat is high, the t-Stat is low, and vice-versa. He further commented that a low P-stat means a good fit for the structural variable under analysis.

The OLS regression statistically equates to the following:

---

[3] O. Adetunji is the Statistics Lecturer at the University of the Witwatersrand for the Masters Programme

$$SP = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

where:

SP	is the expected selling price
α	is the intercept determining the regression
β	is the variable coefficient as determined by the regression
X	is the significant candidate variables
ε	is the residual error term

The above model cannot accurately predict the selling price of the observations under review and a residual error is therefore created. The residual error is based on the sample mean and is used as an estimate of the population mean. The residual exists as an unexplained value that relates to other factors not included in the model (Wikipedia, 2009).

According to O. Adetunji (2009) the residual error, denoted by ε_i can be traced to 'white noise'. He mentioned that if the residual error is low, then the model fits well, but if the residual error is high, then the model unfortunately does not capture reality. Therefore, if a multiple regression indicates a high R^2 (i.e. indicating a very good fit) but the residual error is also high, the independent variables individually do not explain the relationship between x and y very well, but together fit the model well.

The actual selling price is therefore given by:

$$\text{Actual SP} = \text{Predicted SP} + \varepsilon_i$$

With this theoretical background, the study followed a multifaceted regression approach, where certain variables were isolated in different regressions to best display the structural relationships with the selling price of a residential house. The following regressions were run:

1. A general multi-variated regression, including all observations and structural variables as independent variables, was run. This was done to observe the relationships (or lack thereof) existing amongst the structural independent variables and house price.

2. Firstly, all the extreme observations were removed from the data set to allow for a more accurate regression. Statistical and valuation techniques were employed to highlight and delete the outliers.

3. Secondly, the individual suburbs comprising the West Rand were grouped into four areas (NW, NE, SE and SW) to allow for the elimination of explanatory noise if all suburbs were to be included as individual dummy variables. The four larger areas were then run against selling prices to determine the statistical explanatory power of areas on prices. Thirdly, in order to generate a more complete model for the explanation of structural and locational effects on house prices, another regression was run to include all structural variables and suburb groupings (dummy variables) as independent variables.

4. Finally, an isolated regression was run on the structural variables and house prices for a bunched group of suburbs in the West Rand alone, as these suburbs held the most observations (largest sample). This was also deemed necessary to try to eliminate the statistical noise of locational variables in the regression.

3.5 Regression application

Given the limited quantity of sales data currently available for the West Rand, a regression application on the selected pocket of suburbs was unfortunately not possible. It was therefore decided to apply the practical regression analysis on all 203 properties that Property24 currently had for sale in the West Rand, by computing their structural variables into the model to observe their predicted house price.

These predicted house prices were then observed alongside their listed price and the relevant conclusions drawn as to whether or not the property was 'overpriced', 'under priced', or 'well priced'. If any obscurities were to be noticed, could they be explained by means not accounted for by the hedonic price model?

4. RESEARCH DATA, EMPERICAL RESULTS AND ANALYSIS

4.1. Data capture

The data from the hedonic price model was coded and then analysed using a structured medium in the form of Microsoft Excel and Number Crunching Statistical System (NCSS). This allows the data to be analysed using quantitative methods. After the data had been analysed, through the use of statistical inferences, namely a multi-variate linear regression model and correlation matrix, it was interpreted, the results assessed and a composite, statistically significant valuation profile formulated for future use.

4.2. Data adjustment

The original data provided included sales from January 1996 to January 2009 for all suburbs comprising the West Rand. To account for inflation and general increases in housing prices over the 13-year sample period, the selling prices were adjusted accordingly to 2009 property prices, using a property index calculated from the data. As highlighted by Ball (1973), it is better to include individual house prices that have been escalated accordingly, as opposed to calculated averages.

The data presented in Addendum 2 does follow a logical format except for Table 4.7. This table illustrates the rough data and the clumsiness surrounding the format and layout. This table was used to run the first regression.

Table 4.8 is the cleaned data set. This table has been presented in alphabetical order by suburb, with the adjusted selling price listed alongside the original selling price.

The data in Table 4.9 is presented by each properties location with respect to the four quadrants comprising the West Rand. The table begins with the North East Region and ends with the South West Region. The adjusted selling price is listed alongside the original selling price.

Table 4.8 presents the data making up a selected pocket of suburbs. The sales have been grouped by suburb with the adjusted selling price alongside the original selling price.

Table 4.1 gives a summary of the total sales value and volume for each year, along with each year's particular property index in relation to the 2009 base year for the original data set.

Table 4.1 Summary of West Rand house growth data

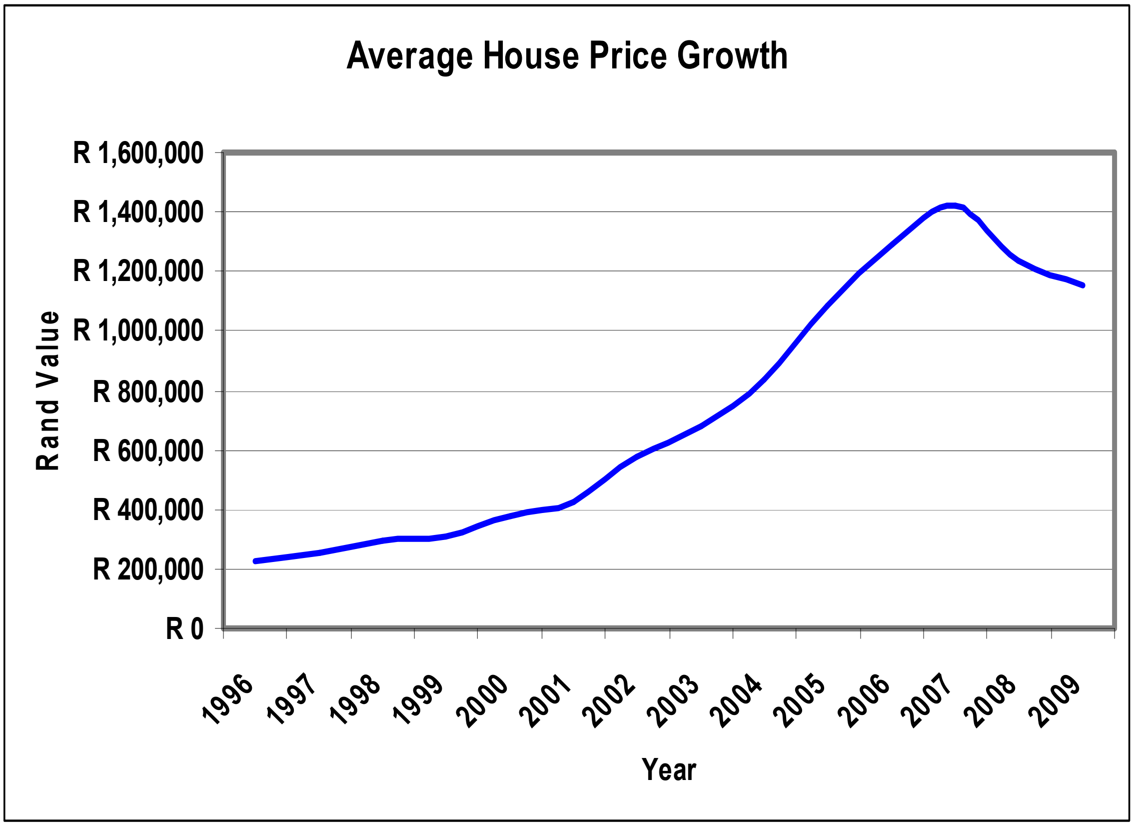
Year	Sum	Count	Average	Percentile
1996	R 8,366,200	37	R 226,114	5.10
1997	R 15,836,929	62	R 255,434	4.52
1998	R 25,215,554	86	R 293,204	3.94
1999	R 51,970,342	168	R 309,347	3.73
2000	R 74,940,580	198	R 378,488	3.05
2001	R 89,808,645	210	R 427,660	2.70
2002	R 97,325,281	169	R 575,889	2.00
2003	R 135,831,000	200	R 679,155	1.70
2004	R 205,807,807	245	R 840,032	1.37
2005	R 192,648,155	178	R 1,082,293	1.07
2006	R 228,426,738	177	R 1,290,547	0.89
2007	R 183,087,275	129	R 1,419,281	0.81
2008	R 90,478,999	73	R 1,239,438	0.93
2009	R 15,004,000	13	R 1,154,154	1.00

Authors own calculations using data from Property24.com

Of particular interest is the gradual increase in the count of property transactions and average sales value that occurred towards the end of the property boom, which came to an end during 2007. After this period, a sharp decline in the sales count and average sales value (marred by the credit crunch) is noticeable.

Figure 4.1 illustrates the gradual increase and sharp decline in property values over the thirteen-year sample period.

Figure 4.1 Average house price growth for the West Rand



Authors own calculations using data from Property24.com

In order to account for general inflation and real house price growth over the thirteen-year sample, a composite property index was formed. The leveling out of house values from the base year helps align the sample data for more accurate results when running the correlation and regression analysis.

This echoes the sentiments described by Hill and Melser (2008), who commented that in order for a price index to be significant, it must compare the sales of equivalent houses from one period to the next.

According to many property experts, South Africa has probably experienced its biggest property boom for many years to come. This is evident in *Figure 4.1* above which indicates a constant increase in property values from 1996 to about 2007. This period was considered the height of the property market, and when this 'property bubble burst', bank lending tightened and individuals struggled to keep up with mortgage bond repayments. This, amongst many other economic and financial factors, helped contribute to the sharp decline in property values as evident in *Figure 4.1* post 2007.

To help limit skewed or biased data pertaining to property locations, only suburbs pertaining to the West Rand were selected for the final analysis. The data was also adjusted to exclude any vacant land, commercial, farming and retail sales, as the study focuses purely on single residential sales.

The original data set included 1965 observations, and many numerous outliers existed. The following adjustments were made to the outliers:

- Sales outside the range R340 000 – R4 500 000 were excluded.
- Sales of any erven where the size was smaller than 250m², and bigger than 10 000m².
- Any records that had no entry pertaining to 'erf size' or 'house size'.
- In order to ensure that each recorded sale was a residential house, all entries that did not have a bedroom were excluded.
- All agricultural holdings were removed from the data.

4.3. Descriptive statistics

The total number of useable observations was 1 548.

4.3.1. Quantitative variables

Table 4.2 gives a summary of the descriptive statistics performed on the candidate variables of the regression.

Table 4.2 Summary statistics of regression variables

	Mean	Min	Max	Std dev
SP	R 1,044,865	R 352,015	R 4,478,867	R 501,862
BED	3.00	1.00	8.00	0.70
BTH	2.00	1.00	8.00	0.60
REC	3.00	0.00	7.00	0.85
STU	0.00	0.00	4.00	0.56
GAR	2.00	0.00	8.00	0.86
DRM	1.00	0.00	6.00	0.57
FLL	0.00	0.00	1.00	0.36
POOL	1.00	0.00	1.00	0.50
Brick	0.00	0.00	1.00	0.49
Plaster	1.00	0.00	1.00	0.49
SOH	250	100	900	114.68
SOE	1144	250	10000	898.89

Authors own calculations using data from Property24.com

SP = subject property BED = bedroom BTH = bathroom REC = recreational room STU = study GAR = garage CRP = carport DRM = dining room FLL = flatlet POOL = pool BRK = brick PLS = plaster SOH = size of house SOE = size of erf

4.3.2. Qualitative variables

Physical and locational attributes have been included in the model through the use of dummy variables.

Table 4.3 Number of observations for each region comprising the West Rand

North East Region Count	727
North West Region Count	251
South East Region Count	222
South West Region Count	348

Authors own calculations using data from Property24.com

The above table shows the number of sales for each suburb within the data set. The data was divided up into four even locationally distributed quadrants of the West Rand.

Table 4.4 shows the number of occurrences of each of the physical qualitative factors within the data set.

Table 4.4 Number of observations for each qualitative factor

Qualitative variable	**Count**
Pool	1037
Flatlet	229
Wall Type	
Brick	620
Plaster	896

Authors own calculations using data from Property24.com

4.4. Correlation analysis

The correlation matrix in Table 4.5 presents the quantitative candidate regression variables for the West Rand. Correlations greater than 50% have been highlighted in bold.

Table 4.5 Correlation matrix of structural variables for West Rand

SP	BED	BTH	REC	STU	GAR	CRP	DRM	FLL	POOL	BRK	PLS	SOH	SOE
1.00													
0.44	1.00												
0.60	**0.52**	1.00											
0.48	0.36	0.44	1.00										
0.32	0.19	0.29	0.30	1.00									
0.38	0.23	0.29	0.30	0.21	1.00								
-0.04	0.12	0.03	0.07	0.04	-0.25	1.00							
0.25	0.30	0.15	0.25	0.21	0.19	0.07	1.00						
0.17	0.12	0.13	0.12	0.06	0.05	0.16	0.03	1.00					
0.23	0.18	0.23	0.26	0.15	0.17	0.06	0.16	0.13	1.00				
0.03	0.03	-0.01	0.04	0.01	0.05	0.03	0.05	0.05	-0.04	1.00			
-0.03	-0.03	0.01	-0.03	-0.02	-0.04	-0.04	-0.05	-0.06	0.05	-0.99	1.00		
0.82	**0.53**	**0.62**	**0.56**	0.38	0.44	0.05	0.37	0.26	0.28	0.05	-0.05	1.00	
0.47	0.30	0.28	0.29	0.20	0.21	0.07	0.31	0.18	0.13	0.05	-0.05	0.49	1.00

Authors own calculations using data from Property24.com

'Selling price' (SP) shows two moderately correlated factors (greater than 50% correlation), being the 'number of bathrooms' (BTH) and the 'number of bedrooms' (BED), as well as four moderate to high correlated factors consisting of the 'size of house’ (SOH), the ‘number of bedrooms’ (BED), the ‘number of bathrooms’ (BTH) and the ‘number of recreational rooms'.

All of the other candidate variables have a weak correlation with selling price below 40%, except for ‘number of bedrooms’ (BED), ‘number of recreational rooms’ (REC) and ‘size of erf’ (SOE). This indicates that these factors are likely to have a high explanatory power in a multiple regression model. From a statistical point of view, certain concerns could be raised, as multi-collinearity exists.

Most of the variables, except for carport (CRP), flatlet (FLL), brick (BRK) and plaster (PLS) are significantly positively correlated to the selling price. It can therefore be anticipated that each of the factors is likely to add value and explanatory power to the model.

Table 4.6 represents the correlation matrix of the quantitative candidate variables for the selected pocket of suburbs located in the West Rand, which included Wilgehuewel, Allens Nek, Little Falls and Strubensvallei. Correlations that are greater than 50% have been highlighted in bold.

Table 4.6 Correlation matrix of structural variables for pocket suburbs

	SP	BED	BTH	REC	STU	GAR	DRM	FLL	POOL	BRK	PLS	SOH	SOE
SP	1.00												
BED	**0.59**	1.00											
BTH	**0.66**	**0.61**	1.00										
REC	**0.57**	0.44	0.45	1.00									
STU	0.39	0.25	0.39	0.37	1.00								
GAR	0.46	0.32	0.37	0.34	0.27	1.00							
DRM	0.46	0.39	0.42	0.38	0.21	0.28	1.00						
FLL	0.18	0.13	0.17	0.09	0.09	0.05	0.14	1.00					
POOL	0.29	0.30	0.31	0.23	0.11	0.15	0.28	0.12	1.00				
BRK	0.02	0.08	0.04	0.04	0.03	-0.01	-0.06	0.08	-0.03	1.00			
PLS	0.00	-0.07	-0.03	-0.03	-0.02	0.02	0.07	-0.07	0.05	-0.96	1.00		
SOH	**0.86**	**0.66**	**0.70**	**0.66**	0.44	**0.51**	0.50	0.23	0.32	0.02	-0.01	1.00	
SOE	0.44	0.40	0.30	0.35	0.14	0.20	0.28	0.10	0.15	-0.01	0.00	**0.52**	1.00

Authors own calculations using data from Property24.com

Cognisance of the moderate to high correlations between the selling price and the ‘number of bathrooms’ (BTH) and the number of bedrooms’ (BED), as well as the highly correlated factors of the ‘size of house’ (SOH), ‘number of bedrooms’ (BED), ‘number of bathrooms’ (BTH) and ‘number of recreational rooms’ (REC). These variables are therefore expected to have a high degree of explanatory power in the regression model.

All of the other variables have a weak correlation with selling price (below 40%) except for ‘number of garages’ (GAR), ‘number of dining rooms’ (DRM) and ‘size of erf’ (SOE).

4.5. Regression analysis

The preliminary investigations shown in *Tables 4.5* and *4.6* above, give rudimentary insight into the relationships of structural independents on the selling price of a house. Multi-variate regression analysis can therefore be used to further investigate and understand the complex relationships amongst the observations included within the study.

4.5.1. Structural attributes on the full data set

The original data included 1945 observations. When the regression analysis was run on this data, a very low R^2 was produced indicating a very poor fit. Any number of reasons could have been responsible for such an occurrence but the most noticeable was the extravagant figures used for some of the independent data and the unrealistic selling prices used for the dependent data.

Table 4.7 Regression output: included observations = 1 938

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	101592.1802	2155611.551	0.047129169	0.962415169
BED	-118245.2233	312936.2503	-0.377857225	0.705578318
BTH	110786.6542	396320.7885	0.279537832	0.779862139
REC	-63039.19112	231912.1919	-0.271823532	0.785786859
STU	711099.9543	353901.3976	2.009316604	0.044642911
GAR	502272.8782	231790.9796	2.166921591	0.030363126
DRM	195324.1715	351305.8286	0.555994679	0.578279108
FLL	-116548.5671	570554.6371	-0.204272404	0.838162227
POOL	-724199.9159	420665.0987	-1.721559307	0.085310038
Brick	855126.9439	2010886.177	0.425248805	0.670702841
Plaster	591743.5732	2004635.09	0.295187676	0.767882349
SOH	200.4707052	1096.930187	0.18275612	0.855008624
SOE	42.83013165	119.8815644	0.357270377	0.720928567

Regression Statistics	
Multiple R	0.089696532
R Square	0.008045468
Adjusted R Square	0.001861855
Standard Error	8625353.009
Observations	1938

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	12	1.16157E+15	9.67972E+13	1.301095049	0.210719648
Residual	1925	1.43214E+17	7.43967E+13		
Total	1937	1.44375E+17			

Dependent variable: SP; Method: Least Squares

In order to clean the data, any unreasonable observations in terms of the dependent and independent data were removed. Examples of such data ranged from 21 reception rooms, 9 studies, selling price of R1, and "size of erf" 1m². The statistical technique known as rank and percentile was adopted to help highlight the outliers present in the observations.

This data would obviously skew all chances of formulating statistically significant interpretations of the observations used, and were therefore deleted from the spreadsheet.

Further analysis was employed to help highlight any extreme outliers that statistical techniques could not highlight. This analysis comprised of a common property valuation technique known as “rate adjustment”.

The basis of this theory is that the smaller the size of the dwelling, the higher the property’s selling rate/m^2, and the bigger the property the lower the selling rate/m^2 should be. On this note, it was decided to calculate the property’s selling price against its dwelling size, and that would equate to its selling rate/m^2.

It was found that a large portion of the sales were either way below or way above the average selling price for the full data set. The extreme outliers were then erased to create a more comprehensive and compact data set.

A multi-variate regression was then run on the full 1 548 observations in the West Rand of selling prices from June 1996 until March 2009. The independent variables of structural attributes were regressed against the dependent variable of selling price, in order to determine which factors are significant in influencing house price in the West Rand housing market.

Table 4.8 Regression output: included observations = 1 548

Variables	*Coefficients*	*Standard Error*	*t Stat*	*P-value*
Intercept	59598.23637	96353.20266	0.618539236	0.536311713
BED	-38941.97399	12396.02357	-3.141489185	0.001713003
BTH	135751.6936	15962.20903	8.504568093	4.26038E-17
REC	3793.599923	10222.54075	0.371101472	0.710613164
STU	-2527.600429	13915.40464	-0.181640455	0.85588888
GAR	12335.76944	9278.69542	1.329472397	0.183889708
DRM	-49584.37819	13706.46874	-3.617589558	0.000306993
FLL	-82788.53993	21064.76387	-3.930190742	8.86522E-05
POOL	2540.168218	15310.21952	0.165913246	0.868247072
Brick	-2662.512483	89919.9016	-0.029609824	0.976382079
Plaster	797.1981191	89724.97787	0.008884907	0.992912117
SOH	3162.714067	103.9546196	30.42398769	1.7072E-159
SOE	59.17044437	8.424600655	7.023531061	3.23926E-12

Regression Statistics	
Multiple R	0.833715305
R Square	0.69508121
Adjusted R Square	0.69269748
Standard Error	280406.8134
Observations	1548

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	12	2.75129E+14	2.29274E+13	291.5939407	0
Residual	1535	1.20694E+14	78627980992		
Total	1547	3.95823E+14			

Dependent variable: SP; Method: Least Squares

The structural attributes included are: 'number of bedrooms' (BED), 'number of bathrooms' (BTH), 'number of recreational rooms' (REC), 'number of study rooms' (STD), 'number of garages' (GAR), 'number of carports' (CRP), 'number of dining rooms' (DRM), 'number of flatlets' (FLL), 'presence of pool' (POOL), 'type of house structure' (brick or plaster), 'size of house' (SOH), and 'size of erf' (SOE).

The model achieved a high degree of fit as expressed by the R^2 of 69.50%. This means that 69.50% of the variation in selling price is explained by the structural attributes. The standard error is R280 406. When this is compared to the mean of R1 044 865, it is moderately high at 26.83% and indicates that the linear model's fit is moderately inadequate, limiting its usefulness as a predictive model.

The model itself is however sufficient in explaining the variation in selling prices across the West Rand, but it must be made aware that the standard error is fairly high. This could possibly indicate that locational and environmental factors have a large degree of influence on the variance.

Most of the structural attributes were found to be significant at the 5% confidence interval, except 'number of recreational rooms', 'number of study rooms', 'number of garages', 'the presence of a pool', and 'type of house construction' (brick or plaster).

A surprising statistic is the negative relationship that 'the number of bedrooms' has with the selling price of the house. This goes against common valuation theory that the higher the number of bedrooms present in a house, the higher the selling price. This could possibly indicate that a linear model does not correctly account for the declining marginal utility of a number of bedrooms with selling price.

Another interesting statistic is noted with the garage and pool. Common valuation theory suggests that a pool and garage add significant value in the determination of a house price.

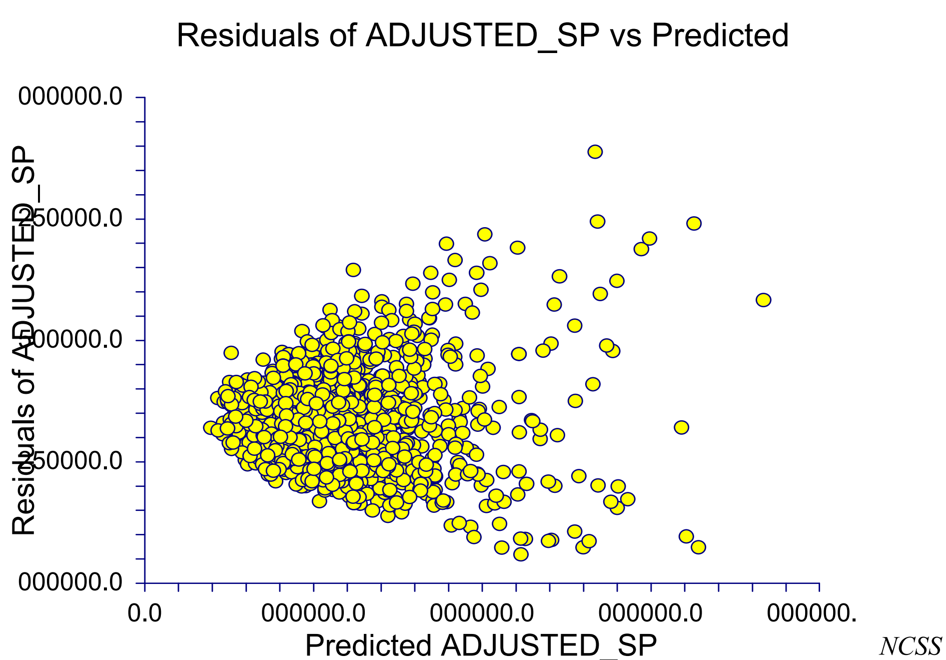
This is especially true in Johannesburg as there is no coastline (therefore the value add attached to a pool is significant), and security threats regarding stolen motor vehicles are high (thus the need for one's car to be locked away in a garage is significant).

For a more reasonable conclusion, it can be considered that the moderate fit of the model as a whole, to the variation in selling prices in the West Rand, has caused the result.

As mentioned earlier, heteroscedasticity and residual error exist in MRA and could potentially affect the P-statistic accepted at the 5% significance level. The test for the above was therefore run on the data used in Addendum 2 Table 4.8 in the statistical package NCSS.

The findings can be viewed in Figure 4.2 below which illustrates the level of heteroscedasticity and residual error.

Figure 4.2 Heteroscedasticity for data set Table 4.8

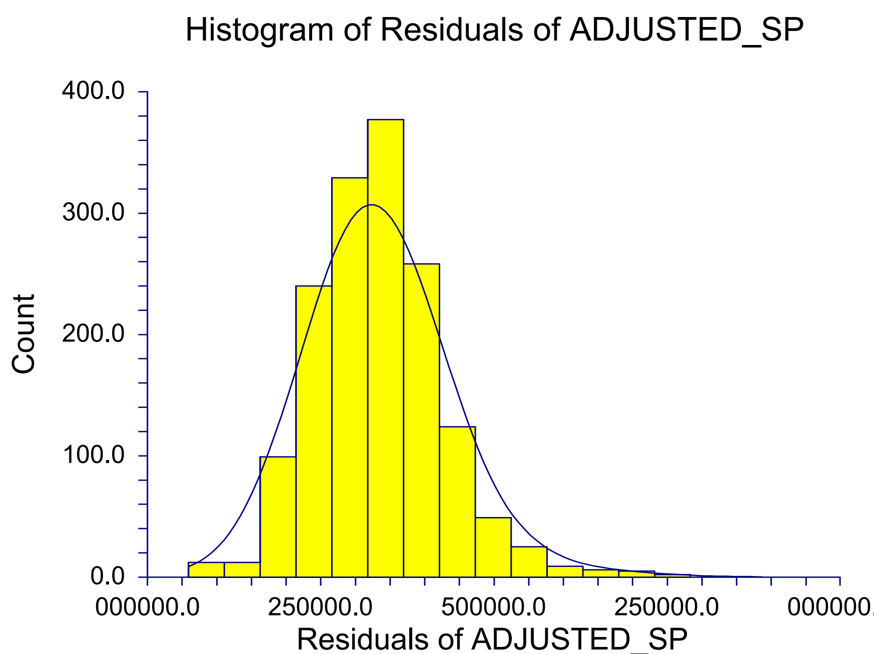


Heteroscedasticity Output

The heteroscedasticity present in the graph above can be interpreted as being fairly large. If a line is drawn down the centre of the graph, the left hand side of the graph does not mirror the image created on the right hand side of the graph. This implies that for properties at the greater end of the price/value range, large variances will exist in the predicted selling price of a house.

If there is large heteroscedasticity, residual error invariably exists. The figure below illustrates the residual error present in regression output run on Addendum 2 Table 4.8.

Figure 4.3 Histogram of residual error for data set Table 4.8

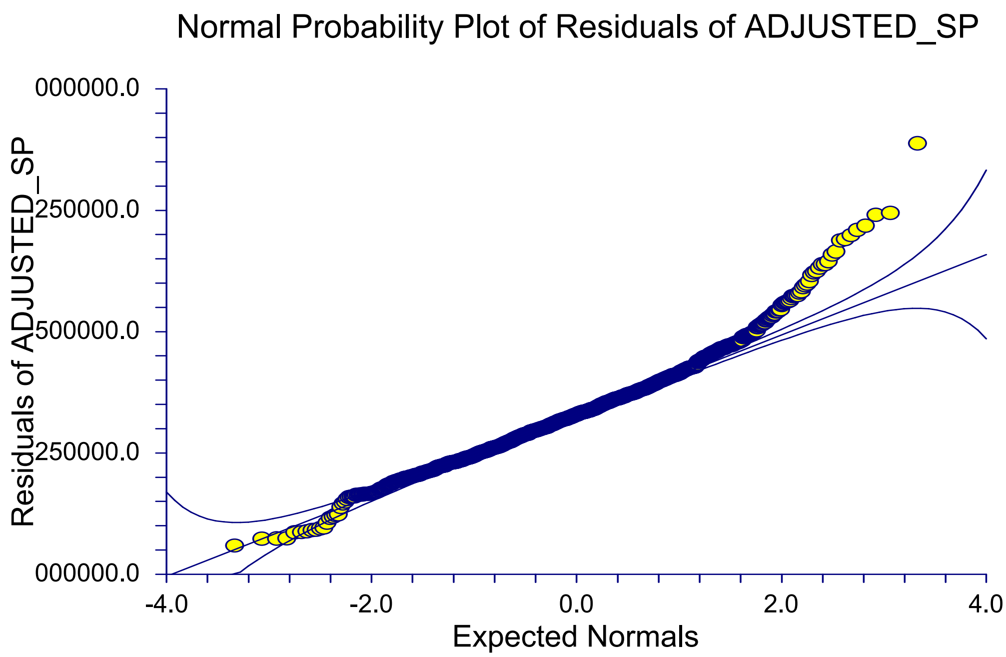


NCSS Histogram of Residual Error

The figure above indicates that a large amount of residual error exists in the regression formulated for data used in Addendum 2 Table 4.8. The figure presents a non-normal distribution which is skewed to the right.

This indicates that there is a much higher degree of fit, or normal fit, at value/prices at the lower end than for prices/values at the higher end of the West Rand property market. The final graph that can be used for testing residual error can be seen below.

Figure 4.4 Normal Probability Plot of residual error for data set Table 4.8



NCSS Normal Probability Plot of Residuals

This figure above plots the residual errors along the 45 degree frontier. As can be seen, a lot of residual error is present at high end property prices compared to the lower end prices. This figure illustrates the level of white noise existing 'behind the scenes' of the R^2 and how this is affecting the overall fit of the model at 69.5%.

The three figures above indicate that the data used in the regression may present too much variance, and the power of the model to predict property prices in the West Rand are not so reliable at the R^2 of 69.5%.

Therefore, in order to account for the heteroscedasticity and residual error present in the regression run on data in Addendum 2 Table 4.8, the inverse log of the adjusted selling price was used as the dependent variable, and regressed against the independent variables. This is similar to the semilogarithmic regression run by Brown (2004). The results of the regression are indicated below in Table 4.9.

Table 4.9 Regression output: Log of Adjusted SP included observations = 1 548

Run Summary Section							
Parameter	**Value**	**Parameter**	**Value**				
Dependent Variable	LOG_ADJ_SP	Rows	1548				
Number Ind. Variables	8						
Weight Variable	None						
R 2	0.6631						
Adj R2	0.6614						
Coefficient of Variation	0.0165						
Mean Square Error	9.84E-03						
Square Root of MSE	9.92E-02						
Ave Abs Pct Error	1.325						
Regression Equation Section							
	Regression	**Standard**	**T-Value**		**Reject**	**Power of**	
Independent	**Coefficient**	**Error**	**to test**	**Prob**	**H0 at**	**Test**	
Variable	**b(i)**	**Sb(i)**	**H0:B(i)=0**	**Level**	**5%?**	**at 5%**	
Intercept	5.6187	0.0114	493.909	0	Yes	1	
BTH	0.0433	0.0054	8.044	0	Yes	1	
DRM	-0.0199	0.0048	-4.159	0	Yes	0.986	
FLL	-0.0231	0.0074	-3.118	0.0019	Yes	0.8765	
GAR	0.0155	0.0033	4.731	0	Yes	0.9972	
REC	0.0081	0.0036	2.276	0.023	Yes	0.6241	
SOE	0	0	3.025	0.0025	Yes	0.8567	
SOH	0.0009	0	26.231	0	Yes	1	
STU	0.015	0.0049	3.052	0.0023	Yes	0.8626	
Analysis of Variance Section							
			Sum of	**Mean**		**Prob**	**Power**
Source	**DF**	**R 2**	**Squares**	**Square**	**F-Ratio**	**Level**	**-5%**
Intercept	1		56227.43	56227.43			
Model	8	0.6631	29.8177	3.727212	378.689	0	1
Error	1539	0.3369	15.14747	9.84E-03			
Total(Adjusted)	1547	1	44.96516	2.91E-02			

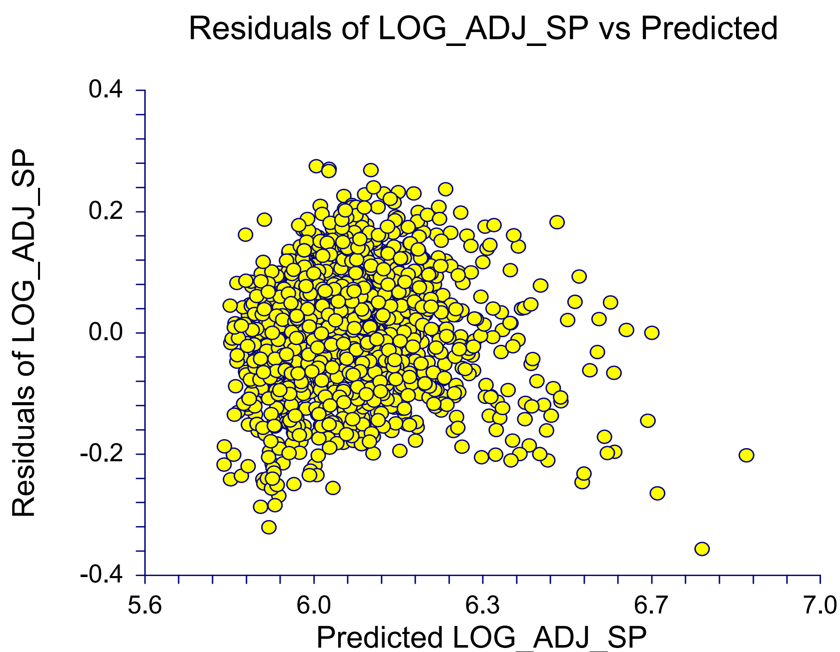
Dependent variable: SP; Method: Least Squares

The R^2 of the model has decreased slightly from 69.5% to 66.31%. An immediate observation would reveal that the model's predictability of property prices in the West Rand has decreased.

Hedonic price analysis of the structural characteristics of residential property in the West Rand

However, the results indicate that 9 independent variables are acceptable at the 5% significance level as opposed to 7 calculated in the previous model. Therefore, the explanatory power of the model as a whole will be more accurate.

Figure 4.5 Heteroscedasticity for data set Table 4.8

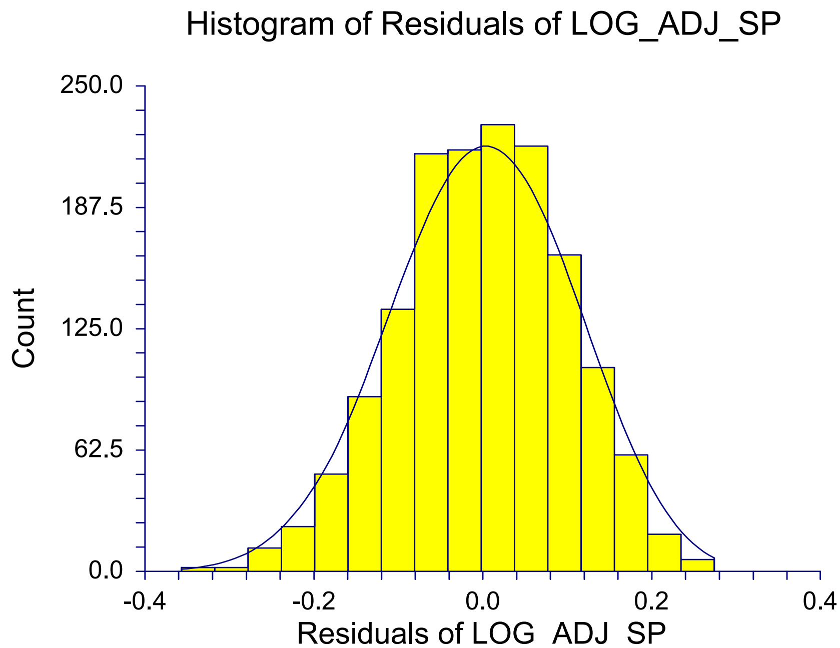


NCSS Heteroscedasticity Output

The figure above indicates a tighter bunched set of residuals than the previous heteroscedacticity results in Figure 4.8. This implies that the model has increased explanatory power and can more accurately predict the selling price of a property.

The histogram below indicates a improved distribution of data which can be considered normal. It must be noticed that there is very limited skewness to the left and right implying a more accurate output than the previous distribution curve illustrated in Figure 4.2. This means that 95% of the time, the results for each coefficient be normally distributed.

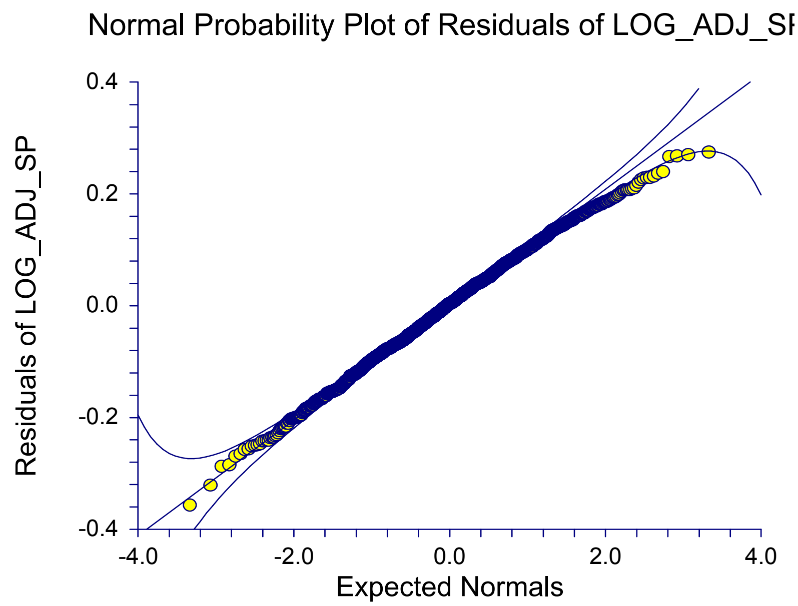
Figure 4.6 Histogram of residual error for data set Table 4.8



NCSS Histogram of Residual Error

The plot of residual errors along the 45 degree value frontier can be seen below. They represent a much better fit along the value frontier indicating improved predictability for higher end property prices.

Figure 4.7 Normal Probability Plot of residual error for data set Table 4.8



NCSS Normal Probability Plot of Residuals

The only down fall of using the log of the adjusted SP is that the coefficient figure calculated gives a proportionate or percentage amount that each significant variable would attribute to a houses selling price instead of a monetary value.

This study seeks to ascertain the monetary value attached to each structural attribute at the 5% significance level and a back transformation from a percentage to a monetary value was not possible.

The log of the adjusted SP was therefore not utilised any further to calculate the multivariate regressions run on the proceeding data sets.

4.5.2. Structural variables and area groupings

Individual suburbs have been grouped together into four suburbs as locational dummies, in order to explain the locational attributes in the multi-variate model.

Table 4.10 Regression output: included observations = 1 548

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	32330.14491	133182.4194	0.242750845	0.808230911
BED	-31428.63322	16844.78735	-1.865777974	0.062263409
BTH	150290.9061	21702.50751	6.925047997	6.39164E-12
REC	9061.855695	13885.33751	0.652620485	0.514098889
STU	25172.42878	18874.31004	1.333687363	0.182504583
GAR	16731.92837	12633.91801	1.324365756	0.185579138
DRM	-46646.82828	19052.12596	-2.448379166	0.014461527
FLL	-40562.6052	28698.06451	-1.413426511	0.157733504
POOL	5229.851035	20819.04686	0.251205114	0.80168922
Brick	23283.83935	122166.4404	0.190591125	0.848871196
Plaster	45878.47886	121833.6564	0.376566543	0.706547901
SOH	3025.063839	141.6247512	21.35971158	8.11889E-89
SOE	56.71331688	11.71454418	4.841273892	1.42038E-06
NE	-6833.09108	25506.5473	-0.267895572	0.788815766
NW	-53930.61711	32753.42743	-1.64656408	0.099852777
SE	-168409.2162	33124.89049	-5.084068616	4.14896E-07

Regression Statistics	
Multiple R	0.759151249
R Square	0.576310618
Adjusted R Square	0.572162224
Standard Error	379723.9382
Observations	1548

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	15	3.00472E+14	2.00315E+13	138.9237659	1.797E-272
Residual	1532	2.20899E+14	1.4419E+11		
Total	1547	5.21371E+14			

Dependent variable: SP; Method: Least Squares

Table 4.11 Regression output: Included observations = 1 548

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1239879.769	29953.89343	41.39294187	1.4078E-252
NE	70478.8879	36424.22391	1.934945493	0.053179425
NW	-227497.9236	46273.23162	-4.916404488	9.75424E-07
SE	-349923.3866	47997.01412	-7.290524068	4.91036E-13

Regression Statistics	
Multiple R	0.274466235
R Square	0.075331714
Adjusted R Square	0.073535079
Standard Error	558782.6362
Observations	1548

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	3.92758E+13	1.30919E+13	41.92933063	4.78117E-26
Residual	1544	4.82096E+14	3.12238E+11		
Total	1547	5.21371E+14			

Dependent variable: SP; Method: Least Squares

As predicted, when the regression included all the structural variables and all of the area groupings (Table 4.10), a R^2 of 57.63% is achieved.

This means that including locational variables, and comparing it to the regression run only on the structural variables, with R^2 of 69.50% (Table 4.8), the explanatory power of the model has decreased substantially. This is an interesting find as Garrod and Willis (1992) indicated in their study that adding dummy variables does not have a substantial affect on the results, but in this case it did.

Bearing the above in mind, a closer review of the four groupings in isolation, (see groupings in Addendum 1) where only the area groupings have been regressed against selling price, a R^2 of only 7.53% has been achieved (Table 4.11). This is consistent with the model in Table 4.10, showing that the locational dummies have very little explanatory power toward the determination of selling prices in houses located within the West Rand.

Table 4.9 in Addendum 1 illustrates how the various suburbs of the West Rand were isolated into four distinct area groupings used in the regression. The regression result for Table 4.9 in Addendum 2 does not include the dummy variable for the SW quadrant as the input data used to run the regression for this quadrant mimics the data input for the SE quadrant. Therefore, in terms of the regression, this SW data inputs are excluded. The data input can be viewed in Addendum 2 Table 4.9.

4.5.3. Structural attributes for selected pocket of suburbs

In order to account for locational attributes and isolate the impact of structural variables on selling price alone, a multi-variate regression was run on 234 observations for the selected pocket of suburbs.

The structural independent variables were regressed against the explanatory variable of selling price in order to determine which factors are significant in influencing house prices in the selected suburbs housing market, as well as determining a theoretical value for these factors. The same structural variables as used in the full data set were also used for the selected pocket.

Table 4.12 Regression output: Included observations = 233

	Coefficients	*Std. Error*	*t Stat*	*P-value*
Intercept	57175.70257	154764.12	0.369437713	0.712156835
BED	1273.926581	33159.67349	0.038417947	0.969389291
BTH	76774.51702	41373.61893	1.855639391	0.064842453
REC	-8569.629751	23932.44341	-0.358075839	0.720629887
STU	5009.447723	36928.35772	0.135653141	0.892219532
GAR	12784.5458	21160.87112	0.604159712	0.546359971
DRM	23929.71031	34380.2459	0.696030807	0.487144189
FLL	-44540.2202	54658.81607	-0.814877149	0.416023878
POOL	4702.33417	34614.66015	0.135848053	0.892065623
Brick	158526.6547	125314.1042	1.265034417	0.207197295
Plaster	159206.1341	123923.3643	1.284714428	0.200243133
SOH	3022.086563	274.8484188	10.99546643	1.06987E-22
SOE	4.848602957	40.18614987	0.120653583	0.904075545

Regression Statistics	
Multiple R	0.867261934
R Square	0.752143262
Adjusted R Square	0.738623803
Standard Error	243101.6732
Observations	233

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	12	3.94547E+13	3.28789E+12	55.63412645	8.63175E-60
Residual	220	1.30017E+13	59098423534		
Total	232	5.24563E+13			

Dependent variable: SP; Method: Least Squares

This model has achieved a relatively high degree of fit, as expressed by the R^2 of 75.21%. This means that the selected independent variables explain 75.21% of the variation in the selling price. When compared to the regressions performed using all of the observations, as well as the regression using only the areas, this model indicates a high degree of fit, and can be considered a powerful model in explaining the selling price of a house.

By isolating these suburbs, the locational element of the regression is removed. This assists in the analysis of the explanatory power of the structural variables on its own, without all of the locational "noise". This highlights the effect that the "noise" has on the regression model, as purely structural factors along with all of the locational variables, reduced the R^2 to 57.63% (Table 4.10). It shows that the addition of locational variables have a negative impact on the model, rather than adding power to its predictability.

This allows a conclusion, based on the fact that the best approximation of a house situated within the West Rand, would exclude the negative "noise" associated with locational variables, and would allow the final model to be focused on one area alone.

The standard error of the model is R243 101 (23.04%), and is slightly high when compared to the mean selling price of R1 055 291. This means that the linear model's fit is not excellent and, as an analytical and forecasting tool, is only adequate. This means that as a predictive tool, the selling price of a house in the West Rand may be inaccurate by up to R243 101 (23%), and cognisance of this risk must be taken into account.

The standard error represented in this model is lower than the other models generated, indicating that this model has a better fit. Despite this model's standard error, it can still be used to identify mispricing in the property market. This application is demonstrated later in section 4.6.

An investigation into the coefficients of the observations helps highlight the existence of multi-collinearity between the explanatory variables. The correlation matrix (see table 4.6) of the explanatory variables for the selected suburbs indicated that 'number of bathrooms' (BTH) and the 'number of bedrooms' (BED), as well as four moderate to high correlated factors consisting of the ‘size of house’ (SOH), the ‘number of bedrooms’ (BED), the ‘number of bathrooms’ (BTH) and the ‘number of recreational rooms', were correlated with selling price.

According to Keller and Warrack (1994), correlations that are not greater than the high level of 80% would lead us to believe that multi-collinearity should not be a problem in the calculated model.

Exactly half of the variables were found to be significant, except for the ‘number of recreational rooms”, the “number of study rooms”, the “number of garages”, the “presence of a pool” and the type of house construction “brick or plaster”. These have therefore been excluded from the final model.

Analyses of the coefficients indicate that a positive relationship exists between the significant variables and selling price, except for the “number of bedrooms”, the “number of dining rooms” and the “presence of a flatlet”. An intuitive answer to the positive relationships are that an increase in the “number of bathrooms”, the “size of the house” and the “size of the erf” will add significant value to a house.

A possible answer to the significant yet negative relationship between the “number of bedrooms” and selling price of a house could be that the selected suburbs are of an old nature, and investors purchasing in the selected areas would look at the size of the house as the overall determining factor when making the purchase, with the view to renovate and include further bedrooms at the expense of other rooms.

From the data, it can be construed that for every additional bathroom present in a house, it will add R139 877 to the overall value of the property. This is consistent with the size of a house and the size of the erf, where each respectively adds R3 234 and R56 per m^2.

The downside to the model is that a reasonable portion (approximately 29%) of the selling price in the selected areas are not explained by the variables included in the model. Factors such as structural integrity, condition and age have not been included in the model as this information was simply not available.

In studies conducted by Cubbin (1970), Kain and Quiqley (1970), Li and Brown (1980), Apps (1971), Wilkinson (1971), and Apps (1971), the age of the property was shown to be highly significant in the determination of a house price. Intuitively, it is presumed that the younger a property, the higher its selling price will be.

The West Rand is a predominately old region meaning the variable relating to age will have a great impact on the determination of house price. This positive bias will therefore inhibit the models ability to accurately predict house prices within the selected area as it will regard all properties sold as being exactly the same age.

4.6. Application of the model for the entire West Rand region

All of the significant variables obtained from the regression analysis run under Table 4.8 have been extracted and used to formulate a pricing model. The insignificant variables have been excluded to eliminate "noise" and allow for a more powerful and accurate model. The following model was formed:

$$\mathbf{Price_i = 59598.24 - 38942[BED_i] + 135751.7[BTH_i] - 9584.4[DRM_i] - 82788.5[FLL_i]}$$
$$\mathbf{3162.714[SOH_i] + 59.17044[SOE_i]}$$

There were 203 property sales in the West Rand and their representative structural attributes were dropped into the model to obtain a theoretical valuation of the property. The predicted and actual values for each of the 203 properties sampled were plotted against each other relative to a fair value frontier (45-degrees from the origin).

To accommodate for the standard error present in the model, an upper and lower standard error bound was put in place. This can be viewed in *Figure 4.8*, as each line is plotted one standard error (approximately R285 503) apart on either side from the fair value frontier.

The chart reveals that the model's predicted value is in line with the agent's valuation on the relevant properties the majority of the time, therefore falling within the upper and lower bounds. It can therefore be concluded that the model has strong explanatory power in predicting housing prices within the selected area, ignoring locational and "noise" factors.

The chart indicates that 32 properties were under priced (lying above the "standard error upper bound") which for potential investors could be viewed as bargain buys.

Of much importance is that the model identifies 62 houses as being over priced (lying below the "standard error lower bound") which for potential investors could be considered as a poor investment.

The discrepancies between the sales price and predicted values could be related to a mismatch in pricing. Reasons could be for factors not included in the model such as tennis courts, alarm systems, under floor heating, aesthetic appeal, building quality or locational factors.

Figure 4.8 illustrates the model's usefulness in determining housing prices of properties located within the West Rand, and presents itself as an important valuation tool that property valuers, banks, real estate agents, and potential buyers can utilise.

This chart illustrates the values ascertained for residential properties located in the West Rand of Johannesburg. It plots the predicted selling price against the sales price and compares them to the 45-degree true value frontier. The standard error, shown by the upper and lower bounds, attempts to show miss-priced properties.

Figure 4.8 Value / price chart for West Rand

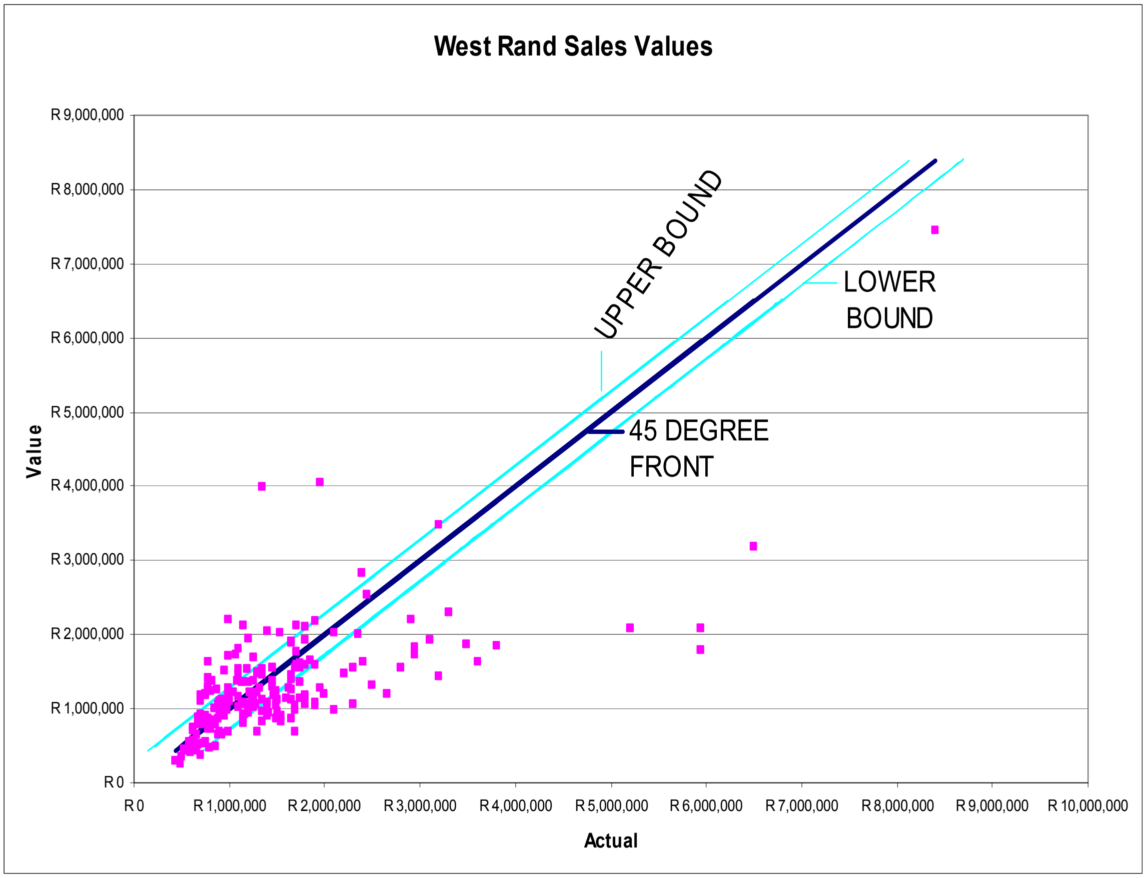


Figure 4.2: Value / price chart for West Rand

5. CONCLUSION

This study has helped indicate the usefulness of employing an automated valuation technique when assessing the potential selling price of properties located in the West Rand of Johannesburg. This technique could therefore be employed by a multitude of individuals from estate agents, property valuers, potential investors, property developers and banks.

The application of this model holds various opportunities for the individuals or organisations listed above. The most noticeable would be the quick determination of potential selling price, the cost and time savings with regard to paying professional property valuers for their services, and the time taken to prepare property valuation reports.

The noticeable disadvantages are that it does not accurately account for locational, aesthetics or building quality of properties that a property valuer would be able to take into account when conducting a physical valuation on the property. Furthermore, it cannot be accurately extended to other Gauteng regions.

This study is however limited in terms of its scope of property appraisal and only focuses on some of the structural determinants through the use of multi-variate regression analysis – which has formerly become known as 'hedonic price modelling' in modern day statistics.

Although this model excludes important micro and macroeconomic influences, locational and aesthetic design of properties which could influence the value, it has still managed to create a satisfactory relationship (a R^2 of 69.50%) between the statistically significant structural variables ('bedroom', 'bathroom', 'dining room', 'flatlet', size of house', and 'size of erf') and the selling price of single residential properties in the West Rand.

This study did not exclude locational attributes in its entirety and attempted to illustrate the effect such a variable would have on potential selling prices. It was therefore decided to include such an aspect in terms of a dummy variable and observe the effects. As expected, the relationship amongst the independent variables and the selling price of properties dropped significantly to a R^2 of 57.7%.

In an attempt to review the statistical "noise" present in the model highlighted by the locational regression, a regression model concentrating on a tight-knit pocket of suburbs was run. The results revealed a R^2 of 75.2% indicating a high degree of fit, and that locational factors play a big part in the determination of properties located in the West Rand.

Sales for the entire West Rand area were then pulled from Property24's database for sales dated between November 2009 and January 2010 and used in a practical application of this model. The screened regression was then run on these sales for analysis. The graph indicated that there was a good degree of fit for properties run parallel to the 45-degree fair value frontier, and that the model could be used for analysing potential sales of properties located within the West Rand.

As mentioned earlier, research within a South African context on hedonic house price data is very limited. It is hoped therefore that the conventional call for more research has been justified, and that the direction of future research is clear.

From the hedonic model created, the implicit prices for attributes of houses can be established. This information can be used to improve the planning, development, construction and management of properties in the West Rand area.

6. RECOMMENDATIONS

The availability of property data within South Africa is limited. Added to this, the quality and reliability of any information available is unfortunately very poor. It is presently noted therefore, that a huge gap in the property market exists, where property companies could prosper from the collection and proper recording of all information relating to property transactions.

Given that current studies within a South African context are limited, it is noted that possible areas for research could include the effects that locational, environmental, aesthetics and building quality have on property prices. Further examples would include the following:

- Distance from CBD
- Population density
- School density
- Distance from shopping centers
- Distance/density of greenbelts/parks
- Pollution measurement
- Colour of roof/walls
- Type/construction of perimeter wall

7. REFERENCES

Abelson, P. Joyeux, R. Milnovich, G and Chung, D. 2005. Explaining house prices in Australia: 1970-2003. *The Economic Record*, 81(255):96-103.

Apps, P. 1971. An approach to modelling residential demand. *CES Seminar* given, December 1971.

Ball, M. 1973. Recent empirical work on the determinants of relative house prices. *Urban Studies*, 10:213-233.

Bover, O & Velilla, P. 2002. Hedonic house prices without characteristics: The case of new multiunit housing. *European Central Bank*, 117:5-6, January.

Boyle, M and Kiel, K. 2001. A survey of house price hedonic studies of the impact of environmental externalities. *Journal of Real Estate Literature*, 9(2)177-144. [Abstract from EBSCOhost].

Brown, C. 2004. Hedonic regression models using in-house and out-of-house data. *Monthly Labour Review*, December: 25-38. [Abstract from EBSCOhost].

Creswell, J. 2003. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (2nd ed). Sage Publications, pp153 & 181.

Cubbin, J.1970. A hedonic approach to some aspects of the Coventry housing market. *Warwick Economic Research Papers 14.*

Fik, T. Ling, D and Mulligan, G. 2003. Modelling spatial variation in housing prices: A variable interaction approach. *Real Estate Economics*, 31(4):623-646.

Fletcher, M. Gallimore, P and Mangan, J. 2000. Heteroscedasticity in hedonic house price models. *Journal of property research*, 17(2)93-108. [Abstract from EBSCOhost database].

Hanke, J and Reitsch, A. 1994, *Understanding business statistics*, Richard D. Irwin Inc, USA.

Garod, G and Willis, K. 1992. The environmental economic impact of woodland: a two stage hedonic price model of the amenity value of forestry in Britain. *Applied Economics,* 24:715-728.

Goodman, A. 1978. Hedonic Prices, Price Indices and Housing Markets. *Journal of Urban Economics*, 5:471-484. [Abstract from EBSCOhost database].

Gressel, J. Milon, J. & Mulkey, D. 1984. Hedonic amenity valuation and functional form specification. *Land Economics*, 60(4):378-387. [Abstract from JStor database].

Hill, R & Melser, D (2008). *Hedonic price indexes for housing across regions and time: The problem of substitution bias*. [Online]. Available:

http://www-cgi.uni-regensburg.de/Fakultaeten/WiWi/sebastian/cms/upload/Hill.pdf [26 March 2008].

Kain, J. and Quigley, J.1970. Evaluating the Quality of Residential Environment. Environment and Planning.

Kathleen, P. Bell, Kevin, and J Boyle. 2006. *Economics of Rural Land Use Change*. England: Ashgate.

Keller, G and Warrack. B, 2003, *Statistics: for management and economics*, Curt

Henricks, Pacific Grove.

Keng, T (1997). *A hedonic model for house prices in Malaysia*. [Online]. Available: www.prres.net/.../tan_an_hedonic_model_for_house_prices_in_malaysia.pdf

Kumar, R. (2005), Research Methodology, A step-by-step guide for beginners, Second Edition. Sage Publications. pp. 130-131.

Leong, C (2002). *A critical review of literature on the hedonic price model and its application to the housing market in Penang*. [Online]. Available: www.kreaa.org/AsRES/doc/Chin%20Tung%20Leong(D3).doc

Li, M & Brown, J. 1980. Micro-neighbourhood externalities and hedonic housing prices. *Land economics*, 56(2):125-141. [Abstract from JStor database].

Mark, J and Goldberg, M. 1988. Multiple Regression Analysis and Mass Assessment: A Review of the Issues. *The Appraisal Journal*, 89-109.

Mueller, V and Sheriff, G. 2010. On Hedonic Valuation of urban Amenities Using Unbalance Data. *Land Economics,* 86(3):545-551. [Abstract from EBSCOhost database].

Meese, R and Wallace, N. 1997. The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches. *Journal of Real Estate Finances and Economics,* 14:51-73. [Abstract from EBSCOhost database].

Neuman, W (2003), Social Research Methods, Qualitative and Quantitative Approaches, Fifth Edition. Pearson Education. pp. 145,148.

Ortalo-Magne, A. 2002. *Bargaining over residential real estate: Evidence from England*.

[Online]. Available:
http://www.ifo.de/pls/guestci/download/CESifo%20Working%20Papers%202002/CESifo%20Working%20Papers%20September%202002/cesifo_wp778.pdf [26 March 2009].

Rappaport, J. 2007. A guide to aggregate house price measures. *Economic Review*, 2:41-71, April.

Wilkinson, R.K. 1971. The determinants of relative house prices. *CES Urban Economics Conference,* Keele, July, 1971.

Yin, R. (2003), Applications of Case Study Research, Second Edition. Sage Publication. pp.33.

Van Rensburg, P and Priilaid, D. (2004). An Econometric Model for Identifying Value in South African Red Wine. *International Journal of Wine Marketing*, 16(1):53-75. [Abstract from JStor database].

Wikipedia. 2009. *Regression analysis*. Available:
http://en.wikipedia.org/wiki/Regression_analysis [24 November 2009]

Data set obtained from Property24.com courtesy of Christo Wiid