

## **AN ANALYSIS OF LAND PRICES: A STRUCTURAL TIME-SERIES APPROACH**

**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" <marko.hannonen@hut.fi>*

Received 21 March 2005; accepted 10 June 2005

**ABSTRACT.** This paper analyses spatio-temporal variation of land prices in two single localities by means of structural time series modelling formalism that combines the flexibility of a time series model with that of the interpretation of a regression analysis. The extension of conventional hedonic models by introducing unobserved components for trend and cycle resulted to significant improvements in their post-sample predictive accuracy. In predictive testing, for most model formulations the unobserved component approach generated only a marginal average prediction error when compared to the orthodox hedonic models, which, in contrast, yielded to a considerable amount of systematic prediction error. It therefore seems that the structural time-series modelling paradigm offers a more viable alternative to the hedonic analysis of land prices than the conventional approach based on least squares estimates. The effect of slope component in the trend specification was found to be statistically insignificant, which implies that the elementary local level model would be the most adequate description of the long-term land price movements.

**KEYWORDS:** Hedonic prices; Unobserved components; Trend; The Kalman filter

### **1. INTRODUCTION**

As a factor of production land is a central tenet for all businesses. 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 pre-defined sub-market and changes that they cause in prices in that market area. The cross-sectional or spatial variation stems from the differentiated nature of a land depending on certain characteristics such as location, lot size, intended use and planning status of a land parcel. This cross-sectional variation is usually divided into the concepts of spatial heterogeneity and spatial dependence. 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 (Francke & Vos, 2004).

The time-series or temporal variation, by contrast, 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 & Werwatz, 2004). An empirical model of land prices has to recognize these two different, yet closely related sources of variations.

This paper emphasises the time-varying aspects of land prices in single localities that, in current practice, are seldom captured by some dynamic model, nor is the time evolution of land prices allowed to depend on the relevant attribute variables. According to Schulz & Werwatz (2004): “the treatment of the over-time variation of the common price component is usually rather unsophisticated in hedonic regression; time dummies are used to capture the law of the motion of the general price trend”.

The conventional or static hedonic regression approach that is based on ordinary least squares (e.g. Hiltunen, 2003) is an appropriate modelling context 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 these methods, 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 & Vos, 2004).

A (univariate) time series is a set of observations  $p_1, \dots, p_t$  ordered in time (Durbin & Koopman, 2002, p. 9); time interval need not be equispaced. In the land market observable time series on price and its attributes rarely constitutes any stationary process, which is a fundamental assumption in the traditional approach (the Box-Jenkins ARIMA methodology) to time series modelling. Instead, they are *nonstationary* or transient meaning that the distributions of random variables depend on time<sup>1</sup>. Furthermore, if multiple submarkets are analysed simultaneously, observations on price information through the passage of time are *multivariate*, which imposes further restrictions to the methodology to be applied.

Given the special characteristics of land markets one natural solution to the dual problem of hedonic modelling caused by spatio-temporal variation is to 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 interpretable components of trend, cycle, seasonal, and regression term plus error. They provide more information for the analyst about different stochastic features underlying the series than the traditional regression analysis. In essence, structural time series models can be thought of as a certain type of generalized regression models in which explanatory variables are functions of time and the parameters are time-varying (Harvey, 1997).

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 & Shephard, 1993; Harvey, 1997; Durbin & 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 (transitory parts of

<sup>1</sup> More precisely, nonstationarity denotes the general sense of processes whose first two moments (conditional expectation and the variance of its error distribution) are not constant over time. The dynamic nature of data-generating processes is attributable to changes in economic environments, technological progress, political shifts, cultural movements, etc. A more detailed account of the fundamental sources of change in economics can be found, e.g. (Day, 1992).

the model specification) 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, e.g. impulse and level interventions, 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.

## Research Problem

The land market is one of the most important sectors of the whole economy; in particular, land itself forms a necessary factor of production for all human activities (housing, industry, business, etc.). However, systematic research about spatio-temporal movements in these markets has typically been scarce, inflexible and insufficient. In this paper, structural time series methodology due to (Harvey, 1989; Harvey and Shephard, 1993; inter alia), i.e. the unobserved component approach, is used for *simultaneous modelling of spatial and temporal variation of land prices*. Unobserved component models can 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. The practical significance of this paper may be found in its flexible management of trend and cyclical behaviour of land prices that would help, with more accurate land price information, governments to make better policies and property investors to create more effective property management strategies. For any valuation method to have a sufficient degree of validity it must produce an accurate prediction of the most probable market price of property<sup>2</sup>. The adequacy of estimated unobserved component models is, therefore, evaluated by post-sample predictive assessment, where forecasts of those models are compared to predictions of some traditional static econometric models.

## Previous Related Research

The importance of structural time series modelling and the associated state space form is only recently understood in the real estate markets. Franke & Vos (2004) applied structural time series models to estimation of selling prices of individual houses and for constructing general and sub-market level price indices. They introduced a specific structural model, the so-called *Hierarchical Trend Model*, in which parameters vary over time, location and house type. General and

---

<sup>2</sup> Most probable transaction price is the market value of a property, which is defined as (IVSC, 2001): “the estimated amount for which an asset should exchange on the date of valuation between a willing buyer and willing seller in an arm’s length transaction after proper marketing wherein the parties each acted knowledgeably, prudently and without compulsion”.

cluster-level trends were modelled as stochastic trends using a random walk scheme for sub-market trends and a random walk with a drift scheme for the general trend. However, the regression coefficients that were not related to attributes on house type and location were kept constant over the study period.

The approach due to (Franke & Vos, 2004) can be described as a dynamic hedonic price modeling, which is an extension of the usual, the so-called covariance analysis methodology, that permits time varying coefficients for any attribute if needed. The unobserved component model was put into the state space form and the Kalman filtering and smoothing techniques were used to estimate the hedonic prices. Franke & Vos (2004) conclude that their Hierarchical Trend Model “provides more up-to-date, detailed and reliable results than using standard hedonic methods and the simple-weighted [median] method”. The Hierarchical Trend model is currently operational in Amsterdam and several other Dutch cities for the determination of local real estate taxes.

Chen *et al.* (2004) examined the time-series properties of house prices in four Asian Markets (Hong Kong, Singapore, Tokyo and Taipei) using structural time-series methodology. Two types of trend models, a stochastic trend model and a deterministic trend model, were applied to study the long-run behaviour of house prices. The trend description was most accurate when a smooth trend was used, which also indicated that Taipei and Singapore had the most pressure for house prices to increase over time. The markets were also found to have three stochastic cycles: around one year, two to four years and seven to ten years. The stochastic nature of trend and cycle components implied that the markets are not in steady state but are still changing.

Schulz (2003) and Schulz & Werwatz (2004) investigated hedonic prices of single-family houses for some local areas of Berlin with an explicit statistical model for common component of prices. A specific two equations system based on the present value model of asset prices was studied, which was subsequently put in the state space form and the Kalman filtering and smoothing recursions were applied. The estimated coefficients were plausible in sign and magnitude and consistent with the assessment of professional appraisers and provided interpretable figures of the premiums of different house characteristics.

Schwann (1998) investigated a time series based method for estimating real estate price indexes in the thin (local) markets. Schwann argues that the core of the estimation problem in local market level is the lack of degrees of freedom, since the conventional hedonic methods involve the use of an extensive set of time dummy variables along with the regressors. Schwann proposed a hedonic model, in which an autoregressive formulation is used in linking current observations to preceding observations with a stochastic structure for the parameter evolution. The log-likelihood function was derived via prediction error decomposition and the diffuse Kalman filter due to (de Jong, 1989 & 1991) was applied. Schwann concluded that his price index construction exhibits superior performance as compared to the standard hedonic indices.

## 2. UNOBSERVED COMPONENT MODELS OF THE STUDY

The classical time series analysis, i.e. the Box-Jenkins ARIMA (autoregressive integrated moving average), assumes that stochastic processes are stationary or can be transformed to stationary series by the process of differencing them (possible multiple times). In the structural analysis of the series neither the concept of stationary process nor the process of differencing plays any fundamental role (Harvey & Shephard, 1993).

Structural time series models can effectively separate trend, cyclical, seasonal and irregular components from exogenous or predetermined variables. It is thus expected that the parameters for exogenous variables can be estimated more accurately using these models. In a regression

context, the principal *structural time series models* are extended hedonic models in which explanatory variables are functions of time and the parameters are time-varying (Harvey 1989, p. 10; Harvey & Shephard, 1993; Harvey, 1997).

### Trend Models

The term *trend* can be understood as that part of the series when extrapolated gives the clearest indication of the future long-term movements. The simplest choice of a trend would be a deterministic linear time trend, but this usually is too restrictive, unless the time period is very short (Harvey, 1997). In a structural model an explicit stochastic trend is assumed in which the level and slope coefficients are allowed to evolve over time.

### Local Level Model

The local level model or the *random walk plus noise* is an elementary, yet effective, structural trend model, which regards an observation on price  $p_t$  at time  $t$  as being made up of an underlying level  $\mu_t$  and an irregular disturbance  $\varepsilon_t$  (Koopman *et al.*, 1999; Durbin & Koopman, 2002, p. 44-45)<sup>3</sup>:

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

The underlying level  $\mu_t$  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  $\eta_t$ . The effect of  $\eta_t$  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  $\mu_0$ .

The signal-to-noise ratio  $q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$  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.

### Local Linear Trend Model

Local linear trend model is a generalisation of the local level model encompassing a stochastic slope  $v_t$  (which itself follows a random walk) allowing the slope coefficients to change. The local linear trend model can be specified as:

$$\begin{aligned} p_t &= \mu_t + \varepsilon_t, & \{\varepsilon_t\} &\sim NID(0, \sigma_\varepsilon^2) \\ \mu_t &= \mu_{t-1} + v_{t-1} + \eta_t, & \{\eta_t\} &\sim NID(0, \sigma_\eta^2) \\ v_t &= v_{t-1} + \xi_t, & \{\xi_t\} &\sim NID(0, \sigma_\xi^2), \end{aligned} \quad (2)$$

where different disturbance terms are mutually independent. If  $\sigma_\xi^2 = 0$ , the trend reduces to a

<sup>3</sup> In this study, observations of different time series are modelled as distinct ones, which simplifies the analysis.

random walk with a drift, whereas for  $\sigma_\eta^2 = 0$ , the trend reduces to an integrated random walk or a smooth trend model.

The use of a smooth trend is sometimes suitable since a stochastic slope might be too sensitive to changes in the series. In particular, combining a smooth trend with a cycle component tends to lead to a more attractive decomposition of the series (Koopman *et al.*, 1999). The reduced form of local linear trend model is ARIMA(0,2,2) with certain restrictions on the parameter space.

### Cycle Models

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 usually cycles can be effectively presented as a mixture of sine and cosine waves with two parameters  $\theta_1$  and  $\theta_2$ . If  $\psi_t$  is a cyclical function of time with frequency  $\lambda_c$  that is measured in radians, then (Harvey & Shephard, 1993):

$$\psi_t = \theta_1 \cos \lambda_c t + \theta_2 \sin \lambda_c t, \quad (3)$$

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

$$\begin{pmatrix} \psi_t \\ \psi'_t \end{pmatrix} = \rho \begin{pmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi'_{t-1} \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa'_t \end{pmatrix}, \quad (4)$$

where  $\kappa_t$  and  $\kappa'_t$  are mutually uncorrelated with a common variance  $\sigma_\kappa^2$ .  $\rho \in [0, 1]$  is a damping factor. Stationary models corresponds to situations where  $\rho$  is strictly less than one. A first-order autoregressive process is an important limiting case of a stochastic cycle when a frequency  $\lambda_c$  is equal to 0 or  $\pi$ .

These different trend and cycle formulations are estimated and tested in the empirical section of the study. They are combined with the two different common hedonic representations (linear and double-log models) for regression effects, which are subsequently put into state-space form and the Kalman filtering and smoothing algorithms (of the section 3) are applied for estimation of trends, cycles and hedonic prices. Only the results of the most data-congruent specification are documented in depth.

### 3. STATE SPACE FORM

Many dynamic hedonic models could usefully be represented in a state-space form, which presents a convenient summary of the system's dynamics. In particular, the state space form allows a unified statistical treatment of structural time series models and thus provides an appropriate calculation method for estimating the hedonic prices and unobserved components.

A state-space form describes the relation between the observations and the unobservable components. The state vector evolves through a stochastic difference equation (a first-order autoregressive process) depending on unknown parameters<sup>4</sup>. The main objective of the state space

<sup>4</sup> State-space models were originally developed in control engineering by Kalman (1960) for describing dynamic systems that involve unobserved state variables, but they are also highly important in describing various economic activities such as consumption behaviour, indices of economic activity and financial markets (Kim & Nelson, 1999, p. 29). Recently there has been a growing interest in the real estate economics in these methods to be used in hedonic modelling (see Schulz 2003; Schulz & Werwatz, 2004; Chen *et al.*, 2004; Francke & Vos, 2004).

analysis is to infer the relevant properties of the unobserved components over time from the knowledge of the observations. The estimation of unobservable states as each new observation becomes available can be carried out with the *Kalman filter*, which plays the same role for structural time series models in a state-space form as the ordinary least squares for common hedonic models. The *Kalman smoother* is used for obtaining the best estimate of the state at any point within the sample (Harvey, 1989, p. 100-167; Harvey, 1993, p. 82-98; Harvey & Shephard, 1993; Durbin & Koopman, 2002, p. 1 & 11).

### Measurement Equation

The state space form consists of a *measurement* or *observation equation* and a *transition* or *state equation*.<sup>5</sup> The measurement equation, which describes the relation between observed variables and unobserved state variables, with regression effects, can be represented by (Harvey & Shephard, 1993; Kim & Nelson, 1999, p. 29; Durbin & Koopman, 2002, p. 121-122):

$$\mathbf{p}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad t=1, \dots, T, \quad (5)$$

which relates  $N \times 1$  vector of observed prices  $\mathbf{p}_t$  at time  $t$  to an  $m \times 1$  *state vector*  $\boldsymbol{\alpha}_t$  and to an  $(p+k) \times 1$  vector of unknown regression coefficients  $\boldsymbol{\beta}$  that are assumed to be constant<sup>6</sup>.  $\mathbf{Z}_t$  is a nonstochastic  $N \times m$  matrix of cycle and trend components,  $\mathbf{X}_t$  is a nonstochastic  $N \times (p+k)$  matrix of observations on explanatory variables and  $\boldsymbol{\varepsilon}_t$  is an  $N \times 1$  vector of serially uncorrelated measurement errors with zero mean and covariance matrix  $\mathbf{H}_t$ , i.e.  $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$  and  $Var(\boldsymbol{\varepsilon}_t) = \mathbf{H}_t$ .

### Transition Equation

The transition equation describes the dynamics of the state vector; the unobservable components  $\boldsymbol{\alpha}_t$  are generated by a first-order Markov process (Harvey & Shephard, 1993; Hamilton, 1994a):

$$\boldsymbol{\alpha}_t = \boldsymbol{\Pi}_t\boldsymbol{\alpha}_{t-1} + \mathbf{W}_t\boldsymbol{\beta} + \mathbf{R}_t\boldsymbol{\eta}_t, \quad t=1, \dots, T. \quad (6)$$

Now  $\boldsymbol{\Pi}_t$  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  $\boldsymbol{\eta}_t$  is a  $g \times 1$  vector of serially uncorrelated error terms with mean zero and covariance matrix  $\mathbf{Q}_t$ , i.e.  $E(\boldsymbol{\eta}_t) = \mathbf{0}$  and  $Var(\boldsymbol{\eta}_t) = \mathbf{Q}_t$ .

The following further assumptions complete the state space system specification:

1. The initial conditions are given by:  $E(\boldsymbol{\alpha}_0) = \boldsymbol{\alpha}_0$  and  $Var(\boldsymbol{\alpha}_0) = \mathbf{S}_0$ ;
2. The system and measurement disturbances are uncorrelated, i.e.  $E(\boldsymbol{\varepsilon}_t\boldsymbol{\eta}_s') = \mathbf{0}$  for all  $s, t=1, \dots, T$  and  $E(\boldsymbol{\varepsilon}_t\boldsymbol{\alpha}_0') = \mathbf{0}$ ,  $E(\boldsymbol{\eta}_t\boldsymbol{\alpha}_0') = \mathbf{0}$  for  $t=1, \dots, T$ .

Matrices  $\mathbf{Z}_t$ ,  $\mathbf{X}_t$ ,  $\mathbf{H}_t$ ,  $\boldsymbol{\Pi}_t$ ,  $\mathbf{W}_t$ ,  $\mathbf{R}_t$ ,  $\mathbf{Q}_t$  are called *system matrices* and assumed to be non-stochastic. Specifically,  $\mathbf{Z}_t$  and  $\boldsymbol{\Pi}_t$  are state system matrices,  $\mathbf{X}_t$  and  $\mathbf{W}_t$  are regression system matrices and others can be regarded as error system matrices. If the system matrices do not

<sup>5</sup> Here is described a typical *linear* state space representation. There are different representations for state-space form depending, e.g. treatments of (diffuse) initial conditions, regression effects and error process assumptions.

<sup>6</sup> 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.

depend on time, the model is said to be *time invariant* or *time homogenous*.<sup>7</sup>

### Kalman Filter

*Filtering* is aimed at updating our knowledge of the system as each new observation  $\mathbf{p}_t$  becomes available. Unlike smoothing (signal extraction) it is based solely on the previous observations. *The Kalman filter* primarily consists of a set of vector and matrix recursions that are optimal, in a sense of a minimum mean square, for calculating the moments of normal distribution of the state vector  $\boldsymbol{\alpha}_{t+1}$  conditional on the currently available information set  $\Omega_t = \{ \mathbf{p}_1, \dots, \mathbf{p}_t; \mathbf{x}_1, \dots, \mathbf{x}_t \}$ , i.e.  $\boldsymbol{\alpha}_{t+1} = E(\boldsymbol{\alpha}_{t+1} | \Omega_t)$  and  $\mathbf{S}_{t+1} = Cov(\boldsymbol{\alpha}_{t+1} | \Omega_t)$ , for  $t=1, \dots, T$ . (Harvey, 1989, p. 104-106).

The importance of the Kalman filter is based on (Koopman *et al.*, 1999):

- Computation of one-step ahead predictions for observations and state vectors with the associated mean square errors;
- Diagnostics via one-step ahead prediction errors;
- Computation of likelihood function by means of prediction error decomposition;
- Smoothing that uses the outcome of the Kalman filter.

There are a variety of forms to the basic Kalman filter, but if we start at time zero with a knowledge  $\boldsymbol{\alpha}_0 \sim N(\boldsymbol{\alpha}_0, \mathbf{S}_0)$ , the filter can be described by four *prediction equations* and two *updating equations* (Harvey, 1993, p. 85-86; Harvey & Shephard, 1993; Kim & Nelson, 1999, p. 29-37).

### Prediction

The prediction equations in the *contemporaneous filtering form* are given by (Harvey, 1993, p. 85-86; Harvey & Shephard, 1993; Durbin & Koopman, 2002, p.67-68)<sup>8</sup>:

$$\boldsymbol{\alpha}_{t|t-1} = \boldsymbol{\Pi}_t \boldsymbol{\alpha}_{t-1} + \mathbf{W}_t \boldsymbol{\beta} \quad , \quad (7)$$

$$\mathbf{S}_{t|t-1} = \boldsymbol{\Pi}_t \mathbf{S}_{t-1} \boldsymbol{\Pi}_t' + \mathbf{R}_t \mathbf{Q} \mathbf{R}_t' \quad , \quad (8)$$

$$\mathbf{v}_t = \mathbf{p}_t - \mathbf{Z}_t \boldsymbol{\alpha}_{t|t-1} - \mathbf{X}_t \boldsymbol{\beta} \quad , \quad (9)$$

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{S}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t \quad (10)$$

with an estimator of  $\mathbf{p}_t$  being:

<sup>7</sup> As an illustrative example, consider a univariate time series model (local linear trend model) with two explanatory variables and an intervention (and without a cycle term). Its state space formulation is given by:

$$p_t = (1 \ 0) \boldsymbol{\alpha}_t + (x_{1t} \ x_{2t} \ 0) \boldsymbol{\beta}_x + (\sigma_\varepsilon \ 0 \ 0) \mathbf{u}_t$$

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \boldsymbol{\alpha}_{t-1} + \begin{pmatrix} 0 & 0 & w_t \\ 0 & 0 & 0 \end{pmatrix} \boldsymbol{\beta}_x + \begin{pmatrix} 0 & \sigma_\eta & 0 \\ 0 & 0 & \sigma_\xi \end{pmatrix} \mathbf{u}_t$$

where  $\mathbf{u}_t = (\varepsilon_t \ \eta_t \ \xi_t)'$ . Different hedonic models lead to different state space formulations, which are omitted here, as they are not the main focus in the paper.

<sup>8</sup> The derivation of the Kalman filter is straightforward relying on the standard results of multivariate normal theory, see e.g. (Hamilton, 1994a)



$$\hat{\mathbf{p}}_{t|t-1} = \mathbf{Z}_t \boldsymbol{\alpha}_{t|t-1} + \mathbf{W}_t \boldsymbol{\beta} \quad . \quad (11)$$

$\mathbf{v}_t$  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.

### Updating

The updating equations in the *contemporaneous filtering form* are given by:

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t|t-1} + \mathbf{S}_{t|t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} (\mathbf{p}_t - \mathbf{Z}_t \boldsymbol{\alpha}_{t|t-1} - \mathbf{W}_t \boldsymbol{\beta}) \quad , \quad (12)$$

$$\mathbf{S}_t = \mathbf{S}_{t|t-1} - \mathbf{S}_{t|t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{S}_{t|t-1} \quad . \quad (13)$$

### Augmented Kalman Filter

The basic Kalman filter is usually supplemented (as in this research) by a set of complementary vector and matrix recursions. In particular, when non-stationary components and/or fixed regression effects are present an *augmented Kalman filter* is usually applied (Koopman *et al.*, 1999; Durbin & Koopman, 2002, p. 115-120):

$$\mathbf{V}_t = -\mathbf{Z}_t \mathbf{A}_{t|t-1} - \mathbf{X}_t \mathbf{B} \quad (14)$$

$$\mathbf{A}_{t+1|t} = \mathbf{\Pi}_t \mathbf{A}_{t|t-1} + \mathbf{W}_t \mathbf{B} + \mathbf{K}_t \mathbf{V}_t \quad (15)$$

$$(\mathbf{m}_t, \mathbf{M}_t) = (\mathbf{m}_{t-1}, \mathbf{M}_{t-1}) + \mathbf{V}'_t \mathbf{F}_t^{-1} (\mathbf{v}_t, \mathbf{V}_t) \quad (16)$$

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 matrix  $\mathbf{B}$ .  $\mathbf{K}_t = \mathbf{\Pi}_t \mathbf{S}_{t|t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1}$  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

$$\boldsymbol{\alpha}_{t|t-1}^* = \boldsymbol{\alpha}_{t|t-1} + \mathbf{A}_{t|t-1} \mathbf{M}_{t-1}^{-1} \mathbf{m}_{t-1} \quad (17)$$

$$\mathbf{S}_{t|t-1}^* = \mathbf{S}_{t|t-1} + \mathbf{A}_{t|t-1} \mathbf{M}_{t-1}^{-1} \mathbf{A}'_{t|t-1} \quad . \quad (18)$$

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

$$\mathbf{v}_t^* = \mathbf{v}_t + \mathbf{V}_t \mathbf{M}_{t-1}^{-1} \mathbf{m}_{t-1} \quad (19)$$

$$\mathbf{F}_t^* = \mathbf{F}_t + \mathbf{V}_t \mathbf{M}_{t-1}^{-1} \mathbf{V}'_t \quad (20)$$

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

### Prediction Error Decomposition

When the disturbances and known initial  $\alpha_0$  and  $S_0$  are normally distributed, the Kalman filter can be used to compute the exact Gaussian log-likelihood function via the *prediction error decomposition* (Harvey 1993, p. 91; Kim & Nelson, 1999, p. 11&19):

$$\log L(\boldsymbol{\psi}) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T (\log |\mathbf{F}_t| + \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t) \quad (21)$$

where  $\boldsymbol{\psi}$  is a set of unknown parameters for a specific statistical model. Typically a modification of the exact Gaussian log-likelihood function is applied. For univariate unobserved component models *concentrated diffuse likelihood function* can be constructed via prediction error decomposition where the component to be concentrated out of the likelihood function is the one possessing the largest standard deviation. In the case of multivariate structural time series models prediction error decomposition can be used to obtain a *diffuse likelihood function*.

### Initialisation of Kalman Filter

For stationary processes, the initial conditions are obtained by unconditional mean and variance of the Kalman filter (see Kim & Nelson, 1999, p. 29; Durbin & Koopman, 2002, p. 99-120). When the transition equation is non-stationary, the diffuse Kalman filter due to de Jong (1989, 1991) can be applied. Using diffuse initial conditions lead to starting values being formed from the observations.

### Kalman Smoother

Estimating  $\alpha$  with the complete set of observations using a backward recursion is known as *smoothing*, which enables the inferences to be based on the entire sample. The importance of smoothing is based on (Koopman *et al.*, 1999):

- Signal extraction, detrending and seasonal adjustment;
- Diagnostics for separating the pieces of information between outlying observations and structural changes via auxiliary residuals;
- EM-algorithm for initial estimation of parameters;
- Calculation of the score vector.

There are several smoothing algorithms; the most basic one with socio-economical data is the *classical fixed-interval smoothing* algorithm, which is an off-line technique. The fixed-interval smoothing algorithm consists of a set of backward recursions starting with the final estimates  $\alpha_T$  and  $S_T$ , i.e. (Harvey, 1993, p. 87; Hamilton, 1994b, p. 394-397):

$$\alpha_{i|T} = \alpha_i + \tilde{S}_i (\alpha_{i+1|T} - \Pi_{i+1} \alpha_i - \mathbf{W}_i \boldsymbol{\beta}) \quad (22)$$

$$S_{i|T} = S_i + \tilde{S}_i (S_{i+1|T} - S_{i+1|T}) S_i' \quad (23)$$

where  $\tilde{S}_i = S_i \Pi_{i+1}' S_{i+1|T}^{-1}$  and  $\alpha_{T|T} = \alpha_T$  with  $S_{T|T} = S_T$ . More efficient versions of this algorithm are typically applied (e.g. Durbin & Koopman, 2002, p. 70-73).

### Functional Representation of Structural Time Series Models

A structural time series model generalises the hedonic model structure in variety of ways. For example, a single equation structural time series model can incorporate exogenous explanatory variables, lagged values of dependent and explanatory variables, and intervention variables, as well as unobserved components such as trend and cycle. Therefore a single equation structural time series model can be expressed as:

$$p_t = \mu_t + \psi_t + v_t + \sum_{\tau=1}^p \phi_\tau p_{t-\tau} + \sum_{\tau=0}^q \sum_{i=1}^k \beta_{\tau i} x_{t-\tau,i} + \sum_{j=1}^h \lambda_j w_{t,j} + \varepsilon_t \quad (24)$$

where  $x_{ij}$  is an exogenous variable,  $w_{ij}$  is an intervention variable,  $v_t$  is a first-order autoregressive component for the irregular disturbance term  $\varepsilon_t$ ,  $\mu_t$  is the trend,  $\psi_t$  is the cycle and  $\phi_\tau$ ,  $\beta_{\tau i}$  and  $\lambda_j$  are unknown parameters. As can explicitly be seen, *an unobserved component model opens up a wide range of possibilities for dynamic modelling*. A classical ordinary least squares regression emerges as a special case in which there exist no stochastic components apart from the irregular disturbance term.

In this study, the following functional representation for the hedonic model is used:

$$p_t = \mu_t + \psi_t + v_t + \sum_{i=1}^k \beta_i x_{t,i} + \sum_{j=1}^h \lambda_j w_{t,j} + \varepsilon_t \quad (25)$$

where  $\mu_t$  and  $\psi_t$  are stochastic and defined in the section 2;  $v_t$  is, in fact, a special case of  $\psi_t$  (see section 2).  $\varepsilon_t$  is irregular term, which is assumed to be white noise.  $\beta_i$  denotes the conventional unknown hedonic price of a land parcel and  $x_{t,i}$  is the  $i^{\text{th}}$  attribute at time  $t$ .  $\lambda_j$  represents the separate influence of an outlying observation, which is captured by the impulse intervention variable  $w_{t,j}$ . In the classical linear model structure for regression effects  $p_t \equiv p_t$  and  $x_{t,i} \equiv x_{t,i}$ , whereas for the common double-log model  $p_t \equiv \ln(p_t)$  and  $x_{t,i} \equiv \ln(x_{t,i})$ . These are subsequently put in the state-space form and the Kalman filtering and smoothing algorithms of (7)-(23) are used to derive the unknown coefficients.

## 4. STRUCTURAL ESTIMATION OF LAND PRICES

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. For example, in the modelling of land prices, there exists a significant amount of uncertainty concerning the choice of a proper response variable. Should the modeller use the original, total sales prices or the unit prices of a land parcel as the proper response vector and should these be analysed in a logarithmic or untransformed scale?<sup>9</sup>

<sup>9</sup> In this section only square or *unit prices* are analysed in both logarithmic and untransformed scales. This choice is based on previous empirical experimentation of the behaviour of land prices in the chosen submarkets. Furthermore, the use of unit prices is helpful in the model evaluation phase, which allows the calculation of some important weighted measures of goodness-of-fit statistics.

In modelling the time dimension of land prices there is always the vexed question of whether we should work with observations in levels or differences. Differenced values, albeit susceptible to losing important long-term price information, might sometimes be as such of direct interest, e.g. the rate of land prices could, in some cases, be defined as the first difference of the logarithm of prevailing price level. In certain situations it may be useful to consider lagged values of dependent price variable and even lagged values of explanatory variables, and if so differenced observations typically lead to greater numerical stability in estimation process since they tend to suffer less from multicollinearity problem. In this study, however, these kind of dynamic responses are not considered.

To improve theory-dependency of analysis, this empirical section considers estimation of two common econometric models of land prices - (i) classical linear model and (ii) multiplicative form of double-log model, see e.g. (Hannonen, 2005) - in the context of structural time-series modelling formalism. These econometric models are combined with the unobserved component models presented in the section 2; their hedonic prices are calculated and the associated statistical significance tested as usual applying the common Neyman-Pearson methodology. The goal is to conclude with a parsimonious model, which has orthogonal regressors and satisfies the necessary conditions for both congruence and encompassing. Furthermore, post-sample predictive accuracy of those models is analysed in section 6 and these results are compared to the predictions of their static analogues.

### Sample Data

Empirical studies of land prices tend to exhibit significant sensitivity to changes in data. Majority of that variation is explained by spatio-temporal movements: functional forms and parameters tend to vary with location and are not homogeneous throughout the data set, whereas temporally changing market conditions cause data-generating processes to evolve over time. To reduce the sample dependency, i.e. to improve the invariance of empirical study, the paper examines two data sets that are located in different submarkets and associated with partially non-overlapping time frames. The land type (land use) is, in contrast, fixed in order to reduce unnecessary heterogeneity of land prices. It represents *undeveloped land* not yet reached its highest and best use: vacant sites without a local detailed plan that are reserved for residential housing purposes.

The first sample data involve observations on land prices and the associated characteristics in the municipality of Espoo, a highly polycentric city, which lies inside the Helsinki metropolitan area with circa 225 000 inhabitants; 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, 1990 to December, 2001 with total number of observations of 400 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 39) are held back for post-sample predictive testing; a choice which is a somewhat arbitrary and mainly dictated by practical valuation concerns. In terms of quality, this data set is preferable over the second sample, since it has been pre-checked and screened by National Land Survey of Finland to ensure the sample data only include genuine arm's length transactions.

In Table 1 are documented some standard sample statistics for the study variables in the case of Espoo. The choice of relevant attribute variables is based on previous empirical knowledge of the phenomenon and could be termed here as, more or less, a fixed set covering only the most important ones in order to enchain model parsimony; some local distance measures, not to mention different metrics or inverse measures of proximity, complex interactions, many qualitative

**Table 1.** Some common sample statistics for the municipality of Espoo

Variable (unit)	Arithmetic mean	Minimum	Maximum	Std. Deviation
Total price (€)	59126.40	3027.00	756846.00	61976.88
Square price (€/m <sup>2</sup> ) (unit price)	22.99	0.24	127.55	19.09
Parcel size (m <sup>2</sup> )	4207.49	1000.00	28400.00	4613.75
Distance to CBD of Helsinki, L <sub>2</sub> -metric (km)	17.22	7.61	27.29	4.34
Quarterly price index of single-family houses	154.06	116.80	187.30	22.35
Parcel type (=0 if whole site; 1 otherwise)	-	0	1	-

features, neighbourhood attributes etc. are ignored as they are not the main focus here. The square price represents the unit price of a land parcel in euros per square meter, which is a variable derived from total sales price of a land parcel, and which is subsequently used as a proper regressand in the empirical study. Parcel size denotes the lot area measured in square metres. Distance to the CBD of Helsinki represents the straight-line distance (in L<sub>2</sub>-metric) from the central business district of Helsinki. Quarterly price index of single-family houses is a relative measure of quarterly price changes that is constructed from the sale prices of single-family houses in the Helsinki metropolitan area only. Parcel type is an indicator variable that receives a value of one, if the land parcel is sold as a whole and independent unit, and a value null otherwise.

The second sample contains observations of land prices and the associated attributes on 793 unbuilt land parcels without a local detailed plan sold in the period spanning from January, 1985 to March, 2004 in the municipality of Nurmijärvi, which lies just outside the Helsinki metropolitan area with approximately 36 000 inhabitants and three distinctive population centres (the parish village, Klaukkala and Rajamäki). Nurmijärvi has also recently witnessed several years of rapid expansion. The last 50 observations were left aside for evaluating the predictive accuracy of estimated local models outside the estimation sample, which corresponds approximately to a one-year period from the late February 2003 to the early March 2004. In terms of quantity, this data set offers more opportunities to flexible trend modelling and analysis of cyclical variation of land prices than the Espoo case. However, it is more pronounced to any errors (e.g. recording errors), since it has not been screened for specific hedonic modelling purposes. In Table 2 are documented some standard sample statistics for the study variables in the Nurmijärvi case. The study variables are otherwise the same as in the Espoo case, but there are also three additional local distance measures relating to the distances, respectively, from the population centres of parish village, Klaukkala and Rajamäki.

**Table 2.** Some common sample statistics for the municipality of Nurmijärvi

Variable (unit)	Arithmetic mean	Minimum	Maximum	Std. Deviation
Total price (€)	22 019.03	673.00	479 336.00	21 262.25
Square price (€/m <sup>2</sup> ) (unit price)	3.80	0.34	22.83	3.27
Parcel size (m <sup>2</sup> )	7387.36	1 000.00	30 000.00	4 651.66
Distance to CBD of Helsinki (km)	33.11	22.28	44.91	5.22
Distance to parish village of Nurmijärvi (km)	8.55	0.32	16.06	3.37
Distance to population center of Klaukkala (km)	9.33	0.50	21.44	5.00
Distance to population center of Rajamäki (km)	11.86	0.43	20.92	4.58
Quarterly price index of single-family houses	154.30	100.00	226.00	32.00
Parcel type (=0 if whole site; 1 otherwise)	-	0	1	-

## 5. RESULTS OF HEDONIC MODEL ESTIMATION

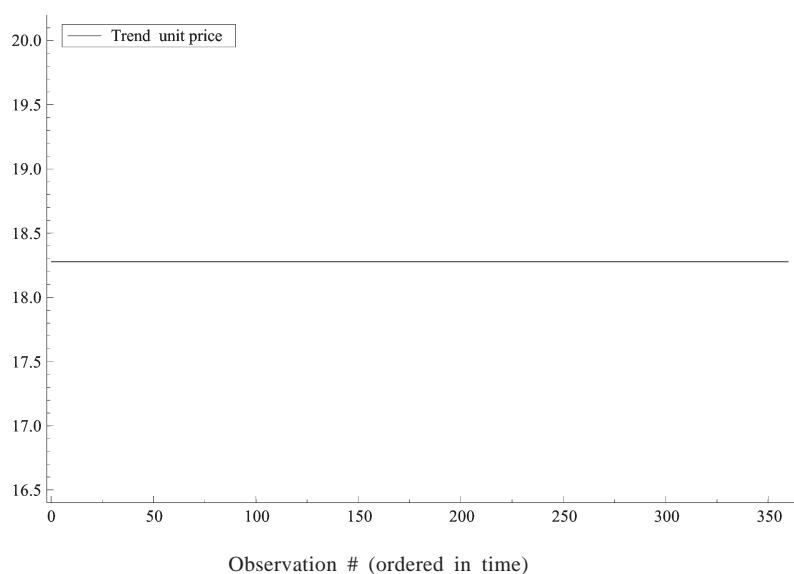
### Case Espoo

#### Linear Model

In this subsection estimation results for the classical linear model of regression effects are presented, which are simultaneously combined with different unobserved component models of the section 2 (local level model, local linear trend model and different cycle models). Kalman filter and smoother algorithms of the section 3 are then applied to estimate an appropriate structural time series model, which is a mixture of time series and regression. Only the results for the best-fitting model structure are documented.

After some empirical experimentation, a simple special case of *local level model* yielded the most adequate description for the trend component of the series; the slope coefficient is statistically insignificant. This local level model possesses an easily interpretable *fixed trend* around the value of 18.25 €/m<sup>2</sup>, which appears plausible constituting about 80% of the arithmetic average of the series (see Figure 1). Fixed trend means that *there exist no price fluctuations* in the trend component as time evolves, and that the slope coefficient is zero producing a deterministic line around the value of 18.25 €/m<sup>2</sup>. Three different sub-cycles are needed for reliably capturing the salient features of the temporal variation in conjunction with a first-order autoregressive process for irregular disturbances. This autoregressive process has the largest estimated variance (q-ratio is circa 0.38), whereas cycle components have estimated variances around 0.01-0.03. Three different impulse interventions are estimated and included to the final model, which represent the influence of discrepant data points. The results are highly sensitive to the effect of these outlying points, e.g. using only two of the most three significant intervention variables leads to a model structure with a small downward trend.

Table 3 documents the estimated hedonic prices and unobserved components. In essence, the effect of fixed trend is statistically relative small (t-value of 3.62), when compared to the statistical significance of the parcel size (t-value of -10.22), distance to the Helsinki CBD (t-value of -12.41)



**Figure 1.** Estimated trend for the linear model

**Table 3.** Estimated unobserved components and hedonic prices

Variable	Coefficient	r.m.s.e	t-value	p-value
Level	18.28	5.06	3.62	0.0003
AR(1)	-2.46	4.60	NA	NA
Cycle 1 (comp. #1)	-1.88	2.83	NA	NA
Cycle 1 (comp. #2)	0.86	2.90	NA	NA
Cycle 2 (comp. #1)	-4.78	2.65	NA	NA
Cycle 2 (comp. #2)	-0.97	2.75	NA	NA
Cycle 3 (comp. #1)	-0.041	3.00	NA	NA
Cycle 3 (comp. #2)	1.63	3.44	NA	NA
Parcel size	-0.0014	0.00013	-10.22	0.0000
Distance	-1.90	0.15	-12.41	0.0000
House price index	0.28	0.032	8.58	0.0000
Intervention #1	68.93	11.00	6.27	0.0000
Intervention #2	51.34	11.00	4.67	0.0000
Intervention #3	46.40	10.97	4.23	0.0000

and the house price index (t-value of 8.58). All the estimated hedonic coefficients appear plausible in sign and magnitude. The relatively small effect of level component is, at least partly, explained by the inclusion of house price index variable, which effectively captures much of the same temporal variation as the estimated trend component. The house price index variable, however, appears to be a genuine explanatory variable in this model structure as without it the trend component would oscillate at too high a level.

The effects of cycle components are even smaller, but they are needed for the estimation of a proper model structure. Cycles 1 & 2 represent high-frequency components of temporal variation, and as such are not very informative, whereas cycle 3 captures a lower frequency variation (indicating larger oscillations in the beginning of the series, which sounds reasonable). The influence of parcel type is statistically insignificant at conventional risk levels and thus is not included to the final model.

### Goodness-of-fit Statistics

Table 4 presents some fundamental information of the relevant goodness-of-fit statistics. PEV is the prediction error variance, which corresponds to the variance of standardised one-step ahead prediction errors, or generalised least squares residuals if regression effects are present, in steady state, or if steady state is not obtained a finite PEV is reported. PEMD is the prediction error mean deviation.  $R^2$  and  $R_d^2$  are, respectively, the classical coefficient of determination statistic

**Table 4.** Goodness-of-fit statistics for the linear model

Goodness-of-fit statistic	Value
$R^2$	0.61
$R_d^2$	0.77
AIC	4.91
BIC	5.12
PEV	123.13
PEMD	91.87

and the modified coefficient of determination statistic; the latter measure compares the fit with a simple random walk plus drift model. In essence, the  $R^2$  does not reach the usual critical value of 0.70 applied in Finnish land valuation studies and, furthermore, the scaled square root of PEV is circa 0.48, which exceeds the commonly used target value of 0.30.

CUSUM and CUSUMSQ statistics indicate that there is no change in the mean process, but significant fluctuations in the variance of the process. Correlogram along with Box-Ljung statistic indicate that the estimated model is not plagued with any serious autocorrelation. QQ-plot and the histogram indicate, moreover, that residuals are closely normally distributed, which, however, confounds the information given by the formal tests of normality (normal-BS and normal-DH indicators) that indicate departures from normality.

### Double-log Model

Table 5 documents the estimated hedonic prices and unobserved components that are obtained using Kalman filtering and smoothing algorithms of the section 3. The multiplicative form of double-log model is used to capture the regression effects and, at the same time, a local level model is chosen by empirical experimentation for the trend component. Two different cycles and a first-order autoregressive process of error terms are also included into the final hedonic model in order to accurately represent the joint spatio-temporal variation of land prices.

The effect of level is now much more pronounced (t-value circa 28), albeit the slope is still statistically insignificant (and thus not incorporated into the model of Table 5). This means that the most adequate description of the salient features of the trend variable is still the simplest unobserved component model: the local level model. This time the local level model produces a highly curvilinear trend, which shows a clear downward and upward movements over time. This is illustrated in the Figure 2. It is noteworthy that now the trend fluctuates at much lower level, approximate in the range of [11.75; 13.25], which represents only 50-60% of the arithmetic average of the series. This is unusually low, albeit it is a well-known phenomenon that arithmetic average tends to exaggerate the price level due to problem of outlying observations that typically causes price distribution to be highly skewed to the right. There is a pronounced dip in the series, which corresponds to the first quarter of 1995 after which series has grown upward for circa five years. The dominant peak occurs at the beginning of the series: in the second half of the year 1990.

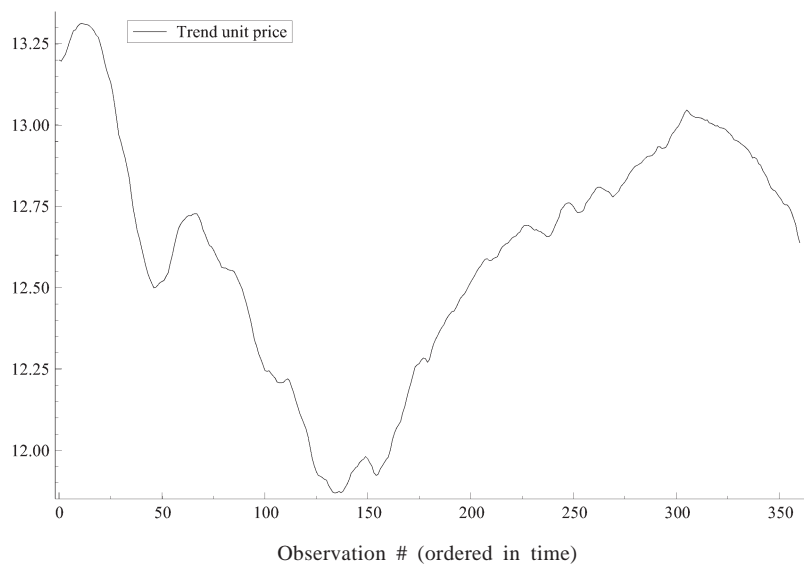
Cycles tend to be an integral part of the temporal structure; two different cycle components are needed, which both, more or less, represent high-frequency part of cyclical variability. The

**Table 5.** Estimated unobserved components and hedonic prices

Variable	Coefficient	r.m.s.e	t-value	p-value
Level	12.64	0.45	27.97	0.0000
AR(1)	-1.77	0.20	NA	NA
Cycle 1 (comp. #1)	0.032	0.033	NA	NA
Cycle 1 (comp. #2)	0.084	0.033	NA	0.0216
Cycle 2 (comp. #1)	-0.20	0.12	NA	NA
Cycle 2 (comp. #2)	-0.069	0.13	NA	NA
Parcel size	-0.76	0.039	-19.60	0.0000
Distance	-1.31	0.14	-9.59	0.0000
Parcel type	-0.31	0.061	-5.15	0.0000
Intervention #1	-2.70	0.54	-4.95	0.0000



other cycle is deterministic with a corresponding p-value of circa 0.02. AR(1) is also important element of the temporal dimension, which now has the highest estimated variance among all unobserved components (q-ratio is 1.00). The effect of parcel size is concave and statistically much higher (t-value of -19.60) than in the case of linear model (almost twice as much). The estimation of the aggregate price index variable failed in a sense that it produced an implausible description for the unobserved components, and was therefore removed from the final model structure. The influence of parcel type (t-value of -5.15) is statistically highly significant and the effect, as expected on the grounds of urban economic theory, of straight-line distance to the Helsinki CBD is convex (with t-value of -9.59). The estimated signs and magnitudes of all hedonic coefficient seem intuitively plausible.



**Figure 2.** Estimated trend for the double-log model

**Goodness-of-fit Statistics**

Table 6 collects some fundamental information of the relevant goodness-of-fit statistics. Overall they show relatively good fit; both coefficient of determination statistics exceed the usual cut-off rate of 0.70 applied in the Finnish land valuation, albeit the deviatation-based measures clearly exceed the typical target value of 0.30.

**Table 6.** Goodness-of-fit statistics for the linear model

Goodness-of-fit statistic	Value
$R^2$	0.70
$R_d^2$	0.83
AIC	-1.12
BIC	-0.97
PEV	0.29
PEMD	0.22

CUSUM and CUSUMSQ statistics indicate that there is no change in the mean and the variance of land price process. This is as it should be. Correlogram along with Box-Ljung statistic indicate that the estimated model is not plagued with autocorrelation. QQ-plot and the histogram indicate minor departures of the normality assumption, which is further evidenced by the formal tests of normality (normal-BS and normal-DH indicators).

## Case Nurmijärvi

### Linear Model

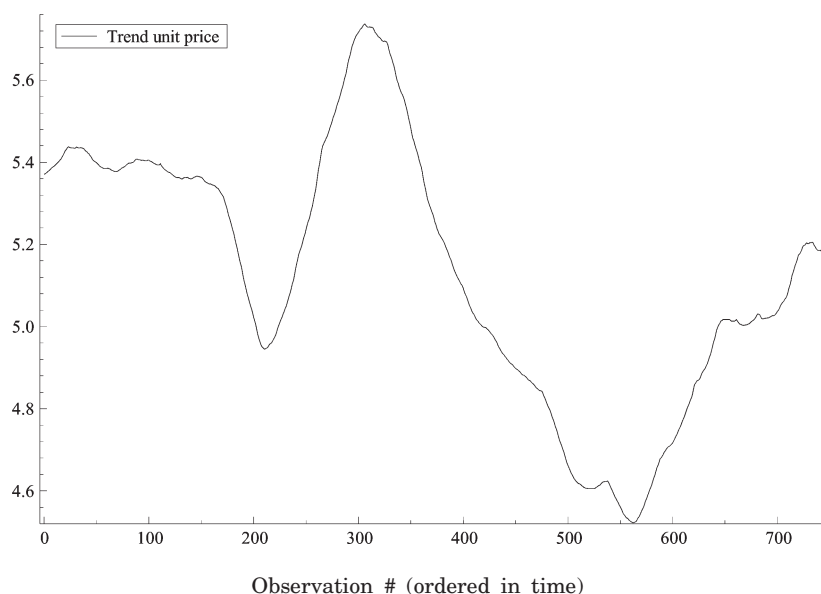
Estimated hedonic prices and unobserved components are documented in the Table 7 for the classical linear model structure, i.e. for the untransformed land price and attribute values. Different unobserved components models (local level model, local linear trend model and different cycle models of the section 2) with regression effects are put into the state-space form and then estimated by Kalman filtering and smoothing algorithms of the section 3.

Some empirical experimentation reveals that the local level model captures most adequately the salient features of the trend component, in which a stochastic level term is subsumed. The estimated trend, whose statistical significance is relatively low with the associated t-value of 3.26, is depicted in the Figure 3. There is an evident peak in the middle of the series, which corresponds to the second half of the year 1990, after which the whole Finnish economy slid into the great depression; the figure shows circa eight years of solid and rapid downward movement. The new upturn occurred in the last quarter of 1998. The price level in the figure is surprisingly high, about 120-150% of the arithmetic average value of the series in that period. It should be stressed again that the estimated p-value for that level component is quite high (0.0012), which increases uncertainty related to the inferences. This is explained partly by the inclusion of price index variable, which effectively reflects different temporal movements.

One relatively low-frequency cycle component is included into the estimated model, which shows expanding and contracting oscillations of prices over time. Also a first-order autoregressive component is included, which has the largest estimated variance (q-ratio of 1.00). The most important explanatory variable, statistically speaking, is the parcel size variable with the t-value of

**Table 7.** Estimated unobserved components and hedonic prices

Variable	Coefficient	r.m.s.e	t-value	p-value
Level	5.18	1.59	3.26	0.0012
AR(1)	-1.96	1.01	NA	NA
Cycle 1 (comp. #1)	-0.31	0.53	NA	NA
Cycle 1 (comp. #2)	0.13	0.55	NA	NA
Parcel size	-0.00028	1.86e-005	-14.97	0.0000
Distance 1	-0.12	0.016	-7.53	0.0000
Distance 2	-0.096	0.025	-3.83	0.0001
Parcel type	-0.75	0.18	-4.18	0.0000
Price index	0.038	0.0072	5.30	0.0000
Intervention #1	13.37	2.27	5.88	0.0000
Intervention #2	13.74	2.27	6.03	0.0000
Intervention #3	11.68	2.27	5.13	0.0000
Intervention #4	13.03	2.27	5.73	0.0000
Intervention #5	11.84	2.27	5.20	0.0000



**Figure 3.** Estimated trend for the linear model

-14.97. Two distance measures are included into the final hedonic model, where distance 1 denotes a straight-line distance to the CBD of Helsinki and distance 2 represents a straight-line distance to the centre of parish village of Nurmijärvi (This seems be the only relevant local distance measure here; if other local distance measures are added to the model, the multicollinearity problem prevents much of the meaningful subsequent analysis). Both distance measures seem plausible in sign and magnitudes with the associated p-values of -7.53 and -3.83, respectively. In contrast to the results of the Table 3, the classical linear model for the Espoo case, now the parcel type is statistically highly significant variable with the t-value of -4.18. Five outlying observations were detected and their influences were captured using impulse intervention variables; in sum, their total effect on land prices is very high.

**Goodness-of-fit Statistics**

Table 8 collects some fundamental information of the relevant goodness-of-fit statistics. Overall they show relatively poor fit; the standard coefficient of determination measure  $R^2$  is clearly below usual cut-off rate of 0.70 applied in many land valuation assignments and the deviatation-based measures clearly exceed the typical target value of 0.30.

**Table 8.** Goodness-of-fit statistics for the linear model

Goodness-of-fit statistic	Value
$R^2$	0.46
$R_d^2$	0.70
AIC	1.68
BIC	1.79
PEV	5.16
PEMD	3.54

CUSUM and CUSUMSQ statistics indicate that there is no change in the mean process, but oscillations in the variance of land price process. Correlogram along with Box-Ljung statistic indicate that the estimated model is not plagued with autocorrelation. QQ-plot and the histogram indicate departures from the normality assumption, which is further evidenced by the formal tests of normality (normal-BS and normal-DH indicators).

### Double-log Model

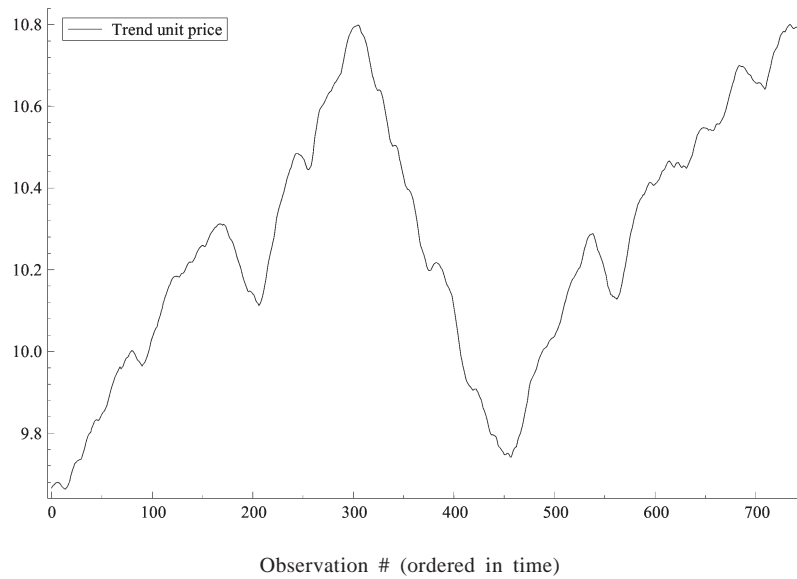
Table 9 documents the estimation results of common hedonic prices and unobserved components for the multiplicative form of double-log model, i.e. for the conventional power function approach. These regression effects are simultaneously included with different unobserved component models of the section 2 and then estimated using Kalman filter and smoothing recursions of the section 3. Some empirical investigation indicates that the effect of trend component is most accurately encapsulated by the local level model (as in all the previous cases).

The estimated trend is depicted in the Figure 4; the estimated trend now fluctuates at a notoriously high level when compared to the average value of the series in that period (circa 2.5 higher than the arithmetic average). There is the same peak discernible as in the Figure 3, which corresponds to the second half of the year 1990; otherwise the behaviour of these trends is somewhat different. The dominant dip of the series takes place about the turn of the years 1995 and 1996, after which the series has moved upward about seven years reaching the high price level of 1990. This seems to be, however, the most valid model structure here among the contenders. If the house price variable (that is excluded from the model) would be included as a proper explanatory variable, the estimated trend would be much more coherent with respect to average value of the series. However, in that case all the slope coefficients would also be statistically insignificant, which is undesirable.

Two cycle components were included to the model, of which the other captures high-frequency component and the other picks up lower-frequency element of price movements over time. All the estimated hedonic prices of attribute variables are concave. Parcel size and level component are clearly statistically the most important determinants of land price fluctuations with t-values of -18.92 and -18.87, respectively. Other explanatory variables have significantly (almost three times) lower t-values. Distance 1 is a straight-line distance to the Helsinki CBD and distance 2 denotes

**Table 9.** Estimated unobserved components and hedonic prices

Variable	Coefficient	r.m.s.e	t-value	p-value
Level	10.78	0.57	18.87	0.0000
Cycle 1 (comp. #1)	0.014	0.087	NA	NA
Cycle 1 (comp. #2)	0.026	0.087	NA	NA
Cycle 2 (comp. #1)	-0.0051	0.054	NA	NA
Cycle 2 (comp. #2)	0.018	0.054	NA	NA
Parcel size	-0.62	0.032	-18.92	0.0000
Distance 1	-0.93	0.12	-7.29	0.0000
Distance 2	-0.22	0.042	-5.12	0.0000
Parcel type	-0.21	0.043	-5.02	0.0000
Intervention #1	-2.34	0.54	-4.28	0.0000
Intervention #2	-2.81	0.54	-5.11	0.0000



**Figure 4.** Estimated trend for the double-log model

a straight-line distance to the centre of the parish village of Nurmijärvi. The separate effects of two outlying observations were estimated by impulse intervention variables.

**Goodness-of-fit Statistics**

Table 10 reports some fundamental information of the relevant goodness-of-fit statistics. Overall they show relatively poor or moderate fit that mimics quite closely the performance of the corresponding linear model for the Nurmijärvi case; the standard coefficient of determination measure  $R^2$  is clearly below usual cut-off rate of 0.70 applied in land valuation and the deviation-based measures clearly exceed the typical target value of 0.30. AIC and BIC statistics are, however, lower indicating a better fit.

CUSUM and CUSUMSQ statistics indicate that there is no change in the mean process and only minor oscillations in the variance of land price process, which is an improvement over the model of the Table 8. Correlogram along with Box-Ljung statistic indicate that the estimated model is not plagued with autocorrelation. QQ-plot and the histogram indicate departures from the normality assumption, which is further evidenced by the formal tests of normality (normal-BS and normal-DH indicators).

**Table 10.** Goodness-of-fit statistics for the linear model

Goodness-of-fit statistic	Value
$R^2$	0.46
$R_d^2$	0.68
AIC	-1.15
BIC	-1.06
PEV	0.30
PEMD	0.23

## 6. PREDICTIVE ASSESSMENT OF THE HEDONIC MODELS

### Measures of Predictive Accuracy

Forecasting accuracy is a critical factor in the evaluation of empirical validity of the chosen hedonic model. 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. However, not all measures can be considered equally informative in the context of hedonic modelling of land prices: *of paramount importance are those criteria underlying the mean prediction error and the strength of association between model's predictions and observed land prices.* Variance-based indicators, albeit commonly utilised in studies, are secondary for most practical land valuation purposes.

Mean prediction error is evaluated in this study by (1) arithmetic average prediction error, (2) median prediction error, (3) error centre of gravity and (4) by arithmetic average and median percentage errors. If unit land prices are analysed (as in this paper), the necessary weights are directly obtained by using parcel size to give the total mass of errors, which obviates the need of estimating the unknown density function. The use of average prediction errors is often, in land value studies, a more plausible alternative than median-based indicators of central tendency of prediction errors, since latter measures down-weight too heavily the influence of unusual prediction errors as compared to typical prediction errors since they ignore relevant information about distances between errors. Also median-based mean error indicators tend to produce, in many cases, highly similar and indecisive results between different model candidates. The major problem of using arithmetic average prediction error or median prediction error, and their common weighted analogues, is that these are absolute measures, which depend on the unit of measurement. Therefore, the relative versions of the basic indicators, i.e. arithmetic average and median percentage errors, are also documented for comparing mean prediction errors between models, where responses are measured on a different scale.

Three 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 error terms and when the focus is on the co-variation of errors. 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, tend to lead to highly similar standard deviations between different model alternatives. Consequently, calculated correlation coefficients tend to produce not only invalid but also indecisive results. The use of scaled conventional inner product (or some weighted version of it, but this seems to add very little new information) is preferable, which measures the strength of prediction in the direction of actual out-of-sample land prices scaled by the total uncentered variation of land prices. It generates more valid and decisive results that are not strongly dependent on any particular distributional assumptions. Also the gravity (see McMillen, 2001) is reported, which seems to be a viable measure of strength of association; it usually produces results that parallel the use of scaled inner product.

Root mean squared error is the most commonly used measure of success of numeric prediction, which mainly controls the reliability or variability of predictions, not the actual predictive validity or predictive unbiasedness. This statistic is also overly 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 valuation studies). This sensitivity or lack of robustness to extreme errors can cause root mean square error to produce results that are

indecisive. Mean absolute error is generally a more appropriate indicator of predictive variability, and is especially suitable in cases of outlying prediction errors. Another useful measure of predictive variability is simply the predictive range expressed in terms of minimum and maximum predictions; shorter the interval, less variability is expected in predictions. Widely used measure of predictive variability is mean absolute percentage error (see e.g. Makridakis and Hibon, 2000), which is reported here with the associated robust version of it: median absolute percentage error. They have also been criticized for the problems of asymmetry and instability, when the data is small.

### Case Espoo

Table 11 documents the forecasting performance of conventional (ordinary least squares) and structural modelling approaches for the classical linear model structure, i.e. for the untransformed land prices. Estimated structural time-series model encompasses the conventional hedonic model in most measurable aspects; thirteen (out of fifteen) different measures of predictive accuracy are superior in the structural case. All mean prediction error measures are lower with the estimated unobserved components model and, in particular, three of the most important ones, average prediction error, error centre of gravity and average percentage indicators, are, respectively, circa 48%, 28% and 49% lower than in case of the ordinary least squares' estimate. This implies that structural approach produces results of higher predictive validity. Common variance-based measures of predictive accuracy: mean absolute error, mean absolute percentage error and root mean squared error, are also about 5-18% lower with the structural modelling approach, depending on the measure used. Measures relating to the strength of association between forecasts and actual outcomes - namely the scaled inner product and gravity - are slightly better (5% and 0.6%, respectively) when conventional least squares are applied, but these minor improvements cannot offset an otherwise relatively low performance.

In the Table 12 are presented the corresponding measures of the Table 11 for the multiplica-

**Table 11.** Prediction results for the classical linear models

Measure of predictive accuracy	Modelling Approach	
	Conventional	Structural
Average prediction error	-4.66	-2.44
Median prediction error	-10.44	-8.07
Error centre of gravity	-5.45	-3.95
Average percentage error	17.41	9.11
Median percentage error	64.78	50.04
Predictive range [min, max]	[-32.96; 79.13]	[-30.31; 80.61]
Mean absolute error	15.02	13.80
Median absolute error	13.20	10.97
Mean absolute percentage error	256.56	211.50
Median absolute percentage error	63.59	69.35
Root mean squared error	19.72	18.82
Root median squared error	13.20	10.97
Correlation (Pearson)	0.75	0.78
Scaled inner product	0.78	0.74
Gravity	$3.62 \cdot 10^9$	$3.60 \cdot 10^9$

**Table 12.** Prediction results for the double-log models

Measure of predictive accuracy	Modelling Approach	
	Conventional	Structural
Average prediction error	-0.37	0.15
Median prediction error	-0.14	0.29
Error centre of gravity	-0.37	0.14
Average percentage error	13.84	5.45
Median percentage error	0.14	10.60
Predictive range [min, max]	[-2.45; 0.88]	[-2.00; 1.41]
Mean absolute error	0.51	0.58
Median absolute error	0.25	0.51
Mean absolute percentage error	66.16	56.99
Median absolute percentage error	9.23	17.69
Root mean squared error	0.78	0.72
Root median squared error	0.25	0.51
Correlation (Pearson)	0.85	0.84
Scaled inner product	1.04	0.89
Gravity	4583.33	4210.65

tive form of double-log model, i.e. for the transformed land prices. Now the model assessment is slightly more difficult. The most important mean prediction error statistics - error centre of gravity, average percent error and average prediction error - are significantly lower; all about 60%. The use of median-based mean prediction error indicators gives contrary information, but they are secondary. Variance-based indicators give somewhat indecisive information: mean absolute error is somewhat higher in the unorthodox case of unobserved components, but commonly applied measures of mean absolute percentage error and of root mean squared error are, on the other hand, slightly lower. Moreover, gravity and scaled inner product are slightly better in the case of ordinary least squares (9% and 17%, respectively), which is explained partly, I believe, by use of price index variable in the conventional approach (the variable which was left aside from the final model in the structural case of double-log model). This increases the strength of statistical association in the static model, but only at the expense of biasing the mean prediction errors further. In essence, the structural model seems to be the most adequate description of land prices in the Table 12, albeit it encompasses the conventional model in only a few measurable respects. Specifically, the necessary property of predictive unbiasedness could be improved significantly if unobserved components are used.

### Case Nurmijärvi

Table 13 reports the results of post-sample predictive competition for the classical linear structure. The conventional methodology yields to superior results; the ordinary least squares estimator encompasses the unobserved component method by thirteen (out of fifteen) different out-of-sample measures of predictive accuracy. The three most important mean prediction error statistics - error centre of gravity, average percent error and average prediction error - are significantly lower; about 24-57% depending on the measure. All variance-based measures are also slightly better in the conventional case. Only two statistics of the strength of association are slightly higher when structural modelling paradigm is applied.



**Table 13.** Prediction results for the classical linear models

Measure of predictive accuracy	Modelling Approach	
	Conventional	Structural
Average prediction error	-0.16	-0.37
Median prediction error	-0.81	-1.11
Error centre of gravity	-0.87	-1.14
Average percentage error	2.83	6.59
Median percentage error	19.65	27.16
Predictive range [min, max]	[-5.51; 13.47]	[-5.61; 13.44]
Mean absolute error	2.26	2.40
Median absolute error	1.85	2.05
Mean absolute percentage error	96.60	107.06
Median absolute percentage error	35.88	41.89
Root mean squared error	3.24	3.32
Root median squared error	1.85	2.05
Correlation (Pearson)	0.70	0.69
Scaled inner product	0.76	0.78
Gravity	$3.04 \times 10^9$	$3.11 \times 10^9$

Table 14 documents the relative performance of the conventional and structural formalisms for the multiplicative form of double-log model. In essence, the structural model improves significantly the mean prediction error measures; average prediction and average percentage error indicators are both 99.9% (!) lower in the structural case. Moreover, error centre of gravity is about 85% lower in the structural case. Variance-based indicators are almost the same producing indecisive information. Furthermore, there are only minor departures between different strength of association indicators so that a strict preference cannot be made. It seems quite clear that significant benefits results from using unobserved component model.

**Table 14.** Prediction results for the double-log models

Measure of predictive accuracy	Modelling Approach	
	Conventional	Structural
Average prediction error	-0.052	0.000077
Median prediction error	0.041	0.080
Error centre of gravity	-0.061	-0.0093
Average percentage error	3.57	0.0053
Median percentage error	2.92	5.64
Predictive range [min, max]	[-2.30; 1.05]	[-2.28; 1.16]
Mean absolute error	0.43	0.44
Median absolute error	0.26	0.30
Mean absolute percentage error	382.65	379.14
Median absolute percentage error	18.47	19.42
Root mean squared error	0.64	0.64
Root median squared error	0.26	0.30
Correlation (Pearson)	0.67	0.67
Scaled inner product	0.86	0.83
Gravity	1919.50	1850.68

## 7. CONCLUSIONS

This study has analysed the spatio-temporal variation of land prices in two single localities by means of structural time series modelling formalism that combines the flexibility of a time series model with that of the interpretation of a regression analysis. Empirical investigation evidenced that such trend and cycle specifications could be identified in most model formulations, which resulted significant improvements in their post-sample predictive validity. The clear exception was the case of classical linear model structure for regression effects in the municipality of Nurmijärvi, in which the conventional modelling approach outperformed the structural one. Overall, these improvements in the post-sample predictive accuracy were more pronounced and coherent in the municipality of Espoo, whose data were screened and pre-checked for any errors prior to the analysis.

The local level model seemed to be the most appropriate description of the trend component in all model formulations. The effect of slope component in the trend specification was found to be statistically insignificant. The behaviour of estimated trends was more plausible in the Espoo case, where the price level component oscillated approximately in the range of (12 €, 18 €) in the study period. In the submarket of Nurmijärvi, the estimated models tended to overshoot the influence of trend component; the estimated level fluctuated around (4.5 €, 11 €), which is significantly higher than the average value of the series in that period. One to three different cycle components were identified in all model formulations. These all seemed to capture, more or less, a high-frequency component of land price movements, and as such, may not be very informative. However, some resonance was discernible in some cycle models. The effects of estimated trends and cycles were in most cases highly curvilinear and quite volatile. AR(1) process for irregular disturbances was found to be an integral part of the overall model in three different case.

The data analysed contained many outlying observations in terms of an unusual high value of standardised residual, which distorted the analysis - especially in the case of Nurmijärvi. 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. The price index variable confounded the analysis of unobserved component models, when transformed land prices were analysed. Goodness-of-fit statistics indicated that the conventional double-log model yielded to a more data-congruent specification than the classical linear model. In that model structure, beyond any reasonable doubt, the two most important determinants of land prices were the estimated level term and parcel size variable.

The main implication of this study is that the structural time-series modelling paradigm offers a more viable alternative to the hedonic analysis of land prices than the conventional approach based on ordinary least squares. It provides more accurate estimates of hedonic prices along with the additional information about unobserved components than the orthodox approach. In particular, the use of unobserved component models can broaden our general understanding of the nature of long-term land price movements and short-run land price fluctuations. The occurrence of stochastic trends and/or cycles in the analysed cases suggests that these markets are not in the steady state but continuously evolving, which would seriously undermine the applicability and performance of conventional modelling approaches in the land market.

## REFERENCES

- Case, B., Clapp, J., Dubin, R. and Rodriguez, M. (2004) Modelling Spatial and Temporal House Price Patterns: A Comparison of Four Models. *Journal of Real Estate Finance and Economics*, 29(2), p. 167–191.
- Chen, M-Ch., Kawaguchi, Y. and Patel, K. (2004) An Analysis of the Trends and Cyclical Behaviours of House Prices in the Asian Markets. *Journal of Property Investment and Finance*, 22(1), p. 55–75.
- Day, R. H. (1992) Complex Economic Dynamics: Obvious in History, Generic in Theory, Elusive in Data. *Journal of Applied Econometrics*, 7, p. 9–23.
- 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), p. 179–208.
- Hamilton, J. D. (1994a) *State-Space Models*, Handbook of Econometrics, Vol. 4, Engle, R. F. & McFadden, D. L. (eds.), Elsevier Science B.V.
- Hamilton, J. D. (1994b) *Time Series Analysis*, Princeton University Press.
- Hannonen, M. (2005) On the Recursive Estimation of Land Prices. *Nordic Journal of Surveying and Real Estate Research* (to appear).
- Harvey, A. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
- Harvey, A. C. (1993) *Time Series Models*, Second edition, Harvester Wheatsheaf.
- Harvey, A. C. (1997) Trends, Cycles and Autoregressions. *The Economic Journal*, 107 (January), p. 192–201.
- Harvey, A. C. and Shephard, N. (1993) *Structural Time Series Models*, Maddala, G. S., Rao, C. R. & Vinod, H. D. (eds.), Handbook of Statistics, Vol. 11, Elsevier Science Publishers B.V.
- Hiltunen, A. (2003) On the Price Formation of Single-Family House Properties: An Analysis of Comparable Sales Prices for Single-Family Houses in Some Regional Areas in Finland, Doctoral Thesis, Helsinki University of Technology, Espoo. (In Finnish).
- IVSC (2001). *International Valuation Standards 2001*, International Valuation Standards Committee.
- de Jong, P. (1989) Smoothing and Interpolation with the State-Space Model. *Journal of the American Statistical Association*, 84(408), p. 1085–1088.
- de Jong, P. (1991) The Diffusive Kalman Filter. *Annals of Statistics*, 19(2), p. 1073–1083.
- Kalman, R. E. (1960) A New Approach to Linear Filtering and Prediction Problems. *Transactions ASME Journal of Basic Engineering*, D82, p. 35–45.
- Kim, Ch-J. and Nelson, Ch. R. (1999) *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, The MIT Press.
- 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.
- Makridakis, S. and Hibon, M. (2000) The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting*, 16, p. 451–476.
- McMillen, D. P. (2001) Nonparametric Employment Subcenter Identification. *Journal of Urban Economics*, 50, p. 448–473.
- Schulz, M. A. R. (2003) Valuation of Properties and Economic Models of Real Estate Markets, Dissertation, 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), p. 37–57.
- Schwann, G. M. (1998) A Real Estate Price Index for Thin Markets. *Journal of Real Estate Finance and Economics*, 16(3), p. 269–287.

**SANTRAUKA****ŽEMĖS KAINOS ANALIZĖ: STRUKTŪRINIS LAIKO EILUČIŲ METODAS****Marko HANNONEN**

Šiame darbe analizuojami erdviniai ir laiko žemės kainų pokyčiai dviejose skirtingose vietovėse, taikant struktūrinį laiko eilučių modelį, derinantį ir šio modelio lankstumą, ir regresinės analizės interpretavimą. Įprastinių hedoninių modelių praplėtimas nestebimais trendo ir ciklo komponentais reikšmingai pagerino prognozavimo tikslumą poatrankinio tyrimo metu. Nestebimų komponentų naudojimas daugelyje modelių, atliekant prognostinius bandymus, davė tik nedidelę vidutinę prognozavimo paklaidą, o tradiciniai hedoniniai modeliai pateikdavo daug sisteminių prognozavimo paklaidų. Panašu, kad struktūrinių laiko eilučių modelio paradigma yra pranašesnė už hedoninę žemės kainų metodiką, pagrįstą įprastine mažiausiųjų kvadratų analize. Nustatyta, kad nuolydžio komponento efektas trendo specifikacijoje yra statistiškai nereikšmingas, o tai reiškia, kad elementaraus lokalinio lygmens modelis adekvačiausiai aprašytų ilgalaikius žemės kainos pokyčius.