A HEDONIC HOUSE PRICE INDEX FOR THE CITY OF JOHANNESBURG: AN ALTERNATIVE TO MEAN AND MEDIAN HOUSE PRICE INDICES

Lizelle Janse van Rensburg* and Philippe Burger

University of the Free State
* Corresponding author: jvrensep@ufs.ac.za

Abstract: Did Johannesburg suffer a house price bubble during the last decade? To explore this question the paper first estimates a hedonic model using data from real estate transactions in Johannesburg. A hedonic model may capture pure house price inflation more accurately compared to a median house price series (e.g. the ABSA series). The paper uses the hedonic pure price series to estimates a number of Vector Error-Correction Models (VECM). These models highlight the rather significant impact of credit extension on the growth of house prices. This impact is over and above the impact of disposable income and the mortgage rate, two fundamental factors that explain house prices. The results also highlight the need to let the mortgage rate play a bigger role in the allocation of mortgage credit.

JEL: D40; L85

Key words: Hedonic price model, house price inflation

1. Introduction

Like many other countries, South Africa experienced significant increases in house prices since the turn of the millennium. As in most of these countries, this increase came to a halt, and turned negative in real terms in 2007/8. South African house price indices are currently compiled in the private sector by three of the four major commercial banks, namely Absa Bank (ABSA), Standard Bank and First National Bank (FNB). Although the underlying methods of these three house price indices differ, they are all mean or median house price indices that are subject to a number of shortfalls. Not only do these indices treat houses as homogenous units, but by assuming neighbourhood homogeneity they also discount the impact that neighbourhoods have on house prices. The heterogeneous nature of housing and its geographical position precludes price comparisons over space and time simply by taking averages in different locations (Coulson, 2008: 3).

As a result of the heterogeneity of houses, mean house price indices are prone to yield biases. Mean house price indices are unable to distinguish between movements in prices and changes in the combination of dwellings sold from one period to the next. For instance, if house prices do not change, but the proportion of higher quality and hence, high-priced houses sold increase, a mean measure will *ceteris paribus* show an overall increase in house prices. Thus, the mean value can be skewed by the sale of very high- or very low-priced houses during a specific index period. As a result the price movements recorded in the mean index may not be representative of the price movements relevant to all houses in the economy.

To incorporate the heterogeneous nature of housing in house price indices and to overcome the limitations of simple means and medians, property analysts frequently apply constant-quality measures of house prices to standardise, and make comparable over time, the information available in the data. Three main methods are used for this: hedonic regressions; the mixed adjustment method and the repeat-sales method. The hedonic regression procedure quantifies the effect of various housing characteristics and neighbourhood characteristics on house prices. House price indices such as the Halifax and Nationwide indices in the UK and the HPX house price index published by Hypoport Group in Europe are indices compiled through the hedonic procedure (Thwaites and Wood, 2003:43; Hypoport, 2009).

Empirically, the hedonic technique uses regression analysis to explain the variation in market values using property characteristics (Goodman and Thibodeau, 1995:25). The key advantage of the general hedonic formulation is that it controls for the heterogeneous nature of houses, something not done by the mean house price series (Hansen, 2006:6). An added benefit of the hedonic methods is that it also can yield estimates for neighbourhood effects on property prices. Thus, this article estimates a hedonic model for the City of Johannesburg to distil a pure house price inflation series for the period 1996Q1 to 2010Q2. The paper also presents quantile regression models to establish whether or not the pure house inflation series differs across the distribution of housing prices.

In making a micro-macro link the performance of the pure house price series is then compared in a VECM context to that of the ABSA house price series for Greater Johannesburg. This is done to consider the role of disposable income, the mortgage rate and credit in the determination of house prices. The ABSA house price series for Greater Johannesburg is closest in coverage to the sample used for the hedonic model for the City of Johannesburg. In addition, the ABSA mean house price series is currently the most comprehensive and most used house price series in South Africa.

2. The hedonic model

To incorporate the heterogeneous nature of housing in house price indices, the hedonic procedure is generally used to quantify the effect of various housing and neighbourhood characteristics on house prices. Hedonic price models focus on the utility derived from individual characteristics of a house (*cf.* Rosen 1974). Empirically, the technique uses regression analysis to explain the variation in market values using property characteristics (Goodman and Thibodeau, 1995:25). The key advantage of the general hedonic formulation, and the reason why it is used in this article, is that it provides direct estimates of pure price changes (house price inflation) and can, in principle, control for housing characteristics, changes in the composition of the mean index and quality of houses sold (Hansen, 2006:6). Thus, the hedonic specification includes trend dummies that capture the effect of pure price changes over time after controlling for the heterogeneity of houses. This article uses the parameters of these trend dummies to construct a pure price time series that is then substituted for the ABSA mean house price series in the VECM estimations. The general specification for a hedonic house price equation is:

$$P_t = \beta_{0t} + \beta_{1t}S + \beta_{2t}L + \beta_{3t}N + \beta_{4t}T + \varepsilon_t \tag{1}$$

where P is the natural log of the observed sale price of the property; S denotes a class of variables describing structural characteristics (number of various rooms and bathrooms, number of storeys, number of garages and carports, the presence of a pool or not, etc.); L is a dummy variable that is 1 if the property is classified as a free-hold unit (full-title unit) and 0 if it is a sectional-title or cluster unit. N denotes a class of neighbourhood dummies; and T represents a quarterly time trend in house prices. It should be noted that this is a pooled rather than panel data regression, as the number of observations and the specific housing units sold vary with each time period t.

To obtain an estimated hedonic price function given the available South African data, the estimates in this article uses the Sold Property Guide database of Property 24 for the residential properties located inside the Johannesburg municipal boundaries based on the demarcation done by the Municipal Demarcation Board in 2000. Property 24 is supported by the Property Association, the MLS multi-listing organisation and over 12 000 real estate agents (Property 24, 2011). The data contain unit record information entered into the database by estate agents of sold housing units that are categorised as free-hold units or sectional title units. In addition, the data also include information on house sale price, location (street address, suburb) and structural attributes (number of rooms, bathrooms, recreational areas, garages, carports etc). With regard to the attributes the record shows either the number of rooms, bathrooms, recreational areas, garages, carports, and domestic rooms or whether or not an attribute is presents (i.e. a storey, pool, flatlet and second house). With respect to the latter, according to Property24, it is a widespread convention for estate agents to complete entries to the database by checking and completing only the affirmative categories. Thus, 'no' and 'unreported' entries were all coded as zero. The transaction-based data in the dataset required a moderate amount of "cleaning" that includes the removal of double entries and incomplete entries (entries with no price or number of bedrooms). The final Property24 dataset covers 39 289 house sales over a fourteen and half year period from 1 January 1996 to 30 June 2010. The sample period is selected on the basis of the availability of data.

The hedonic model, Model A, includes neighbourhood dummies that capture the fixed-effects of predefined geographical submarkets and is estimated over four different time periods. The rationale for modelling different time periods in the dataset is to establish the consistency (stability) of the parameters. Thus, Model A(0) includes trend dummies that extends over the full time period from 1 January 1996 to 30 June 2010. To test for robustness the article also presents three models estimated for subsample periods (Model A(1), Model A(2) and Model A(3)). Specifically Model A(1) is estimated for the period (and thus include quarterly trend dummies) 1 January 1996 to 31 December 2001, Model A(2) is estimated for the period 1 October 2001 to 31 December 2006 and Model A(3) for the period 1 October 2006 to 30 June 2010. Note that the first and last quarters included in Model A(1) and Model A(2) overlap and again with Model A(2) and Model A(3). The overlapping trend dummies included in the Models enables the analysis to draw a pure house price inflation index over the full period by linking the time dummies of the three models.

Table 1 present the estimates for Model A using the Property24 dataset. The house characteristics grouped in the first neighbourhood dummy are located in Northcliff (Johannesburg), and forms the base for the neighbourhood-dummy variables.[1] To obtain a robust variance estimate that adjusts for within-cluster correlation, the estimation procedure for each model controls for clustering.

Table 1: Hedonic house price function with neighbourhood fixed effects as submarket variables

	MODEL A(0)		*MODEL A(1)*		*MODEL A(2)*		*MODEL A(3)*	
LN Price	*Coef.*	*P>t* [1]	*Coef.*	*P>t* [1]	*Coef.*	*P>t* [1]	*Coef.*	*P>t* [1]
Constant	11.402	0.000	11.367	0.000	11.460	0.000	12.305	0.000
Story [2]	0.038	0.000	0.092	0.000	0.058	0.000	-0.014	0.236
Type [3]	0.127	0.000	0.170	0.000	0.219	0.000	0.053	0.000
Bedrooms	0.194	0.000	0.254	0.000	0.209	0.000	0.197	0.000
Bathrooms	0.164	0.000	0.123	0.002	0.150	0.000	0.167	0.000
Recreational areas	0.088	0.000	0.059	0.000	0.078	0.000	0.108	0.000
Studies	0.074	0.000	0.061	0.000	0.065	0.000	0.087	0.000
Garages	0.118	0.000	0.101	0.000	0.114	0.000	0.139	0.000
Carports	0.021	0.000	0.018	0.000	0.027	0.000	0.014	0.003
Domestic rooms	0.052	0.000	0.031	0.000	0.050	0.000	0.043	0.001
Second house on stand [2]	0.094	0.000	0.009	0.773	0.081	0.003	0.186	0.000
Flatlet [2]	0.117	0.000	0.088	0.000	0.121	0.000	0.113	0.000
Pool [2]	0.039	0.000	0.050	0.000	0.040	0.000	0.029	0.001
Bedrooms-sq	-0.011	0.001	-0.027	0.000	-0.016	0.000	-0.009	0.000
Bathroom-sq	-0.011	0.000	0.002	0.813	-0.008	0.000	-0.011	0.000
Recreational area-sq	-0.004	0.000	-0.002	0.000	-0.004	0.000	-0.004	0.000
Garages-sq	-0.007	0.001	-0.004	0.000	-0.006	0.047	-0.019	0.000
Adj R-sq	0.874		0.796		0.882		0.786	
Nr of obs	39289		6342		20804		12143	
Nr of clusters	456		205		361		405	
RESET test	88.910	0.000	75.910	0.000	119.860		2.090	0.100

[1] p-values are obtained from t-statistics calculated from heteroskedasticity robust standard errors clustered by neighbourhoods.
[2] Yes = 1, No = 0 (for instance if a house has a 'story', it takes a value of 1, and zero otherwise)
[3] Type: Full title = 1, Sectional title = 0

[1] The results of the neighbourhood variables are available on request.

Empirical applications of the hedonic price model rarely involve rigorous specification testing. According to Kuminoff *et al* (2008) this is largely due to an influential simulation study by Cropper *et al* (1988) who found, among other, that simpler linear specifications outperformed more flexible functional forms even in the face of omitted variables. The findings by Cropper *et al* (1988) was recently confirmed by Kuminoff *et al* (2008) who obtained results from Monte Carlo simulations that suggest that the addition of spatial fixed effects to linear models such as semi-log and double-log functional forms improve their performance in cases where all attributes are observed as well as cases where some attributes are unobserved. Moreover, their analysis showed that by adding spatial fixed effects to more flexible specifications such as quadratic and quadratic Box-Cox specifications do not improve model performance. Therefore, the analysis in this article relies on the findings of Cropper *et al* (1988) and Kuminoff *et al* (2008) and consider linear models, specifically with a semi-log functional form with additional spatial fixed effects.

According to the diagnostic tests (Table 1) the variables of all the Models account for roughly 80%-90% of the variation in the sold price of the residential units as presented by the adjusted R-square.[2] Using a Wald Test the analysis also compares the results of the three subsamples (Models A(1)-A(3)). Table 2 shows that there are statistically significant differences between the parameters estimated for the various subsamples, indicating that linking the time trends of the three subsample models might present a more robustly

[2] The RESET test of Model A(3) indicates that the analysis fails to reject the null hypothesis of no misspecification. Although the Ramsey RESET indicates that functional form of Model A(0), Model A(1) and Model A(2) is incorrectly specified, the literature suggests (Micheal *et al*, 2000: 287; Malpezzi, 2003:20; Triplett, 2004: 232; Fletcher *et al*, 2004: 195; Selim, 2008: 65-76) that the semi-log model is to be preferred to the simple linear model (which was also estimated, but yielded significantly poorer results - results are not reported for brevity, but are available from the authors on request). (Refer to Malpezzi (2003:20) for a discussion of the reasons for estimating a semi-log hedonic function as opposed to other functional forms.) In addition, because of the nature of the data double-log models cannot be applied. This leaves the semi-log model as the only functional form available to estimate the hedonic model. In doing so the article follows general practice in hedonic model estimation, as can be seen in Els and Fon Fintel (2010: 418-436) and Zietz and Newson (2001: 247) who also estimated semi-log models despite weak RESET test results. Thus, the analysis proceeds cautiously on the assumption that the functional forms of all the above models are correct.

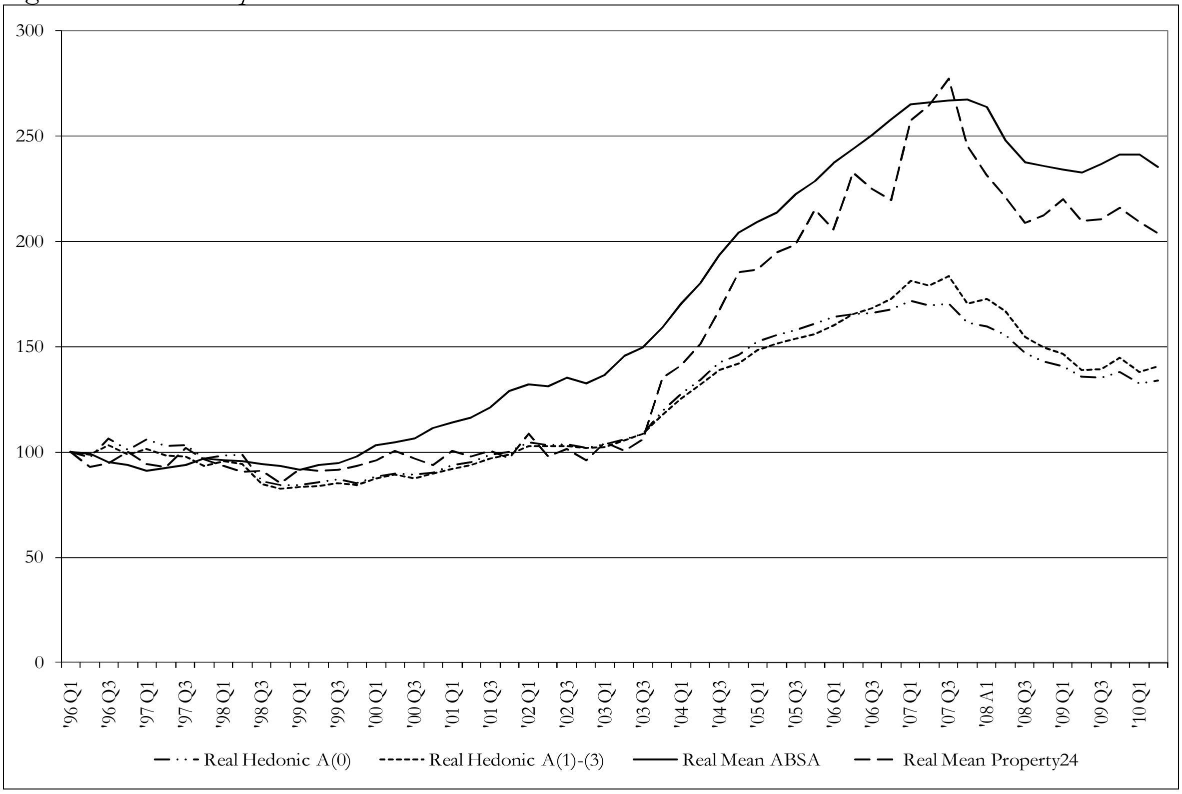
estimated pure house price index than the model estimated for the full sample period (however, estimates using both are presented in the macro section of the paper).

Table 2 – Wald Tests to test equality of parameters of housing characteristics in Model A(1)-(3)

	Prob > F		
	1 vs 2	*1 vs 3*	*2 vs 3*
Storey	*0.059**	*0.000****	*0.000****
Type	*0.378*	*0.000****	*0.000****
Bedrooms	*0.238*	*0.059**	*0.218*
Bathrooms	*0.345*	*0.164*	*0.197*
Recreational areas	*0.008****	*0.000****	*0.000****
Studies	*0.164*	*0.001****	*0.025***
Garages	*0.075**	*0.000****	*0.049***
Carports	*0.000****	*0.356*	*0.000****
Domestic rooms	*0.000****	*0.073**	*0.042***
Second house	*0.003****	*0.000****	*0.002****
Flatlet	*0.040***	*0.223*	*0.131*
Pool	*0.119*	*0.007****	*0.242*

In all the models trend coefficients capture pure house price inflation in nominal terms while the model controls for the characteristics of each housing unit. To graphically compare the consistency of the price trends the house price inflation index distilled from the three shorter-period models are presented in conjunction with the house price inflation distilled from Model A(0), estimated for the full time period (Figure 1) (both indices start at a 100). The indices are presented in real terms (deflated by the Consumer Price Index).

Figure 1 – Real house price indices



In addition to the two indices distilled from the hedonic formula, Figure 1 also presents in real terms the quarterly mean values of the Property24 dataset used in the analysis (expressed as an index starting at 100) and compares it to the real ABSA house price series for Greater Johannesburg (also expressed as an index starting at 100). From Figure 1 it can be seen that all series display the same pattern. The ABSA series peaks at 267.47 in the fourth quarter of 2007, while the mean of the Property24 dataset peaks at 277.66 in the third quarter of 2007. The two indices distilled from the hedonic formula peak at 183.45 in the third quarter of 2007 (Models A(1)-(3)) and 171.66 in the first quarter of 2007 (Model A(0)).

Note that the large increase in the mean Property24 index (which is similar to the rather large increase in the ABSA series) contrasts with the more dampened increase in the two hedonic house price series distilled from Model A. This finding is similar to that of Els and Von Fintel (2010:418-436), who also suggests that the house price inflation distilled from the hedonic procedure, is considerably smaller than the house price increases reported by property analysts using mean house price indices in both nominal and real terms.

The difference between the unconditional house price inflation as measured by the Property24 mean index and the conditional house price inflation distilled from the hedonic models that are estimated with the Property24 data, point to either changes in the number of characteristics (e.g. the number of bedrooms) or changes in the combination of dwellings sold from one period to the next. Table 3 shows the change in the mean of the characteristics entering the hedonic model for the first and second half of the sample period.

Table 3 – Change in the mean of the characteristics

	Storey	*Type*	*Bedrooms*	*Bathrooms*	*Recreational areas*	*Studies*
1st half	*0.03*	*0.76*	*2.94*	*1.69*	*2.29*	*0.26*
2nd half	*0.07*	*0.58*	*2.78*	*1.71*	*2.04*	*0.20*
	Garages	*Carports*	*Domestic rooms*	*Second house*	*Flatlet*	*Pool*
1st half	*1.19*	*0.76*	*0.56*	*0.01*	*0.09*	*0.33*
2nd half	*1.07*	*0.72*	*0.32*	*0.01*	*0.09*	*0.38*

With the exception of pools, storeys and bathrooms, the average of all characteristics decreased from the first to the second half of the sample period (Table 3). Thus, the difference between the unconditional house price inflation as measured by the Property24 mean index and the conditional house price inflation distilled from the hedonic models that are estimated with the Property24 data cannot be ascribed to an increase in the number of characteristics. This point to a possible change in the combination of dwellings sold. In other words the unconditional house price inflation as measured by the Property24 mean index may be skewed by the sale of very high or very low-priced houses during a specific index period when compared to another period. To determine whether or not the proportion of high-priced dwellings increased over time, the analysis ranked the neighbourhood coefficients of Model A(0) to establish which neighbourhoods sort in the upper 20% according to its coefficients and hence price range. Next the analysis determines the frequency of observations for the upper 20% of neighbourhoods over the three subsample periods covered in Model A(1), Model A(2) and Model A(3). The frequency of observations in the upper 20% of neighbourhoods as a percentage of all the observations (full sample) for the first period was 0.88%, 3.99% for the second and 3.64% for the third period. Thus, the proportion of houses in higher priced neighbourhoods increased with almost 3% from the first to the second period where after the proportion of higher priced houses slightly decreased again with around 0.3%. These changes in the proportion of higher priced houses sold causes unconditional mean house price indices to be skewed during a specific index period and explain why the unconditional house price series displays a much steeper increase in the early to mid 2000s.

3. The quantile regressions of the hedonic model

As mentioned in the introduction the paper also presents quantile regression models to establish whether the pure house inflation series differs across the distribution of housing prices. Recall that the pure house price series is distilled once the regression controls for housing characteristics. However, different consumers may value housing characteristics differently (Zietz *et al*, 2007:2). Zietz *et al* (2007) examined forty studies that applied the hedonic formula with submarket variables and found that almost half of the studies show a negative or non-significant result for the number-of-bedrooms variable. This contradicts *a priori* expectations that the number of bedrooms would have a positive effect on the house price (Zietz *et al*, 2007: 2). Zietz *et al* (2007) suggest that there may be marked differences in the elasticity of house prices with respect to housing characteristics across the distribution of housing prices. For example, if a particular housing characteristic is priced differently for houses in the upper-price range as compared to houses in the lower-price range, the typical OLS regression may not provide useful information for either price range since it is based on the mean of the entire price distribution. The use of pooled quantile regression is offered as solution. Where OLS is constrained to explaining characteristics at the mean of the dependent variable, quantile regressions aim to explain the dependent variable at any point in the dependent variable's distribution, and not just at the mean. Thus, quantile regressions allow one to statistically examine the extent to which housing characteristics are valued differently across the distribution of housing prices.

Although this method seem similar to segmenting the data into groups based on property prices, it avoids the truncation bias of running separate regressions for different house price brackets by utilising the entire set of data (Els and Von Fintel 2010:426; Zietz *et al* 2007:5; Heckman 1979). Els and Von Fintel (2010:426) suggest that a quantile regression would control for some neighbourhood effects as houses in similar price categories are found in similar geographical areas, and share other characteristics. However, the estimation does not include neighbourhood dummy variables due to the strong correlation between the house-price quantiles and their geographic areas. As opposed to an OLS regression that minimises the sum of squared residuals, quintile regression minimizes a weighted sum of the absolute deviations. The standard errors of the coefficient estimates are estimated using bootstrapped standard errors (Gould 1992; 1997) and are statistically more robust to heteroskedasticity.

(For more detail on the quantile regression technique, see Koenker and Hallock, 2000; Hao and Naiman, 2007 and Ziets *et al*, 2007).

Table 4 presents the results for the 0.25, 0.5 and 0.75 quantiles, while Table 5 presents the results for the Wald Tests to test whether or not the parameters for each variable estimated for the different quantiles are statistically significantly different from each other. The Wald test shows that there are differences for several of the parameters of the characteristics, indicating that various characteristics might be valued differently across the distribution of housing prices.

Table 4 – Quantile regressions results

	Pseudo (.25)	P>t[1]	Pseudo (.5)	P>t[1]	Pseudo (.75)	P>t[1]
Constant	10.369	0.000	10.741	0.000	11.031	0.000
Story[2]	0.048	0.000	0.032	0.000	0.024	0.010
Type[3]	-0.074	0.000	-0.030	0.000	0.014	0.011
Bedrooms	0.235	0.000	0.161	0.000	0.084	0.000
Bathrooms	0.420	0.000	0.370	0.000	0.373	0.000
Recreational areas	0.114	0.000	0.119	0.000	0.115	0.000
Studies	0.113	0.000	0.117	0.000	0.146	0.000
Garages	0.222	0.000	0.186	0.000	0.137	0.000
Carports	0.041	0.000	0.036	0.000	0.026	0.000
Domestic rooms	0.012	0.006	0.000	0.990	-0.005	0.314
Second house on stand[2]	0.076	0.048	0.077	0.003	0.114	0.000
Flatlet[2]	0.066	0.000	0.068	0.000	0.074	0.000
Pool[2]	0.180	0.000	0.148	0.000	0.119	0.000
Bedrooms-sq	-0.027	0.000	-0.018	0.000	-0.005	0.139
Bathroom-sq	-0.042	0.000	-0.024	0.000	-0.021	0.000
Recreational area-sq	-0.006	0.000	-0.006	0.000	-0.005	0.006
Garages-sq	-0.020	0.000	-0.010	0.005	0.005	0.203
Pseudo R-sq	0.512		0.545		0.551	

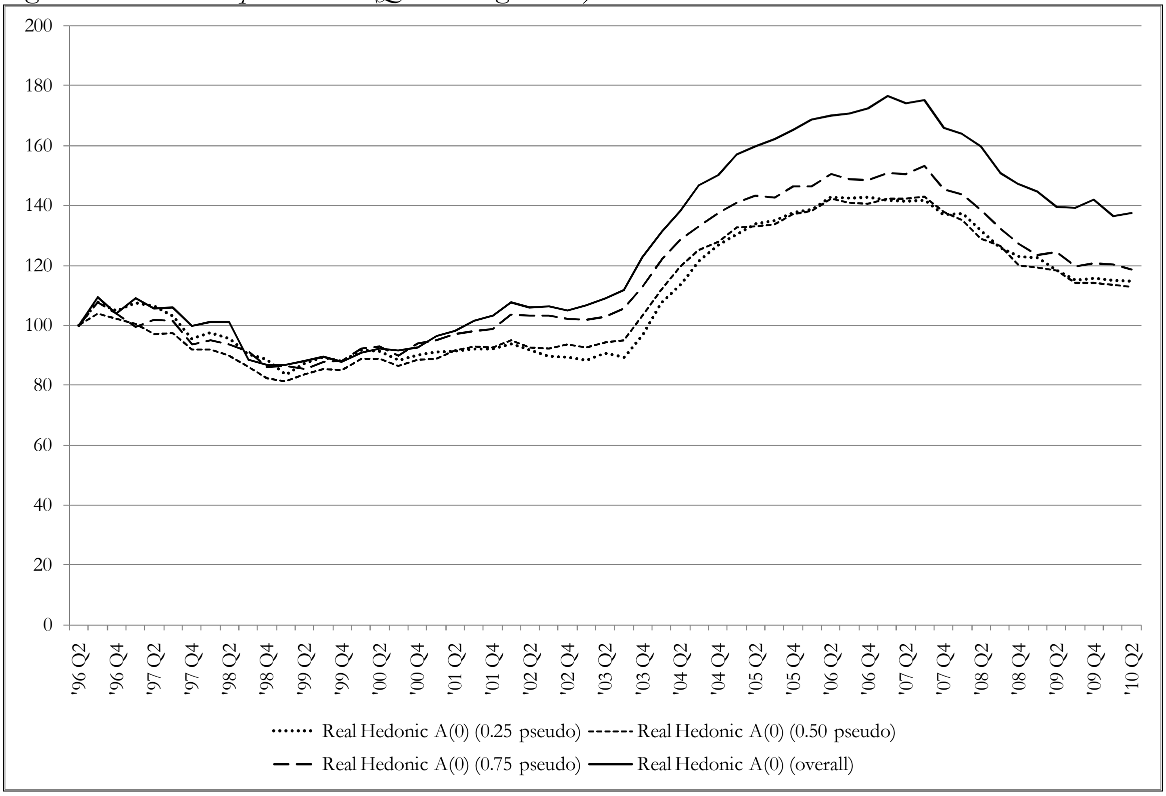
[1] p-values are obtained from t-statistics calculated from Bootstrap standard errors. The Bootstrap is based on 20 replications.

Table 5 – Wald Tests to test equality of parameters of the quantiles

	Prob>F		
	(.25) vs (.5)	(.25) vs (.75)	(.5) vs (.75)
Story[2]	0.040	0.024	0.351
Type [3]	0.000	0.000	0.000
Bedrooms	0.003	0.000	0.000
Bathrooms	0.033	0.232	0.903
Recreational areas	0.367	0.920	0.643
Studies	0.428	0.000	0.000
Garages	0.000	0.000	0.001
Carports	0.002	0.000	0.000
Domestic rooms	0.002	0.002	0.174
Second house on stand [2]	0.987	0.468	0.318
Flatlet [2]	0.764	0.364	0.330
Pool [2]	0.000	0.000	0.000
Bedrooms-sq	0.046	0.000	0.000
Bathroom-sq	0.007	0.063	0.632
Recreational area-sq	0.976	0.761	0.697
Garages-sq	0.000	0.000	0.019

Figure 2 shows the pure house price inflation distilled from the three quantiles together with the pure price series distilled from the model estimated for the full sample period (Model A(0) above). It clearly shows that when higher priced houses are included, the pure house price index displays a higher level of inflation during the early to mid 2000s. By sample end (2010Q2) the price index estimated with the full sample period (Model A(0) above) is in real terms 18.22% higher than the price index estimated for the 0.25 quantile and 14.75% higher than the price index estimated for the 0.75 quantile. This is an indication of an increasing wealth gap between higher and lower priced houses.

Figure 2 – Real house price indices (Quantile regressions)



4. The macro model

To assess the usefulness of the hedonic price series estimated above, this article uses the hedonic price series, distilled from the two micro models, in the estimation of macro-level models. More specifically, the model seeks to explain house prices on a macroeconomic level. This is done by estimating a cointegrated relationship between the hedonic house price series and possible determinants of house prices. In addition, the analysis also presents estimated models using the available and standard ABSA series of the Greater Johannesburg. According to the ABSA residential property market database the South African residential property market is categorised into three major segments: luxury houses, middle-segment houses (subdivided into large -, medium - and small houses) and affordable houses.[3] To model residential property prices in South Africa the article employs house prices that are categorised in the middle segment of the Greater Johannesburg residential property market.

According to literature disposable income is the key variable explaining house price movements, with the mortgage rate also playing an important role (McQuinn O'Reilly 2006:4-5; Leishman 2003:137; ECB 2003:22; Case, *et al* 2000:142; Muellbauer and Murphy 1997:1705). In addition, mortgage credit can also be added as a possible determinant of house prices. The inclusion of mortgage credit is predicated on the notion that the interest rate does not necessarily equilibrate the supply and demand for mortgages since the allocation of credit is done by administrative means on the part of the mortgage providers. Thus, the basic model is:

$$P_t = \beta_{0t} + \beta_{1t} RDY_t + \beta_{2t} RMOR_t + \beta_{3t} RCREDIT_t + \varepsilon_t \quad (2)$$

where *P* denotes the log of real house prices, *RDY* is the log of real disposable income, *RMOR* denoting the real mortgage rate and *RCREDIT* is real mortgage credit. Based on the literature reviewed above and the data available this article uses the predominant rate of new mortgage loans extended by the banking sector for dwellings and disposable income of households to explain long-run house prices. For credit the article uses the mortgage credit extended by financial institutions.

---

[3] The luxury segment refer to houses valued between R2.6 million and R9.5 million; Middle segment housing refer to houses that are valued up to R2.6 million and are further divided into sub segments according to the size of the property. Small houses extend over an area of $80m^2$ to $140m^2$, medium houses $141m^2$ to $220m^2$ and large houses $221m^2$ to $400m^2$. Affordable housing refers to houses of $40m^2$ to $79m^2$ and priced at R193 000 or less. Due to the increase in house prices, the cut-off prices for the classification of houses were adjusted in 2006. Prior 2006 the price bracket for luxury housing was between R2.2 million and R8.2 million. Middle segment housing were priced up to R2.2 million and affordable housing were priced up to R193 000.

To estimate the long-run relationship as captured by Equation (2), or versions thereof, this section uses throughout a Johansen vector error-correction framework (Johansen, 1991, 1992), specified as:

$$\Delta X_t = \Pi X_{t-1} + \sum_{t=i}^{k} \Gamma_i \Delta X_{t-i} + \varepsilon_{kt} \tag{3}$$

where X_t = (containing the variables specified in equations (2)). These are usually the I(1) variables. Γ_i are nxn short-run coefficient matrices (where n is the number of lags identified by the information criteria); ε_{kt} are normally and independently distributed error terms. The Trace test is used throughout to determine the number of cointegrated vectors, while information criteria are used to establish the number of lags to include in the analysis.

To test for the presence of unit roots the analysis employs the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) (Kwiatkowski *et al* (1992). The KPSS test is selected because standard tests such as the ADF and PP tests can suffer from lack of power as they tend to accept the null of a unit root too frequently against a stationary alternative (Masih and Masih, 1996: 321). Note that the null of the KPSS is that the series is stationary. Table 6 shows that according to the KPSS test all the variables are I(1).

Table 6 – KPSS tests for mean stationarity.

	I(0)	*I(1)*
Hedonic A(0)	*0.698*	*0.169*
Hedonic A(1)-(3)	*0.734*	*0.177*
ABSA	*0.870*	*0.217*
YD	*0.906*	*0.145*
Credit	*0.863*	*0.240*
Mortgage rate	*0.660*	*0.063*

Critical values: 1%, 5% and 10% levels respectively 0.739, 0.463 and 0.347

Following the stationarity tests, the article presents estimates of Equation (2). Note that the estimation experiments with various combinations of the variables entering Equation (2). Thus, it first presents estimates of Equation (2) containing all three variables, before dropping either mortgage credit or the mortgage rate, or both, but retaining disposable income in all.

None of these models containing disposable income yielded satisfactory results, whether estimated with the ABSA series, or the hedonic price series estimated above (the results are reported in Appendix 1). Either there was no cointegration, or if cointegration was present according to the Trace test, the error-correction term (i.e. the α-parameter) that measures the reaction of house prices to shocks in the long-run relationships is statistically insignificant or has the wrong sign. Normalising on the other variables does not yield better results.

Subsequently the VECM model is re-estimated without disposable income and includes the real house price series, credit and the real mortgage rate. Again the model alternately uses the ABSA series and the two pure house price series distilled from the hedonic models.

Table 7 – Cointegration tests

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	*Eigenvalue*	*Trace Statistic*	*0.05 Critical Value*	*Prob.***
Hedonic A(0)				
*None **	*0.317*	*37.296*	*35.193*	*0.029*
At most 1	*0.203*	*15.964*	*20.262*	*0.176*
At most 2	*0.056*	*3.254*	*9.165*	*0.534*
Hedonic A(1)-(3)				
*None **	*0.337*	*37.495*	*35.193*	*0.028*
At most 1	*0.194*	*14.456*	*20.262*	*0.259*
At most 2	*0.041*	*2.367*	*9.165*	*0.704*
ABSA				
*None **	*0.308*	*36.028*	*35.193*	*0.041*
At most 1	*0.2*	*15.397*	*20.262*	*0.205*
At most 2	*0.051*	*2.921*	*9.165*	*0.596*
DY				
*None **	*0.306*	*28.374*	*25.872*	*0.024*
At most 1	*0.132*	*7.9*	*12.518*	*0.26*

Trace tests indicate 1 cointegrating equation at the 0.05 level for all three models

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

By using both the ABSA and the hedonic pure house price series in separate VECMs that include real house prices, real mortgage credit and the real mortgage rate, the analysis first found cointegration in all three cases (see Table 7, top three panels). Subsequently the analysis estimated the cointegrating relationship for all three house price series (see Table 8). Table 8 indicates that both real mortgage credit and the real mortgage rate explain house price movements in all three models (i.e. the Hedonic A(0), Hedonic A(1)-(3) and ABSA models). However, note that the model estimated with the ABSA mean house price series does not performs satisfactory, given that its error-correction term (i.e. the α-parameter) is statistically insignificant. This leaves the models with the hedonic price series as the only satisfactory models.

Table 8 – Cointegration – House prices, credit and the mortgage rate

		HedonicA(0)			*HedonicA(1)-(3)*			*ABSA*	
Houseprice		*1*			*1*			*1*	
Credit		*-0.161*			*-0.270*			*-0.734*	
		(-2.063)			*(-3.985)*			*(-11.634)*	
Mortgage rate		*3.503*			*3.049*			*3.850*	
		(3.077)			*(3.050)*			*(4.110)*	
C		*-2.180*			*-0.533*			*-3.137*	
		(-1.876)			*(-0.526)*			*(-3.323)*	
Error Correction:	*D(House price)*	*D(Credit)*	*D(Mort rate)*	*D(House price)*	*D(Credit)*	*D(Mort rate)*	*D(House price)*	*D(Credit)*	*D(Mort rate)*
α	*-0.102*	*0.008*	*-0.032*	*-0.114*	*0.012*	*-0.036*	*0.013*	*0.036*	*-0.044*
	(-2.385)	*(0.546)*	*(-2.627)*	*(-2.634)*	*(0.779)*	*-2.798)*	*(0.375)*	*(2.066)*	*(-3.057)*
Weak exogeneity test (prob)	*0.040*	*0.647*	*0.079*	*0.019*	*0.493*	*0.045*	*0.752*	*0.166*	*0.047*
Adj. R-squared	*0.16*	*0.72*	*0.25*	*0.21*	*0.72*	*0.27*	*0.46*	*0.75*	*0.29*
Autocorrelation	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
LM (prob) at	*0.08*	*0.39*	*0.03*	*0.16*	*0.61*	*0.07*	*0.11*	*0.31*	*0.10*
lags 1 to 6	*4*	*5*	*6*	*4*	*5*	*6*	*4*	*5*	*6*
	0.14	*0.14*	*0.86*	*0.06*	*0.94*	*0.33*	*0.13*	*0.67*	*0.54*

According to the Hedonic A(0) model (see Table 8) an increase of one percentage point in the real mortgage credit extended, leads to an increase in the real house price series of approximately 16.1%, while in the case of the Hedonic A(1)-(3) model it is 27%. (Note that the long-run component of Table 8 is stated in vector format, meaning that a minus (such as the minus in front of credit) should be interpreted as a plus, while a plus should be interpreted as a minus). Furthermore, an increase of one percentage point in the real mortgage rate leads to a decrease of approximately 3.5% and 3.1% in real house price series, respectively for the Hedonic A(0) and Hedonic A(1)-(3) models. Therefore, the signs of the credit and mortgage rate variables accord with *a priori* expectations. The error-correction terms (i.e. the α-parameter) of house prices in the Hedonic A(0) and Hedonic A(1)-(3) models are -10.2 and -11.4 and both are statistically significant (see Table 8). The LM test for autocorrelation indicates a general absence of autocorrelation.

Table 9 – Variance decomposition test – House prices, credit and the mortgage rate

	Variance explained by		
Variance explained	*HedonicA(0)*	*Credit*	*Mortgage rate*
HedonicA(0)	*19.83*	*79.57*	*0.60*
Credit	*0.16*	*88.06*	*11.79*
Mortgage rate	*12.92*	*4.21*	*82.87*
Cholesky ordering: Mortgage rate, credit and HedonicA(0)			
	Variance explained by		
Variance explained	*HedonicA(1)-(3)*	*Credit*	*Mortgage rate*
HedonicA(1)-(3)	*22.13*	*77.82*	*0.05*
Credit	*0.98*	*85.84*	*13.18*
Mortgage rate	*15.04*	*4.02*	*80.94*
Cholesky ordering: Mortgage rate, credit and HedonicA(1)-(3)			

The weak exogeneity test confirms the endogeneity of the hedonic house price series, which indicates the legitimacy of normalising the regression on the house price series. This is further supported by the variance decomposition analysis contained in Table 9. Table 9 indicates that house prices explain only 19.83% and 22.13% of their own variance in the Hedonic A(0) and Hedonic A(1)-(3) models respectively, while credit explains 79.57% and 77.82% of the variance of the house price series in the Hedonic A(0) and Hedonic A(1)-(3) models respectively. (Both credit and the mortgage rate explains the majority of their own variances, which is an indication of their exogeneity in this model). Also note that in the Hedonic A(0) and Hedonic A(1)-(3) models the real mortgage rate explains less than 1% of the variance of the house price series. It also explains only 11.79% and 13.18% of mortgage credit in the Hedonic A(0) and Hedonic A(1)-(3) models respectively. Lastly, the VECM Granger causality test indicates that in both the Hedonic A(0) and Hedonic A(1)-(3) models mortgage credit Granger causes house prices and the mortgage rate (but not *vice versa*), which is an indication that the interest rate might not play a large role in equilibrating the credit market (Table 10).

Table 10– VEC Granger Causality/Block Exogeneity Wald Tests

	HedonicA(0) (prob)	*HedonicA(1)-(3) (prob)*
Excluded		
Dependent variable: D(House price)		
D(Credit)	*0.002****	*0.001****
D(Mortgage rate)	*0.366*	*0.458*
All	*0.003****	*0.001****
Dependent variable: D(Credit)		
D(House price)	*0.937*	*0.661*
D(Mortgage rate)	*0.869*	*0.883*
All	*0.984*	*0.900*
Dependent variable: D(Mortgage rate)		
D(House price)	*0.272*	*0.164*
D(Credit)	*0.030***	*0.017***
All	*0.092**	*0.050***

****, ** and * level of significance at 1%, 5% and 10%*

The finding that credit explains house prices, while disposable income seems unable to do so, might be an indication that the changes in house prices observed were the result of credit extension policies of financial institutions that did not take sufficient cognisance of the income of their clients. In short, the significant increases in house prices during the sample period might therefore be ascribed to liberal credit extension policies pursued by financial institutions.

However, it is often considered common knowledge that banks use disposable household income as the main determinant in deciding the amount of credit to extend to households. For instance, the mortgage premium that a household has to repay, should, according to this common knowledge, not exceed one third of that household's disposable income. However, it has also been common knowledge that at times some households were able to extend the loan maturity (say from 25 to 30 years) so as to reduce their monthly premium, or to increase the amount that they borrowed by simply registering a second mortgage. This would relax the constraint that disposable income places on the amount of mortgage credit extended. Furthermore, as house prices increased and the perceived collateral of their owners increased, some banks might have been willing to focus more on the perceived safety of the collateral (even though the collateral value might have been inflated by the very loans that the banks extend), than the actual ability of households to service their loans from their disposable income. Furthermore, as households often use mortgage credit to finance other household expenditure, mortgage credit might become a determinant of consumption and, through consumption, a determinant of (disposable) income and GDP. All of the above might serve to eliminate or reverse the direction of causality between disposable income and mortgage credit extended.

Table 11 – Cointegration – Disposable income and credit

YD		*1*	
Credit		*-0.174*	
		(-5.048)	
Trend		*-0.008*	
		(-8.862)	
C		*-9.758*	
Error Correction:		*D(YD)*	*D(Credit)*
a		*-0.311*	*0.014*
		(-4.313)	*(0.194)*
Weak exogeneity test (prob)		*0.001*	*0.870*
Adj. R-squared		*0.373*	*0.726*
Autocorrelation	*1*	*2*	*3*
LM (prob) at	*0.05*	*0.20*	*0.61*
lags 1 to 6	*4*	*5*	*6*
	0.02	*0.90*	*0.11*

To consider this link between real mortgage credit extended and real disposable income, the analysis also shows, again using a VECM, that there is a long-run relationship between real disposable income and real mortgage credit (see Table 7, bottom panel for the Trace test results and Table 11 for the estimated model). However, the weak exogeneity test indicates that causality in this relationship runs from credit to income and not *vice versa* (again see Table 11). This is further supported by the VECM Granger causality test (see Table 12) and the variance decomposition analysis (see Table 13). The VECM Granger causality test indicates that in the short run mortgage credit Granger causing disposable income, but not *vice versa*. The variance decomposition analysis indicates that real mortgage credit extended explains 92.18% of the variance of real disposable income and 99.93% of its own variance.

Thus, whereas one would expect that income is an explanatory variable for mortgage credit extended, the reverse seems to be true. That is in line with findings that durable consumption expenditure by household, usually itself driven by household credit, is a leading indicator and possibly a driver of GDP (and hence disposable income) (cf. Burger 2008).

Table 12 – VEC Granger Causality/Block Exogeneity Wald Tests

	YD
Excluded	
Dependent variable: D(YD)	
D(Credit)	*0.000****
Dependent variable: D(Credit)	
D(YD)	*0.489*

Table 13 – Variance decomposition test – YD and Credit

	Variance explained by	
Variance explained	*YD*	*Credit*
YD	*7.82*	*92.18*
Credit	*0.07*	*99.93*
Cholesky ordering: Credit and YD		

5. Conclusion

Given that this article compares models estimated with mean house prices and hedonic house prices, and given that the latter performed better, demonstrates the value of using hedonic house prices. As such, the article recommends that serious consideration should be given to estimate and use hedonic house price series on a continuous basis.

The hedonic model also indicates that pure house price inflation in Johannesburg has been much more dampened compared to the rate calculated with the ABSA mean series or the unconditional series calculated using the mean values of the Property24 dataset. More specifically, the hedonic models indicate that the sharper increase of the unconditional price series generated with the mean values of the Property24 dataset can be ascribed to a shift in the weight of houses in higher priced neighbourhoods in the sample over time. Once the model controls for the shift in weight, house price inflation seems much more moderate.

Furthermore, the quantile regression analysis shows that pure house price inflation over the sample period caused a wedge between the price indices of higher and lower priced houses. This might point to a larger housing wealth gap.

The macro analysis indicates the rather significant role that credit extension played in the growth of the house prices in South Africa. The one overarching conclusion that the above analysis yields is the importance of mortgage credit in explaining house price movements over and above the impact of what many would consider to be the fundamental factors explaining house prices, such as disposable income and the mortgage rate. Given the impact of household wealth on consumption and given that residential property constitutes the major component of that wealth, should highlight the role and importance of mortgage credit extension in the economy. Thus, it also highlights the harmful impact that a too liberal credit extension policy might have on the economy – a situation that seems to have been the case in South African (and elsewhere).

Given the relatively small role of real disposable income and the even smaller role of the real mortgage rate in explaining real house prices, indicates the institutional, non-price allocative mechanism at play in the housing market. The small role of the mortgage rate may also be indicative of the difficulty that monetary policy at times has to combat inflation by way of interest rate increases.

Appendix 1

Table 1A shows the combination of the house price series, disposable income, mortgage credit and the mortgage rate tested for cointegration, with the first model containing all four variables, the second dropping mortgage credit, the third dropping the mortgage rate, while the fourth model drops both.

Table 1A – Cointegration test results

House price, disposable income, credit and mortgage rate	*Cointegration present*
HedonicA(0)	*Yes*
HedonicA(1)-(3)	*No*
ABSA	*Yes*
House price, disposable income and mortgage rate	*Cointegration present*
HedonicA(0)	*Yes*
HedonicA(1)-(3)	*Yes*
ABSA	*Yes*
House price, disposable income and credit	*Cointegration present*
HedonicA(0)	*Yes*
HedonicA(1)-(3)	*No*
ABSA	*Yes*
House price and disposable income	*Cointegration present*
HedonicA(0)	*Yes*
HedonicA(1)-(3)	*Yes*
ABSA	*No*

Table 2A shows the various long-run relationships estimated (if no cointegration was found, no relationship was estimated). (Note that the long-run component of Table 2A is stated in vector format, meaning that a minus (such as the minus in front of credit) should be interpreted as a plus, while a plus should be interpreted as a minus.

Table 2A – Cointegration – House prices, disposable income, credit and the mortgage rate

	HedonicA(0)			*HedonicA(1)-(3)*			*ABSA*		
House price	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
YD	*0.412*	*-0.333*	*4.514*	*-0.223*	*-0.336*	*-0.120*	*-0.525*	*-1.048*	*-3.194*
	(4.161)	*(-22.531)*	*(4.030)*	*(-7.900)*	*(-19.652)*	*(-3.479)*	*(-3.932)*	*(-74.048)*	*(-6.790)*
Credit	*-0.685*		*-2.721*				*-0.459*		*0.640*
	(-7.489)		*(-4.922)*				*(-3.716)*		*(2.733)*
Mortgagerate	*-0.755*	*7.346*			*9.851*		*4.139*	*11.385*	
	(-1.015)	*(3.302)*			*(3.830)*		*(4.638)*	*(5.301*	
C			*-26.568*						*20.616*
a	*-0.125*	*-0.018*	*0.046*	*0.005*	*-0.011*	*-0.003*	*-0.003*	*0.006*	*-0.035*
	(-1.522)	*(-1.568)*	*(1.954)*	*(0.710)*	*(-1.211)*	*(-1.228)*	*(-0.098)*	*(0.517)*	*(-0.684)*
Adj.R-squared	*0.299*	*0.228*	*0.257*	*0.412*	*0.204*	*0.204*	*0.445*	*0.438*	*0.451*

7. Bibliography

BURGER, P. (2008) The changing volatility of the South African Economy. *South African Journal of Economics,* 76**,** 335-355.

CASE, K. E., GLAESER, E. L. & PARKER, J. A. (2000) Real Estate and the Macroeconomy. Brookings Papers on Economic Activity, 2000, 119-162.

COULSON, E. (2008) Monograph on Hedonic Estimation and Housing Markets. Department of Economics, Penn State University.

CROPPER, M., DECK, L. B. & MCCONNELL., K. E. (1988) On the Choice of Functional Form for Hedonic Price Functions. Review of Economics and Statistics, 70, 668-75.

ECB. (2003) Structural Factors in the EU housing markets. *European Central Bank,* March 2003**,** 1-56.

ELS, M. & FINTEL, D. V. (2010) Residential property prices in a submarket of South Africa: Separating real returns from attribute growth. South African Journal of Economics, 78, 418-436.

FLETCHER, M., MANGAN, J. & RAEBURN, E. (2004) Comparing hedonic models for estimating and forecasting house prices. Property Management, 22, 189-200.

GOODMAN, A. C. & THIBODEAU, T. G. (1995) Age-Related Heteroskedasticity in Hedonic House Price Equations. Journal of Housing Research, 6, 25-42.

GOULD, W. W. (1992) Quantile Regression with Bootstrapped Standard Errors. Stata Technical Bulletin, 9, 19-21.

GOULD, W. W. (1997) Interquantile and Simultaneous-Quantile Regression. Stata Technical Bulletin, 38, 14-22.

HAO, L. & NAIMAN, D. Q. (2007) Quantile Regression, SAGE Publications Inc.

HECKMAN, J. J. (1979) Sample Selection Bias as a Specification Error. Econometrica, 47, 153-161.

HYPOPORT, A. (2009) Hypoport converts to hedonic house price calculation. Berlin.

JOHANSEN, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica, 59, 1551-1580.

JOHANSEN, S. (1992) Testing weak exogeneity and the order of cointegration in UK money demand data. Journal of Policy Modeling, 14, 313-334.

KOENKER, R. & HALLOCK, K. F. (2001) Quantile Regression. Journal of Economic Perspectives, 15, 143?156.

KUMINOFF, N. V., PARMETER, C. F. & POPE, J. C. (2008) Hedonic Price Functions: Guidance on Empirical Specification. American Agricultural Economics Association Annual Meeting. Orlando, Florida.

KWIATKOWSKI, D., PHILLIPS, P., SCHMIDT, P. & SHIN, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics, 54, 157-178.

LEISHMAN, C. (2003) Real Estate Market Research and Analysis, New York, Palgrave Macmillan.

MALPEZZI, S. (2003) Hedonic Pricing Models: A Selective and Applied Review. IN (EDS.), O. A. K. G. (Ed.) Housing Economics: Essays in Honor of Duncan Maclennan. Malder MA, Blackwell.

MASIH, R. & MASIH, A. M. (1996) Stock-Watson dynamic OLS (DOLS) and error-correction modelling approaches to estimating long- and short-run elasticities in a demand function: new evidence and methodological implications from an application to the demand for coal in mainland China. Energy Economics, 18, 315-334.

MCQUINN, K. & O'REILLY, G. (2006) Assessing the Role of Income and Interest Rates in Determining House Prices. Central Bank and Financial Services Authority of Ireland, 15/RT/06.

MICHEAL, H. J., BOYLE, K. J. & BOUCHARD, R. (2000) Does the measurement of environmental quality affect implicit prices estimated from hedonic models? Land Economics, 76, 283-298.

MUELBAUER, J. & MURPHY, A. (1997) Booms and Busts in the UK Housing Market. The Economic Journal, 107, 1701-1727.

PROPERTY24 (2011) About Property24.com. www.property24.com

ROSEN, S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. The Journal of Political Economy, 82, 34-55.

SELIM, S. (2008) Determinants of house prices in Turkey: a hedonic regression model. Doğu#? Űniversitesi Dergisi, 9, 65-76.

THWAITES, G. & WOOD, R. (2003) The measurement of house prices. Bank of England Quarterly Bulletin, Spring, 38-46.

TRIPLETT, J. E. (2004) Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products. STI Working Paper Series. OECD.

ZIETZ, J. & NEWSOME, B. (2001) A Note on Buyer's Agent Commision and Sale Price. Journal of Real Estate Research, 21, 245-253.

ZIETZ, J., ZIETZ, E. N. & SIRMANS, G. S. (2007) Determinants of House Prices: A Quantile Regression Approach. DEPARTMENT OF ECONOMICS AND FINANCE.