Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP04/2013

Econometric Modeling and Estimation of Theoretically Consistent Housing Price Indexes

Alicia N. Rambaldi and D.S. Prasada Rao

Date:
July 2013

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Econometric Modeling and Estimation of Theoretically Consistent Housing Price Indexes*

Alicia N. Rambaldi† and D.S. Prasada Rao‡

Abstract

Recent developments in the economic theory behind hedonic price models and price index numbers have shown that the preferred combination is one where hedonic imputed price indexes (HI) are computed using predictions from time-varying hedonic functions. This paper proposes a spatial time series model as the econometric model consistent with the theoretical developments. In addition, the paper deals with issues relating to HI index numbers including weighting systems, seasonality in housing sales data, and the construction of annual and monthly chained indexes.

Keywords: state-space, spatial time series, hedonic imputation, housing prices, Tornqvist, Fisher.

JEL: C43; C23; C53

---

*A much earlier version of some of this work appeared as: Rambaldi, A.N. and D.S. Prasada Rao, "Hedonic Predicted House Price Indices Using Time-Varying Hedonic Models with Spatial Autocorrelation," School of Economics Discussion Paper No. 432. July 2011, School of Economics, The University of Queensland. Australia.

†Corresponding Author: School of Economics, The University of Queensland. St Lucia, QLD 4072. Australia. Tel.:+61 7 33656576; Fax.: +61 7 33657299. Email: a.rambaldi@uq.edu.au

‡School of Economics, The University of Queensland. St Lucia, QLD 4072. Australia.

1 Introduction

Recently derived results in the price index literature (Silver and Heravi (2007), Hill and Melser (2008)) show that the class of indexes known as *hedonic imputed (HI) price indexes* are consistent with the theoretical foundations of the "hedonic approach" to price index construction. This approach requires hedonic coefficients, which are implicit prices attached to characteristics, to vary over time. The HI price index method is more flexible than the traditional time dummy hedonic method and it is related to the standard matched-model methodology of constructing a price index. The assets sold and priced in period 0 are also priced in period 1 using estimated regression coefficients (and vice versa) in order to construct the index. The repeat sales method is a matching method; however, there are a number of well documented disadvantages with the use of repeat sales methods stemming from the need to use only a small sub-set of the observed transactions [1].

Imputing a price in period 1 for an asset sold in period 0 requires an econometric model based prediction. Consequently the first and foremost recommendation is against using regressions with fixed parameters to construct the predictions. Triplett (2004) argued for the use of a "two period adjacent" (or "rolling two-period) estimation of the hedonic model used to impute prices necessary for HI and labeled it best practice. Hill and Melser (2008) show that indexes based on fixed parameter hedonic regression models are biased and that HI indexes suffer from bias unless the hedonic parameters are allowed to vary over time and over heterogeneous regions. In fact, there are economic theoretic reasons to support this argument as well. Diewert (2003), in an excellent paper, provides a detailed economic theoretic framework for hedonic regressions based on standard consumer theory incorporating vectors of characteristics of the products consumed. Following on from the work of Silver (1999) and Pakes (2001), Diewert recommends the use of a time varying functional form for the hedonic regression model so that the shadow prices of hedonic characteristics can vary over time.

These utility theoretic and price index number related recommendations have led to the advocacy and adoption of Triplett (2004)'s two-period adjacent approach to the estimation of the regression using a rolling window. This recommendation is not new as the use of a rolling window approach to the estimation of the hedonic model dates back to suggestions made by Court (1939) and Griliches (1961) in their attempts to make allowance for the time-varying nature of the hedonic utility function. However, the rolling window approach is conceptually inconsistent as a strict implementation of the method should yield constancy of the hedonic coefficients over all the years, and, as discussed in the next section, it produces two estimates of the parameters for each time period.

The objectives of this paper are twofold. The first is to propose an econometric framework consistent with a time-varying hedonic utility function. The paper also provides price predictions obtained through an optimum estimator of the pricing model. A second objective is to make two innovations in the direction of the actual construction of housing price indexes based on the imputed prices from the estimated hedonic models.

In this paper the rolling window approach is used as the base to compare our proposed modeling and estimation framework, specially in relation to prediction performance. The predictions from the proposed econometric modeling are used to compute monthly hedonic imputed property price indexes. The main recommendations of Hill and Melser (2008) are used in that we focus mainly on the Fisher and Tornqvist index number formula. However, we make two innovations in this area. The first is that we consider two alternative types of index weights, *plutocratic weights* based on value shares of houses sold and *democratic weights* which are based on the number of houses sold. Recognizing the difference between housing price index numbers and the standard cost-of-living index numbers, we argue that both types of weights are valid. An important feature of property sales prices is the presence of seasonality in the mix of properties sold and its influence on central tendency measures such as the median of observed property prices. In accounting for seasonality, our paper constructs year-on-year monthly price index numbers using hedonic imputations and compares these with annual time-dummy hedonic price indexes. The method and procedures proposed in the paper are illustrated using multiple sales data from the city of Brisbane, Australia, and it is expected that these methods would be equally applicable and results similar when applied to residential data from any location.

The paper is organized as follows: Section 2 proposes an econometric framework consistent with both economic and price index theory. Section 3 presents the HI used in this study and discusses how issues such as composition of sales and seasonality in the sales are dealt with in the construction of the indexes. Section 4 presents empirical results and Section 5 concludes.

2 Model Specification and Econometric Estimation

To compute the HI index imputations of the price of house h at time periods t and $t+s$ are required from the specified hedonic model. As discussed earlier, the theoretical fundations imply a functional form that allows for movements in shadow prices over time. This requirement is accomodated in the price index literature through the use of a rolling window approach (explained shortly). In this paper we propose an alternative to this approach which is statistically sound and performs better in terms of mean square prediction error. To establish notation, consider a hedonic model for time period $t = 1, ...T$ (where t could be an index for a month, a quarter or a year and we return to this issue in Section 3).

$$y_t = \mu + X_t\beta + \epsilon_t \quad (1)$$

where,

$y_t - N_t \times 1$ vector of observations of the dependent variable, typically the log of sale price (P), $y_t = \ln P_t$;

$N_t -$ number of properties sold in period t;

$\mu -$ intercept;

$\beta - K \times 1$ vector of unknown slope parameters;

$X_t - N_t \times K$ matrix of independent variables (property attributes) ;

$\epsilon_t - N_t \times 1$ vector of random errors.

Following Triplett recommendations, (2004, pp 71.), the standard procedure in the housing price index literature is the use of *adjacent period regressions* to construct hedonic imputed indexes. The approach consists of estimating (1) for each period t in a *rolling window* (RW), each time with data from two adjacent time periods, t and $t + 1$. This means that the slope coefficients/hedonic coefficients are kept constant within two adjacent periods, t and $t + 1$, but allowed to vary across subsequent rolling-windows. Under the rolling window approach there are two alternative estimates of the slope coefficients at a given period, t. One estimate is obtained using price data for the periods $t - 1$ and t and another based on t and $t + 1$. In this paper we use the estimate obtained using the data for periods $t - 1$ and t to compare to those from our approach. Though this specification is intuitive and practical, there is a logical inconsistency in the approach in that if parameters are the same for periods t and *t+1* and then for periods *t+1* to *t+2* it should then imply that parameters in periods t and *t+2* are identical and following this argument should lead to a constant parameter model as intercepts are also equal. Notwithstanding this problem, we simply follow the literature and implement the RW model.

In this paper two alternative specifications of ϵ_t are considered, namely, the errors are spherical ($\epsilon_t \sim (0, \sigma^2 I_{N_t})$), and alternatively the errors are spatially correlated with,

$$\epsilon_t = \rho W_t \epsilon_t + u_t \tag{2}$$

where,

$u_t - N_t \times 1$ vector of independently and identically distributed errors;

$W_t - N_t \times N_t$ matrix of spatial weights with spatial weights as functions of distance between houses in the sample;

$\varepsilon_t - N_t \times 1$ vector of spatially correlated errors;

$\rho - scalar$ spatial autocorrelation parameter, $|\rho| < 1$.

Inclusion of spatial errors in model (1) results in the standard Spatial Error Model (SEM) used in the spatial econometrics literature. In the context of hedonic models to explain property sale prices, the spatial error is designed to take explicit account of the role of unobserved location characteristics in determining house prices. This model is expected to be particularly useful when there are no location characteristics (such as distances to landmarks) included as regressors in the hedonic model.

The spatial weights matrix, W_t, has elements $0 \leq w_{ij} \leq 1$ for i and j, $i \neq j$ and $\sum_{j=1}^{N} w_{ij} = 1$ for each i, and $w_{ij} = 0$ otherwise. In this paper we use spatial weights by identifying the nearest neighbours and assign using information on the latitude and longitude of the properties sold. The nearest neighbors are identified using a *Delaunay triangulation*. A detailed exposition of Delaunay triangulation method can be found in Section 4.11 of LeSage and Pace (2009). Note that when a spatial weights matrix, W_t, is derived using Delaunay triangles, it represents the nearest m neighbors, and thus W_t^2 represents neighbors to neighbors, and so on. A spatial weights matrix could also be obtained by defining contiguity or nearest neighbors in terms of Euclidean distances; however, it can be computationally burdensome, and thus computational geometry is preferred as a more elegant and computationally less intensive approach.

It is customary for the rows of W_t to add up to one (in which case W_t is row (or right) stochastic), which also implies $|\rho| < 1$ (see Ord, 1965 and Krämer, 2005)[2] which guarantee the existence of $(I - \rho W_t)^{-1} = \sum_{j=0}^{\infty} \rho^j W_t^j = I + \rho W_t + \rho^2 W_t^2 + \rho^3 W_t^3 + ...$ and thus $\mathrm{E}(\varepsilon_t \varepsilon_t')$ is proportional to $(I - \rho W_t)^{-1} (I - \rho W_t)^{-1\,\prime}$.

We will denote the rolling window version of the model in (1) along with the error structure in (2) as RW_SEM (rolling window spatial errors model).

In this paper we propose an alternative to the RW idea. Instead of allowing the hedonic parameters to vary overtime through re-estimation as is the case in the RW method, we specify a model where hedonic parameters vary over time through a specific stochastic process. This extension is in line with economic theory, and in addition, a minimum mean square estimator[3] can be used to estimate the parameters of the regression model, making it the preferred approach from a statistical perspective. The model proposed here denoted as TV Model allows for a different stochastic process for the time-varying intercept parameter compared to that used to model the slope coefficients, and is specified as follows:

$$y_t = \mu_t + X_t \beta_t + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2 I_{N_t}) \tag{3}$$

$$\begin{aligned} \beta_t &= \beta_{t-1} + \eta_t \qquad \eta_t \sim N\left(0, \sigma_\eta^2 I_k\right) && (4) \\ \mu_t &= \mu_{t-1} + \xi_t \qquad \xi_t \sim N(0, \sigma_\xi^2) && (5) \\ E\left(\epsilon_t \eta_t'\right) &= 0 && (6) \end{aligned}$$

$$\text{for } t = 1, 2, ..., T$$

where,

N_t- number of houses sold at time t;

$X_t - (N_t \times K)$ matrix of independent (hedonic) characteristics;

μ_t is a time-varying intercept

β_t is the vector of time-varying hedonic slope coefficients

This model is known as a local level model with explanatory variables in the state-space literature (see for example Durbin and Koopman, 2012). Here, μ_t, known as the local level follows a stochastic process different from that of the slope parameters (hedonic attribute parameters, β_t). The basic idea is that movements in the local level represent secular trends in housing prices resulting from macroeconomic shocks independent of the movements in the hedonic regression coefficients. In practice, movements in the local level tend to dominate house prices movements over time. The time-varying model is also extended to allow for the presence of spatial correlation in the disturbance term resulting in a new specification denoted by TV_SEM where in (3), ϵ_t is assumed to follow the model in (2). The following assumptions are made:

$$E(\epsilon_t \eta_t') = 0,\ E(u_t \eta_t') = 0,\ E(u_t \xi_t') = 0,\ E(u_t u_{t-s}') = 0 \text{ for } s \neq 0.$$

The spatial correlation parameter, ρ, is assumed to be fixed over time. We note that $\epsilon_t = (I_{N_t} - \rho W_t)^{-1} u_t$ in (2). Assuming $\epsilon_t \sim N(0, H_t)$, it follows that:

$$H_t = \sigma_u^2 (I_{N_t} - \rho W_t)^{-1} (I_{N_t} - \rho W_t)^{-1\prime} \tag{7}$$

It is easily seen that if ρ is zero the error term $\epsilon_t = u_t$, and the model reduces to (3). The estimation of the TV and TV_SEM models can be achieved through likelihood or Bayesian estimation approaches (details provided below).

2.1 Econometric Estimation

To estimate (1) in a rolling window framework, OLS is used when the assumption is that there is no-spatial correlation in the errors and we use maximum likelihood, using the *sem.m* Matlab function developed by LeSage and Pace (see `http://www.spatial-econometrics.com/`), when errors are assumed to follow (2). Computationally efficient estimation of spatial models (both in a classical or Bayesian context) require evaluation of log-determinants that are functions of W_t. Chapter 4 of LeSage and Pace (2009) presents a detailed account of a number of important results about log-determinants that can be used to simplify the estimation of unknown parameters in spatial models. The function *fdelw2.m* was used to construct the spatial weight matrices for each of the spatial models in the paper.

TV and TV_SEM are state-space models and are estimated using Kalman filtering algorithms and, in this paper, using maximum likelihood to obtain estimates of unknown covariance parameters. In order to explain how they are estimated it is convenient to write (3),(4),(5), and (2) in state-space representation shown in (8) and (9) below.

$$y_t = Z_t \alpha_t + \epsilon_t \tag{8}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t \tag{9}$$

where,
$Z_t = \begin{bmatrix} i & X_t \end{bmatrix}$; i is an $N_t \times 1$ vector of ones
$\alpha_t' = \begin{bmatrix} \mu_t' & \beta_t' \end{bmatrix}$
$\zeta_t = \begin{bmatrix} \xi_t \\ \eta_t \end{bmatrix}$, $E(\zeta_t \zeta_t') = Q_t$, $Q_t \sim N(0, \begin{bmatrix} \sigma_\xi^2 & 0 \\ 0 & \sigma_\eta^2 I_K \end{bmatrix})$, and
$E(\epsilon_t \epsilon_t') = H_t$ in equation (7).

Kalman filtering algorithms provide an estimate of α_t and its Mean Squared Error Matrix *given* estimates of the parameters $\psi = \{\rho, \sigma_u^2, \sigma_\xi^2, \sigma_\eta^2\}$ (or $\psi = \{\sigma_u^2, \sigma_\xi^2, \sigma_\eta^2\}$ if error are not assumed spatially correlated). In both cases the parameters in ψ were estimated using numerical maximization of the conditional likelihood function and the Kalman filter algorithm. The models were estimated using code specially written by the first author (available upon request).

A brief sketch of the estimation algorithm is presented below:

1. Given an initial guess for $\psi = \{\rho, \sigma_u^2, \sigma_\xi^2, \sigma_\eta^2\}$, ψ_0, the equations of the Kalman filter are used to obtain a value of the conditional likelihood function which is defined as

$$\ln L(y_t;\psi) = -\frac{1}{2}\sum_{t=1}^{T} N_t \ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n} ln|F_t| - \frac{1}{2}\sum_{t=1}^{n} \nu_t' F_t^{-1} \nu_t \quad (10)$$

where, the *prediction error* is given by $\nu_t = y_t - \tilde{y}_{t|t-1}, \;\; t = 1,...,T$; with covariance matrix $E(\nu_t \nu_t') = F_t$, and from the measurement equation (8), $\tilde{y}_{t|t-1} = Z_t a_{t|t-1}$where $a_{t|t-1}$ is a Kalman filter estimate of α_t given $Y_{t-1} = \{y_{t-1,}y_{t-2}, ...y_0\}$. Assuming a starting distribution of the state vector $(\alpha_0 \sim N(a_0, \Omega_0))$[4], this is easily set in an iterative Newton type numerical optimization algorithm to find the maximum likelihood estimates of the hyperparameters (further details are presented in the appendix), given by:

$$\hat{\psi} = argmax_\psi \ln L(y_t|\psi)$$

2. Given $\hat{\psi}$, estimates of the covariances Q_t and H_t , $\hat{Q}_t$ and $\hat{H}_t$, are now available. These are used to obtain the estimates of α_t, $\tilde{\alpha}_{t|T}$, $t = 1,...,T$ and its Mean Squared Error matrix, $\Omega_{t|T}$, by running the state-space model through the equa-

tions of the Kalman filter (see Appendix equation (28) for the form of the Kalman filter estimator, a_t) and smoother (see Harvey (1989) or Durbin and Koopman (2012) for details on smoothing) with initial state vector a_0. Standard errors of $\tilde{\alpha}_{t|T}$ are given by the square root of the diagonal elements of $\Omega_{t|T}$

2.2 Prediction and Imputation of the Sale Price

The hedonic model is log-linear and thus the prediction of the sale price at time t of properties $h = 1, ..., N_t$ sold at time t, is given by

$$\hat{P}_t^h(x_t^h; \hat{\mu}_t, \hat{\beta}_t) = exp(\hat{y}_t^h) \tag{11}$$

while the imputation of the sale price for period $t+s$ of properties $h = 1, ..., N_t$ sold at time t is given by

$$\hat{P}_{t+s}^h(x_t^h; \hat{\mu}_{t+s}, \hat{\beta}_{t+s}) = exp(\hat{y}_{t+s}^h) \tag{12}$$

It is well known that (11), the naïve predictor, is a biased predictor of the conditional mean; however, it is not a biased predictor of the conditional median. As property prices are asymmetrically distributed and have a long tail in the positive direction, the median is the preferred estimator of the centre of the distribution (for further discussion see for instance Greene (2012) pp. 123).

To obtain the predictions ($\hat{y}_t^h$) and imputations ($\hat{y}_{t+s}^h$) when the model is assumed to have spherical errors, we use (13) and (14),

$$\hat{y}_t^h = \hat{\mu}_t + X_t^h \hat{\beta}_t \tag{13}$$

$$\hat{y}^h_{t+s} = \hat{\mu}_{t+s} + X^h_t \hat{\beta}_{t+s} \tag{14}$$

where the notation $\hat{\mu}_t(\hat{\mu}_{t+s})$ and $\hat{\beta}_t(\hat{\beta}_{t+s})$ is used here to denote parameter estimates obtained using either a rolling window approach (RW) or the estimates $\tilde{\alpha}_{t|T}$ of TV.

When the model is assumed to have spatial errors, a feasible predictor is given by (15) following Goldberger (1962) derivation of an optimal predictor, where $\hat{\mu}^s_t$, $\hat{\beta}^s_t$ and $\hat{\rho}$ are the estimates from RW_SEM or TV_SEM, $w_{hj,t}$ is the jth element of the hth row of W_t, and $\hat{\epsilon}_{j,t} = y^j_t - (\hat{\mu}^s_t + X^j_t \hat{\beta}^s_t)$, $j = 1, ..., N_t$.

$$\hat{y}^h_t = \hat{\mu}^s_t + X^h_t \hat{\beta}^s_t + \hat{\rho} \sum_{j=1}^{N_t} w_{hj,t} \hat{\epsilon}_{j,t} \tag{15}$$

To obtain an equivalent imputation, we would compute,

$$\hat{y}^h_{t+s} = \hat{\mu}^s_{t+s} + X^h_t \hat{\beta}^s_{t+s} + \hat{\rho} \sum_{j=1}^{N_t} w_{hj,t} \hat{\epsilon}_{j,+s} \tag{16}$$

however, this is not feasible as $\hat{\epsilon}_{j,t+s} = y^j_{t+s} - (\hat{\mu}^s_{t+s} + X^j_t \hat{\beta}^s_{t+s})$ cannot be computed given no observed sale price is available in period $t + s$ for properties sold in period t. Thus, in this study we use the *truncated* predictor to obtain both the prediction and the imputation of the price of each property in the sample, (17) and (18), and these are used to compute the indexes which are presented in the next section.[5]

$$\hat{y}^h_t = \hat{\mu}^s_t + X^h_t \hat{\beta}^s_t \tag{17}$$

$$\hat{y}^h_{t+s} = \hat{\mu}^s_{t+s} + X^h_t \hat{\beta}^s_{t+s} \tag{18}$$

3 Hedonic Imputed Housing Price Indexes

In this paper we report several sets of hedonic imputed price index numbers for residential housing. Hill and Melser (2008) discuss of a range of index number formulas that are based on different sets of weighting systems and on different sets of imputed prices. The general conclusion by Hill and Melser (2008) is that it is best if imputed prices are used for both current and base periods instead of using imputed prices only for the current period. In addition they recommend that value shares used should be based on actual sale prices instead of imputed prices. Our approach follows these suggestions with several modifications. As a deviation from the general practice in this area, we construct price index numbers with *plutocratic* and *democratic weights.* Plutocratic weights are proportional to the sales prices of properties and higher priced properties are accorded higher weights in the price index construction. These are essentially value shares of different properties sold at a given point of time. In contrast, the democratic weighting system gives the same weight to each house sold in the market at any given point of time. Therefore, the use of democratic weights leads to unweighted arithmetic or geometric averages of imputed price relatives. The use of democratic weights essentially stems from the use of a stochastic approach where the properties sold in any given area is taken as a random sample and therefore the price observations are assumed to have the same variance. Price indexes with democratic weights can be extended to accommodate stratification of houses by quality and/or locational characteristics.

Let P_t^h represent the sale price of house h in period t. Then the value share of the house h, w_t^h is given by:

$$w_t^h = \frac{P_t^h}{\sum\limits_{n=1}^{N_t} P_t^n} \tag{19}$$

where,

P_t^h is the observed sale price of house h and N_t is the number of houses sold in period t. Typically in our case t refers to a particular month as we are making use of monthly sales data. Construction of annual indexes is described in Section 3.3.

We define next the two types of indexes used in the study.

3.1 Plutocratic Indexes

These indexes are weighted indexes where weights represent the relative value of each of the properties included in the sample. We use the Fisher and Tornqvist variants of this index.

All HI price indexes are presented for period t, with period $t-s$ as the base.

The plutocratic **Fisher index** (F) is defined as:

$$F_{(t-s),t}^{P} = \sqrt{L_{(t-s),t}^{P}\, P_{(t-s),t}^{P}} \tag{20}$$

where $L_{t-s,t}^{p}$ and $P_{t-s,t}^{p}$ are respectively the plutocratic Laspeyres and Paasche index numbers with the following definitions:

$$L_{(t-s),t}^{p} = \sum_{h=1}^{N_s} w_{(t-s)}^{h} \left(\frac{\hat{P}_t^h(x_{(t-s)}^h)}{\hat{P}_{(t-s)}^h(x_{(t-s)}^h)} \right) \tag{21}$$

$$P_{(t-s),t}^{p} = \left[\sum_{h=1}^{N_t} w_t^h \left(\frac{\hat{P}_{(t-s)}^h(x_t^h)}{\hat{P}_t^h(x_t^h)} \right) \right]^{-1} \tag{22}$$

where,

$\hat{P}_{(t-s)}^h(x_t^h)$ for $s \geq 0$ is an *imputation* of the price of property h, sold at time t with characteristics x_t^h, using a vector of shadow prices for time period $t-s$ and the value shares defined as in (19).

We note here that if the shares were based on predicted prices instead of observed prices then the Laspeyres and Paasche indexes, defined in (21) and (22) simply turn out to be *ratios of the value of the stock of houses* in periods $t-s$ and t respectively evaluated using the hedonic price models in these periods. Thus the indexes in (20), (21) and (22) simply measure the change in the value of the housing stock due to changes in prices as reflected in the hedonic model of prices.

The plutocratic **Tornqvist indexes** are defined similar to equations (20), (21) and (22) using geometric averages in the place of the arithmetic mean. Following Hill and Melser (2008), we define these indexes as follows:

$$T^{P}_{(t-s),t} = \sqrt{GL^{p}_{(t-s),t} \times GP^{p}_{(t-s),t}} \tag{23}$$

where $GL^{p}_{(t-s),t}$ and $GP^{p}_{(t-s),t}$ are the plutocratic geometric Laspyeres and geometric Paasche indexes which are defined as:

$$GL^{P}_{(t-s),t} = \prod_{h=1}^{N_{(t-s)}} \left[\frac{\hat{P}^{h}_{t}(x^{h}_{(t-s)})}{\hat{P}^{h}_{(t-s)}(x^{h}_{(t-s)})} \right]^{w^{h}_{(t-s)}} \tag{24}$$

$$GP^{P}_{(t-s),t} = \prod_{h=1}^{N_{t}} \left[\frac{\hat{P}^{h}_{t}(x^{h}_{t})}{\hat{P}^{h}_{(t-s)}(x^{h}_{t})} \right]^{w^{h}_{t}} \tag{25}$$

These indexes are "plutocratic" as they are influenced by houses with large price tags. Despite this, the Fisher and Tornqvist indexes in (20) and (23) measure the changes in the housing stock values that can be attributable exclusively to price changes, and therefore provide useful information.

The use of plutocratic weights along with a Laspeyres type index (as in equation 21) measures the price change by comparing the total value of the housing stock in the base period and current period using hedonic imputations[6]. Similarly the Paasche index compares the housing stock of the current period at the base and current period prices. However, the geometric indexes like the Tornqvist indexes cannot be interpreted along the same lines. All the indexes discussed in Hill and Melser (2008) are essentially plutocratic indexes.

We now deviate from the Hill and Melser (2008) approach and define democratic indexes which are statistically based measures of price changes.

3.2 Democratic Indexes

Consistent with the use of a log-price hedonic model, we focus on the democratic geometric Laspeyres, Paasche and Tornqvist indexes. These are defined as:

$$\begin{aligned} T^D_{(t-s),t} &= \sqrt{GL^D_{(t-s),t}\, GP^D_{(t-s),t}} \\ &= \sqrt{\left[\prod_{h=1}^{N_{t-s}} \left(\left[\frac{\hat{P}^h_t(x^h_{(t-s)})}{\hat{P}^h_{(t-s)}(x^h_{(t-s)})}\right]^{\frac{1}{N_{t-s}}}\right)\right]\left[\prod_{h=1}^{N_t}\left(\left[\frac{\hat{P}^h_t(x^h_t)}{\hat{P}^h_{(t-s)}(x^h_t)}\right]^{\frac{1}{N_t}}\right)\right]} \end{aligned} \tag{26}$$

where N_t and N_{t-s} are respectively the number of houses sold in periods t and $(t-s)$.

The democratic index provides a measure of price change that is consistent with the distribution of price relatives. The distribution of the prices is likely to be skewed and the use of a geometric mean is consistent with a general log-normal distribution of price relatives[7]. When the geometric Tornqvist index is computed, we explicitly recognize the unequal numbers of properties sold in the two periods and define a geometric mean of the geometric Laspeyres and Paasche indexes (see equation 26).[8] The use of democratic weights is appropriate if the principal aim is to generate a statistically sound estimator of the central tendency of the distribution of property prices. Given that the expenditure weights used in hedonic imputed price indexes do not have the same theoretical basis as the expenditure shares used in the construction of the consumer price index, the choice between the plutocratic and democratic weights should be really driven by the main objective behind the housing price index construction.

3.3 Seasonality and the Construction of Annual Price Indexes

As seasonality affects both sales prices and volumes of properties sold, we consider the use of chained indexes. From an index number perspective, chaining may be undesirable when it leads to index drift. Szulc (1983) made the point that when prices or quantities oscillate ('bounce'), chaining can lead to considerable index drift: that is, if after several periods of bouncing, prices and quantities return to their original levels, an index that chains month-to-month changes will not normally return to unity. Hence, the chaining of noisy monthly or quarterly series is not recommended.

In view of the drift caused by chaining in the presence of oscillations, the presence of seasonality in the sales of properties and the types of properties sold, it is more meaningful to compute year-on-year monthly housing price indexes (eg January in year t to January in year $t+1$ etc.) and combine them to yield an annual price index. Under this approach, first indexes measuring the price change in observed prices over one year from the given month are computed. The Yule(1921)'s method from chapter 22 of the ECE-ILO Manual on the Consumer Price Index (International Labour Office, 2004) is used for the task. The method consists of two steps. First, compute the year-over-year monthly index for each month using a standard index number formula. In our case we can use Fisher and Tornqvist indexes with plutocratic or democratic weights. Second, compute the annual index as a simple unweighted geometric mean of the year-on-year monthly indexes.

4 Empirical Results

4.1 Data

The data used in this study refer to sales of residential detached properties (i.e. houses on blocks of land) and excludes units, terraces, townhouses and duplexes in the Brisbane (Australia) metropolitan area. There are 65,239 sales transactions over the period of 252 months from 1985:1 to 2005:12. The data are from one of the leading providers of property sales information services in Australia, 'RP Data Ltd' (`www.rpdata.com`), and each data point (transaction) includes, the date (month and year) of sale, sale price, geocode (latitude, longitude), the post code, the size of the land (lot) in m^2(AREA), the number of bedrooms (BED), the number of Bathrooms (BATH), the number of car spaces (lock-up garages and carports) (CARLUG). These data were first collected by Cominos (2006). Further filtering of the data was conducted by Svetchnikova (2007)[9] and the resulting data set is used in this study.

The distribution of transactions over the sample period is important as it might have an impact on the accuracy of some of the results. Figure 1 plots the number of transactions per month in the dataset. The number of recorded transactions has risen substantially since the mid 1990s. While the actual number of transactions is likely to have risen due to rapid population growth in the city of Brisbane in the last 20 years, it is also the case that the market for electronic databases was not established in the earlier part of the period, and therefore it is possible that some non-trivial number of transactions were never included in the electronic database for the earlier years.

[Figure 1. Number of transactions per month in the dataset]

4.2 Hedonic Model Results

We estimate the two versions of the model (with and without spatial errors) using sales data of 65,239 transactions over 252 months using the methodology discussed in Section (2.1). To obtain the RW and RW_SEM results we pool *two months* and overlap one as we move through the sample. That is, the parameter estimates for 1985:2 is obtained from a regression of pooled data for 1985:1 and 1985:2; 1985:3 is obtained from data for 1985:2 and 1985:3, and so on. TV and TV_SEM are obtained by estimating the state space with monthly transactions. We present the estimates of the local level, BED and BATH shadow price parameters to illustrate the difference in the estimates under the rolling window and the time-varying specifications.

The estimate of ρ from TV_SEM is $\hat{\rho} = 0.48$ with a $p-value = 0.000$[10]. Therefore, there is strong evidence to suggest that errors are spatially correlated. Given that the hedonic characteristics included in the model do not provide any measure for the location of the property, the spatial error is liklely to be capturing this unobserved dimension. When a rolling window estimation is used, it is the case that not only the estimates of the hedonic coefficients (β) will change over time, but also the estimate of ρ will change across estimation windows, unlike the case of TV_SEM where ρ is time invariant. The estimates of ρ obtained through RW_SEM change substantially over the 20 years; the minimum estimate is 0.372 and the maximum is 0.608.

4.2.1 Comparative Performance of Alternative Models

To study the in-sample predictive performance we compute an aggregate measure for the sample ($T = 252$ months), the root mean square prediction error (RMSPE) of each estimation approach based on the prediction of the log of sale price for each transaction. The results are presented in Table 1.

where,

$$RMSPE = \sqrt{\frac{1}{N}\sum_t \sum_h (y_{ht} - \hat{y}_{ht})^2}$$

y_{it}= the log of observed sale price for house h at time period t

$\hat{y}_{it}$ =a prediction of the log of sale price for house h at time period t

$N = \sum_{t=1}^{T} N_t$

[Table 1 here]

The results are presented for both spatial and non-spatial models, although as mentioned above there is strong evidence that errors are spatially correlated. Comparing the RW and TV, the results are that the time-varying parameters model produces a negligible reduction in RMSPE of 0.03%. In contrast, for the models where spatial effects are included, to capture the otherwise omitted location information, the results provide some surprises. By implementing an adjacent period rolling window in a model with spatial errors (RW_SEM) the RMSPE increases (although marginally) over the RW alternative (as well as TV's). Basically this result implies that RW_SEM not only lacks predictive power, but it might also be an incorrect specification. In contrast, the combined effect of spatial errors in a time-varying hedonic parameters model (TV_SEM) results in a larger reduction in RMSPE (9.92%) over the RW_SEM model's performance[11].

The improvement achieved through the combination of time-varying coefficients and spatially correlated errors makes the TV_SEM most desirable in terms of housing price predictions, and it should be adopted in preference to the rolling window approach. However, as hinted above, this aggregate measure might be covering the extent to which the number and composition of sales in each time period affect predictions. This is explored further in the next sub-section

4.2.2 Coefficient Estimates from Time Varying and Rolling Window Models

We use as illustration estimates of three parameters, the local level (Figure 2), and the BATH and BED coefficients (Figure 3).

In Figure 2 the estimates from RW_SEM and TV_SEM (and ±2 standard error bands) are presented. An important point to note is that the rolling window estimates are much more volatile than those obtained from the time-varying model. This is an important result as it illustrates clearly the undesirable effect that a small number of transactions joint with different compositions of properties sold can have on the estimates of the shadow prices.

[Figure 2. Local Level (intercept coefficient), μ_t]

In Figure 3 the estimates from RW_SEM and TV_SEM (and ±2 standard errors bands) for the parameters of bathrooms and bedrooms are presented. As in the case of the local level, it is extremely clear that the rolling window approach produces highly volatile estimates and it is likely that its performance is dismal over shorter than 252 time periods.

[Figure 3. BATHROOMS and BEDROOMS Coefficients]

4.2.3 The Evolution of Prices

The model with the lowest RMSPE is the TV_SEM. We produce price predictions for the properties sold in a particular period using the following model (see the discussion in Section 2.2):

$$\widehat{\ln P_t^h} = \hat{\mu}_t + \hat{\beta}_{1t} \ln Land_t^h + \hat{\beta}_{2t} BED_t^h + \hat{\beta}_{3t} BATH_t^h + \hat{\beta}_{4t} CARLUP_t^h \tag{27}$$

where,

$\hat{\mu}_t$ and $\hat{\beta}_{kt}$, $k = 1, ..., 4$ are the Kalman smoothed estimates from the TV_SEM model

In Figure 4 we compare median monthly price of properties sold in each time period to its prediction which we compute as follows:

$$\hat{P}_t^m = median(exp(\widehat{\ln P_t^1}), exp(\widehat{\ln P_t^2}), ..., exp(\widehat{\ln P_t^{N_t}}))$$

.

where $\widehat{\ln P_t^h}$ is obtained from equation (27).

For each period, we compute the median of the TV_SEM predicted prices of all the houses sold in that period ($h = 1, ..., N_t$). This is slightly different, and conceptually superior, to the normal practice of computing the predicted price of a property using median values of hedonic characteristics (land, bedrooms and bathrooms) in different months. The observed and predicted median property prices are closely aligned over the period which provides reassurance of the predictive ability of TV_SEM .

[Figure 4. Prediction of median sale prices over the sample period]

Figure 4 provides an interesting profile of prices of properties sold in Brisbane over the last two decades. Treating the median prices as an observed time series from 1985 to 2005, we can see that there are several structural breaks in the price series. After a relatively stable period until early 1987, a surge in property prices is evident over the two year period until July 1989. Over the next decade from January 1990 until January 2001 there was a steady increase in median prices from just above 100,000 dollars to 175,000 dollars. A sharp rise in property prices occurred from January/July 2001 until January 2004 where prices more than doubled. It is possible that structural breaks have occurred over this period; however, we have not formally tested for this possibility.

4.3 Housing Price Indexes

In this section we present both monthly and annual housing price indexes for the sample period, 1985-2005, computed using the formulas described in Section 3.

4.3.1 Annual Chained Indexes

In Figures 5 a and b, we present chained annual housing price index numbers using the procedure outlined in Section 3.3. We also present a chained annual index computed using the observed median price in each year as well as a conventional TDH index computed by exponentiating the estimates of the time-dummy coefficients from a spatial errors hedonic model estimated with the complete sample. These serve as reference for comparison purposes. All the indexes are computed using 1985 as the base year. For the HI indexes we compute the plutocratic and democratic versions of the Fisher and Tornqvist indexes (see Sections 3.1 and 3.2 for formula). As the Fisher and Tornqvist indexes are both superlative[12] and in most empirical studies tend to be numerically close we expect the same result in our case. For the 1985-1995 period we observe that the TDH index is above the median (of observed prices for period t) index for a large portion of the period while the HI indexes are consistently below the median based index. For the 1996-2005 period, all the chained indexes are below the median-based chain index.

[Figure 5 Annual indexes - Chained, 1985-1995 (1985=1) and 1996-2005 (1985=1)]

4.3.2 Monthly Chained indexes

From the annual price indexes we now turn to chained price indexes constructed using month-to-month price indexes. Here we compute the price indexes from one month to the next and accumulate the changes over time through chaining. In this section we mainly focus on the plutocratic and democratic weighted price indexes computed using the Fisher and Tornqvist index numbers. The median (of observed prices) based price index is also presented as a reference.

Figure 6 presents chained HI monthly indexes with January 1985 as the base. Both plutocratic and democratic weighted indexes are presented. By the end of the study period, there is a substantial divergence that has accumulated between the median index and the hedonic price index numbers of the magnitude of 20 to 30 percent. With the median based index higher than the Fisher and Tornqvist indexes. The democratic weighted Tornqvist index is uniformly higher than the plutocratic weighted index but the percentage difference between them is much smaller than that shown by both in relation to the median based price index.

As with the annual chained price indexes, we note the presence of three episodes of acceleration in the housing prices. There is evidence of seasonal fluctuations in the indexes but we do not notice any major drift in the indexes. In order to facilitate visual examination of the differences, Figures 7 and 8 present the indexes over two sub-periods, 1985 to 1995 and 2001 to 2005. These represent two periods of accelerated increases in property prices in Brisbane; however, the underlying dynamics of the market are markedly different. From Figure 7 we observe that there are several periods during which the trends in the index of median prices and the hedonic index are in the opposite direction which is a clear indication of the influence of the mix in property sales in different periods which a median based index does not control for. However, all the indexes are much more closely aligned during the period 2001 to 2005 (Figure 8) and there are no appreciable differences between the median and hedonic price indexes. This sharp contrast between these two periods is likely due to the fact that over the 2001-2005 period there was a price boom largely attributed to a large influx of population into the Brisbane and greater Brisbane areas coupled with a lag in the supply of new land for residential construction (see Smith and Rambaldi, 2010), and therefore, price increases were large driven by land prices. If this was the case, the increase in property prices would have been more uniform across all types of properties sold, which in turn implies that the mix of houses sold (and the hedonic quality adjustment) might play a much smaller role in this instance, with the local level trend driving the index. This is an aspect that requires further analysis.

[Figures 6. Chained Monthly indexes. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 1985:1 to 2005:12]

[Figures 7. Chained Monthly indexes. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 1985:1 to 1995:12]

[Figures 8. Chained Monthly indexes. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 2001:1 to 2005:12]

5 Conclusions

The paper has dealt with several important issues relating to theoretically consistent hedonic modeling of housing prices and their use in the construction of housing price index numbers. First, the paper focuses on the issue of econometric specification and stresses the need to model the time-varying nature of the hedonic coefficients in order to be consistent with theory. The importance of making optimal use of the information on the influence of locational characteristics available in the form of spatially correlated errors is highlighted. The next issue is the problem of choosing the best specification. We use the root mean squared error of prediction of (log) prices of houses sold in different periods as a criterion to judge the predictive performance of various models. The second objective of the paper is to examine the effect of using various hedonic models on the resulting housing price index numbers obtained. We also focus on the influence of the index weights to account for the composition of sales, and discuss the role of plutocratic versus democratic weights in property price indexes. Finally, we examine chained annual indexes as well as month-to-month housing price indexes. The empirical analysis of the paper is based on housing price data from the city of Brisbane in Australia for the period 1985 to 2005. The analysis clearly demonstrates the superior predictive power of the time-varying hedonic model with spatially correlated errors over the rolling window approach recommended in the hedonic price index literature. We find the annual indexes computed using the time dummy hedonic method tend to be closer to the median than the HI imputed versions. The theoretical literature shows that time-dummy hedonic price index numbers are biased. In the application used in this study, the median of observed sale prices is well above the price indexes from all the other approaches with the possible exception of periods of rapid and uniform price increases, where land prices are likely to dominate the trend in property prices. For the city of Brisbane we find that the median housing price index significantly diverges during the period 1985 to 1995 but seems to align quite well with the hedonic price indexes during the period 2001-2005 which is a period of rapid population increase and low release of new land for urban construction which resulted in a price boom in the Brisbane housing market. We attribute this feature to the possibility that property price increases were uniform across different types of dwellings and different locations during the real estate market boom. Trends in the chained annual price indexes as well as chained monthly price indexes clearly show three phases of housing price acceleration during the study period. These periods are consistent with the anecdotal evidence on house prices in Brisbane during this period. Thus, the main conclusion of the paper is that the theoretically consistent time-varying hedonic models with spatially correlated disturbances possess superior predictive power and thus future studies of housing prices should endeavor to use the approach outlined in this paper.

Notes

[1]Readers interested in a comparison of methods to construct housing price indexes are referred to recently materialized Handbook of Property Price Indexes (HPPI) commissioned by EuroStat (de Haan and Diewert, 2011), `http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/owner_occupied_housing_hpi/rppi_handbook`, which accomplishes this task very competently and also provides a comprehensive set of references to relevant literature and a number of empirical comparisons between currently used methods such as the repeat sales method (see Chapters 6, 10 and 11) and hedonic based methods (see Chapters 5, 10 and 11) using data drawn for a number of countries.

[2]This follows because ρ is strictly bounded by the inverse of the eigenvalues of W_t.

[3]minimum linear mean square estimator if normality of logarithmic prices is not assumed.

[4]The distribution of the initial state vector, α_0 is assumed to be normal with a diffuse mean square error matrix, $\Omega_0 = \kappa I$ (where κ is a very large number and I is the identity)

[5]we use the term 'truncated predictor' in the sense used by Kouassi et al (2011), in that it is based on efficient estimates of the model parameters; however, it ignores the contribution of the spatial correlation to the predictor.

[6]This type of interpretation holds exactly when the expenditure share weights are also based on imputed prices, and it may be considered approximate when actual prices are used.

[7]Analogously if the prices are normally distributed then a simple arithmetic mean would be used in the place of a geometric mean.

[8]It is possible to consider a more sophisticated approach after stratifying the sample into different regions and by the types of properties.

[9]For details on the steps carried out to clean and check the data and descriptive statistics, see Cominos (2006) and Svetchnikova (2007).

[10]This estimate is very close to that obtained by estimating a conventional time-dummy hedonic model with a spatial error using LeSage and Pace *sem.m* routine. The estimate is $\hat{\rho} = 0.46$, $p-value = 0.000$. This is very reassuring as the two models and estimators differ substantially.

[11]We note that we estimated two conventional fixed parameter hedonic models (with time-dummy

intercepts) using the complete sample and computed the RMSPE as above. A time-dummy hedonic (TDH) model without spatial errors has a RMSPE of 0.4224, while the RMSPE of the TDH with spatial errors (TDH_SEM) has an RMSPE of 0.4214.

[12]See Diewert (1976) for more details on *exact* and *superlative indexes.*

References

Anselin, L. and Bera, A., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, Amman and Giles, David E.A., editors, Handbook of Applied Economic Statistics, pages 237–289. Marcel Dekker, New York.

Bailey, M.J., R.F. Muth and H.O. Nourse (1963), "A Regression Method for Real Estate Price Construction", *Journal of the American Statistical Association*, 58, 933-942.

Cominos, H., 2006. "Estimation of House Prices and the Construction of House Price Index Numbers: A new methodology applied to the Brisbane Metropolitan Area," University of Queensland, A Thesis submitted to the School of Economics in partial fulfillment for the Degree of Bachelor of Economics (Honours), in the field of Econometrics.

Court, A. T., 1939. "Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*, 99-117. New York: General Motors Corporation.

de Haan, J. and W.E. Diewert, 2011. *Handbook on Residential Property Price indexes.* B. Balk (Statistics Netherlands) project coordinator, de Haan and Diewert *ed* `http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/owner_occupied_housing_hpi/rppi_handbook`

Diewert, W.E., 1976. “Exact and Superlative Index Numbers”, *Journal of Econometrics*, 4, 114-145.

Diewert, W.E., 2003. “Hedonic Regressions: A consumer Theory Approach” in R.C. Feenstra and M. D. Shapiro (eds), Scanner Data and Price Indexes, NBER Conference on Research in Income and Wealth. Studies in Income and Wealth, vol. 64. Chapter 10, pp. 318-348. The University of Chicago Press.

Durbin, J. and S. J. Koopman, 2012. Time Series Analysis by State Space Methods, second ed. Oxford Statistical Science Series 38. Oxford University Press Inc.

Greene, W., 2012. Econometric Analysis, seveth ed. Pearson Education Ltd.

Griliches, Z., 1961. "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change. In *Price Indexes and Quality Chan*ge, ed. Z. Griliches, 55-87. Cambridge, Massachusetts: Harvard University Press.

Harvey, A. C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge.

Hill, R. J. and D. Melser, 2008. “Hedonic Imputation and the Price Index Problem: An Application to Housing,” *Economic Inquiry*, 46:4, pp. 593–609.

International Labour Office, 2004. Consumer Price Index Manual. Theory and Practice. ISBN 92-2-113699-X, xxxi+537 pages, ILO Publications, International Labour Office 4, Geneva, Switzerland.

Kouassi, E., J. Sango, J.M. Bosson Brou, F. N. Teubissi and K. O. Kymn, 2011. Prediction from the Regression Model with Two-Way Error Components. *Journal of Forecasting*, 30, pp. 541–564. DOI: 10.1002/for.1176.

Krämer, W., 2005. "Finite sample power of Cliff Ord type tests for spatial disturbance correlation in linear regression," *Journal of Statistical Planning and Inference*, 128, pp. 489-496.

Lane, W. and M. L. Schmidt. 2006. "Comparing U.S. and European inflation: the CPI and the HICP," *Monthly Labor Review*, 129:5, pp. 20-27.

LeSage, J. and Pace, K., 2009. *Introduction to Spatial Econometrics*, Boca Raton: Chapman & Hall, 2009.

Ord K., 1965. "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, pp:120-126.

Pace, R. Kelley, James P. LeSage, and Shuang Zhu, 2009. "Impact of Cliff and Ord on the Housing and Real Estate Literature," *Geographical Analysis*, 41:4, pp. 418-424.

Pakes, A., 2001. "Some Notes on Hedonic Price indexes, With An Application to PCs," Paper presented at the NBER Productivity Program Meeting. 16, March, Cambridge, Massachusetts.

Silver, M., 1999. "An Evaluation of the Use of Hedonic Regressions for Basic Components of Consumer Price indexes," *The Review of Income and Wealth*, 45:1, pp. 41-56.

Silver, M. and S. Heravi, 2007. "The Difference Between Hedonic Imputation indexes and Time Dummy Hedonic indexes", *Journal of Business & Economic Statistics*, 2007, 25, 239-246.

Smith, S. and A.N. Rambaldi, 2010. "An Econometric Analysis of the Determinants of SEQ Housing Prices," Final Report. Local Government

Association of Queensland, http://www.lgaq.asn.au/c/document_library/get_file?uuid=3e6718c4d9f44c0387b8ef4ba7ce5bb2&groupId=10136.

Standard and Poor's, 2009. S&P/Case-Shiller Home Price Indices; Index Methodology, New York: Standard and Poor's.

Svetchnikova, D., 2007. Spatial-Temporal Modeling of the Real Estate Market, University of Queensland, A Thesis submitted to the School of Economics in partial fulfillment for the Degree of Bachelor of Economics (Honours), in the field of Econometrics.

Szulc, B., 1983. "Linking Price Index Numbers" in Diewert and Montmarquette (eds.) Price Level Measurement, Statistics Canada, Ottawa, Canada.

Triplett, J., 2004. "Handbook on Hedonic indexes and Quality Adjustments in Price indexes: Special Application to Information Technology Products", OECD Science, Technology and Industry Working Papers, 2004/9, ISBN-92-64-02814-5, OECD Publishing, `http://www.oecd.org/dataoecd/37/31/33789552.pdf`

Wenzlick, R., 1952. "As I See the Fluctuations in the Selling Prices of Single Family Residences", *The Real Estate Analyst,* 21, 541-548.

Wyngarden, H., 1927. "An Index of Local Real Estate Prices", *Michigan Business Studies,* 1:2, Ann Arbor: University of Michigan.

A Computing the value of the likelihood function using the equations of the Kalman Filter

To obtain a value of the $\ln L$, the prediction error, ν_t, and its covariance matrix, F_t, for $t = 1, ..., T$ are required.

The prediction error is obtained using the parameter estimates at $t-1$, $\nu_t \sim (0, F_t)$

$$\nu_t = y_t - Z_t a_{t-1}$$

and the covariance matrix is given by,

$$\begin{aligned} F_t &= Z_t \Omega_{t|t-1} Z_t' + H_t \\ &= Z_t \Omega_{t|t-1} Z_t' + \sigma_u^2 \left(I_{N_t} - \rho W_t\right)^{-1} \left(I_{N_t} - \rho W_t\right)^{-1\prime} \end{aligned}$$

where, a_t and Ω_t are the Kalman filter estimates of α_t and its mean square prediction matrix which are obtained using the equations of the Kalman filter,

$$a_t = a_{t-1} + G_t \nu_t, \qquad \text{for } \ t = 1, ..., T \tag{28}$$

with covariance matrix, Ω_t, given by

$$\Omega_t = \Omega_{t|t-1} - M_t F_t^{-1} M_t', \qquad \text{for } \ t = 1, ..., T \tag{29}$$

$G_t = M_t F_t^{-1}$ is known as the Kalman gain. Captures the information gain from $t-1$ to t.

$M_t = \Omega_{t|t-1} Z_t$

$$\Omega_{t|t-1} = \Omega_{t-1} + Q = \begin{bmatrix} (\sigma_\xi^2 + \omega_{t-1}^{[1,1]}) & \omega_{t-1}^{[1,2]} & \omega_{t-1}^{[1,3]} & \cdots & \omega_{t-1}^{[1,K-1]} & \omega_{t-1}^{[1,K]} \\ \omega_{t-1}^{[2,1]} & (\sigma_\eta^2 + \omega_{t-1}^{[2,2]}) & \omega_{t-1}^{[2,3]} & \cdots & \omega_{t-1}^{[2,K-1]} & \omega_{t-1}^{[2,K]} \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ \omega_{t-1}^{[K-1,1]} & \omega_{t-1}^{[K-1,2]} & \cdots & & (\sigma_\eta^2 + \omega_{t-1}^{[K-1,K-1]}) & \omega_{t-1}^{[K-1,K]} \\ \omega_{t-1}^{[K,1]} & \omega_{t-1}^{[K,2]} & \cdots & \cdots & \omega_{t-1}^{[K,K-1]} & (\sigma_\eta^2 + \omega_{t-1}^{[K,K]}) \end{bmatrix}$$

$\omega_{t-1}^{[i,j]}$ is the $(i,j)^{th}$ element of Ω_{t-1}

$\alpha_0 \sim N(a_0, \Omega_0)$ is the initial distribution

For further details the reader is referred to Harvey (1989) or Durbin and Koopman (2012)

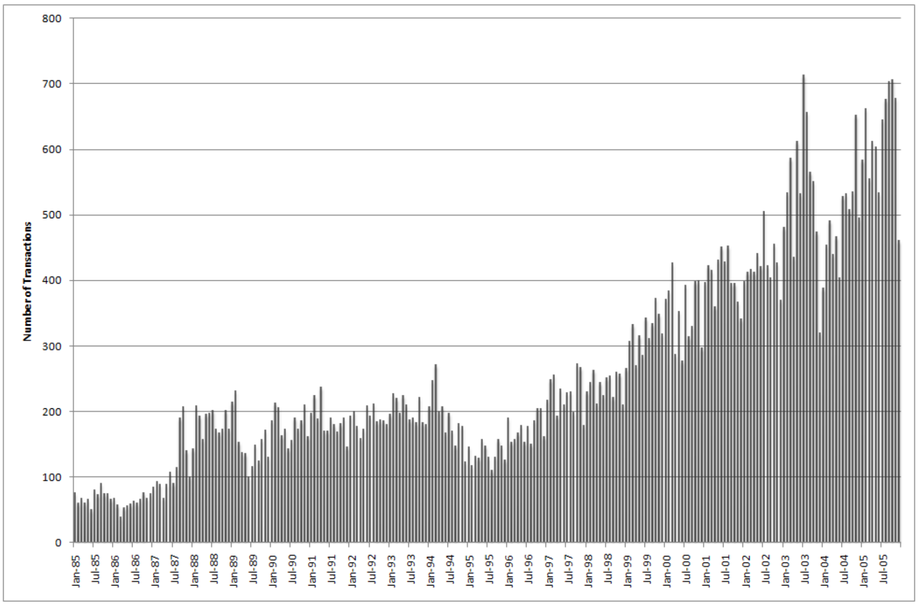


Figure 1: Number of transactions per month in the dataset

Table 1: Root Mean Square Prediction Error (in-sample)- Prediction of ln(Sale Price) over the period 1985-2005

Model	No Spatial Effects		Spatial Effects	
	Rolling Window (RW)	Time Varying (TV)	Rolling window (RW_SEM)	Time Varying (TV_SEM)
Root MSPE	0.4154	0.4153	0.4196	**0.3780**
reduction from base model	BASE	0.03%	BASE	**-9.92%**

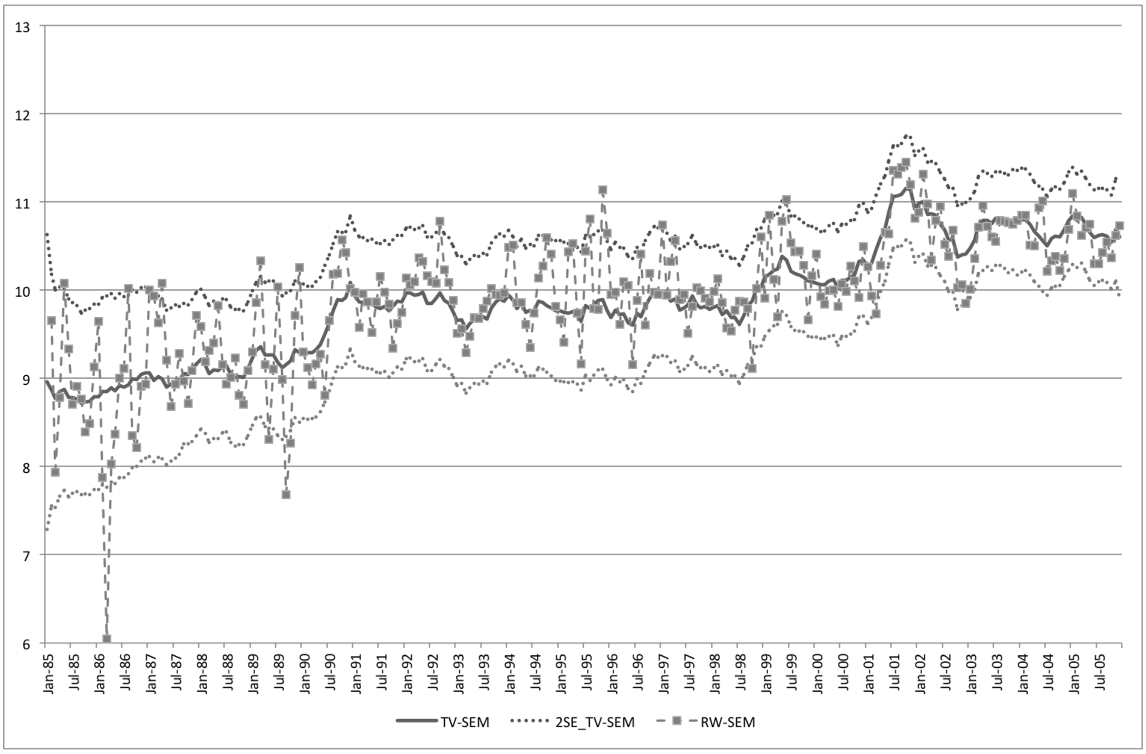


Figure 2: Local Level (intercept coefficient), μ_t

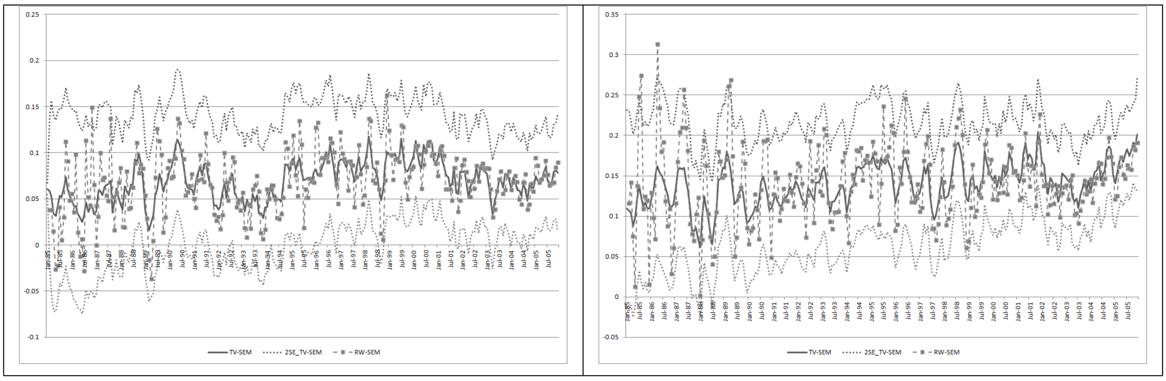


Figure 3: BATH and BED Coefficients

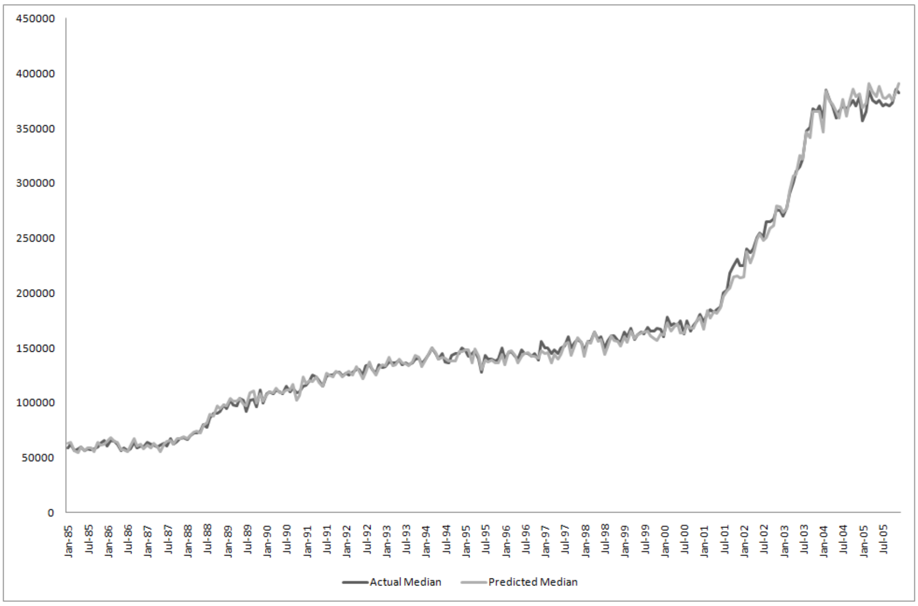
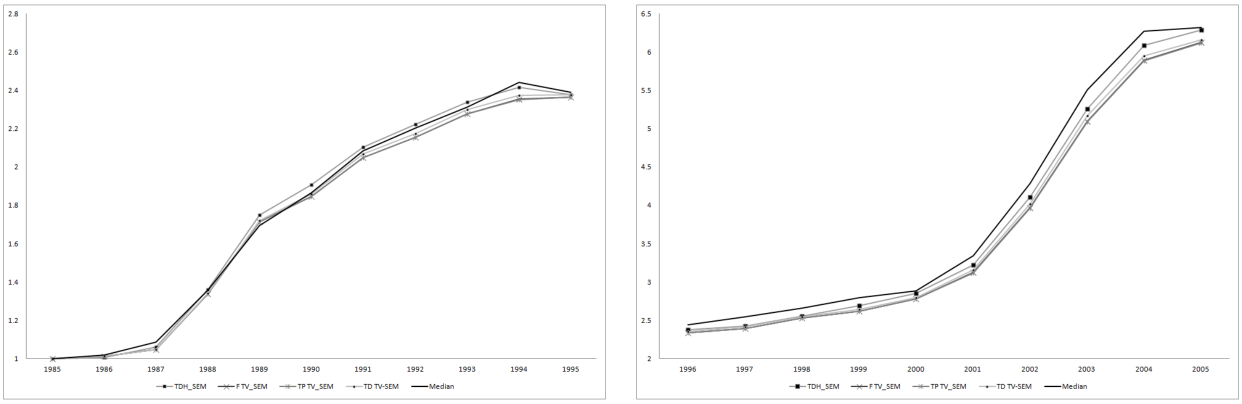


Figure 4: Prediction of median sale prices over the sample period



a. 1985-1995 (1985=1)

b. 1996-2005 (1985=1)

Figure 5: Annual Indices - Chained

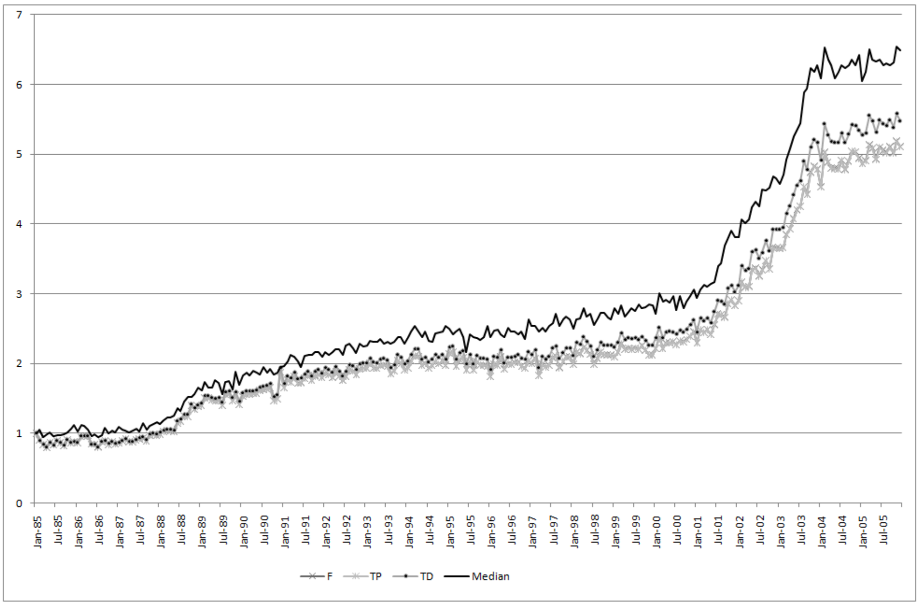


Figure 6: Chained Monthly Indices. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 1985:1 to 2005:12

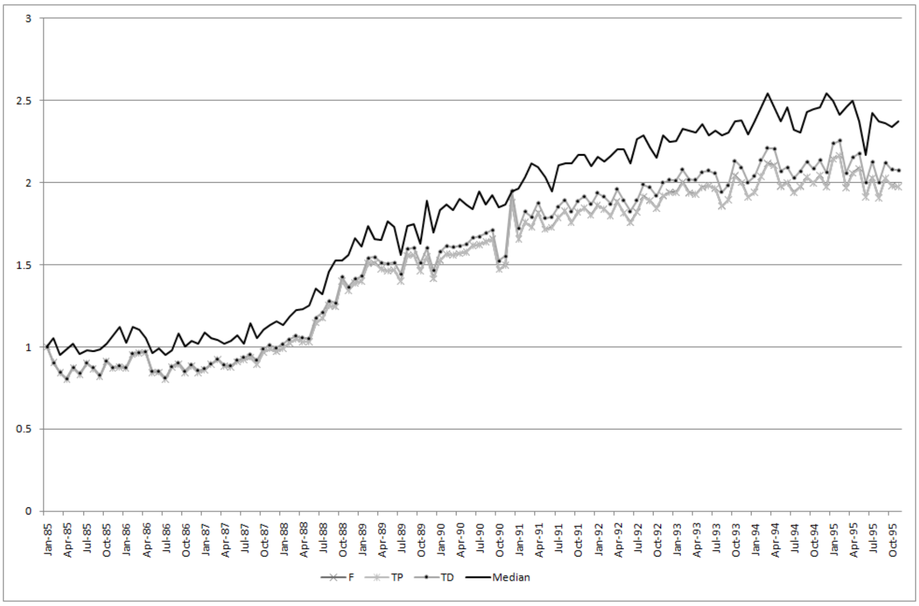


Figure 7: Chained Monthly Indices. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 1985:1 to 1995:12

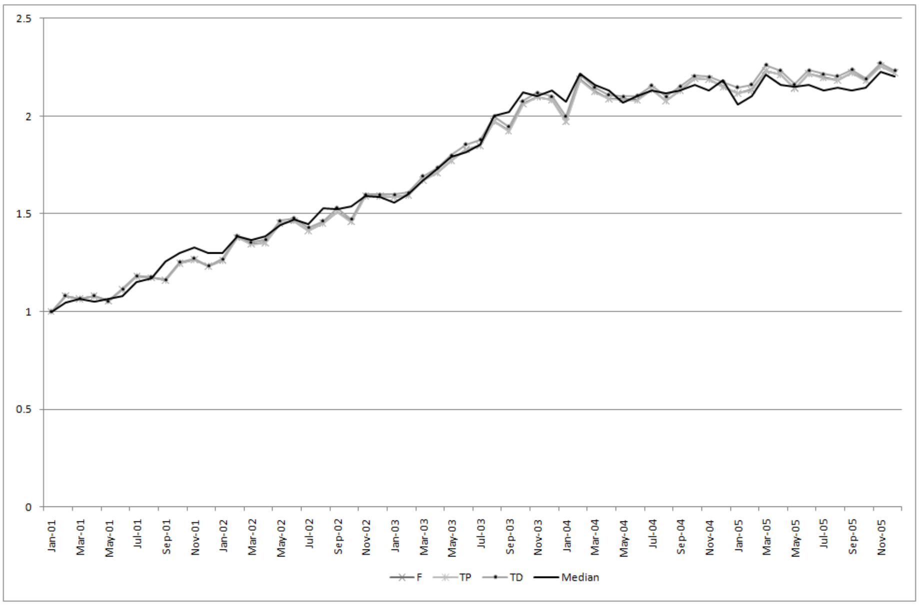


Figure 8: Chained Monthly Indices. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 2001:1 to 2005:12