Nonparametric Estimation of a Hedonic Price Model: A South African Case Study

M Du Preez, DE Lee and M Sale

ERSA working paper 379

October 2013

Nonparametric estimation of a hedonic price model: A South African case study

M Du Preez, DE Lee and M Sale*

October 11, 2013

Abstract

Parametric regression models of hedonic price functions suffer from two main specification issues: the identification of appropriate dependent and independent variables, and the choice of functional form. Although the first issue remains relevant with the use of nonparametric regression models, the second issue becomes irrelevant since these models do not presume functional forms *a priori*. We estimate a linear parametric model via OLS, which fails a common specification test, before showing that recently developed nonparametric regression methods outperform it significantly. In addition to estimating the models, we compare the out-of-sample prediction performance of the OLS and nonparametric models. Our data reveals that the nonparametric models provide more accurate predictions than the parametric model.

1 Introduction

Hedonic pricing theory is not new (see Rosen, 1974) and has been used for some time now to investigate how the attributes of a good influence its price, especially in the housing market (Haab and McConnell, 2002; Palmquist, 2005). In most of these cases, parametric techniques were used to estimate the hedonic price functions. The use of these techniques (such as the ordinary least squares (OLS) method) requires the analyst to select appropriate dependent and independent variables and to determine the appropriate functional form governing the variables and the associated parameters (Pace, 1993; 1995).

In terms of variable selection, previous research does provide some guidance on the selection of appropriate dependent and independent covariates (see for example, Sirmans, Macphearson and Zietz, 2005), but economic theory provides very little guidance on functional form selection (Cropper, Deck and McConnell,

*Department of Economics, PO Box 77000, Nelson Mandela Metropolitan University, Port Elizabeth, 6031, South Africa

1988; Haab and McConnell, 2002; Palmquist, 2005). To a limited extent, Box-Cox transformations[1] of the functional form provide guidance on the form that best fits the data (Williams, 2008). However, specification errors associated with parametric estimation techniques may still adversely affect estimator performance (Pace, 1993; 1995).

A possible solution to the problems associated with specification error can be found in nonparametric estimation techniques (Pace, 1993; 1995; Anglin and Gencay, 1996; Bin, 2004; Palmquist, 2005). Nonparametric estimation techniques still require the analyst to select the appropriate variables but they do not impose an *a priori* parametric specification (Pace, 1993; 1995). In other words, "nonparametric estimators produce their inferences free from a particular functional form (Pace, 1995)". Popular nonparametric techniques include kernel density estimation, spline smoothing, series approximators and nearest neighbours (Anglin and Gencay, 1996).

A number of international hedonic studies have employed semiparametric and/or nonparametric estimation techniques. The nonparametric methods employed in these studies show nonlinear relationships that conventional parametric methods are unable to capture. In addition, the nonparametric methods produce much smaller prediction errors compared to their parametric counterparts. Meese and Wallace (1991) employed several different parametric specifications of a hedonic function and compared these to a regression model specified nonparametrically. The results of the study show that strong assumptions regarding functional form selection can be avoided through the use of a nonparametric regression model. In addition, the latter model provides robustness to the impacts associated with unusual observations. Coulson (1992) employed a nonparametric technique, known as spline smoothing, to investigate the relationship between housing price and floorspace size. The results of this study support the use of nonparametric models instead of parametric ones since the former are able to avoid misspecification issues. Pace (1993, 1995) applied the kernel nonparametric regression estimator to two different residential housing data sets. The study found that the nonparametric estimator outperformed the parametric estimator (OLS estimator) based on comparisons of the R^2, root mean squared error, and the mean absolute error (Pace, 1993; 1995). Anglin and Gencay (1996) estimated a benchmark parametric model (which passed several common specification tests) and compared the results to a semiparametric model. The study concluded that the semiparametric model outperformed the parametric model since it provided more accurate mean predictions (Anglin and Gencay, 1996). Bin (2004) compared the predictive powers of a conventional parametric model with a semiparametric model. In terms of predictive capability, the study found that the parametric regression was inferior to the semiparametric regression for both in-sample and out-of-sample price predictions (Bin, 2004). Racine and Parmeter (2012), however, argue that the use of *one* sample in the Bin (2004) study is entirely incorrect when assessing the out-of-sample prediction perfor-

[1]For Y^((λ)), a basic Box – Cox transformation on a single variable, the transformation can be defined as: Y^((λ)) = (Y^((λ))-1)λ for λ ≠0 or Y^((λ)) = lnY for λ=0'.

mance of two models. Anglin and Gencay's (1996) partially linear specification was recently challenged by Parmeter, Henderson and Kumbhakar (2007) who proposed the use of a nonparametric approach instead. The Parmeter *et al.* (2007) study found the following: compared to its nonparametric counterpart, the partially linear model failed a correct specification test; the within-sample fit measure applied to the partially linear model, as estimated by Anglin and Gencay (1996), was overstated; a loss of efficiency may have been caused by the inclusion of discrete variables as continuous ones into the unknown function. Racine and Parmeter (2012) examined the revealed performance of Anglin and Gencay's (1996) parametric and partially linear specifications and Parmeter et al.'s (2007) fully nonparametric specification. In contrast to Anglin and Gencay's (1996) study, Racine and Parmeter's (2012) test results show that the linear model outperforms both the semiparametric (Anglin and Gencay, 1996) and fully nonparametric specifications (Parmeter et al., 2007), while the nonparametric and semiparametric models perform equally well.

There is, however, a paucity of South African studies that have employed nonparametric estimation in a hedonic price context. This study aims to fill this gap. More specifically, this study uses a unique dataset (the Du Preez and Sale (forthcoming) one) to investigate whether recently developed nonparametric (kernel) methods provide any additional insight into hedonic pricing parametric analysis. The Du Preez and Sale (forthcoming) study was conducted in order to determine whether social housing development projects lead to the deterioration of surrounding residential property values. The locus of the study is a single neighbourhood (the Walmer one in Port Elizabeth, South Africa) and the adjacent Gqebera Township[2] (a proxy for a social housing development).

In what follows, Section 2 presents the methodology for the OLS (parametric) and kernel density (nonparametric) estimators as well as the data used for analysis, Section 3 applies the OLS and nonparametric estimators to the Du Preez and Sale (forthcoming) dataset, and finally, Section 4 concludes the paper.

2 Methodology

2.1 Parametric regression

In terms of regression analysis, the dependent variable, y, comprises a systematic component, $E(y|x)$, which varies with the covariates, x, a random component, and an error term, ε:

$$y = E(y|x) + \varepsilon \tag{1}$$

Parametrically, $\mathrm{E}(y|x)$ is estimated by applying a two-stage procedure: first, $E(y|x)$ is modelled as a function of the parameters, β, and second, β is estimated (Pace, 1993; Cameron and Trivedi, 2005). A linear interpretation[3] of Equation

[2] The Gqebera Township is also known as the Walmer Township.
[3] Traditionally, researchers have favoured a linear specification of the hedonic price model.

(1), for example, assumes the conditional expectation of y is modelled to depend on the parameters, β, in a linear fashion:

$$E(y|x) = x_0\beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_p\beta_p \quad (2)$$

where: $p =$ the number of non-constant variables
$x_0 =$ a constant vector which acts as the intercept variable (Pace, 1993).

The estimation of Equation (1), assuming the dependency of y on x is linear, via the OLS estimator is very popular because the estimator involves very little computational work and it has all the desired properties under the classical linear regression model assumptions. Parametric regression, regardless of the estimator selected[4], demands the researcher carefully considers both the specification of Equation (1) and its estimation. Specification issues relate to the identification of the appropriate dependent variable and explanatory covariates, as well as the relevant functional form.

One of the main concerns regarding the specification of a hedonic price function revolves around omitted variable bias[5]. Hedonic price functions specified in the traditional manner do not address the issue of omitted variable bias (Brasington & Hite, 2005). This can limit the ability of the hedonic regression to accurately measure the implicit prices for housing characteristics. Neighbourhood characteristics that matter to households but have not been observed by the researcher may well be correlated with the variables of interest (or other independent variables). Examples of potential omitted variables include air pollution, the presence of shopping centres, the crime rate, the quality of landscaping and the presence of highways.

Evaluating the implications of omitted variables in hedonic price studies is challenging because the market clearing process determines housing values and consumer welfare simultaneously (Rosen, 1974). In order to meet this challenge, Cropper et al. (1988) developed a theoretically consistent framework for imitating hedonic equilibria. It was found that traditional functional forms (linear, semi-log and double-log) performed best in the presence of omitted variables. More flexible functional forms were preferred when all variables were included in the model. This is one of the reasons why the vast majority of hedonic price studies use traditional functional form specification (to avoid the risk of omitted variable bias).

Possible remedies to overcome omitted variable bias include zoning in on small geographical areas and collecting as much data as possible (Brasington & Hite, 2005). An alternative method for addressing this issue is to incorporate a spatial autoregressive term in the traditional hedonic price function, as this term captures the influence of omitted variables (Brasington & Hite, 2005) – this study employed this method. These unmeasured influences help to determine the value of neighbouring houses which, in turn, are related to the subject house.

[4] A popular alternative to OLS estimation is maximum likelihood estimation.
[5] We would like to thank an anonymous referee for pointing this out.

2.2 Nonparametric regression

2.2.1 Kernel density estimation

Although the identification of the appropriate dependent variable and explanatory covariates remains relevant in the case of nonparametric regression, the distribution for the unobserved random element and the functional form of Equation (1) becomes less important (Pace, 1993; Haab and McConnell, 2002; Cameron and Trivedi, 2005). Since nonparametric regression estimators estimate $E(y|x)$ directly, the specification of variable interactions are circumvented (Pace, 1993). Integrating y times its conditional probability density function (*pdf*) produces $E(y|x)$, where the *pdf* is the result of dividing $pdf(y,x)$ by $pdf(x)$:

$$E(y|x) = \int y.pdf(y,x)/pdf(x)dy \tag{3}$$

where: $pdf(y,x)$ = the joint density of y and x

$pdf(x)$ = the marginal density of x (Pace, 1993; Cameron and Trivedi, 2005).

Thus, a nonparametric estimate of $E(y|x)$ can be derived if nonparametric estimates of $pdf(y,x)$ and $pdf(x)$ can be estimated. Traditionally, nonparametric estimates of densities can be obtained by using histograms[6] (Pace, 1993; Cameron and Trivedi, 2005; Racine, 2008). The latter, however, produce discontinuities. To avoid these, the kernel method[7] of nonparametric density estimation can be employed (Pace, 1993; Racine, 2008). This method employs a function, K(·), called a kernel, whose value fluctuates inversely and smoothly with u_i:

$$g(x_{s0}) = [1/(hn)][\sum_{i=1}^{n} K(u_i)] \tag{4}$$

where: n = the number of observations

h = the window width for a variable[8]

$u_1 = (x_{0j} - -x_{ij})/h_j$

A direct estimate of $E(y|x)$ can, thus, be produced by employing a kernel function to produce a smooth estimate of the appropriate densities (Pace, 1993; Racine, 2008). If the function is extended to include multiple variables then the Nadaraya-Watson estimator[9] (also called the local constant estimator)

---

[6] The construction of a histogram is relatively simple. The first step is to construct a series of bins (Racine, 2008). For positive and negative integers m, the bins are intervals $[x_0 + mh, x_0 + (m+1)h)$. Formally, the histogram is defined as:

$$\hat{f}x = 1/n.(\#\text{ of } X_i \text{ in the same bin as } x)/(\text{width of bin containing } x)$$

$$= 1/nh \sum_{i=1}^{n} 1(X_i \text{ is in the same bin as } x)$$

where: $\mathbf{1}(A)$ = an indicator function which takes on the value of 1 if A is true, zero otherwise (Racine, 2008). The researcher usually selects an origin x_0 and bin width h by employing rules-of-thumb. Both these choices have an impact on the resulting estimate (Racine, 2008).

[7] Rosenblatt (1956) introduced this method.

[8] Hedonic price theory, as applied to the housing market, provides a logical explanation for h: it controls how many houses actually affect y_i.

[9] This estimator was first formulated by Nadaraya (1964) and Watson (1964).

results (Pace, 1993; Cameron and Trivedi, 2005; Racine, 2008). Given a specific instance of x, x_0, this estimator produces the dependent variable's (y_i's) predicted value, y_0, by linearly weighting[10] it:

$$'g(x_0) = E(y|x) = \sum_{i=1}^{n} y_i w_i \tag{5}$$

where: $w_i = K(u_i)/\sum_{i=1}^{n} K(u_i)$

The multivariate kernel function is obtained by multiplying the kernel function of each variable together:

$$K(u_i) = \prod_{j=1}^{p} K(u_{ij}) \tag{6}$$

The Nadaraya-Watson estimator[11] is a local constant estimator since it assumes $E(y|x)$ is equivalent to a constant in the local neighbourhood of x_0 (Cameron and Trivedi, 2005; Racine, 2008). Alternatively, if one assumes $g(x)$ is linear in the x_0 neighbourhood, then $x_0 = a_0 + b_0(x - -x_0)$ in the x_0 neighbourhood. In order to obtain $\hat{g}(x_0)$ under these circumstances, $\sum_i K((x_i - x_0)/h)(y_i - g_0)^2$ must be minimized with respect to g_0 (Cameron and Trivedi, 2005; Racine, 2008). The local linear regression[12] minimizes

$$\sum_{i}^{n} K((x_i - x_0)/h)(y_i - a_0 - b_0(x_i - x_0))^2 \tag{7}$$

with respect to a_0 and b_0 (Cameron and Trivedi, 2005; Racine, 2008). This means that $\hat{g}(x)$ is equal to $\hat{a}_0$ $+b_0(x - x_0)$ in the x_0 neighbourhood, and at exactly x_0, the estimate is given by $\hat{g}'(x) = \hat{a}_0$, and $\hat{b}_0$ can be interpreted as an estimate of the first derivative $\hat{g}'(x_0)$ (Cameron and Trivedi, 2005; Racine, 2008). A major difference between the local constant estimator and the local linear estimator is that the former has the property that irrelevant variables can be smoothed out totally, whereas the latter does not, which may lead to excessive variability (Racine, 2008).

2.2.2 Kernel coefficient estimation

Implicit price and welfare estimation in hedonic price studies rely on the use of coefficient estimates for the covariates of interest. With a parametric model, the coefficient, β, for a specific variable, x_j, can be estimated by differentiating the regression function, $E(y|x)$, with respect to x_j:

$$\beta_{j0} = \partial E(y|x)/\partial x_j \tag{8}$$

The analogue to β in Equation (8), in the case of the Nadaraya-Watson estimator, is the amorphous partial derivative, $\hat{\beta}_{j0}$. It is defined as "the numerical

[10] Positive kernels, such as the normal *pdf*, have weights that vary between 0 and 1.
[11] Regression based on joint densities only holds for the Nadaraya-Watson estimator.
[12] This is also known as the local polynomial method.

derivative of the kernel estimator predictions with respect to the independent variable" (Ullah, 1988; Pace, 1993), and depends upon $\partial w_i/\partial x_j$:

$$\partial E(y|x)/\partial x_j = \sum_{i=1}^{n} y_i(\partial w_i/\partial x_j) \tag{9}$$

Note, $\partial w_i/\partial x_j$ in Equation (9) can be separated into the sum of two quantities if the product rule is invoked:

$$\partial w_i/\partial x_j = K_{1i} + K_{2i} \tag{10}$$

where: $K_{1i} = K'(u_i)/\sum_{i=1}^{n} K(u_i)$
$K_{2i} = K'(u_i)/[\sum_{i=1}^{n} K(u_i)]^2.[\sum_{i=1}^{n} K(u_i)]$
$K'(u_i)$ = the partial derivative of the kernel function with respect to x_j (Ullah, 1988; Pace, 1993).

Substituting Equation (10) into Equation (8) produces the amorphous partial derivative estimator:

$$\tilde{\beta}_{j0} = \sum_{i=1}^{n} y_i(K_{1i} + K_{2i}) \tag{11}$$

As was mentioned in Section 3.3.1, $\hat{b}_0$ can be interpreted as the local linear model's partial derivative estimate.

2.2.3 Kernel density estimation with discrete and continuous data[13]

Conventional hedonic price models very often contain both discrete and continuous variables as predictors of house value. This does not pose a problem in parametric models, but does require special treatment in the case of nonparametric ones[14]. Traditionally, estimation of frequency nonparametric models, which include qualitative covariates, necessitated splitting the data into subsets containing only the continuous covariates of interest. Smooth nonparametric regression models, which smooth only the continuous data, would then be constructed for each of these subsets. This procedure may, however, lead to a loss of efficiency. The most popular modern approach[15] to dealing with a mix of discrete and continuous data entails the use of the concept of "generalised product kernels" (Racine and Li, 2004; Racine, 2008). This approach avoids sample splitting, which means that sound nonparametric estimation can take place using the full sample of observations. In the case of continuous variables, standard continuous kernels (denoted by $W(\cdot)$) are employed, whereas in the

[13] A very thorough discussion of nonparametric estimation of regression functions with both categorical and continuous data is provided by Racine and Li (2004).

[14] In earlier times, the presence of categorical regressors necessitated the use of semiparametric models (see Robinson, 1988; Stock, 1989). This was before Li and Racine (2004) and Racine and Li (2004) developed a nonparametric kernel regression technique which could smooth out both ordered and unordered categorical data.

[15] Traditionally, the "frequency approach" was adopted by researchers to deal with mixed data (Racine, 2008). According to this approach, the continuous data were broken up "into subsets according to the realizations of the discrete data ("cells")" (Racine, 2008).

case of an unordered discrete variable, $\tilde{x}^d$, Aitchison and Aitken's (1976) kernel is employed:

$$\bar{l}(\bar{X}_i^d, \bar{x}^d) = \{ \begin{matrix} 1-\lambda, \ if \ \bar{X}_t^d, \bar{x}^d, \\ \frac{\lambda}{c-1}, otherwise., \end{matrix} \quad (12)$$

Wang and Van Ryzin's (1981) kernel can be employed for an ordered discrete variable, $\tilde{x}^d$:

$$\tilde{\imath}(\tilde{X}_i^d, \tilde{x}^d) = \{ \begin{matrix} 1-\lambda, \ if \ \tilde{X}_i^d, = \tilde{x}^d \\ \frac{\lambda-1}{2}, \lambda|\tilde{X}_i^d - \tilde{x}^d|, if \tilde{X}_i^d \neq \tilde{x}^d. \end{matrix} \quad (13)$$

One can define a generalised product kernel for one continuous, one unordered, and one ordered variable as follows:

$$K(\cdot) = W(\cdot) * \tilde{\imath}(\cdot) * \tilde{\imath}(\cdot) \quad (14)$$

The product kernel employed above is the same kernel used in the Nadaraya-Watson and local linear estimators (though for the local linear estimator only the continuous predictors appear in the polynomial component). A joint probability/density function defined over mixed data can now be estimated by employing the generalised product kernels (Racine, 2008). The kernel estimator of the *pdf* for one unordered discrete variable, $\tilde{x}^d$, and one continuous variable, x^c, is as follows:

$$\hat{g}(\bar{x}^d, x^c) = 1/n^h x^c \sum_{i=1}^{n} \bar{l}(\bar{X}_i^d, \bar{x}d)(\frac{x_i^c - x^c}{W(^h x^c)}) \quad (15)$$

2.2.4 Parametric versus nonparametric specifications[16]

Three measures are frequently employed to assess the performance of parametric models compared to their nonparametric counterparts, namely goodness-of-fit values, out-of-sample prediction performance, and revealed performance tests.

Goodness-of-fit comparison

Conventionally, the within-sample goodness-of-fit measure for the parametric model is compared to the one for the nonparametric model. To carry out this comparison a nonparametric R^2 measure comparable to the parametric R^2 is required. This measure can be formally defined as follows by letting Y_i represent the outcome and $\hat{Y}_i$ the fitted value for observation i:

$$R^2 = [\sum_{i=1}^{n} (Y_i - \bar{y})(\hat{Y}_i - \bar{y})]^2 / \sum_{i=1}^{n} (Y_i - \bar{y})^2 \sum_{i=1}^{n} (\hat{Y}_i - \bar{y})^2 \quad (16)$$

The R^2 defined above will always lie in the range [0,1] where 1 implies a perfect fit and 0 implies no predictive power above that provided by the unconditional mean of the target (Racine, 2008). If, for example, the nonparametric

[16] The quality of the data impacts heavily on the performance of conventional estimators, such as OLS (Pace, 1993). Not unlike a nonparametric estimator, the OLS's predicted value of the dependent variable is a weighted mean of the observations. But unlike the nonparametric estimator, the OLS weights can fall below zero and exceed one. Thus, the quality of the data and the presence of outlying observations affect the OLS estimator more than the nonparametric one (Pace, 1993).

model has a better goodness-of-fit compared to the parametric one then it is appropriate to investigate whether the improvement is due to overfitting or whether it reflects the fact that the nonparametric model is more faithful to the underlying data generating process (DGP).

Out-of-sample prediction performance

The primary aim of statistical valuation models, such as the hedonic price one, is out-of-sample prediction. In order for a nonparametric estimator to outperform the standard OLS estimator, it must show superior out-of-sample prediction capabilities (see Racine and Parmeter, 2012). OLS and nonparametric estimators could compete on the basis of out-of-sample errors, which are estimated via cross-validation (see Li and Racine, 2004). More specifically, the original sample is split into two independent training/evaluation samples, namely n_1 and n_2. Each model is then fitted on the n_1 training observations and evaluated on the n_2 hold-out observations via the predicted squared error (PSE) criterion, namely $n_2^{-1}\sum_{i=1}^{n2}(y_i - \hat{y}_i)^2$, where the y_is are the house price values for the hold-out observations and the $\hat{y}_i$s are the predicted values (Li and Racine, 2004). This process is repeated M times after which the distribution of the PSEs for the models are compared.

Revealed performance tests

In addition to the goodness-of-fit comparison and cross-validation exercise, Racine and Parmeter's (2012) revealed performance ('RP') test could be performed to test whether there is any significant difference among the parametric and nonparametric models in terms of their expected performance on unseen data. More specifically, a simple (paired) test of differences in means for the distributions of the models' PSEs is carried out.

The test of revealed performance carried out in this study tests two distinct null hypotheses. First, we test if the local constant nonparametric and linear parametric models have equal PSEs, and second, we test if the local linear nonparametric and linear parametric models have equal PSEs. For both tests our alternative hypothesis is that the less general model has a greater PSE.

2.3 Data collection

The housing market data in this study comes from the suburb of Walmer, Port Elizabeth. Walmer comprises of a total of 2625 properties and a total of 1326 transactions took place from 1995 – 2009 (excluding repeat sales) (South African Property Transfer Guide, 2011). The dependent variable of the empirical model is the actual recorded sales price of a house. The Absa house price index was used to inflate the sales price of each individual house to constant 2009 rands. Data on 170 houses that have been traded at least once during the past 15 years in Walmer were collected, resulting in a sample size of 13%. Data on the following variables[17] were collected, namely sales price, number of bathrooms, number of bedrooms, the presence of a swimming pool, the presence of air conditioning, the presence of a garage, the presence of an electric fence, the number

[17] Research by Sirmans *et al.* (2005) was consulted in order to select appropriate structural and neighbourhood characteristics.

of stories, the size of the erf, age, distance to a social housing development (Walmer Township), and distance to a school (Clarendon primary school). In addition, an autoregressive term[18] was also estimated to accommodate the fact that the transaction price of a house is determined not only by its structural and neighbourhood characteristics, but also by transaction prices of prior sales within its vicinity (Can and Megbolugbe, 1997).

3 Results and Discussion

3.1 Parametric Hedonic Model Summary

We first estimate a linear hedonic model using OLS, and then subject the model to some specification analysis. The hedonic price model used in this study is assumed to be separable in the unobservables, which implies that the effects of the different observables on the price are homogenous.[19] Summary results (parameter estimates) are summarized in Table 1.

The results from the linear hedonic regression generally conform to *a priori* expectations - the number of stories, the size of the erf, the presence of a pool and the presence of an electric fence all have significant, positive effects on property values in the sample. In addition, a significant positive relationship exists between house prices and the variable of interest, namely distance from the Gqebera Township (i.e. DISTWAL)[20]. The coefficient of the latter covariate can be interpreted as its implicit price (or marginal value): the average price of a house in the Walmer neighbourhood increases by R228.85 for every metre further away from the low-cost housing development. According to Table 1, the model estimated explains approximately 59 percent of the variation in the dependent variable (actual sales price).

We test whether the parametric model is consistent with the data being analyzed using Ramsey's (1969) test. According to this test, the parametric model is rejected by the data (the p-value for the null of correct specification

[18] This spatial relationship is appropriate because an individual will often base his/her offer bid after having researched the prior transaction prices in the surrounding area (Brasington and Hite, 2005). This practice, known as "comparable sales" is often employed by real estate experts when trying to estimate the market value of a specific property (Can and Megbolugbe, 1997). The presence of spatial autocorrelation may lead to unsatisfactory estimate of the coefficients. In order to address this issue in this study, a spatial autoregressive term was included in order to determine the impact of prior transactions on the sale price of the selected house.

[19] We would like to thank an anonymous referee for pointing this out.

[20] An anonymous referee pointed out that the geographical distance from the township may have remained the same throughout the 15 years covered by the data, but the perceived nuisance of the proximity to the township would likely be a dynamic factor, which cannot be captured by a simple geographical measure. The referee suggested that a possible solution to this problem would be to weight the latter by, for example, a crime perception measure. The authors acknowledge that this may be a complication with the use of a simple geographical measure. They were, however, forced to use data spanning 15 years simply because of insufficient house sales data. Unfortunately, no data is available to compile a crime perception measure for that period.

is 0.006119). Hence, we investigate recently developed nonparametric methods (i.e. kernel methods) capable of handling the mix of categorical and continuous data types.

3.2 Nonparametric Hedonic Models

Having rejected the parametric hedonic model, we next turn to the estimation of two nonparametric hedonic price models, namely a local constant one and a local linear one. The bandwidths for both models were selected via least-squares cross-validation[21]. Since the hedonic price model contains a mix of continuous and categorical predictors (see Section 2.3), we adopt the nonparametric approaches described in Li and Racine (2004) and Racine and Li (2004). For our analysis, we begin with a local constant specification having the capability to automatically remove irrelevant regressors, both continuous and categorical (see Appendix A, Table 2). A local linear model, which also contains all the regressors, is estimated for comparative purposes (see Appendix A, Table 3). We then retain the variables that were not smoothed out by the local constant estimator, and estimate both the local constant and local linear[22] models on the subset of regressors (see Appendix A, Tables 4 and 5) – these reduced models are deemed the preferred ones since irrelevant regressors have been eliminated.

We display partial regression plots in Figures 1 and 2 for the local constant and local linear nonparametric models, respectively. These plots represent a 2D plot of the outcome y versus one explanatory variable x_j when all other variables are held constant at their respective medians/modes (Racine, 2008). In addition, we also plot bootstrapped variability bounds – these are frequently more desirable compared to those obtained through the asymptotic approximations (Racine, 2008) – Figures 3 and 4 present the partial response plots along with their bootstrapped error bounds for the local constant and local linear models, respectively.

Figure 1 reveals that (holding other regressors constant at their median/mode), houses with more than two bedrooms sell for higher prices than those that only have two. Houses also sell for more if they have a swimming pool, one or two air conditioners instead of none, and two stories instead of one. Houses with an electric fence, however, sell for less than ones that have no fence. Figure 2 shows very similar plots for the discrete variables, except for the electric fence variable - houses with an electric fence sell for more than ones that have no fence. In terms of the continuous variable plots, house prices fall the further away a house from a primary school (DistClarendon), house prices rise the greater the size of the erf, and house prices rise the further away houses are located from the Gqebera Township (see Figure 1). The plots of continuous variables in Figure 2 are generally similar to those in Figure 1. Figures 3 and 4 below show the first derivative of each of the plots in Figures 1 and 2 above with respect to

[21] Four other general bandwidth selection methods exist, namely reference rules-of-thumb, plug-in methods, likelihood cross-validation and bootstrap methods.

[22] The local linear specification has improved finite-sample properties though it lacks the ability to automatically remove continuous variables.

the x-axis variable (i.e. marginal effects, the rate of change) while those for the categorical are finite differences (all holding of x-axis variables constant at their median/modes).

According to the local constant estimation (see Appendix A, Table 2), the following regressors are irrelevant, namely Bath, GARAGEDUMMY and Age – the bandwidth, h, for the continuous regressor is very large and the bandwidth, λ, for the discrete regressors is equal to unity (the upper bound value of λ is one)[23]. The nonparametric models explain substantially more variation in the dependent variable compared to the parametric model, viz. the parametric model has an R-squared of 0.5906 (see Table 1 above), the local linear (on the restricted subset) has an R-squared of 0.874, and the local constant (on the restricted subset) an R-squared of 0.9842 (see Appendix A, Tables 4 and 5).

In addition, the performance of the parametric model versus that of the local linear model is judged by looking into the relative size of the coefficient on DISTWAL, the variable of interest in this study (the local linear estimator automatically provides smooth estimates of the first-order partial derivatives while the local constant does not – see Section 2.2.2). For the parametric model it is 228.8465, while the derivative associated with DISTWAL for the non-parametric local linear model, evaluated at the median value of DISTWAL (median=1800) holding all other regressors constant at their median/mode, is 235.1758 – see Figures 3 and 4, DISTWAL panel. As mentioned above, these coefficients represent the amount by which the price of an average house in the sample increases by for every additional meter in distance away from the Gqebera Township (R228.85 and R235.18, respectively, for the parametric and local linear models). The difference between the coefficient estimates (i.e. implicit prices) is slight and as a result the models are also judged in terms of their out-of-sample predictions (see Racine and Parmeter, 2012).

We split the original sample into independent training/evaluation samples, fit each model on the training and then evaluate on the hold-out sample via predicted squared error (PSE) (Figure 5 shows that the PSE for the local constant nonparametric model is the lowest followed by the local linear nonparametric model and then the linear parametric model).

We repeat these M=10,000 times and then compare the distribution of the PSEs. The relative local linear median PSE is 0.892 (local linear nonparametric/parametric), and the relative mean PSE is 0.9612. In other words, the local linear nonparametric model is approximately 11% more efficient (taking the median PSE) compared to the linear parametric model as measured in terms of performance on independent data. The relative local constant median PSE is 0.7712 (local constant nonparametric/parametric), and the relative mean PSE is 0.7919. Thus, the local constant nonparametric model is approximately 23% more efficient (taking the median PSE) compared to the linear parametric model as measured in terms of performance on independent data.

[23] To obtain consistent estimators when employing kernel methods with continuous and discrete variables, two sets of conditions, namely $h \to 0$ and $\lambda \to 0$, must be met (Racine and Li, 2004).

Next, we test whether there is any significant difference among the models in terms of their expected performance on unseen data using Racine and Parmeter's (2012) revealed performance ('RP') test. The test of Racine and Parmeter (2012) rejects the null that the parametric model has out-of-sample performance better than to or equal to the nonparametric local linear model at all conventional levels (P-value $=1.2635 \times 10^{-11}$), and also rejects the null that parametric model has out-of-sample performance better than to or equal to the nonparametric local constant model at all conventional levels (P-value $= 5.1824 \times 10^{-285}$). We therefore conclude that the improvement in in-sample fit revealed by the nonparametric specifications is not simply an artifact of 'overfitting', rather, it reflects a genuine improvement in the model's fidelity to the underlying data generating process (DGP) since they also outperform the parametric model on independent data drawn from the same DGP as witnessed by the results of Racine and Parmeter's (2012) RP test.

4 Conclusion

This study has considered parametric and non-parametric estimation of a hedonic price model to determine the impact of social cost housing on nearby residential property prices of a neighbourhood located in Port Elizabeth, South Africa. This paper finds that the Gqebera Township (a proxy for a social housing development) has a statistically significant negative impact on the Walmer neighbourhood's property prices, regardless of which regression technique is employed. Not unlike other international studies (see for example Pace (1993), Anglin and Gencay (1996) and Bin (2004)), the study found that the parametric regression was inferior to the nonparametric regression based on out-of-sample price predictions and Racine and Parmeter's (2012) revealed performance test.

Acknowledgements

We would like to thank Prof. Jeffrey Racine for valuable comments, discussions and assistance with estimating the nonparametric models in R.

References

[1] **A**itken, C G G (1976): "Multivariate Binary Discrimination by the Kernel Method" *Biometrika*, 63(3): 413 – 420.

[2] Anglin, P and Gencay, R (1996): "Semiparametric Estimation of Hedonic Price Functions. *Journal of Applied Econometrics*, 11: 633 – 648.

[3] Bin, O (2004): "A Prediction Comparison of Housing Sales Prices by Parametric Versus Semi-parametric Regressions", *Journal of Housing Economics*, 13: 68 – 84.

[4] Brasington, D and Hite, D (2005): *Demand for Environmental Quality: A Spatial Hedonic Analysis*, Working Paper 2005-08, Department of Economics, Louisiana State University.

[5] Cameron, A C and Trivedi, P K (2005): *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

[6] Can, A and Megbolugbe, I (1997): "Spatial Dependence and House Price Index Construction", Journal *of Real Estate Finance and Economics*, 14(2): 203-222.

[7] Coulson, E (1992): "Semiparametric Estimates of the Marginal Price of Floorspace", *Journal of Real Estate Finance and Economics*, 5(1): 73 – 83.

[8] Cropper, M, Deck, L and McConnell, K (1988): "On the Choice of Functional Form for Hedonic Price Functions", *The Review of Economics and Statistics*, 70 (4): 668 – 675.

[9] Du Preez, M and Sale, M C (forthcoming): "The Impact of Social Housing on Nearby Property Prices: A Nelson Mandela Bay Case Study", *The South African Journal of Economics*.

[10] Haab, T and McConnell, K (2002): *Valuing Environmental and Natural Resources: The Econometrics of Non-market valuation*, Edward Elgar, Vermont.

[11] Li, Q and Racine, J (2004): "Cross-validated Local Linear Nonparametric Regression", *Statistica Sinica*, 14(2): 485–512.

[12] Meese, R and Wallace, N (1991): "Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices", *American Real Estate and Urban Economics Association Journal*, 19: 308 – 332.

[13] Nadaraya, E A (1964): "On Estimating Regression", *Theory of Probability and Applications*, 9: 141 – 142.

[14] Pace, R (1993): "Nonparametric Methods with Applications to Hedonic Models", *Journal of Real Estate Finance and Economics*, 7: 185 – 204.

[15] Pace, R (1995): "Parametric, Semiparametric, and Nonparametric Estimation of Characteristics Values within Mass Assessment and Hedonic Pricing Models", *Journal of Real Estate Finance and Economics*, 11: 195 – 217.

[16] Palmquist, R B (2005): "Property Value Models", in Maler, K G and Vincent, J (eds.): *Handbook of Environmental Economics: Valuing Environmental Changes*, Elsevier, Amsterdam.

[17] Racine, J and Li, Q (2004): "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data", *Journal of Econometrics*, 119(1): 99–130.

[18] Racine, J (2008): "Nonparametric Econometrics: A Primer", *Foundations and Trends In Econometrics*, 3(1): 1 – 88.

[19] Racine, J and Parmeter, C (2012): "Data-driven Model Evaluation: A Test for Revealed Per- formance", in Ullah, A, Racine, J and Su, L (eds.): *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press, Oxford.

[20] Ramsey, J (1969): "Tests for Specification Error in Classical Linear Least Squares Regression Analysis", *Journal of the Royal Statistical Society, Series B*, 31: 350–371.

[21] Robinson, P M (1988): "Root-N-Consistent Semiparametric Regression", *Econometrica*, 56(4): 931 – 954.

[22] Rosen, S (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Perfect Competition", *Journal of Political Economy,* 82(1): 34 – 55.

[23] Rosenblatt, M (1956): "Remarks on Some Nonparametric Estimates of Density Function", *Annals of Mathematical Statistics*, 27(3): 832 – 837.

[24] Sirmans, G, Macphearson, D and Zietz, E (2005): "The Composition of Hedonic Pricing Models", *Journal of Real Estate Literature*, 13(1): 1-43.

[25] South African Property Transfer Guide (2010): *South African Property Transfer Guide*. http://www.saptg.co.za [25 March 2010]

[26] Stock, J (1989): "Nonparametric Policy Analysis", *Journal of the American Statistical Association*, 84(406): 567 -575.

[27] Stock, J (1991): "Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits", in Barnett, W, Powell, J and Tauchen, G (eds.): *Nonparametric and Semiparametric Methods in Economics and Statistics: Proceedings of the* 5^{th} *International Symposium in Economic Theory and Econometrics*, Cambridge University Press, New York.

[28] Ullah, A (1988): "Nonparametric Estimation of Economic Functionals", *Canadian Journal of Economics*, 21: 625 – 658.

[29] Wang, M C and Van Ryzin, J (1981): "A Class of Smooth Estimators for Discrete Distributions", *Biometrika*, 68(1): 301 – 309.

[30] Watson, G S (1964): "Smooth Regression Analysis", *Sankhya: The Indian Journal of Statistics, Series A*, 26(4): 359 – 372.

[31] Williams, B (2008): "A Hedonic Model for Internet Access Service in the Consumer Price Index", *Monthly Labour Review*, 131(7): 33-48.

Table1: Summary table for linear regression parameter estimates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-87293.6097	359650.8403	-0.24	0.8086
Bath.L	408322.3894	416112.5574	0.98	0.3281
Bath.Q	-100421.7858	351462.8087	-0.29	0.7755
Bath.C	279126.0358	404325.2250	0.69	0.4911
Bath^4	194021.3003	347375.4407	0.56	0.5774
Bath^5	-92973.8715	347413.0600	-0.27	0.7894
Bath^6	-270698.2779	348347.3308	-0.78	0.4384
Bath^7	-496861.6509	382333.5972	-1.30	0.1959
Bath^8	-555537.3474	392735.2436	-1.41	0.1594
Bath^9	-494281.3030	333877.7317	-1.48	0.1410
Bath^10	-299517.9008	387141.6192	-0.77	0.4404
Bath^11	101728.3256	378272.5798	0.27	0.7884
Bath^12	562428.7237	256995.4739	2.19	0.0303
Bed.L	-146195.6483	430422.4872	-0.34	0.7346
Bed.Q	-248109.5167	380571.6554	-0.65	0.5155
Bed.C	457196.9928	315237.2782	1.45	0.1492
Bed^4	265145.0986	244150.4456	1.09	0.2793
Bed^5	-36806.6807	162899.5151	-0.23	0.8216
Swim.L	235260.7668	82136.0953	2.86	0.0048
Aircon.L	50802.1781	410431.0117	0.12	0.9017
Aircon.Q	-62180.3331	239982.0027	-0.26	0.7959
GARAGEDUMMY.L	14342.9991	84190.3596	0.17	0.8650
ElecFence.L	170664.1376	73879.2691	2.31	0.0223
Stories.L	183955.9863	88875.2734	2.07	0.0403
DISTWAL	228.8465	86.8565	2.63	0.0094
DistClaredon	-28.8512	76.0022	-0.38	0.7048
Size.Erf	637.0159	80.7348	7.89	0.0000
Age	1430.1180	2105.1905	0.68	0.4980
AUTO.REGRESSIVE.TERM	0.0386	0.0690	0.56	0.5769
Adjusted R^2	0.5906			

Figure 1: Partial local constant model plots

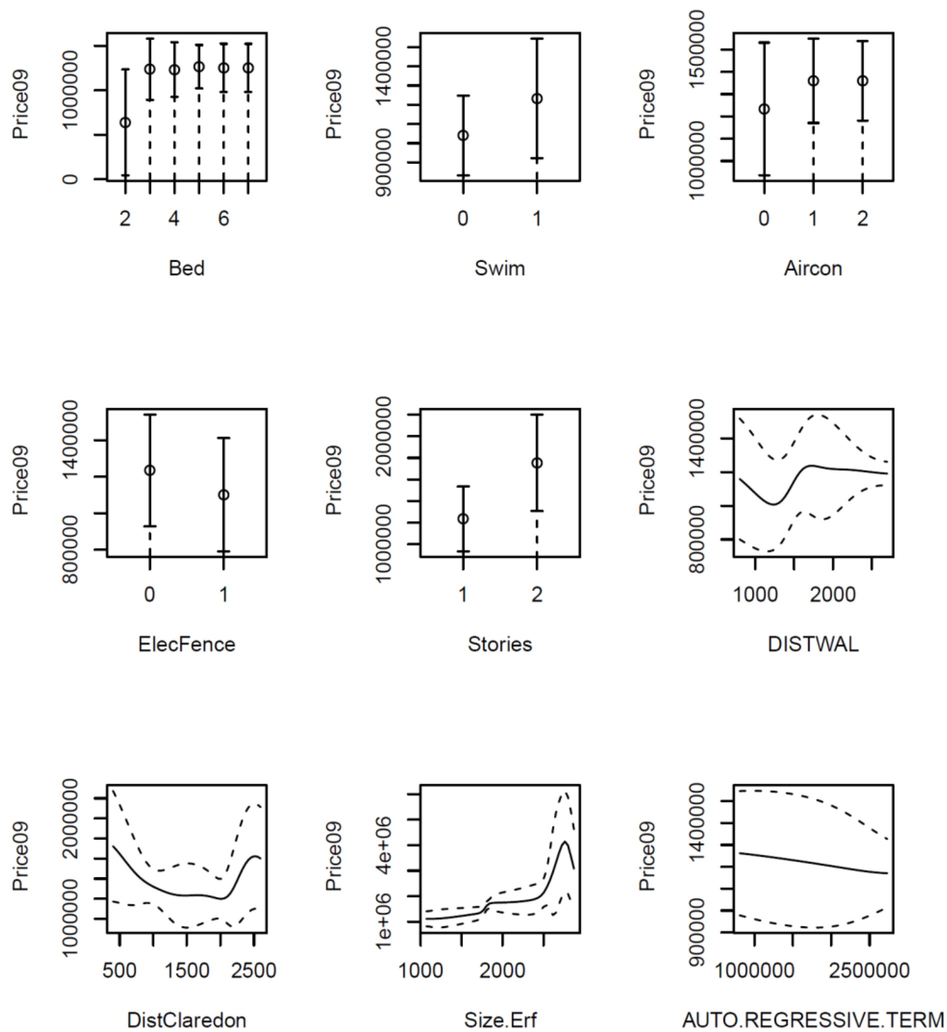


Figure 2: Partial local linear model plots

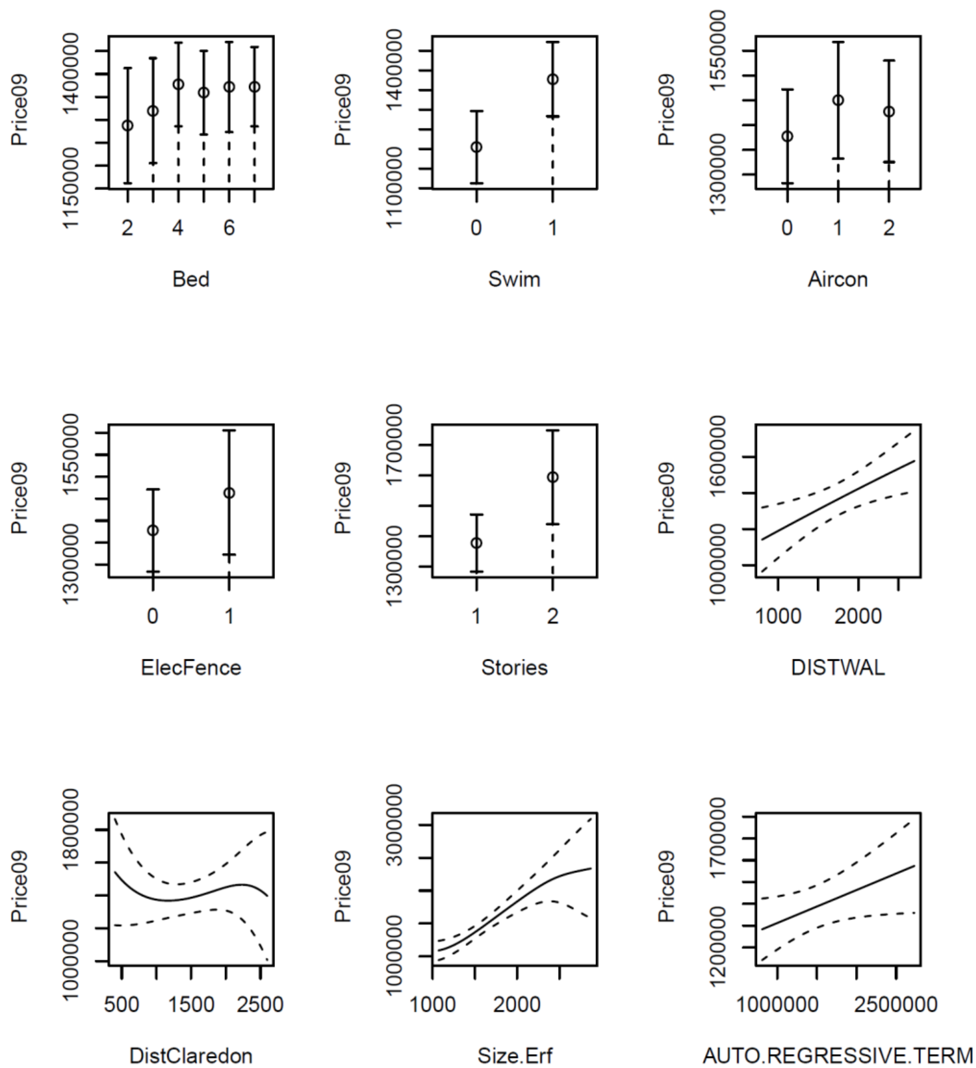


Figure 3: Local constant model first derivative plots

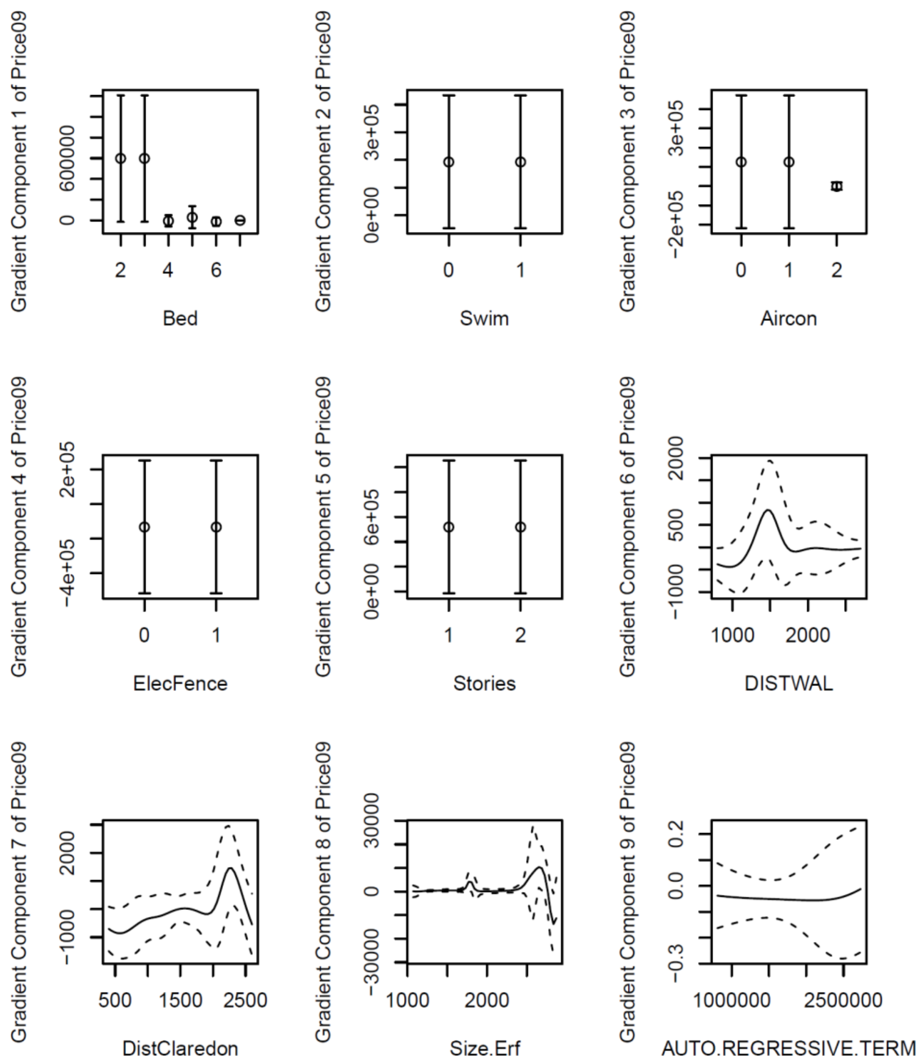


Figure 4: Local linear model first derivative plots

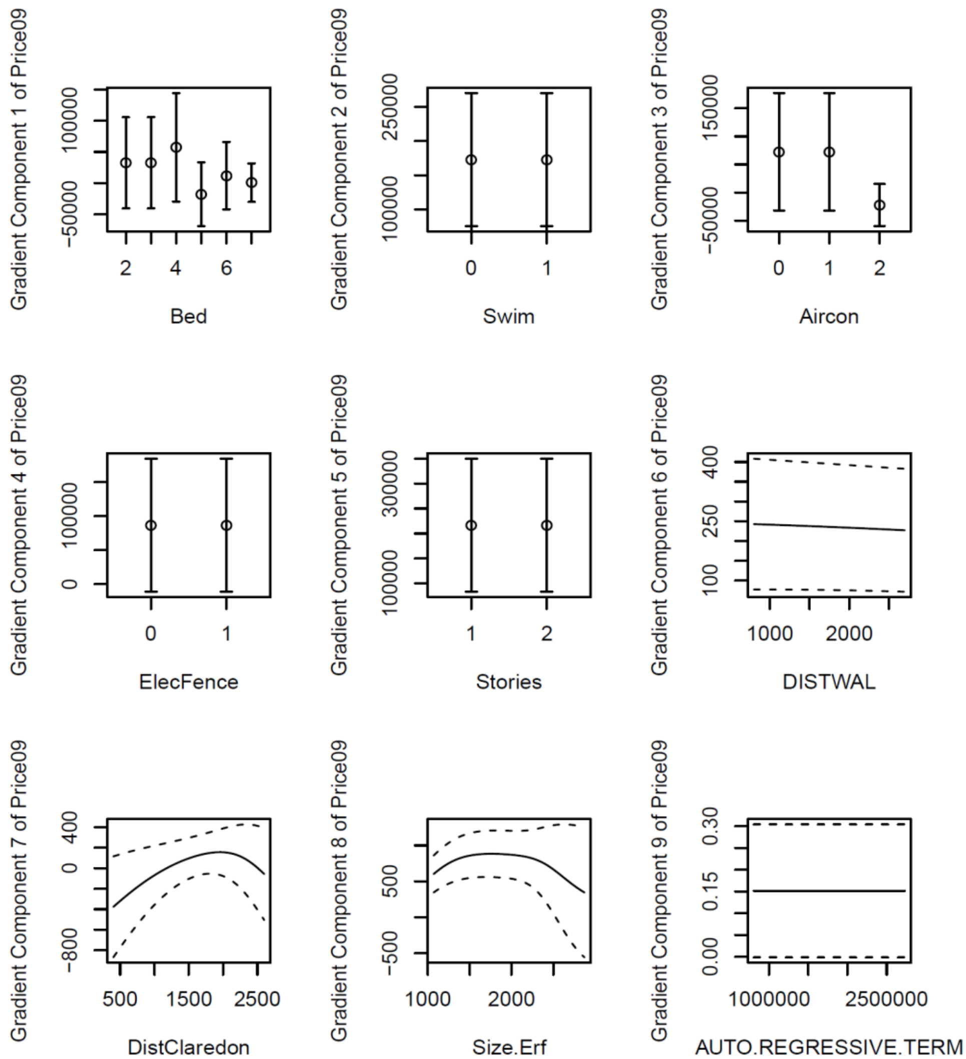
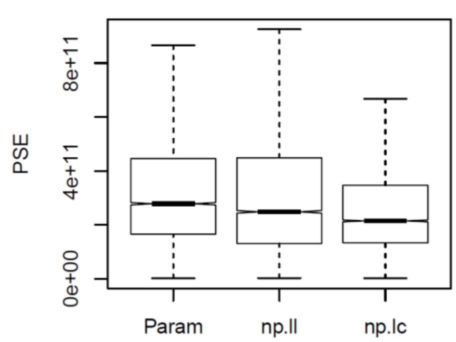


Figure 5: Box plot of PSEs



Appendix A: Model Summaries

Table 2: Local Constant Model Summary

```
Regression Data (170 observations, 12 variable(s)):

Regression Type: Local-Constant
Bandwidth Selection Method: Least Squares Cross-Validation
Formula: Price09 ~ Bath + Bed + Swim + Aircon + GARAGEDUMMY + ElecFence +
Stories + DISTWAL + DistClaredon + Size.Erf + Age + AUTO.REGRESSIVE.TERM
Bandwidth Type: Fixed
Objective Function Value: 2.765e+11 (achieved on multistart 9992)

Exp. Var. Name: Bath                 Bandwidth: 1     Lambda Max: 1
Exp. Var. Name: Bed                  Bandwidth: 0.05726    Lambda Max: 1
Exp. Var. Name: Swim                 Bandwidth: 0.2749     Lambda Max: 1
Exp. Var. Name: Aircon               Bandwidth: 0.0009241  Lambda Max: 1
Exp. Var. Name: GARAGEDUMMY          Bandwidth: 1     Lambda Max: 1
Exp. Var. Name: ElecFence            Bandwidth: 0.03791    Lambda Max: 1
Exp. Var. Name: Stories              Bandwidth: 0.004072   Lambda Max: 1
Exp. Var. Name: DISTWAL              Bandwidth: 223  Scale Factor: 0.6586
Exp. Var. Name: DistClaredon         Bandwidth: 209.3      Scale Factor: 0.5551
Exp. Var. Name: Size.Erf             Bandwidth: 107.8      Scale Factor: 0.3516
Exp. Var. Name: Age                  Bandwidth: 44844766  Scale Factor: 7135978
Exp. Var. Name: AUTO.REGRESSIVE.TERM Bandwidth: 480685  Scale Factor: 1.351

Continuous Kernel Type: Second-Order Gaussian
No. Continuous Explanatory Vars.: 5

Ordered Categorical Kernel Type: Wang and Van Ryzin
No. Ordered Categorical Explanatory Vars.: 7

Regression Data: 170 training points, in 12 variable(s)
Bath                    Bed  Swim  Aircon GARAGEDUMMY ElecFence  Stories
Bandwidth(s):       1 0.05726 0.2749 0.0009241 1  0.03791 0.004072
DISTWAL DistClaredon Size.Erf      Age AUTO.REGRESSIVE.TERM Bandwidth(s):
                    223     209.3  107.8 44844766        480685

Kernel Regression Estimator: Local-Constant
Bandwidth Type: Fixed
Residual standard error: 63837
R-squared: 0.9932

Continuous Kernel Type: Second-Order Gaussian
No. Continuous Explanatory Vars.: 5

Ordered Categorical Kernel Type: Wang and Van Ryzin
No. Ordered Categorical Explanatory Vars.: 7
```

Table 3: Local Linear Model Summary

```
Regression Data (170 observations, 12 variable(s)):

Regression Type: Local-Linear
Bandwidth Selection Method: Least Squares Cross-Validation
Formula: Price09 ~ Bath + Bed + Swim + Aircon + GARAGEDUMMY + ElecFence +
Stories + DISTWAL + DistClaredon + Size.Erf + Age + AUTO.REGRESSIVE.TERM
Bandwidth Type: Fixed
Objective Function Value: 2.984e+11 (achieved on multistart 6110)

      Exp. Var. Name: Bath    Bandwidth: 0.5042 Lambda Max: 1
      Exp. Var. Name: Bed     Bandwidth: 1      Lambda Max: 1
      Exp. Var. Name: Swim    Bandwidth:        Lambda Max: 1
      Exp. Var. Name: Aircon  Bandwidth: 0.4159 Lambda Max: 1
      Exp.     Var.      Name:Bandwidth: 0.04359Lambda Max: 1
      Exp.     Var.      Name:Bandwidth: 1      Lambda Max: 1
      Exp. Var. Name: Stories Bandwidth: 1      Lambda Max: 1
      Exp. Var. Name: DISTWAL Bandwidth: 745.2  Scale    Factor:
      Exp.     Var.      Name:Bandwidth: 550.3  Scale    Factor:
      Exp. Var. Name: Size.ErfBandwidth: 523.3  Scale    Factor:
      Exp. Var. Name: Age     Bandwidth:        Scale    Factor:
Exp. Var. Name: AUTO.REGRESSIVE.TERM Bandwidth: 6.534e+12 Scale Factor: 18369680

Continuous Kernel Type: Second-Order Gaussian
No. Continuous Explanatory Vars.: 5

Ordered Categorical Kernel Type: Wang and Van Ryzin
No. Ordered Categorical Explanatory Vars.: 7

Regression Data: 170 training points, in 12 variable(s)
Bath Bed                    Swim Aircon GARAGEDUMMY ElecFence Stories
Bandwidth(s): 0.5042    1 0.007697 0.4159          0.04359    1     1
DISTWAL DistClaredon Size.Erf      Age AUTO.REGRESSIVE.TERMBandwidth(s):
                745.2    550.3  523.3 14115267     6.534e+12

Kernel Regression Estimator: Local-Linear
Bandwidth Type: Fixed
Residual standard error: 176909
R-squared: 0.9558

Continuous Kernel Type: Second-Order Gaussian
No. Continuous Explanatory Vars.: 5

Ordered Categorical Kernel Type: Wang and Van Ryzin
No. Ordered Categorical Explanatory Vars.: 7
```

Table 4: Local Constant Model (Subset) Summary

```
Regression Data (170 observations, 9 variable(s)):

Regression Type: Local-Constant
Bandwidth Selection Method: Least Squares Cross-Validation
Formula: Price09 ~ Bed + Swim + Aircon + ElecFence + Stories +
DISTWAL + DistClaredon + Size.Erf + AUTO.REGRESSIVE.TERM
Bandwidth Type: Fixed
Objective Function Value: 2.741e+11 (achieved on multistart 5553)

Exp. Var. Name: Bed                  Bandwidth: 0.09719    Lambda Max: 1
Exp. Var. Name: Swim                 Bandwidth: 0.3697     Lambda Max: 1
Exp. Var. Name: Aircon               Bandwidth: 0.004051   Lambda Max: 1
Exp. Var. Name: ElecFence            Bandwidth: 0.05954    Lambda Max: 1
Exp. Var. Name: Stories              Bandwidth: 0.012Lambda Max: 1
Exp. Var. Name: DISTWAL              Bandwidth: 240 Scale Factor: 0.7613
Exp. Var. Name: DistClaredon         Bandwidth: 234.8      Scale Factor: 0.6688
Exp. Var. Name: Size.Erf             Bandwidth: 118 Scale Factor: 0.4134
Exp. Var. Name: AUTO.REGRESSIVE.TERM Bandwidth: 538209   Scale Factor: 1.625

Continuous Kernel Type: Second-Order Gaussian
No. Continuous Explanatory Vars.: 4

Ordered Categorical Kernel Type: Wang and Van Ryzin
No. Ordered Categorical Explanatory Vars.: 5

Regression Data: 170 training points, in 9 variable(s)
Bed                         Swim Aircon ElecFence Stories DISTWAL
Bandwidth(s): 0.09719 0.3697 0.004051       0.05954 0.012 240
DistClaredon Size.Erf AUTO.REGRESSIVE.TERM
Bandwidth(s):          234.8    118           538209

Kernel Regression Estimator: Local-Constant
Bandwidth Type: Fixed
Residual standard error: 97542
R-squared: 0.9842

Continuous Kernel Type: Second-Order Gaussian
No. Continuous Explanatory Vars.: 4

Ordered Categorical Kernel Type: Wang and Van Ryzin
No. Ordered Categorical Explanatory Vars.: 5
```

Table 5: Local Linear Model (Subset) Summary

```
Regression Data (170 observations, 9 variable(s)):

Regression Type: Local-Linear
Bandwidth Selection Method: Least Squares Cross-Validation
Formula: Price09 ~ Bed + Swim + Aircon + ElecFence + Stories +
DISTWAL + DistClaredon + Size.Erf + AUTO.REGRESSIVE.TERM
Bandwidth Type: Fixed
Objective Function Value: 3.351e+11 (achieved on multistart 8820)

   Exp. Var. Name: Bed        Bandwidth:        Lambda Max: 1
   Exp. Var. Name: Swim       Bandwidth:        Lambda Max: 1
   Exp. Var. Name: Aircon     Bandwidth:        Lambda Max: 1
   Exp.     Var.     Name:Bandwidth: 0.82      Lambda Max: 1
   Exp. Var. Name: Stories    Bandwidth:        Lambda Max: 1
   Exp. Var. Name: DISTWAL    Bandwidth: 1958   Scale   Factor:
   Exp.     Var.     Name:Bandwidth: 563       Scale   Factor:
   Exp. Var. Name: Size.ErfBandwidth: 474      Scale   Factor:
   Exp.  Var.  Name:  AUTO.REGRESSIVE.TERM  Bandwidth:  1.1e+10  Scale  Factor:
   33220

Continuous Kernel Type: Second-Order Gaussian
No. Continuous Explanatory Vars.: 4

Ordered Categorical Kernel Type: Wang and Van Ryzin
No. Ordered Categorical Explanatory Vars.: 5

Regression Data: 170 training points, in 9 variable(s)
Bed                     Swim Aircon ElecFence Stories DISTWAL DistClaredon
Bandwidth(s): 0.5552 0.4361 0.3808  0.82  0.4245          1958563
Size.Erf AUTO.REGRESSIVE.TERMBandwidth(s):  474
                   1.1e+10

Kernel Regression Estimator: Local-Linear
Bandwidth Type: Fixed
Residual standard error: 297367
R-squared: 0.874


Continuous Kernel Type: Second-Order Gaussian
No. Continuous Explanatory Vars.: 4

Ordered Categorical Kernel Type: Wang and Van Ryzin
No. Ordered Categorical Explanatory Vars.: 5
```