VGTU LEIDYKLA TECHNIKA

PREDICTING URBAN LAND PRICES: A COMPARISON OF FOUR APPROACHES

Marko HANNONEN

Institute of Real Estate Studies, Department of Surveying, Helsinki University of Technology, Espoo, P.O. Box 1200, FIN-02015 HUT, Finland
E-mail: marko.hannonen@pp.inet.fi; Tel: +358 05 596 6065; Telefax: +358 9 465 077

Received 23 May 2008; accepted 9 July 2008

ABSTRACT. This paper investigates forecasting accuracy of four different hedonic approaches, when vacant urban land prices are predicted in local markets. The investigated hedonic approaches are: 1) ordinary least squares estimation, 2) robust MM-estimation, 3) structural time series estimation and 4) robust local regression. Post-sample predictive testing indicated that more accurate predictions are obtained if the unorthodox methods of this paper are used instead of the conventional least squares estimation. In particular, the predictive unbiassness can significantly be improved when using the unconventional hedonic methods of the study. The paper also studied the structure of urban land prices. The most important attribute variables in explaining land prices were permitted building volume, house price index, northing and easting. The influence of parcel size variable and different indicator variables on land prices were much weaker.

KEYWORDS: Land price; Hedonic model; Prediction; Robustness; Flexibility

1. INTRODUCTION

Hedonic methods are often advocated in complex land valuation assignments in order to objectively minimise the systematic valuation error and in order to produce the necessary quality-adjustments, which stem from the differentiated nature of separate land parcels, validly and reliably. However, the use of hedonic models is plagued with some fundamental problems imposing serious threats to their empirical adequacy. These fundamental dilemmas include: (1) the temporal variability of land prices, (2) the spatial variability of land prices, (3) the model specification dilemma and (4) outlying and influential observations.

When investigating the temporal dimension of land prices it is important to understand that the behaviour of land prices is generally nonstationary. This is a typical characteristic of many economic time series, which means that the data-generating process that produces the observables is itself transient in time. The effect of time is also multidimensional: Often we can legitimately separate from each other the price trend, the price cycle, seasonal variation and random variation. Traditionally, when modelling temporal land price movements, the effect of time has been tried to reduce to the variation of cost-of-living index or house price index, which have subsequently been used as explanatory variables in a hedonic regression. Also the indicator variable technique (i.e. by

International Journal of Strategic Property Management
ISSN 1648-715X print / ISSN 1648-9179 online
http://www.ijspm.vgtu.lt
DOI: 10.3846/1648-715X.2008.12.217-236

using yearly time dummy variables) has been a very popular approach when analysing the temporal dimension of land prices. These approaches contain problems mainly because the influence of time can only be estimated in a manner, which is not very accurate in practice. Structural time series models, on the other hand, usually provide a more accurate description about temporal movements.

The spatial variation of land prices can be divided to the spatial heterogeneity and spatial dependency. Spatial heterogeneity implies that functional forms and parameters vary with location and are not homogeneous throughout the data set, whereas spatial dependence implies that the variation is a function of distance. The spatial dependency problem can usually be solved by including location or some distance variables into a hedonic regression as explanatory variables. The spatial heterogeneity problem is usually more problematic: One natural solution would be to narrow the analyses into reasonably small submarkets, which homogenises the data. However, in practise this operation is not typically feasible due to the scarcity of observations for the hedonic modelling purposes. Adaptive modelling techniques, such as local regression, usually provide a better solution to the spatial heterogeneity problem in that they possess a spatial adaptation property and thus explicitly address the spatial heterogeneity problem.

The model specification dilemma can be solved by three different ways: (1) parametrically, (2) semiparametrically and (3) nonparametrically. Parametric modelling is the classical approach in the hedonic modelling of land prices, which is theory-laden because pre-specified functional forms are used in the analysis. Nonparametric techniques are on the other hand data-driven, very flexible tools and semiparametric techniques combine features from parametric and nonparametric approaches. The exact research problem determines what approach should be used. Generally, nonparametric methods are useful when associations between variables are complex (i.e. highly nonlinear) and theoretically unknown. Parametric models apply well to a less complex setting where there exists valid prior knowledge about model's functional form. Irrespective of a chosen approach the model specification dilemma contains the choice of a hedonic model's functional form, the selection of relevant study variables and an error distribution assumption. And it should be noted that the result depends on the chosen scale, which is often, however, implicit.

Parametric models that represent data modelling culture (Breiman, 2001) have formed the conventional dogma of hedonic pricing methods in land price studies, where prespecified global models are estimated by means of ordinarily least squares or some modification thereof. Benefits of parametric approaches undeniably include: simplicity, interpretability, parsimony and comprehensive statistical theory. The fundamental obstacle, however, under-lying the general use of parametric models is their inflexibility, i.e. inability to learn genuine structure about the hedonic relationship from the evidence in such decision-making settings, where theoretically unknown nonlinearity is expected. This is the typical case when the effects of variables representing location and time are considered (McMillen and Thorsnes, 2003). The conventional result is that even the best parametric model tends to impose restrictions that substantially reduce the explanatory and predictive power of hedonic equation (Pace, 1993 and 1995; Anglin and Gencay, 1996; inter alia). Unless the theory-laden parametric model coincides with the data-generating process, profound mis-specification errors may result imposing serious threats to their empirical validity.

Semiparametric and nonparametric approaches are representative of algorithmic modelling culture (Breiman, 2001) that emphasise aspects of learning the complex structure

from the available facts and adaptability to the features underlying the data. Semiparametric estimators are, more precisely, an intermediate strategy between theory-laden and data-driven estimators that have restricted learning ability, i.e. semiparametric estimators can approximate functions only within some prespecified classes. Their practical relevance is mainly in balancing the dual goals of low specification error and high efficiency (Pace, 1995; Anglin and Gencay, 1996) and in enchaining the interpretability of results. Nonparametric estimators are by their nature highly flexible and, thus, capable of approximating very general classes of functions (e.g. smooth functions, square integrable functions) that does not require any restrictive, unwarranted prespecification of the functional form of mean response function (nor any specific error distribution assumption). This renders nonparametric estimators to be powerful data-driven tools, albeit highly sensitive to the problem of undersmoothing or overfitting, if local estimation is implemented unduly.

Outlying and influential observations are very common in the land value studies, which may be genuine, faultless values, generated under conditions of some untypical factors or they can contain different errors (such as recording and measurement error; wrong population, etc.). Traditional hedonic modelling techniques, especially the ordinary least squares technique, are sensitive to outlying observations; even a single outlier can drastically change the results and misguide the inferences. In fact, a single sufficiently deviating data point can cause that the least squares estimator breaks down and generates results that are utterly unreliable and uninformative. Robust methods such as MM-estimation, on the contrary, are not sensitive to outliers or influential observations and, therefore, can tolerate a certain amount of bad observations without the fear that the estimator breaks down and produces completely useless results.

2. THE RESEARCH PROBLEM

In this study four different hedonic modelling approaches are empirically compared together when urban land prices are modelled in a local market of Espoo, Finland. The fits are analysed and post-sample predictions are calculated across different modelling schemes. The main research question is: "Which approach produces the most accurate post-sample predictions with the given vacant urban land price data?" The forecasting accuracy is perhaps the most important operation criterion, which determines the much of the utility of the corresponding hedonic model. Generally, for any valuation method to have a sufficient degree of validity it must produce an accurate prediction of the most probable market price of a land parcel.

The four different hedonic approaches that are investigated in this paper consist of:

1) Ordinary least squares estimation.
2) Robust MM-estimation.
3) Structural time series estimation.
4) Robust local regression.

Ordinary least squares estimation and robust MM-estimation represent parametric hedonic methods, structural time series estimation is a semiparametric hedonic method and robust local regression is a nonparametric hedonic method.

Post-sample predictions are analysed using six different predictive accuracy indicators, which are: 1) mean prediction error, 2) mean absolute percentage error, 3) mean absolute error, 4) root mean squared error, 5) correlation coefficient and 6) gravity.

3. PREVIOUS RELATED RESEARCH[1]

Most of the hedonic modelling studies in land markets have been based on the ordinary least squares estimation, yet some nonparametric and semiparametric estimation techniques have also been applied. However, none of the hedonic studies has been focused

on the issue of land price prediction, which is the main focus in this paper.

Shimizu and Nishimura (2007) estimated using ordinary least squares hedonic price equations of commercial and residential land prices in Tokyo for a 25-year period (from 1975 to 1999) and investigated possible structural changes in these price equations. They find that the price structure differed significantly among locations reflecting differences in supplier pricing and end-user preferences. They also found significant structural changes in the underlying price structure, identifying pre-bubble, bubble and post-bubble periods.

Colwell and Munneke (2003) examined urban land prices within a nonparametric framework using piecewise parabolic regression with specific interest in the land price gradient with respect to distance from the inner city. They concluded that the piecewise parabolic regression is an amazingly flexible technique, which can be used to represent very complex land value functions.

Clapp et al. (2001) estimated a hedonic price index equation to determine the value of land under residential structures in Fairfax county, Virginia at various points in time over the 1975 to 1992 time frame. A set of three simultaneous equations explained land value together with changes in population density and the percentage working at home. The method of estimation was ordinary least squares. They stressed the importance of dealing with the double simultaneity issue and found that the land-value surface has changed dramatically over time.

Lin and Evans (2000) investigated the relationship between the price of land and size of plot when plots were small. They used a land price data from the city of Taipei, Taiwan. They found that the price of land per unit of area increases with lot size.

Colwell and Munneke (1999) investigated spatial dimension to the concavity in the total price and parcel size relationship, when the dataset consisted of sales of vacant residential, commercial and industrial land in Cook County, Illinois during the time period of 1986 to 1993. They used ordinary least squares and found that concavity is higher in the rest of Cook County than in the CBD for all three land-use types.

Thorsnes and McMillen (1998) used a semiparametric estimator to analyse the relationship between land values and parcel size in the Portland, Oregon, metropolitan area. The value-size relationship was estimated nonparametrically and a simple log-linear parametric relationship was assumed for the rest of the model. They found that ordinary least squares and semiparametric estimates imply similar results: There was a concave value-size relationship meaning that subdivision costs cause large parcels to trade at a discount.

Colwell (1998) investigated a nonparametric method, a piecewise parabolic regression analysis, for estimating spatial land price functions in the Chicago CBD. The independent variables were barycentric coordinates that uniquely described the location of observations in space. Colwell found that this nonparametric method goes a long toward solving the problem of the spatial correlation of residuals, which affects most hedonic models.

Atack and Margo (1998) investigated using ordinary least squares a simple monocentric urban model of the price of vacant land in Manhattan in the time frame of 1835 to 1900. They also found that vacant land in Manhattan was price elastic with respect to distance from the CBD in 1845 but becomes price inelastic in the post-Civil War period.

Colwell and Munneke (1997) studied the structure of urban land prices in Chicago using data from the sales of commercial, residential and industrial land during the time period of 1986 to 1992. The method of estimation was ordinary least squares (and multinomial logit estimation to control for possible sample selection bias).They found evidence that land prices are non-linear in nature and that land prices are concave in parcel size.

McMillen (1996) analysed locally weighted regression in modelling land prices in Chicago using two different data sets from 1836 to 1990. Two parametric models were estimated: a simple monocentric model and a more flexible spatial expansion model. These fits were compared to local linear regression estimates, which locally estimated the spatial expansion model. McMillen also demonstrated that local regression is useful for both prediction as well as testing hypothesises in land markets. McMillen summarised: "Locally weighted regression is a useful tool for spatial modelling. Nonlinearity is handled directly and simply".

4. HEDONIC METHODS OF THE STUDY

4.1. Ordinary least squares estimation

Ordinary least squares estimation is by far the most applied hedonic method in practice. This is a parametric estimator where the form of hedonic function is specified before seeing the data. The only aspects that are determined from the data are the hedonic prices of different attribute variables.

The conventional hedonic regression approach that is based on ordinary least squares is an appropriate modelling context, strictly speaking, if the interest solely focuses on the cross-sectional variation of the hedonic prices and if the problem due to spatial heterogeneity can adequately be addressed. When temporal aspects are analysed with the ordinary least squares estimator, several problems are encountered. According to Scwann (1998) the core problem in local markets is the lack of sufficient degrees of freedom, since estimation involves an extensive set of time-indexed dummy variables along with other regressors, at least one for each time period. Even if the locality of the markets imposes no dilemma, the major weakness of these methods remains: parameters values in one period do not affect the values of parameters in other periods (Francke and Vos, 2004).

Some nonlinear features can be accounted by the ordinary least squares estimator e.g. by using the double-log model specification. However, many nonlinear features are in any case omitted if the orthodox least squares estimator is applied. As a result, the ordinary least squares estimation produces only a coarse description about the actual dependencies between the regressand and the regressors. Whether this approximation is satisfactory in practise depends largely on the predictive accuracy of the estimated hedonic model.

The time element is estimated by the OLS using house-price index measure. The main idea is then that the temporal variability can be reduced to that of the variability in that index. Time indexed dummy variables were not used, on the one hand, because of high collinearity between the different indicators and, on the other hand, because the use of house price index variable tend to produce a better approximation to temporal movements than by simply using time indicators.

4.2. Robust MM-estimation

The aim of robust statistics is to investigate the behaviour of estimators, when the basic modelling assumptions (linearity, normality, independence, etc.) are not exactly valid but are at most approximations to reality. To put it slightly differently, the basic aims of robust statistics are (Hampel et al., 1986, p. 11):

- To describe a structure best fitting the bulk of the data.
- To identify deviating data points (outliers) or deviating substructures for further treatment.
- To identify and give a warning about highly influential data points (leverage points).
- To deal with unspecified serial correlations, or more generally, with deviations from the assumed correlation structures.

In practice the approximate nature of hedonic models is largely result of the occurrence of gross errors, the empirical character of models and only partial validity of theoretical model-ling assumptions. In general, the hedonic model can be considered as robust if:

- It is reasonably unbiased and efficient.
- Small deviations from the hedonic model assumptions will not substantially impair the performance of the hedonic model.
- Somewhat larger deviations will not invalidate the hedonic model completely.

In this study a very fault tolerant and computationally intensive method, the three-stage MM-estimation, is analysed in the hedonic modelling of land prices. This estimator is parametric in nature, i.e. the model structure is fixed in advance. In the first phase of the MM-estimation is calculated a regression estimate, which is consistent and have a high break-down point, but is not necessarily efficient. In the second phase the scale of errors is estimated, which is based on the residuals of the first phase. In the third phase is calculated the M-estimate of the hedonic prices. The breakdown point the MM-estimator is the highest, i.e. 50% of the data can be corrupted before the estimator provides useless results.

The computational algorithm used to derive the hedonic prices is a variant of iterative re-weighted least squares, which is applied in the M-estimation. Iterative solution is needed because weights depend on residuals, residuals depend on estimated hedonic prices and hedonic prices depend on weights. Lets assume that we have an initial estimate of hedonic prices, $\hat{\beta}_0$, and its deviation measure s^2. Lets define the weights:

$$w_i(\beta) = \frac{\psi_1\left(r_i(\beta)/s\right)}{r_i(\beta)/s}, \tag{1}$$

where ψ_1 is double-weighted objective function, $r_i(\beta)$ are residuals and s is a measure of scale. Then lets define:

$$\mathbf{g}(\beta) = \frac{1}{s^2}\sum_{i=1}^{n} w_i(\beta) r_i(\beta)\mathbf{x}_i = \frac{1}{s}\sum_{i=1}^{n} \psi_1\left(r_i(\beta)/s\right)\mathbf{x}_i, \tag{2}$$

and

$$\mathbf{M}(\beta) = \frac{1}{s^2}\sum_{i=1}^{n} w_i(\beta)\mathbf{x}_i\mathbf{x}_i', \tag{3}$$

where – g(β) is the gradient of residuals sum of squares. Now the applied iterative formula for the derivation of hedonic prices can be written:

$$\beta_{j+1} = \beta_j + \frac{1}{2^k}\Delta(\beta_j), \tag{4}$$

where $\Delta(\beta) = \mathbf{M}^{-1}(\beta)\mathbf{g}(\beta)$. The integer k is chosen so that the left side of the inequality:

$$S\left(\beta_j + \frac{1}{2^k}\Delta(\beta_j)\right) \leq S(\beta_j) - \delta\left(\frac{1}{2^k}\Delta(\beta_j)\right)' \mathbf{g}(\beta_j) \tag{5}$$

is minimised and $0 < \delta < 1$.

The time element is estimated by the MM-method using house-price index measure.

4.3. Structural time series models

The variation of observed land prices is a combination of cross-sectional and time-series variations (Schulz, 2003, p. 58). Besides the spatial characteristics, the selling date is an important attribute in explaining the evolution of market prices through the flux of time which itself is directly an unobservable quantity, i.e. time is a latent variable. What we can observe are different states that occur in a predefined submarket and changes that they cause in prices in that market area. (Francke and Vos, 2004)

The time-series or temporal variation is a result of changing market conditions, which

are driven by, among others, changes in consumers' preferences, investors' expectations and technological advantages. The temporal variation can be understood as representing that part of price variation that is more or less common to all parcels of land in the same submarket (Schulz and Werwatz, 2004). An empirical model of land prices has to recognize these two different, yet closely related sources of variations.

Given the special characteristics of land markets one natural solution to the dual problem of hedonic modelling caused by spatio-temporal variation is combine the flexibility of a time series model with that of the interpretation of a regression. This is the underlying rationale in the structural time series approach: the observations are directly made up of trend, cycle, seasonal, and regression components plus error. In essence, structural time series models can be thought of as regression models in which explanatory variables are functions of time and the parameters are time-varying (Harvey, 1997).

Structural time series methods can also be understood as semiparametric estimators that combine many of the benefits of parametric and nonparametric estimators; temporal variability of land prices is estimated in a nonparametric fashion, which permits the effect of time to be linear, convex and concave in different regions, whereas the hedonic prices of attribute variables are estimated in a parametric manner.

When considering the determination of hedonic prices in land markets and, specifically the temporal dimension, there are several benefits in using the structural time series approach and the associated state space form as compared to the Box-Jenkins ARIMA methodology. These include (Harvey and Shephard, 1993; Harvey, 1997; Durbin and Koopman, 2002, p. 51–53):

- Structural analysis of the problem. Different components that make up the series, including the regression elements, are modelled explicitly when, in contrast, the Box-Jenkins approach is a sort of "black box". A structural model provides not only the forecasts of the series but also presents a set of stylised facts. Also a structural model can be handled within a unified statistical framework that produces optimal estimates with well-defined properties.
- Management of nonstationarity. In a structural model nonstationarity can be handled conveniently by unobserved components without the need of differencing any variables. By comparison, in the Box-Jenkins approach the stationary is assumed, and nonstationary components of the series are usually eliminated by differencing the variables, which results to a potential loss of valuable long-term information. Furthermore, the standard unobserved component models are simple, yet effective, leading to parsimonious representations for the systems.
- Generality. Multivariate observations can easily be handled with structural models, which cover as special cases a wide range of econometric models (including all ARIMA models). Explanatory variables can be introduced into the model structure and the associated regression coefficients (hedonic prices) can be permitted to vary stochastically over time if needed. Different kinds of intervention variables can be specified and lagged values of dependent as well as explanatory variables can be incorporated to a model. Missing observations and varying dimensionality of observations are issues that are straightforward to deal with structural models.

In this study the local level model or the random walk plus noise model is used to capture the underlying trend in the series. The local level model is the simplest, yet effective, structural trend model, which regards an observation on land price p_t at time t as being

made up of an underlying level μ_t and an irregular disturbance ε_t (Koopman et al., 1999; Durbin and Koopman, 2002, p. 44–45):

$$p_t = \mu_t + \varepsilon_t, \quad \{\varepsilon_t\} \sim NID(0, \sigma_\varepsilon^2),$$
$$\mu_t = \mu_{t-1} + \eta_t, \{\eta_t\} \sim NID(0, \sigma_\eta^2). \quad (6)$$

The underlying level μ_t is not directly observable. It is generated by a random walk, i.e. the level term in the current period is equal to the level term in the previous period plus a level disturbance term η_t. The effect of η_t is to allow the level of the trend to shift up and down. It is generally assumed that the level and irregular disturbances are mutually independent and independent of μ_0. The signal-to-noise ratio $q = \sigma_\eta^2 / \sigma_\varepsilon^2$ plays a vital role in determining how observations should be weighted for prediction and smoothing. Basically the higher q is, the greater is the discounting of past observations. The reduced form of local level model is ARIMA(0,1,1) with certain restrictions on the parameter space.

Cycles are characteristic to many economic time series as economy goes from boom to recession and back again. These can be modelled in different ways, but in this study cycles are effectively presented as a mixture of sine and cosine waves with two parameters θ_1 and θ_2. If Ψ_t is a cyclical function of time with frequency λ_c that is measured in radians, then (Harvey and Shephard, 1993):

$$\psi_t = \theta_1 \cos\lambda_c t + \theta_2 \sin\lambda_c t, \quad (7)$$

where the period of the cycle is $2\pi/\lambda_c$, $\sqrt{\theta_1^2 + \theta_2^2}$ is the amplitude and $\tan^{-1}\left(\theta_2/\theta_1\right)$ is the phase. A stochastic cycle can be constructed recursively:

$$\begin{pmatrix} \psi_t \\ \psi_t' \end{pmatrix} = \rho \begin{pmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}' \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t' \end{pmatrix}, \quad (8)$$

where κ_t and κ'_t are mutually uncorrelated with a common variance σ_κ^2. $\rho \in [0,1]$ is a damping factor. Stationary models correspond to situations where ρ is strictly less than one. A first-order autoregressive process is an important limiting case of a stochastic cycle when a frequency λ_c is equal to 0 or π.

The calculations of unobserved components and hedonic prices are done by using the Kalman filtering and smoothing recursions. These can be expressed as:

$$\hat{\alpha}_{t|t-1} = \Pi_t \hat{\alpha}_{t-1} + \mathbf{W}_t\beta, \quad (9)$$

$$\mathbf{S}_{t|t-1} = \Pi_t \mathbf{S}_{t-1} \Pi_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t', \quad (10)$$

$$\mathbf{v}_t = \mathbf{p}_t - \mathbf{Z}_t \hat{\alpha}_{t|t-1} - \mathbf{X}_t\beta, \quad (11)$$

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{S}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t, \quad (12)$$

$$\hat{\alpha}_t = \hat{\alpha}_{t|t-1} + \mathbf{S}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} (\mathbf{p}_t - \mathbf{Z}_t \hat{\alpha}_{t|t-1} - \mathbf{W}_t\beta), \quad (13)$$

$$\mathbf{S}_t = \mathbf{S}_{t|t-1} - \mathbf{S}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{S}_{t|t-1}, \quad (14)$$

where $\mathbf{p}_t$ is a $N \times 1$ vector of observed land prices at time t, α_t is a $m \times 1$ state vector and β is a $(p + k) \times 1$ vector of unknown regression coefficients that are assumed to be constant[3]. $\mathbf{Z}_t$ is a non-stochastic $N \times m$ matrix of cycle and trend components, $\mathbf{X}_t$ is a non-stochastic $N \times (p + k)$ matrix of observations on explanatory variables and ε_t is a $N \times 1$ vector of serially uncorrelated measurement errors with zero mean and covariance matrix $\mathbf{H}_t$, i.e. $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \mathbf{H}_t$. Now $\mathbf{\Pi}_t$ is a $m \times m$ state transfer matrix, $\mathbf{W}_t$ is a $m \times (p + k)$ matrix, $\mathbf{R}_t$ is a $m \times g$ matrix and η_t is a $g \times 1$ vector of serially uncorrelated error terms with mean zero and covariance matrix $\mathbf{Q}_t$, i.e. $E(\eta_t) = 0$ and $Var(\eta_t) = \mathbf{Q}_t$. The estimator of $\mathbf{p}_t$ is $\hat{\mathbf{p}}_{t|t-1} = \mathbf{Z}_t \hat{\alpha}_{t|t-1} + \mathbf{W}_t\beta$. $\mathbf{v}_t$ are one-step ahead prediction errors called innovations, which represents that part of the $\mathbf{p}_t$ that cannot be predicted from the past. $\mathbf{F}_t$ is the conditional variance of the prediction error.

The basic Kalman filtering and smoothing recursions, which are described in the formulas 9-14, are supplemented in this research by a set of complementary vector and matrix re-

cursions, because non-stationary components and fixed regression effects are present. This is called an augmented Kalman filter (Koopman et al., 1999; Durbin and Koopman, 2002, p. 115–120), which is described by the equations:

$$\mathbf{V}_t = -\mathbf{Z}\,\mathbf{A}_{t|t-1} - \mathbf{X}_t\mathbf{B}, \tag{15}$$

$$\mathbf{A}_{t+1|t} = \Pi_t\mathbf{A}_{t|t-1} + \mathbf{W}_t\mathbf{B} + \mathbf{K}_t\mathbf{V}_t, \tag{16}$$

$$(\mathbf{m}_t, \mathbf{M}_t) = (\mathbf{m}_{t-1}, \mathbf{M}_{t-1}) + \mathbf{V}_t'\mathbf{F}_t^{-1}(\mathbf{v}_t, \mathbf{V}_t) \tag{17}$$

with $\mathbf{A}_{1|0} = \mathbf{W}_0\mathbf{B}$ and $\mathbf{B} = (\mathbf{B}_x, \mathbf{B}_i)$ is a square selection matrix of zeros and ones and the subscripts x, i are related to regression and initial effects, respectively. The number of columns of $\mathbf{V}_t$ and $\mathbf{A}_{t+1|t}$ is the same as in the matrix $\mathbf{B}$. $\mathbf{K}_t = \Pi_t\mathbf{S}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}$ is the so-called Kalman gain. Now the one-step ahead prediction of the state vector and the associated mean square error matrix are given by:

$$\hat{\alpha}^*_{t|t-1} = \hat{\alpha}_{t|t-1} + \mathbf{A}_{t|t-1}\mathbf{M}_{t-1}^{-1}\mathbf{m}_{t-1}, \tag{18}$$

$$\mathbf{S}^*_{t|t-1} = \mathbf{S}_{t|t-1} + \mathbf{A}_{t|t-1}\mathbf{M}_{t-1}^{-1}\mathbf{A}'_{t|t-1}. \tag{19}$$

The one-step ahead prediction errors and the associated mean square error matrix are given by:

$$\mathbf{v}_t^* = \mathbf{v}_t + \mathbf{V}_t\mathbf{M}_{t-1}^{-1}\mathbf{m}_{t-1}, \tag{20}$$

$$\mathbf{F}_t^* = \mathbf{F}_t + \mathbf{V}_t\mathbf{M}_{t-1}^{-1}\mathbf{V}_t'. \tag{21}$$

The matrix inversions for $\mathbf{M}_t$ can be evaluated in a manner similar to recursive regressions (de Jong, 1991).

4.4. Robust local regression[4]

Much of the aim of applied hedonic analysis is to produce a reasonable approximation to the generally unknown mean response function. The primary implication of the theoretical literature concerning hedonic prices in the real estate markets is that hedonic relationships are expected to be highly nonlinear due to their locational uniqueness that induces spational heterogeneity of regression surfaces (Wallace, 1996; McMillen and Thorsnes, 2003) that cannot be, in general, specified a priori (Anglin and Gencay, 1996). Nonlinearity indicates locally changing degrees of curvature in the hedonic function with non-constant characteristic values.

Nonlinearity is a fundamental feature that characterise processes in the real estate markets imposing serious threats to empirical validity of hedonic models that in current practise are predominantly used. The complex question of validity underlying hedonic model specification can be divided into three subproblems that involve determining (Pace, 1993 and 1995; Wallace, 1996):

(1) the relevant set of response and attribute variables;
(2) the appropriate functional form between these variables;
(3) the adequate error distribution for inference.

Economic theory and past experience usually provide useful a priori information of what variables should enter the model structure that substantially reduce the threat of omitted variable bias. Phenomena in real estate markets are, however, strongly dependent on the particular submarket, time period and property type and, as a consequence, the selection of proper set of dependent and conditioning variables is partially an empirical question, too.

Economic theory or previous experience rarely provides any specific, valid guidance on the choice of an appropriate functional form of the hedonic model (Pace, 1993 and 1995; Anglin and Gencay, 1996; Gencay and Yang, 1996). A prespecified functional form is, however, the fundamental assumption underlying the use of theory-laden parametric models; a poor choice imposes artificial structure on data and significantly invalidates results of the subsequent analysis[5]. In contrast, nonparametric techniques are data-driven, flexible approaches that can learn much of the genuine struc-

ture from available facts and, therefore, allow greatly reduced attention to the question of which functional form ought to be used.

Local regression techniques can significantly reduce the mis-specification error by letting the data to determine the appropriate functional relationship between the response and a set of attributes. Locally weighted regression adapts locally to changing curvature in the hedonic surface by giving more weight to nearby observations (McMillen, 1996) and, therefore, can account for complex nonlinear patterns. The local adaptation property, which is achieved by parametric localization (Cleveland and Loader, 1996), makes it a highly attractive tool for estimating spationally nonhomogeneous hedonic functions. Furthermore, any specific assumption underlying the error distribution can be relaxed and, in most cases, derived directly from the evidence e.g. by resampling techniques.

Data on land prices are imperfect, which generate difficult problems with conventional parametric approaches. In particular, extreme points, influential and outlying observations, which might represent erroneous data or otherwise reflect unusual market conditions such as non-arm's length transactions, can seriously undermine the performance of parametric estimator. The results of locally weighted regression can be robustified in a straightforward manner by a scheme, which is a variant of M-estimation. This simple régulation robuste typically offers enough protection against unusual or aberrant observations.

The local regression problem can be formalized by using locally weighted least squares (e.g. Ruppert and Wand, 1994; Loader, 2004):

$$Minimize \sum_{i=1}^{n} \mathbf{W_H}(\mathbf{x}_i - \mathbf{x})(p_i - \langle \theta, \mathbf{F}(\mathbf{x}_i - \mathbf{x})\rangle)^2, \quad (22)$$

where θ is the $d + 1$ vector of unknown coefficients and $\mathbf{F}(\cdot)$ is a vector of basis polynomials. $\mathbf{W_H}$ is a multivariate weight function and $\mathbf{H}^{1/2}$ is a bandwidth matrix. The local least squares estimate of the unknown regression function $f(\mathbf{x})$ is then[6]:

$$\hat{f}(\mathbf{x}) = \mathbf{e}_1' \left(\mathbf{X}'\mathbf{W}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{W}\mathbf{p}. \quad (23)$$

For local cubic regression $\mathbf{e}_1$ is a $\left\{1 + d + \frac{1}{2}d(d+1) + \frac{1}{6}d(d+1)(d+1)\right\} \times 1$ vector having 1 in the first entry and all other entries 0. For local quadratic and linear model the dimension of $\mathbf{e}_1$ is, respectively, $\left\{1 + d + \frac{1}{2}d(d+1)\right\} \times 1$ and $\{1 + d\} \times 1$. $\mathbf{p} = [p_1, \ldots, p_n]'$ is a vector of observed land prices and the data matrix $\mathbf{X}$ for the local cubic model is:

$$\mathbf{X} = \begin{bmatrix} 1 & (\mathbf{x}_1 - \mathbf{x})' & vec'\{(\mathbf{x}_1 - \mathbf{x})(\mathbf{x}_1 - \mathbf{x})'\} & (\mathbf{x}_{(1)} \otimes \mathbf{x}_{(1)} \otimes \mathbf{x}_{(1)})' \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (\mathbf{x}_n - \mathbf{x})' & vec'\{(\mathbf{x}_n - \mathbf{x})(\mathbf{x}_n - \mathbf{x})'\} & (\mathbf{x}_{(n)} \otimes \mathbf{x}_{(n)} \otimes \mathbf{x}_{(n)})' \end{bmatrix}, \quad (24)$$

where $\mathbf{x}_{(i)} = (\mathbf{x}_i - \mathbf{x})$ and the *vec*-operator stacks the columns of $(\mathbf{x}_i - \mathbf{x})\,(\mathbf{x}_i - \mathbf{x})'$, each below the previous, but with entries above main diagonal omitted; $\otimes$ is the Knonecker (tensor) product. The local linear and quadratic model uses only the first two and three, respectively, columns of the data matrix. The weight matrix is composed of $\mathbf{W} = diag\left\{\mathbf{W_H}(\mathbf{x}_1 - \mathbf{x}), \ldots, \mathbf{W_H}(\mathbf{x}_n - \mathbf{x})\right\} \equiv diag\left\{w_1(\mathbf{x}), \ldots, w_n(\mathbf{x})\right\}$.

The computational method for estimating the fit at points of evaluation is based on a damped Newton-Raphson algorithm (Loader, 1999, p. 209-211; Loader, 2004):

$$\tilde{\theta}_{k+1} = \tilde{\theta}_k + \frac{1}{2^j}\mathbf{J}^{-1}\mathbf{X}'\mathbf{W}\mathbf{q}. \quad (25)$$

$\tilde{\theta}_{k+1}$ is an estimate of the parameter vector θ at the $k + 1$ iteration. Here j is selected to be the smallest non-negative integer that results in an increase of the local log-likelihood at every step, and $\mathbf{q}$ is the usual score function. Furthermore, the Jacobian matrix can be expressed as $\mathbf{J} = \mathbf{D}^{1/2}\mathbf{P}'\,\Sigma\,\mathbf{P}\mathbf{D}^{1/2}$, where $\mathbf{D}$ is the matrix of diagonal elements of $\mathbf{J} = \mathbf{X}'\mathbf{W}\mathbf{V}\mathbf{X}$ with $\mathbf{P}$ and Σ representing, respectively,

the eigenvectors and eigenvalues of $\mathbf{D}^{1/2}\mathbf{J}\mathbf{D}^{1/2}$. $\mathbf{V}$ is the diagonal observed information matrix. Since the sample sizes and the dimensions of attribute spaces are small in this study, direct evaluation of fit is potentially feasible (although not applied) using the recursion of (25), which would mean that separate weighted least squares regression is estimated for each data point, with more influence given to nearby observations.

Direct evaluation is, however, computationally infeasible for larger data sets[7] and for increased dimensions of attribute space; a general algorithm is needed to perform the selection of evaluation points, where the fit is subsequently estimated. Tree-based structures are popular; in particular, the k-d-trees due to (Friedman, Bentley and Finkel, 1977) or growing adaptive trees (see, inter alia, Loader, 1999, p. 212–218). Different interpolation schemes (e.g. blending functions) can be used to define the fit elsewhere.

The time element is estimated by the local regression using house-price index measure.

5. SAMPLE DATA

The sample data of this study involve observations on urban residential land prices and the associated characteristics in the municipality of Espoo, a highly polycentric city[8], which lies inside the Helsinki metropolitan area with circa 225 000 habitants; its population is the second largest of the cities in Finland, which has experienced a rapid growth in its late history. The study period is from January, 1985 to December, 2007 with total number of observations of 3149 that constitute a judgement sample and cover phases of upward and rapid downward movements of land prices. In that period Finnish economy has experienced a great depression, which has had a major influence of land prices also. The observations from the last year (total of 78) are held back for post-sample predictive testing; a choice which is a somewhat arbitrary and mainly dictated by practical valuation concerns. In Table 1 are documented some standard sample statistics for the study variables.

Table 1. Sample statistics of study variables

Variable	Arithmetic mean	Minimum	Maximum	Std. Deviation
Total sales price	191738.49	631.00	19450000.00	647188.81
Permitted building volume	736.23	30.00	54322.40	1750.24
Parcel size	2286.93	300.00	113000.00	4800.90
Northing	6677724.65	6668110.00	6781565.00	5954.26
Easting	2540238.28	2528820.00	3475399.47	17009.52
House price index	204.20	116.40	350.10	57.78
Presence of a shore indicator	NA	0	1	NA
Presence of a shore NA indicator	NA	0	1	NA
Housing block indicator	NA	0	1	NA
Singly-family house indicator	NA	0	1	NA
Row house indicator	NA	0	1	NA
Municipality indicator	NA	0	1	NA
Private person indicator	NA	0	1	NA

The total sales price is chosen as the proper dependent variable (instead of the unit price) after some empirical experimentation: The goodness-of-fit statistics are much better when total sales price is explained by attribute variables. The unit of total sales price is €. Permitted building volume and parcel size variables are expressed in square meters. Northing and easting represent co-ordinates in the Finnish KKJ-system. The house price index variable is a quality-adjusted measure of house prices in the Helsinki metropolitan area and it is unitless. There are also seven indicator variables in the data set that deserves a mentioning. Presence of a shore indicator receives a value of one if the land parcel is bordered on water system and null otherwise. Presence of a shore NA indicator receives a value of one, if it is not known whether the land parcel is bordered on water system and null otherwise. There existed 327 observations where it was not known whether it was bounded by a water system or not. Housing block indicator receives a value of one, if the intended use of the land in the detailed plan is multistorey apartment block and null otherwise[9]. Single-family house indicator receives a value of one if the intended use of the land in the detailed plan is for single-family houses and null otherwise. Row house indicator receives a value of one, if the intended use of the land in the detailed plan is for row houses and null otherwise. Municipality indicator receives a value of one if the buyer of the land is a municipality and null otherwise. And finally, private person indicator receives a value of one if the buyer of the land is a private person and null otherwise.

6. RESULTS OF HEDONIC MODEL ESTIMATION

6.1. Ordinary least squares estimation

Table 2 summarises the results of the ordinary least squares estimation, in which double-log model specification is used (i.e. all quantitative variables are logarithmised). The standard goodness-of-fit statistics, the coefficient of determination and standard error of regression, indicate the fit is quite good. In particu-

Table 2. Fit and hedonic prices from the ordinary least squares regression

Variable	Coefficient	Standard error	t-value	p-value
Constant	1658.300	152.110	10.90	0.0000
Permitted building volume	0.766	0.0182	42.01	0.0000
Parcel size	0.115	0.0191	6.04	0.0000
Northing	-260.586	10.909	-23.89	0.0000
Easting	165.110	5.998	27.53	0.0000
House price index	1.455	0.0262	55.53	0.0000
Presence of a shore indicator	-0.768	0.143	-5.38	0.0000
Singly-family house indicator	0.0807	0.0269	3.00	0.0027
Row house indicator	0.138	0.0404	3.42	0.0006
Municipality indicator	-0.295	0.0671	-4.40	0.0000
Private person indicator	-0.0619	0.0181	-3.42	0.0006
Presence of a shore NA indicator	0.0848	0.0235	3.61	0.0003
	Coefficient of determination: 0.88 Number of outliers: 74		Standard error of regression: 0.37	

lar, the coefficient of determination statistics is over 0.70, which commonly used target in land valuation. Furthermore, the standard error of regression is below 0.40 indicating that the internal precision is acceptable. Statistically, four most significant attribute variables are, respectively: house price index (t-value is 55.53), permitted building volume (t-value is 42.01), easting (t-value is 27.53) and northing (t-value is -23.89). Furthermore, there are six statistically significant indicator variables and parcel size variable in the hedonic model. However, these remaining seven attribute variables explain the observed variability in land prices much less than the four most significant attribute variables. All explanatory variables are plausible in sign and magnitude. Overall, 74 outliers were dropped from the final hedonic model[10].

6.2. Robust MM-estimation

Table 3 summarises the results of the robust MM-estimation, in which double-log model specification is used. The standard error of regression statistic indicates a slightly better fit than in the case of ordinary least squares[11]. All explanatory variables are plausible in sign and magnitude. The standard error of regression statistic is below 0.40 indicating that the internal precision is acceptable. Statistically, four most significant attribute variables are, respectively: easting (t-value is now increased to 143.62!), house price index (t-value is 56.26), permitted building volume (t-value is 46.19) and northing (t-value is -32.89). Furthermore, there are six statistically significant indicator variables and parcel size variable in the hedonic model. However, these remaining seven attribute variables explain the observed variability in land prices much less than the four most significant attribute variables. No outliers were dropped from this hedonic model. Instead, the influence of aberrant observations was down weighted by using a specific weight function.

6.3. Structural time series models

Table 4 summarises the results of the structural time series estimation, in which double-log model specification is used for regression effects. The standard goodness-of-fit statis-

Table 3. Fit and hedonic prices from the MM-estimation

Variable	Coefficient	Standard error	t-value	p-value
Constant	1589.955	123.018	12.93	NA
Permitted building volume	0.774	0.0168	46.19	NA
Parcel size	0.109	0.0174	6.25	NA
Northing	-257.274	7.823	-32.89	NA
Easting	166.216	1.157	143.62	NA
House price index	1.453	0.0258	56.26	NA
Presence of a shore indicator	-0.867	0.133	-6.51	NA
Singly-family house indicator	0.0803	0.0258	3.11	NA
Row house indicator	0.182	0.0395	4.61	NA
Municipality indicator	-0.313	0.0672	-4.66	NA
Private person indicator	-0.0666	0.0177	-3.77	NA
Presence of a shore NA indicator	0.0665	0.0233	2.85	NA
	Coefficient of determination: NA		Standard error of regression: 0.34	

tics, the coefficient of determination and the standard error of regression, indicate the fit is pretty good. In particular, the coefficient of determination statistics is 0.90 and the standard error of regression is below 0.34 indicating that the internal precision is acceptable[12]. Statistically, three most significant attribute variables are, respectively: permitted building volume (t-value is 36.96), easting (t-value is 28.07) and northing (t-value is -24.79). The statistical significance of the house price index variable is now significantly reduced (the t-value is now only 6.61). The reason for this is that the unobserved components are, in fact, already revealing much of the same information than the house price index variable. Furthermore, there are six statistically significant indicator variables and parcel size variable in the hedonic model. However, these remaining seven attribute variables explain the observed variability in land prices much less than the three most significant attribute variables. All explanatory variables are plausible in sign and magnitude.

Structural time series model also uses unobserved components to account for the temporal variability in the dependent variable. In Table 4 there are three different unobserved components in the model structure: the level term (which is the dynamic version of the constant variable), one cycle term (with two components) and an 1st order autoregressive (AR(1)) process. The data analysed contained many outlying observations in terms of an unusual high value of standardised residual[13]. Instead of removing the outlier its effect was statistically measured by an impulse intervention variable and the influence was subsequently included as part of the overall model specification resulting to no loss of price information. In the final hedonic model there are 73 impulse intervention variables.

Table 4. Fit and hedonic prices from the structural time series estimation

Variable	Coefficient	Standard error	t-value	p-value
Level	1704.400	144.250	11.82	0.0000
Cycle (comp. #1)	0.0178	0.126	NA	NA
Cycle (comp. #2)	-0.00430	0.141	NA	NA
AR(1)	0.0339	0.143	NA	NA
Permitted building volume	0.7050	0.0190	36.96	0.0000
Parcel size	0.179	0.0195	9.19	0.0000
Northing	-256.740	10.358	-24.79	0.0000
Easting	158.11	5.632	28.07	0.0000
House price index	0.927	0.140	6.61	0.0000
Presence of a shore indicator	-1.101	0.146	-7.54	0.0000
Housing block indicator	0.168	0.0378	4.45	0.0000
Singly-family house indicator	0.123	0.0268	4.59	0.0000
Row house indicator	0.178	0.0398	4.46	0.0000
Municipality indicator	-0.358	0.0630	-5.68	0.0000
Private person indicator	-0.0811	0.0171	-4.74	0.0000
	Coefficient of determination: 0.90 Number of interventions: 72		Standard error of regression: 0.34	

6.4. Robust local regression

Table 5 summarises the results of the robust local regression, in which local double-log model specification is used. To avoid the curse of dimensionality, only the six most significant variables from the ordinary least squares estimation were included into final model of local regression[14]. The overall in-sample fit is better than in the former approaches. The coefficient of determination statistic is 0.93 and the standard error of regression statistic is 0.29. Because the local regression is a nonparametric method, there are no coefficient estimates that can be reported[15]. No outliers were dropped from this hedonic model. Instead, the influence of aberrant observations was down weighted by use of M-estimation.

7. MEASURES OF PREDICTIVE ACCURACY

Predictive accuracy is perhaps the single most important operational criterion in the evaluation of performance of chosen hedonic model. The success of hedonic model-based forecast depends on (see, Hendry, 1997):

(1) the existence of structure;
(2) whether such structure is informative about the future;
(3) the proposed method capturing the structure;
(4) the exclusion of irregularities that swamp the structure.

The aspects in (1)-(2) are characteristics of the economic system and the last two of the chosen forecasting method. When structure is understood as a systematic relation between the entity to be forecast and the available information, the conditions in (1)-(4) are sufficient for forecastability.

There are numerous different indicators for post-sample predictive assessment of hedonic models (e.g. Case et al., 2004) and the relative ranking of the performance of various models varies according to the applied accuracy measure. Mean prediction error is evaluated in this study by the arithmetic average prediction error, which measures the predictive unbiassness of the hedonic model. Two measures of strength of the association between predictions and observed out-of-sample land prices are reported. First, the usual correlation coefficient is calculated, which is a useful measure of statistical relation in the case of normally distributed variables and when the focus is on the co-variation of variables. The major problem of using the classical correlation measure in land valuation studies lies in its strong dependency on the normality assumption, which is typically violated by the influence of aberrant error terms, whose effect is squared in the denominator, which, in turn,

Table 5. Fit of the local regression

Variable	Coefficient	Standard error	t-value	p-value
Constant	NA	NA	NA	NA
Permitted building volume	NA	NA	NA	NA
Parcel size	NA	NA	NA	NA
Northing	NA	NA	NA	NA
Easting	NA	NA	NA	NA
House price index	NA	NA	NA	NA
Presence of a shore indicator	NA	NA	NA	NA
	Coefficient of determination: 0.93		Standard error of regression: 0.29	

tend to lead to highly similar standard deviations between different model alternatives. Secondly, the gravity (see McMillen, 2001) is reported that is not strongly dependent on any particular distributional assumptions. Generally, the gravity seems to be a viable measure of strength of association[16].

Root mean squared error (RMSE) is the most commonly used measure of success of numeric prediction, which controls the reliability or variability of predictions. This statistic is very sensitive to outlying observations tending to exaggerate the variance of prediction errors of model choices in which the prediction error is larger than the others (which is typical in land price studies). Mean absolute error (MAE) is generally a more appropriate indicator of predictive variability, and is especially suitable in cases of outlying prediction errors. Widely used measure of predictive variability is mean absolute percentage error (MAPE) (see e.g. Makridakis and Hibon, 2000) which, however, has some problems of asymmetry and instability, when the data are small.

8. FORECASTING ACCURACY OF DIFFERENT HEDONIC APPROACHES

Table 6 summarises the post-sample prediction statistics for the four different approaches (ordinary least squares, MM-estimation, structural time series and local regression). Six different measures of predictive accuracy are reported. First of all, the mean prediction error, which measures the predictive unbiassness, is significantly reduced, when MM-estimation, structural time series or local regression is used instead of the orthodox ordinary least squares. The mean prediction error is 81% smaller, when MM-estimation is used, 64% smaller when structural time series is used and 59% smaller, when local regression is used, instead of ordinary least squares. It therefore seems that predictive validity can be significantly improved when unorthodox methods (MM-estimation, structural time series, local regression) are applied. The mean prediction error is smallest when MM-estimation is used.

MAPE is useful predictive measure and usually in practice it is the measure we looking for. Here all approaches produce an error, which is only a slight over 2%. The unorthodox approaches produce a smaller MAPE that the orthodox approach: Structural time series gives 7% smaller MAPE, MM-estimation generates 5.5% smaller MAPE and local regression produces 4.3% smaller MAPE than in the case of ordinary least squares. MAPE is smallest when structural time series are applied.

The unorthodox approaches all give the same RMSE of 0.18, whereas the ordinary

Table 6. Post-sample prediction statistics for different approaches

Measure of predictive accuracy	Modelling approach			
	Ordinary least squares	MM-estimation	Structural time series	Local regression
Mean predict. error	0.16	0.031	-0.057	0.066
MAPE	2.56	2.42	2.38	2.45
RMSE	0.20	0.18	0.18	0.18
MAE	0.31	0.29	0.28	0.29
Correlation	0.86	0.86	0.87	0.86
Gravity	2714	2936	2971	2891

least squares produces RMSE of 0.20. It means that the unorthodox approaches produce a RMSE that is 10% smaller the one that is obtained by ordinary least squares. MAE is a robust version of RMSE and it is usually more reliable indicator than RMSE. Again MAE is highest when ordinary least squares is used: Structural time series generates 9.7% smaller MAE, MM-estimation and local regression produce 6.5% smaller MAE, when compared to the case of ordinary least squares.

Correlation coefficients are very similar between the approaches, the only exception is the value underlying structural time series, which is 1% higher than in the other cases. When gravity is used there are more differences between the approaches: the highest association is obtained when structural time series is used and the lowest association is obtained when ordinary least squares is used. Specifically, structural time series produces 9.5% higher gravity, MM-estimation generates 8.2% higher gravity and local regression gives 6.5% higher gravity, when compared to the case of ordinary least squares estimation.

9. CONCLUSIONS

This paper has investigated the structure of urban residential land prices and, specifically, the predictive accuracy of hedonic models between four different approaches, when land prices are predicted in a local market. In this study applied hedonic approaches are: 1) ordinary least squares estimation, 2) robust MM-estimation, 3) structural time series estimation and 4) robust local regression. Ordinary least squares and robust MM-estimation are parametric methods, structural time series estimation is a semi-parametric method and robust local regression is a nonparametric method.

Post-sample predictive assessment indicated that more precise predictions are obtained if the unorthodox methods of this study are used instead of the conventional least squares estimation. In particular, the predictive unbiassness can significantly be improved, if we move from the orthodox least squares estimation to using robust MM-estimation, structural time series estimation or robust local estimation. All six different forecasting indicators were better (or at least equal) in the case of the non-standard hedonic methods. Among the four different hedonic approaches, in overall, finest post-sample predictions are produced by the structural time series estimation.

The hedonic estimation revealed that there are four separate attribute variables that have an overriding effect of land prices. These independent variables are: permitted building volume, house price index, northing and easting. The influence of parcel size variable and different indicator variables on land prices were much weaker.

Notes

[1] This section reviews the hedonic price studies in land markets, which are presented in major scientific journals since 1995. Major findings, the data and the modelling methods are documented.

[2] In the study these are obtained by using S-estimation (Rousseeuw and Yohai, 1984).

[3] Regression coefficients can be time-varying, but this the representation used in the empirical section of the study. Here p and k denote the number quantitative and qualitative explanatory variables, respectively.

[4] Local regression means locally weighted regression, in which local polynomial functions are used in estimating the regression surface.

[5] In the parametric modelling context, a common solution to the problem of selecting an appropriate functional form is to consider a set of parametric functions with the objective of finding a model structure that matches the evidence in most measurable respects. However, there is no clear evidence that this practice will be successful in avoiding functional form mis-specification (Anglin and Gencay, 1996; Hannonen, 2005). Specification searches can be highly time-consuming

and the intrinsic power of these specification tests is somewhat questionable.

[6] Assuming, as usual, that $\mathbf{X}'\mathbf{WX}$ is non-singular.

[7] Also visualisation of regression surfaces, variance functions etc. demands that a separate, reduced number of fitting points are selected.

[8] Because of this polycentric nature numerous distance measures are needed to various subcenters. From the hedonic modelling viewpoint this creates problems of multicollinearity, when several distance measures are used. As a solution, no distance measures are used but the co-ordinates describing location are used instead.

[9] Intended use of all sites in this study is for housing so that there does not exist non-residential types of land use.

[10] Outliers are considered here as those observations whose standardised residual is larger than 3.5. This is a typical value in the Finnish practice when hedonic based land analysis is conducted.

[11] The coefficient of determination statistic cannot be calculated in standard manner and thus it is not reported in the case of MM-estimation.

[12] In fact, the standard error of regression statistic is now same as in the case of MM-estimation.

[13] Intervention variables are used for those observations whose standardised residual is larger than 3.5.

[14] In other words, only those attribute variables were included into the local regression from the ordinary least estimation with global double-log specification whose t-values in absolute terms were higher than five.

[15] To be more specific, nonparametric estimators are, in fact, "over-parametric" in a sense that they generate an infinite number of hedonic prices for each attribute depending on the values of that attribute. In practice, for a particular characteristic a single representative hedonic price is often of direct interest and, consequently, some kind of average derivative is usually needed. However, there are problems in the average derivate estimation, so that in this study no average hedonic prices are calculated.

[16] Here gravity is calculated so that the weighted (by the area) inner product of predictions and realisations is divided by the L_2-distance between predictions and realisations.

REFERENCES

Anglin, P. M. and Gencay, R. (1996) Semiparametric estimation of a hedonic price function, *Journal of Applied Econometrics*, 11(6), pp. 633–648.

Atack, J. and Margo, R.A. (1998) Location, location, location! The price gradient for vacant urban land: New York, 1835 to 1900, *Journal of Real Estate Finance and Economics*, 16(2), pp. 151–172.

Breiman, L. (2001) Statistical modeling: The two cultures, *Statistical Science*, 16(3), pp. 199–231.

Case, B., Clapp, J., Dubin, R. and Rodriquez, M. (2004) Modelling spatial and temporal house price patterns: A comparison of four models, *Journal of Real Estate Finance and Economics*, 29(2), pp. 167–191.

Clapp, J. M., Rodriquez, M. and Pace, R. K. (2001) Residential land values and the decentralization of jobs, *Journal of Real Estate Finance and Economics*, 22(1), pp. 43–61.

Cleveland, W. S. and Loader, C.R. (1996) Smoothing by local regression: Principles and methods. In: Härdle, W. and Schimek, M. G. (eds.) *Statistical Theory and Computational Aspects of Smoothing*, Physica-Verlag.

Colwell, P.F. (1998) A primer on piecewise parabolic multiple regression analysis via estimations of Chicago CBD land prices, *Journal of Real Estate Finance and Economics*, 17(1), pp. 87–97.

Colwell, P. F. and Munneke, H. J. (1997) The structure of urban land prices, *Journal of Urban Economics*, 41(3), pp. 321–336.

Colwell, P. F. and Munneke, H. J. (1999) Land prices and land assembly in the CBD, *Journal of Real Estate Finance and Economics*, 18(2), pp. 163–180.

Colwell, P. F. and Munneke, H. J. (2003) Estimating a price surface for vacant land in urban area, *Land Economics*, 79(1), pp. 15–28.

Durbin, J. and Koopman, S. J. (2002) *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series #24, Oxford University Press.

Francke, M. K. and Vos, G. A. (2004) The hierarchical trend model for property valuation and local price indices, *Journal of Real Estate Finance and Economics*, 28(2/3), pp. 179–208.

Friedman, J. H., Bentley, J. L. and, Finkel, R. A. (1977) An algorithm for finding best matches in

logarithmic expected time, *ACM Transactions on Mathematical Software*, 3(3), pp. 209–226.

Gencay, R. and Yang, X. (1996) A forecast comparison of residential housing prices by parametric versus semiparametric conditional mean estimators, *Economic Letters*, 52(2), pp. 129–135.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986) *Robust Statistics: The Approach Based on Influence Functions*, Wiley Series in Probability and Mathematical Statistics.

Hannonen, M. (2005) On the recursive estimation of hedonic prices of land, *Nordic Journal of Surveying and Real Estate Research*, 2(2), pp. 30–56.

Harvey, A. C. (1997) Trends, cycles and autoregressions, *Economic Journal*, 107 (January), pp. 192–201.

Harvey, A. C. and Shephard, N. (1993) Structural Time Series Models. In: Maddala, G. S., Rao, C. R. and Vinod, H. D. (eds.) *Handbook of Statistics*, Vol. 11, Elsevier Science Publishers B. V.

Hendry, D.F. (1997) The econometrics of macroeconomic forecasting, *Economic Journal*, 107 (September), pp. 1330–1357.

de Jong, P. (1991) The diffusive Kalman filter, *Annals of Statistics*, 19(2), pp. 1073–1083.

Koopman, S. J., Harvey, A. C., Doornik, J. A. and Shephard, N. (1999) *Structural Time Series Analysis, Modelling and Prediction using Stamp*, Timberlake Consultants, London.

Lin, T. and Evans, A. W. (2000) The relationship between the price of land and size of plot when plots are small, *Land Economics*, 76(3), pp. 386–394.

Loader, C. (1999) *Local Regression and Likelihood*. Springer Series in Statistics and Computing, Springer-Verlag.

Loader, C. (2004) *Smoothing: Local Regression Techniques*, Handbook of Computational Statistics, (eds.) Gentle, J., Härdle, W. and Mori, Y., Springer-Verlag.

McMillen, D. P. (1996) One hundred fifty years of land values in Chicago: A nonparametric approach, *Journal of Urban Economics*, 40(1), pp. 100–124.

McMillen, D. P. (2001) Nonparametric employment subcenter identification, *Journal of Urban Economics*, 50(3), pp. 448–473.

McMillen, D. P. and Thorsnes, P. (2003) The aroma of Tahoma: Time-varying average derivates and the effect of a superfund site on house prices, *Journal of Business and Economics Statistics*, 21(2), pp. 237–246.

Pace, R. K. (1993) Nonparametric methods with applications to hedonic models, *Journal of Real Estate Finance and Economics*, 7(3), pp. 185–204.

Pace, R. K. (1995) Parametric, semiparametric, and nonparametric estimation of characteristic values within mass assessment and hedonic pricing models, *Journal of Real Estate Finance and Economics*, 11(3), pp. 195–217.

Rousseeuw, P. J. and Yohai, V. J. (1984) Robust Regression by Means of S-estimators. In: Franke J., Hardle, W. and Martin D. (eds.) Robust and Nonlinear Time Series, *Lecture Notes in Statistics*, Springer-Verlag, 26, pp. 256–272.

Ruppert, D. and Wand, M. P (1994) Multivariate Locally Weighted Least Squares Regression, *Annals of Statistics*, 22(3), pp. 1346–1370.

Shimizu, C. and Nishimura, G. N. (2007) Pricing structure in Tokyo metropolitan land markets and its structural changes: Pre-bubble, bubble, post-bubble, *Journal of Real Estate Finance and Economics*, 35(4), pp. 475–496.

Schulz, M.A.R. (2003) *Valuation of Properties and Economic Models of Real Estate Markets*. Dissertation, Berlin: Humboldt University Berlin.

Schulz, R. and Werwatz, A. (2004) A state space model for Berlin house prices: Estimation and economic interpretation, *Journal of Real Estate Finance and Economics*, 28(1), pp. 37–57.

Schwann, G. M. (1998) A real estate price index for thin markets, *Journal of Real Estate Finance and Economics*, 16(3), pp. 269–287.

Thorsnes, P. and McMillen, D. P. (1998) Land value and parcel size: A semiparametric analysis, *Journal of Real Estate Finance and Economics*, 17(3), pp. 233–244.

Wallace, N. E. (1996) Hedonic-based price indexes for housing: Theory, estimation and index construction, *FRBSF Economic Review*, No. 3, pp. 34–48.

SANTRAUKA

ŽEMĖS KAINŲ MIESTUOSE PROGNOZĖS: KETURIŲ METODŲ PALYGINIMAS

Marko HANNONEN

Nagrinėjama, kokiu tikslumu keturi skirtingi hedonistiniai metodai prognozuoja laisvų žemės plotų kainas vietinėse miestų rinkose. Nagrinėti tokie hedonistiniai metodai: 1) mažiausiųjų kvadratų metodas, 2) daugybinių modelių vertinimas, 3) struktūrinių laiko eilučių vertinimas, 4) lokalinė regresinė analizė. *Post-sample* prognostinis testas parodė, kad tikslesnės prognozės gaunamos taikant netradicinius šiame darbe nurodytus metodus, o ne įprastą mažiausiųjų kvadratų metodą. Taikant netradicinius hedonistinius tyrimo metodus, gali gerokai padidėti prognozių nešališkumas. Darbe nagrinėta ir žemės kainų mieste struktūra. Aiškinant žemės kainas iš būdingų kintamųjų svarbiausi buvo leidžiamas pastato dydis, būsto kainų indeksas, sklypo padėtis. Sklypo dydžio kintamasis ir įvairių rodiklių kintamieji žemės kainoms turėjo daug mažesnę įtaką.