Hedonic Predicted House Price Indices Using Time-Varying Hedonic Models with Spatial Autocorrelation.

Alicia N. Rambaldi*and D.S. Prasada Rao

July 23, 2011

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

Abstract

Hedonic housing price indices are computed from estimated hedonic pricing models. The commonly used time dummy hedonic model and the rolling window hedonic model fail to account for changing consumer preferences over hedonic characteristics and typically these models do not account for the presence of spatial correlation in prices reflecting the role of locational characteristics. This paper develops a class of models with time-varying hedonic coefficients and spatially correlated errors, provides an assessment of the predictive performance of these compared to the commonly used hedonic models, and constructs and compares corresponding price index series. Alternative weighting systems, plutocratic versus democratic, are considered for the class of hedonic imputed price indices. Accounting for seasonality in house sales data, monthly chained indices and annual chained indices based on averages of year-on-year monthly indexes are presented. The empirical results are based on property sales data for Brisbane, Australia over the period 1985 to 2005. On the basis of root mean square prediction error criterion the time-varying parameter with spatial errors is found to be the best performing model and the rolling-window model to be the worst performing model.

Keywords: state-space, hedonic imputation, Tornqvist, Fisher.

*Corresponding author. Email: a.rambaldi@uq.edu.au; Tel.:+61 7 33656576

1 Introduction

Compilation and publication of housing price index numbers is considered critical in the assessment of the general economy and is an important input into monetary policy including the setting of interest rates. In the past median house prices were used in measuring price changes but it is generally recognised that median prices are unduly affected by the mix of houses sold and can exhibit strong patterns of seasonality. A review of the most commonly used methods to construct housing price indices can be found in Hansen (2009). An overview of hedonic based methods is presented in Hill and Melser (2008). However, the most comprehensive collection of alternative methods to compute housing price indices to date is the recently released Handbook on Residential Property Price Indices, Version 3.0 (Statistics Netherlands and EuroStat , 2011) where chapters are dedicated to available methods, each of which are used by different statistical agencies, central banks and other organisations that regularly construct housing price indices. The general consensus in the price index literature is that hedonic based indices are superior, this is not only for the case of housing (see for instance Triplett (2004) for a discussion relating to information technology products). This is because indices based on hedonic regression models measure price changes after accounting for quality changes in the purchased product. As detailed in Silver and Heravi (2007), there are two types of hedonic based indices, namely, the time dummy hedonic (TDH) indices and the hedonic imputed (HI) indices. Silver and Heravi conclude that HI indices are preferred when there is parameter instability and that superlative formulations (for definition and properties of superlative indices see Diewert (1976)) such as a Törnqvist HI index are 'well grounded in index number theory and considered preferable to the TDH index which constrains the parameters to be the same.' The reference to parameters being the same refers to the assumption that shadow prices of hedonic characteristics do not change over time.

Hill and Melser (2008) present the theoretical foundations that underpin the use of hedonic models as well as the multitude of index number formulae that are available for the construction of housing price index numbers. The traditional hedonic based method to compute price indices is based on a regression where the slope parameters (the hedonic coefficients) are fixed overtime but the model includes time-dummy effects. The estimates of the parameters of the time dummies measure the quality-adjusted price change between periods 0 and $t = 1, 2, ...T$[1] leading to the TDH. In contrast, hedonic imputation indices (HI) are computed using predictions of prices from estimated hedonic models. Hill and Melser (2008) show that TDH is biased and that HI suffer from bias unless the hedonic parameters are allowed to vary

[1]Improved estimators for the time-dummy hedonic model have been proposed in the literature (see Hill et al (1997))

over time and over heterogeneous regions. Hill and Melser (2008) label this "substitution bias"[2] and advocate the estimation of a separate regression for each time period and region. A particular case of this approach uses rolling-window regressions, which consist of estimating a fixed parameter model over adjacent two-period housing prices so that the hedonic coefficients are only held constant for two periods, as opposed to the entire sample period. Triplett (2004) argues that this is a more "benign constraint" because coefficients would usually change less between two adjacent periods than over extended intervals, and hence labels the rolling-window approach as best practice among TDH. In addition, the standard hedonic regression models discussed in Hill and Melser (2008) and Triplett (2004) also make the implicit assumption[3] that prices of houses sold are independent of each other and depend only on the hedonic characteristics. This assumption is not consistent with the popular notion that when it comes to sales prices of houses location is a major factor.

If hedonic regressions have a theoretical foundation along the lines discussed in Diewert (2001) and Hill and Melser (2008) then one would expect the hedonic regression coefficients to evolve over time. In addition, the assumption of independence of house prices has long been abandonded in the real estate literature where it has become common practice to model housing prices with models that include spatial interaction effects (see Pace et al (2009)).

The main objective of this paper is to develop an econometric approach that captures the notion of smooth evolution of parameters and at the same time includes spatial effects in the hedonic models for housing. The model is written in a state-space representation and predictions from this model are used to compute HI indices. The paper concentrates on the econometric aspects of the specification, efficient estimation of parameters of the hedonic models and the subsequent computation of housing price indices. Both fixed and time-varying parameter models with and without spatial effects are considered. The paper considers the performance of a fixed parameter hedonic model with time dummy effects, a hedonic model estimated as a rolling window (RW), and time-varying parameter models. The paper details the estimation procedure for the time-varying parameter model with spatial errors. As the hedonic regressions are an intermediate step in the computation of housing price index numbers, the second part of the paper is devoted to the compilation of housing price index numbers. The paper draws on the main recommendations of Hill and Melser (2008) and focuses mainly on the Fisher and Tornqvist index

[2]This description hinges on the interpretation of hedonic coefficents as shadow prices associated with various price determining characteristics of houses. Otherwise, it is difficult to interpret the bias induced by ignoring the time-varying nature of the hedonic regression parameters.

[3]The assumption is implicit in the specification that the random disturbance terms in the housing price regression model are independently and identically distributed.

number formulae. The paper significantly deviates from the Hill and Melser (2008) approach and considers both plutocratic weights based on value shares of houses sold and democratic weights which are based on the number of houses sold. Recognising the difference between the housing price index numbers and the standard cost-of-living index numbers, the paper argues for the use of both types of weights. An important feature of the housing price sales is the presence of seasonality in the mix of houses sold and its influence on the median house prices. In accounting for seasonality, the paper constructs year-on-year monthly price index numbers using hedonic imputations and compares these with annual hedonic price index numbers.

The paper is organised as follows. Section 2 is devoted to a description of the econometric specification of various hedonic regression models considered in the paper. Details of the econometric estimation and hedonic imputation are also discussed in the section. Section 3 describes the housing sales price data for the Brisbane metropolitan area (Australia) used in the study. Section 4 presents estimated hedonic coefficeints from different models. The root mean squared prediction error is used in assessing the performance of the hedonic models and the results are discussed in the section. Estimates of hedonic models with special focus on the temporal movements of the hedonic coefficients relating to land, number of bedrooms and the number of bathrooms and the predicted price series from competing specifications are presented. Section 5 focuses on hedonic imputed indices for housing. Index number formulae with democratic and plutocratic weighting systems are presented along with a discussion of the respective merits and applications. Annual chain based hedonic price indices are presented and contrasted with the commonly used median based price index series. A few concluding observations are made in Section 6.

2 Econometric Modelling and Estimation

The housing price indices reported in the paper are based on suitably specified hedonic functions. As one of the objectives of the paper is to evaluate the in-sample prediction performance of competing econometric formulations, a comprehensive range of hedonic models are considered in the study. The base model is a *time dummy hedonic model* (TDH) which is the most commonly used model in the construction of housing price index numbers. The TDH model includes dummy variables representing time and assumes constancy of the slope parameters over time. The TDH model is first generalised to accommodate the presence of spatially correlated error term and denoted by TDH_SEM. In the spatial econometrics literature there are two main approaches to deal with spatial interactions; a spatial lag or spatial autoregressive (SAR) process, typically modelling an equilibrium outcome of a spatial or social interaction, and the spatial error

(SEM) which is not necessarily based on a theoretical model but is a special case of a non-spherical error covariance matrix (see Anselin and Bera (1998) and many references there in). The SEM is of particular relevance to hedonic modelling as omitted hedonic characteristics will end up in the error term of the model and these are likely to be related to the location of the property (such as proximity to ammenities, views, etc).

The main extension of the TDH model is to allow hedonic parameters to vary over time through a specific stochastic process denoted as the TV model. The model proposed here allows for time-varying parameters and it also allows for a different stochastic process for the time-varying intercept parameter compared to that used to model the slope coefficients. The basic idea here is that movements in the intercept term represent secular trends in prices independent of the movements in the hedonic regression coefficients such as macroeconomic shocks. In practice, these trends tend to dominate the house price movements over time. The TV model is extended to allow for the presence of spatial correlation in the disturbance term resulting in a new model denoted by TV_SEM. The spatial correlation parameter is assumed to be fixed over time and space. The estimation of the TV and TV_SEM models can be achieved through likelihood or Bayesian estimation approaches. In this paper, an additional estimation approach is used, namely a *rolling window estimator* (RW). In this method, the TDH regression is estimated with data from two adjacent time periods, and the estimation is a "rolling window" with the end result that parameters are allowed to vary over time. The reason for inclusion of the RW is that there has been some discussion and support in the housing price indices literature for the use of adjacent period regressions (see Triplett (2004)). The shortcoming of this approach is that it is not immediately clear how to obtain standard errors for the parameter estimates. A two-period rolling window regression with a model that includes spatial errors is included and denoted by RW_SEM.

The specifications used in the study are described below.

2.1 Time Dummy Hedonic (TDH) Model

The TDH model is a multiple regression model where the dependent variable is typically the log of the sale price and the explanatory variables are hedonic characteristics (attributes) of the houses in the sample. The model includes time dummy intercepts which under the assumption of fixed hedonic attributes are proper price indices.

$$y = \mu + \gamma D + X\beta + \varepsilon \tag{1}$$

where,

$N-$ Number of observations in the sample, that is, the total number of houses sold over the sample period 1,2,..., T;

$y-N \times 1$ vector of observations of the dependent variable, typically the log of sale price (p), $y = \ln p$;

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

$\mu-$ intercept

$X-N \times K$ matrix of independent variables (house attributes) ;

$D-$ is a $N \times (T-1)$ matrix of $T-1$ year time-dummy variables used in estimating price indices

$\gamma-(T-1) \times 1$ vector of unknown parameters associated with the time-dummies; and

$\varepsilon-N \times 1$ vector of random errors.

This is the traditional model used in the literature where year time dummy variables are included. The model has essentially $T-1+K$ parameters to be estimated. As an intercept μ is included in the regression only $T-1$ year time dummy variables are introduced.

2.1.1 Time Dummy Hedonic Model with Spatial Errors (TDH_SEM)

An extension of (1) to include a spatial correlation structure through the error term is given by:

$$\begin{aligned} y &= \mu + \gamma D + X\beta + \varepsilon \qquad (2) \\ \varepsilon &= \rho W \varepsilon + u \end{aligned}$$

where,

$u-N \times 1$ vector of independently and identically distributed errors;

$W-N \times N$ matrix of spatial weights (that is, it is only a function of distance between houses in the sample);

$\varepsilon-N \times 1$ vector of correlated errors;

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

Inclusion of spatial errors is designed to take explicit account of the role of unobserved locational characteristics in determining house prices. This model is expected to be particularly useful when the hedonic model does not include location variables in the regression model.

The matrix W has the following characteristics

- $w_{ii} = 0$ for all i
- w_{ij} weight representing the 'neighbour strength' of the ith house with the jth house.
- W is a row-stochastic matrix with row sums equal to unity.

That is,

$$w_{ij} = \begin{cases} \leqslant 1 & if\, i \neq j\, are\, neighbours \\ = 0 & otherwise \end{cases}, \text{ and } \sum_{j=1}^{N} w_{ij} = 1$$

In this study we assume that housing prices are influenced by the prices of the nearest neighbours. The year of sale does not enter the construction of W in this model.

In order to identify the nearest neighbours, we make use of information on the latitude and longitude of the houses sold. To form a spatial weights matrix based on contiguity or nearest neighbours in terms of Eucledian distances is intuitive; however, it can be computationally burdensome. A more elegant and computationally less intensive approach is to use computational geometry. In this paper the nearest neighbours are identified using a Delaunay triangulation. A detailed exposition of Delaunay triangulations can be found in LeSage and Pace (2009) Section 4.11. Note that when a spatial weights matrix, W, is derived using Delaunay triangles, it represents the nearest m neighbours, and thus W^2 represents neighbours to neighbours, and so on.

It is customary for the rows of W to add up to one (in which case W is (right or) row stochastic), but this also implies $\rho < 1$ (see Ord (1965) and Krämer (2005))[4]. It then follows that $(I - \rho W)^{-1} = \sum_{j=0}^{\infty} \rho^j W^j = I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \ldots$ and $E(\varepsilon\varepsilon')$ is proportional to Ω, where $\Omega = (I - \rho W)^{-1} (I - \rho W)^{-1\,\prime}$.

2.2 Rolling Window Spatial Errors Model (RW_SEM)

This is a regression model with fixed parameters and a spatial error structure. However, the parameters are allowed to vary over time through the re-estimation of the model using data from two consecutive years each time. This is often referred to as the *rolling window* (RW) model[5].

[4]This follows because ρ is strictly bounded by the inverse of the eigenvalues of W.

[5]Though this model is intuitive and practical and a method recommended by Tripplet (2004), 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 TDH model. Notwithstanding this problem, we simply follow the literature and implement the RW model.

$$y_\tau = \mu_\tau + X_\tau \beta_\tau + \varepsilon_\tau \qquad (3)$$
$$\varepsilon_\tau = \rho W_\tau \varepsilon_\tau + u_\tau$$

where,

$\tau = t + (t+1)-$ two consecutive years of pooled observations of houses sold;

$y-(N_t + N_{t+1}) \times 1$ vector of observations of the dependent variable, typically the log of sale price (p), $y = \ln p$;

$\beta_\tau - K \times 1$ vector of unknown parameters;

$X_\tau-(N_t + N_{t+1}) \times K$ matrix of independent variables (house attributes);

$u_\tau-$ vector of i.i.d random errors;

$W_\tau - (N_t + N_{t+1}) \times (N_t + N_{t+1})$.

We use the SEM specification accounting for possible unobserved location characteristics. This flexible form of (2) is given by estimating (3) through a rolling window. For instance, in a two-adjacent period overlapping window, estimates for a pooled sample of the first two periods is obtained first, periods two and three are then pooled together, three and fourth and so on. Two estimates of each period (except for the initial and last periods) are obtained through this procedure. In this paper we present the average value between the two estimates of each time period. As mentioned, a drawback of this approach is that it is difficult to obtain standard errors for the estimates which are themselves averages of estimates from two adjancent rolling window regressions..

2.3 Time Varying Parameter Models (TV)

Now we consider a more general specification where parameters are allowed to vary over time. If all the parameters are allowed to vary without a structure, the model is underidentified as there will be more parameters than observations. Further, it is intuitive to consider the case when parameters move through time in a systematic manner and we use a random walk model where the parameters in any period are a small (random) perturbation from the parameter values of the previous period. In the specification of the model, we make a distinction between the intercept and the slope parameters. The model is specified as:

$$y_t = \mu_t + X_t \beta_t + \epsilon_t, \;\; \epsilon_t \sim NID(0, \sigma^2_\epsilon I_t) \qquad (4)$$

$$\beta_t = \beta_{t-1} + \eta_t \qquad \eta_t \sim NID\left(0, \sigma_\eta^2 I_k\right) \tag{5}$$

$$\mu_t = \mu_{t-1} + \xi_t, \quad \xi_t \sim NID(0, \sigma_\xi^2) \tag{6}$$

$$E\left(\epsilon_t \eta_t\right) = 0 \tag{7}$$

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

where,

N_t number of houses sold at time t.

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

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

μ_t is a local level process

β_t is the vector of time-varying hedonic characteristics

This model is known as a local level model with explanatory variables in the state-space literature (see for example Durbin and Koopman (2001)). The local level, μ_t, follows a separate stochastic process from the slope parameters (hedonic attribute parameters, β_t). We denote this model by TV and note that it is in the form of a state-space model with state vector $\alpha_t = [\mu_t, \beta_t]$. Therefore the estimation of this model is straightforward using Kalman filtering and smoothing algorithms.

2.3.1 Time Varying Hedonic Model with Spatial Errors (TV_SEM)

A variation of the time-varying parameter model is the model where errors are assumed to be spatially correlated.

$$y_t = \mu_t + X_t \beta_t + \epsilon_t, \;\; \epsilon_t \sim NID(0, H_t) \tag{8}$$

$$\epsilon_t = \rho W_t \epsilon_t + u_t \qquad u_t \sim NID\left(0, \sigma_u^2 I_{N_t}\right) \tag{9}$$

$$\beta_t = \beta_{t-1} + \eta_t \qquad \eta_t \sim N\left(0, \sigma_\eta^2 I_k\right) \tag{10}$$

$$\mu_t = \mu_{t-1} + \xi_t, \quad \xi_t \sim NID(0, \sigma_\xi^2) \tag{11}$$

$$E(\epsilon_t, \eta_t) = 0, \; E(u_t, \eta_t) = 0, \; E(u_t, \xi_t) = 0, \; E(u_t, u_{t-s}) = 0 \;\text{ for } s \neq 0.$$

where,

u_t–$N_t \times 1$ vector of uncorrelated errors;

W_t–$N_t \times N_t$ row-stochastic matrix of spatial weights (that is, it is only a function of distance between houses sold in period t);

ϵ_t-$N \times 1$ vector of correlated errors with covariance H_t;

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

We note here that the parameter ρ is assumed to be the same for all time periods, t, and the form of H_t is shown below. Similar to the case in (4), (5) and (6), this is also a state-space model. Although the error term of the measurement equation (8), ϵ_t, is spatially correlated, it is assumed to be uncorrelated over time and Gaussian and therefore satisfies the assumptions necessary to use the Kalman algorithms. That is, the measurement and state equations both have linear Gaussian forms. To show this, we incorporate equation (9) into (8) and we can easily write the TV_SEM model as,

$$
\begin{aligned}
y_t - \begin{bmatrix} i & X_t \end{bmatrix} \alpha_t &= \rho W_t \left(y_t - \begin{bmatrix} i & X_t \end{bmatrix} \alpha_t \right) + u_t \\
\left(I_{N_t} - \rho W_t\right) y_t &= \left(I_{N_t} - \rho W_t\right) \begin{bmatrix} i & X_t \end{bmatrix} \alpha_t + u_t \\
\widetilde{y}_t &= \widetilde{X}_t \alpha_t + u_t \qquad (12)
\end{aligned}
$$

where $\widetilde{y}_t = \left(I_{N_t} - \rho W_t\right) y_t$, $\widetilde{X}_t = \left(I_{N_t} - \rho W_t\right) \begin{bmatrix} i & X_t \end{bmatrix}$, i is an $N_t \times 1$ vector of ones and $\alpha_t' = \begin{bmatrix} \mu_t & \beta_t \end{bmatrix}$. We obtain (12), a linear Gaussian form since $u_t \sim NID\left(0, \sigma_u^2 I_{N_t}\right)$.

Alternatively, setting $\epsilon_t = \left(I_{N_t} - \rho W_t\right)^{-1} u_t$, in (8) also shows a linear Gaussian form since $\epsilon_t \sim N\left(0, H_t\right)$ where

$$H_t = \sigma_u^2 \left(I_{N_t} - \rho W_t\right)^{-1} \left(I_{N_t} - \rho W_t\right)^{-1\prime}. \qquad (13)$$

and, it is easily seen that if ρ is zero in (9), the error term $\epsilon_t = u_t$, and (8) is (4).

Given ρ, a convenient state-space representation is;

$$\tilde{y}_t = \tilde{X}_t \alpha_t + u_t \qquad (14)$$

$$\alpha_t = \alpha_{t-1} + \zeta_t \qquad (15)$$

where,

$\zeta_t = \begin{bmatrix} \xi_t \\ \eta_t \end{bmatrix}$ with $Q_t = E(\zeta_t \zeta_t')$, $Q_t \sim N(0, \begin{bmatrix} \sigma_\xi^2 & 0 \\ 0 & \sigma_\eta^2 I_K \end{bmatrix})$.

when ρ is unknown and must be estimated a more convenient state-space representation is given by (16) and (17):

$$y_t = Z_t \alpha_t + \epsilon_t \tag{16}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t \tag{17}$$

where,

$Z_t = \begin{bmatrix} i & X_t \end{bmatrix}$

$E(\zeta_t \zeta_t') = Q_t$, and $E(\epsilon_t \epsilon_t') = H_t$

Estimates of the parameters $\psi = \rho, \sigma_u^2, \sigma_\xi^2, \sigma_\eta^2$ are required for the Kalman filtering algorithms to provide an estimate of α_t and its Mean Squared Error Matrix. Estimation is discussed next.

2.4 Econometric Estimation

The TDH model is estimated using least squares. The TDH_SEM, and RW_SEM are estimated by maximum likelihood using the *sem.m* Matlab function developed by LeSage and Page (see `http://www.spatial-econometrics.com/`). Computationally efficient estimation of spatial models (both in a classical or Bayesian context) require evaluation of log-determinants that are functions of W. LeSage and Pace (2009), Chapter 4, 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 and *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 maximum likelihood approaches. The models were estimated using code specially written by the first author. In both cases hyperparameters (that is, constants of proportionality, σ_i^2,$i = \epsilon, \eta, ..$, and the spatial parameter ρ) were estimated using the standard state-space approach based on the numerical maximization of the conditional likelihood function (see Harvey (1989), Chapter 3 or Durbin and Koopman (2001), Chapter 7 for example) and the use of the Kalman filter and smoothing algorithms.

To obtain the likelihood function, the definition of a conditional probability density function is used

$$L(y; \psi) = \prod_{t=1}^{T} p(y_t | Y_{t-1}) \tag{18}$$

where $p(y_t|Y_{t-1})$ denotes the distribution of y_t conditional on the information set at time $t-1$, that is $Y_{t-1} = \{y_{t-1}, y_{t-2}, ..., y_1\}$.

Using the measurement equation (16), the prediction of the conditional distribution of y_t, $(N_t \times 1)$ is given by

$$E_{t-1}(y_t) = \tilde{y}_{t|t-1} = Z_t a_{t|t-1}$$

where $a_{t|t-1}$ is an estimate of α_t given Y_{t-1}. A prediction error is given by $\nu_t = y_t - \tilde{y}_{t|t-1}, \quad t = 1, ..., T$; with covariance matrix $E(\nu_t \nu_t') = F_t$

Therefore for a Gaussian model, the log likelihood function can be written as:

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

Newton type numerical optimisation methods are used to find the values of ψ. Choosing a set of starting values for the parameters $\psi = \psi_0$, an iterative algorithm provides a value of lnL at given value of ψ by running the state-space model through the equations of the Kalman Filter to obtain a value of ν_t and F_t at each iteration. The maximum likelihood estimates of the hyperparameters is then:

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

Given $\hat{\psi}$, estimates of the covariances Q_t and H_t , $\hat{Q}_t$ and $\hat{H}_t$, are now available and the estimates of α_t and its Mean Squared Error matrix are obtained by running the state space through the equations of the Kalman filter and smoother.

3 Data

The data used in this study refer to residential property sales in the Brisbane (Australia) metropolitan area for the period 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). These data were first collected by Cominos (2006). Further filtering of the data was conducted by Svetchnikova (2007) and the resulting data set is used in this study. The empirical work for the study is limited to price data for residential houses on blocks of land and excludes units, terraces, townhouses and duplexes[6]. The dataset used in the

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

study contains 65,239 single transactions over the sample period. Each data point (transaction) includes, the date (month and year) of sale, sale price, geocode (latitude, longitude), the postcode, 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) combined into one series (CARLUG).

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 period.

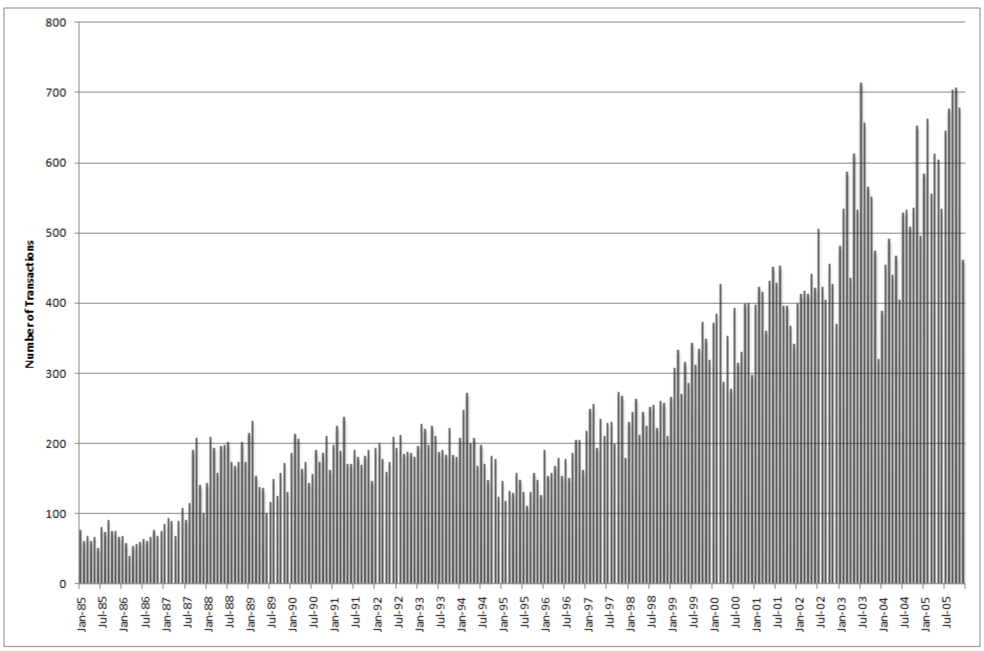


Figure 1: Number of transactions per month in the dataset

4 Hedonic Model Results

For the estimation of TDH and TDH_SEM the sample is pooled over the period 1985-2005. These models are not expected to perform well in prediction; however, they will serve as base models for the purpose of comparison. To study the in-sample predictive performance we compute the Root Mean Square Prediction Error (Root MSPE) of each alternative model based on the prediction of individual log transform of sale price. For the RW-SEM we pool two years and overlap one year as we move through the sample (that is

1985 and 1986, 1986 and 1987,..).The estimated coefficients for 1986 would be an average of the coefficeints from the 1985 and 1986 data and from 1986 and 1987 data. The TV and TV_SEM are estimated with monthly transactions ($T = 252$ for the period 1985:1 to 2005:12). The estimates of ρ from the TDH_SEM and TV_SEM are very close, $\hat{\rho} = 0.46$ for the first and $\hat{\rho} = 0.48$ for the TV_SEM. This is very reassuring as the spatial correlation parameter is assumed to be constant across time and space.

4.1 Comparative Performance of Alternative Models

From the fitted models we compute the Root Mean Square Prediction Error (*RMSPE*) below, and the results are in presented in Table 1.

$$RMSPE = \sqrt{\frac{1}{N}\sum_{i}\sum_{t}(y_{it} - \hat{y}_{it})^2}$$

where,

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

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

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

Model	No Spatial Effects		Spatial Effects		
	Time Dummy (TDH)	Time Varying (TV)	Time Dummy (TDH_SEM)	Rolling window Model (RW_SEM)	Time Varying (TV_SEM)
Root MSPE	0.4224	0.4153	0.4214	0.4263	0.3780
reduction/ increase from base model	base	-1.7%	base	1.2%	-11.5%

In Table 1 we present the models in two groups depending on whether spatial effects were considered in the modelling and estimation. The TDH is the base model for models without spatial effects, and we see that allowing the hedonic parameters to vary over time results in a marginal reduction in RMSPE of 1.7%. Among the models with spatial effects, the base model is TDH_SEM. We note that TDH_SEM has a lower RMSPE than TDH. However, a surprising finding is that relaxing the fixed parameters assumptions by implementing an adjacent period rolling window (RW_SEM) results in an increase and not a decrease in RMSPE. Basically this result implies that RW_SEM lacks predictive power and at the same time has specification and conceptual issues with its formulation. The compounding effect of spatial errors and time-varying hedonic parameters in TV_SEM results in a large reduction in RMSPE (11.5%) over the TDH_SEM model's performance.

These results indicate that there are gains to be made by using time-varying parameters; but, using adjacent periods regression might not result in any gains in predictions. 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.

4.2 Coefficient Estimates from TV_SEM and Rolling Window (two years) RV_SEM Models

In this section we compare the estimates of hedonic coefficients associated with number of bathrooms, bedrooms and land. We also present estimates of the local level TV_SEM as well as the rolling window SEM. The general observation is that the estimated coefficients from the RW_SEM do not perform well. We present our estimates from these two models in the following figures.

Figure 2: TV_SEM. Local Level (intercept), μ_t

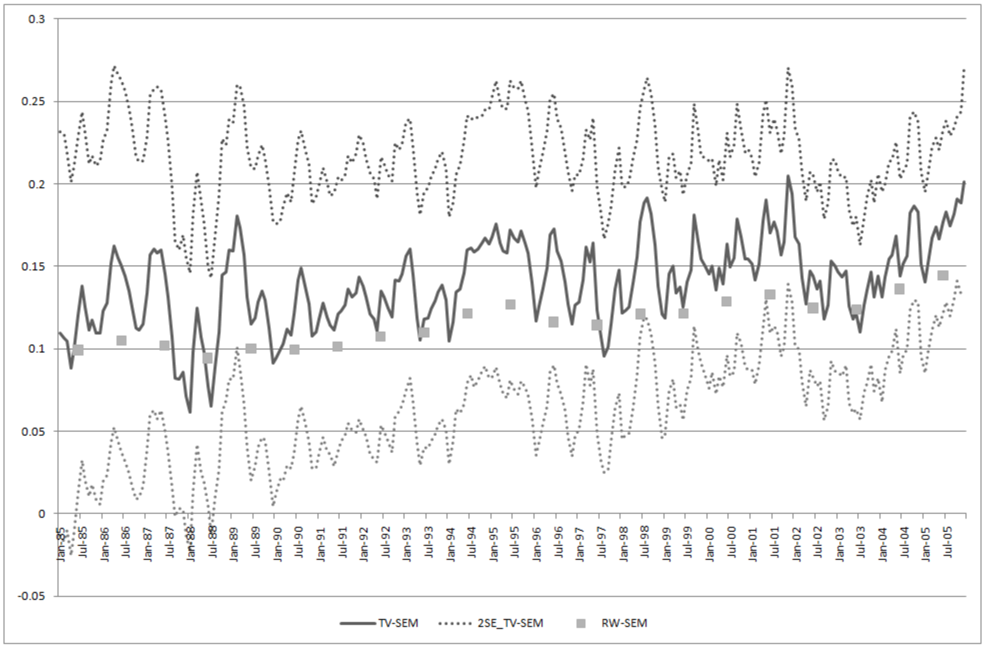


Figure 3: TV_SEM. BATH Coefficient

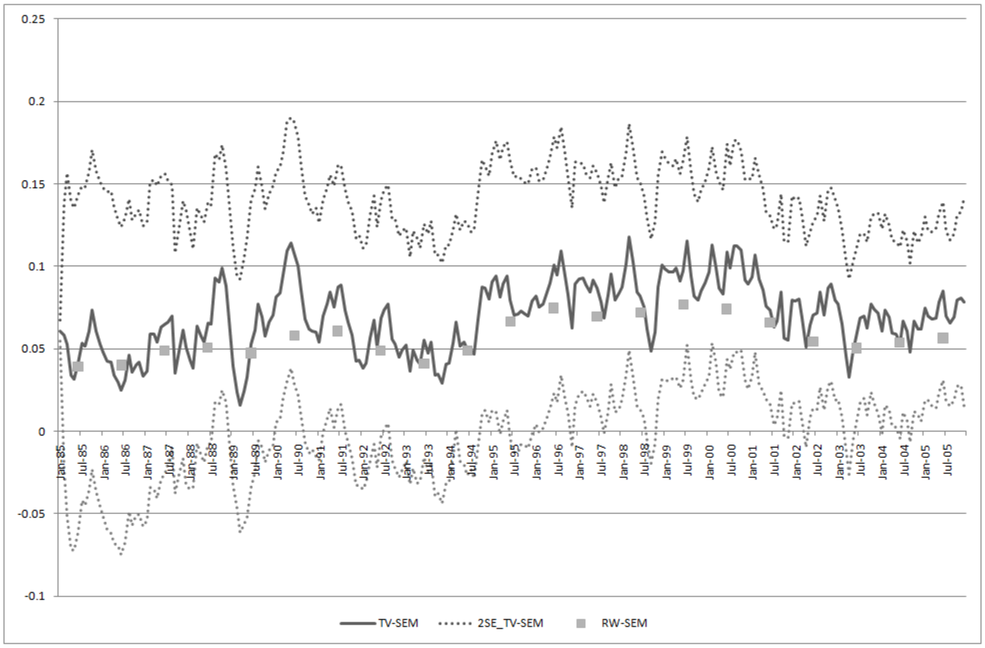


Figure 4: TV_SEM. BED Coefficient

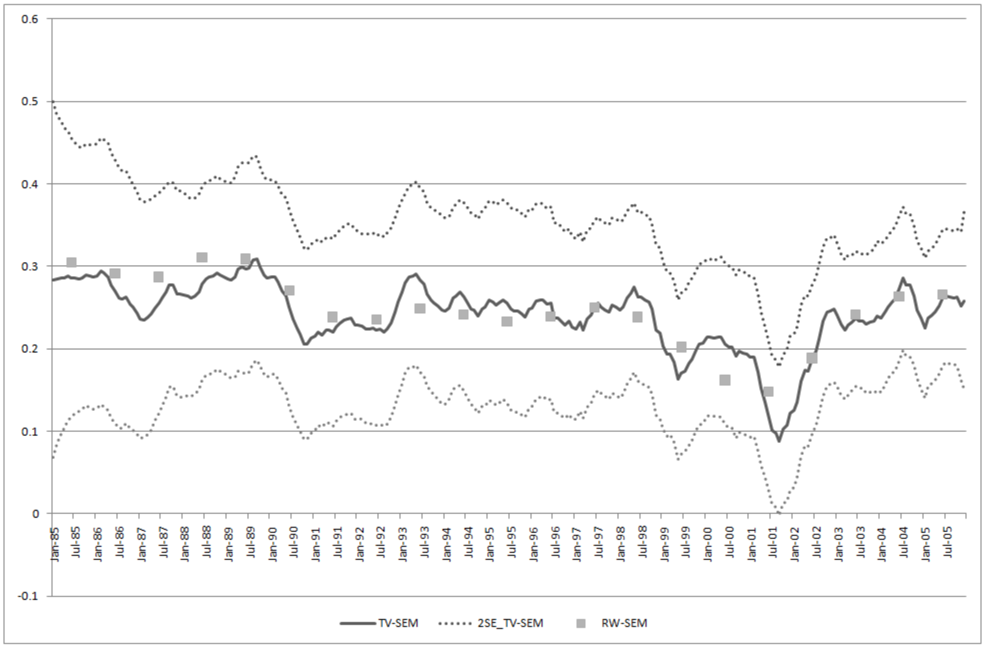


Figure 5: TV_SEM. Land Coefficient

From all the figures it is clear that the RW_SEM parameter estimates lie within the 2-standard error interval around the TV_SEM model. However, we find serious discrepancies between the two sets of estimates. In Figure 2 we plot the estimated local level component from the TV_SEM model together with a 2-standard error band and the corresponding estimates of the intercept from the RW_SEM estimation While the RW_SEM estimates of the intercept appear to track the more general and conceptually superior TV_SEM model there are certain periods (July 93 to July 2000) when the RW_SEM intercepts are higher than those derived using the TV_SEM model. Given the relative magnitude of the intercept coefficient (as the model is log-linear), it is clear that these differences will have a significant impact on the predictive performance of these two models. From Table 1 we can see that that RW_SEM produces predictions with the highest root mean square prediction error. The poor performance of the RW_SEM may also be attributable to estimates of parameters of the slope coefficients from the RW_SEM. Estimates of hedonic coefficients for the bathrooms and bedrooms, in Figures 3 and 4 respectively, from the TV_SEM model are a lot more volatile than the coefficients from the RW_SEM which is consistent with the RW_SEM approach which implies relative constancy of parameters through time. From Figure 5 we find that house prices are relatively inelastic with respect to the land size. This is a somewhat surprising result as the cost of land is a major component of the price of the house. A careful examination shows that a possible reason for this result is the lack of variability in the size of land. In fact, most of the blocks of land on which dwellings in the sample are found are around 670 square meters. With a few exceptions, the RW_SEM appears to perform reasonably well with respect to the land coefficient for most of the periods.

4.3 The Evolution of Prices

The model with the lowest RMSPE is the TV_SEM. We produce house price predictions for the houses sold in a particular period using the model:

$$\widehat{\ln P_t} = \hat{\mu}_t + \hat{\beta}_{1t} \ln Land_t + \hat{\beta}_{2t} BED_t + \hat{\beta}_{3t} BATH_t + \hat{\beta}_{4t} CARLUP_t + \hat{\rho} W_t \hat{\epsilon}_t \tag{19}$$

where,

$\hat{\rho}$ is the estimate of ρ in the TV_SEM model obtained by numerical maximization of the conditional likelihood.

$\hat{\mu}_t$ and $\hat{\beta}_{kt}$, $k = 1, ..., 4$ are the Kalman smoothed estimates from the TV_SEM model in Sectoin 3.3.1

$\hat{\epsilon}_t$ is the residual from the estimated TV_SEM model.

$\widehat{\ln P_t}$ in (19) is the best linear predictor of the $\ln P_t$ in the presence of spatially correlated disturbances.

Due to the log-log nature of the model in terms of the land area and the lack of information on the age of the structure[7], it is not possible to separate the land from the "structure" components of the price of a property. This is an important and topical issue and an in-depth discussion can be found in Diewert et al (2011). Therefore, we focus on the predicted price of the bundle. In Figure 6 we present an estimate of the median monthly price of properties sold in each time period. The estimates are obtained as follows:

$$\hat{P}_t^m = median(exp(\widehat{\ln P_t}))$$

For each period, we compute the median of the predicted prices of all the houses sold in that period computed using our preferred TV_SEM. This is slightly different, but conceptually superior, to the normal practice of computing predicted price of a house using median values of hedonic characteristics (land, bedrooms and bathrooms) in different months. As expected, the use of median values of characteristics produced a much more volatile series of median prices. Given the superior predictive performance of the TV_SEM, the observed and predicted median house prices are closely aligned over the period.

[7]In Australia there are two main commercial providers of multiple listing transaction data on sale price of real estate products. Although the quality of the data available is improving there are two important hedonic characteristics that are not available through either of the commercial providers for the majority of the submarkets in the country; namely, year the structure was built and size of the structure. Only through merging commercial with government datasets it is possible to have information on these very important hedonic characteristics.

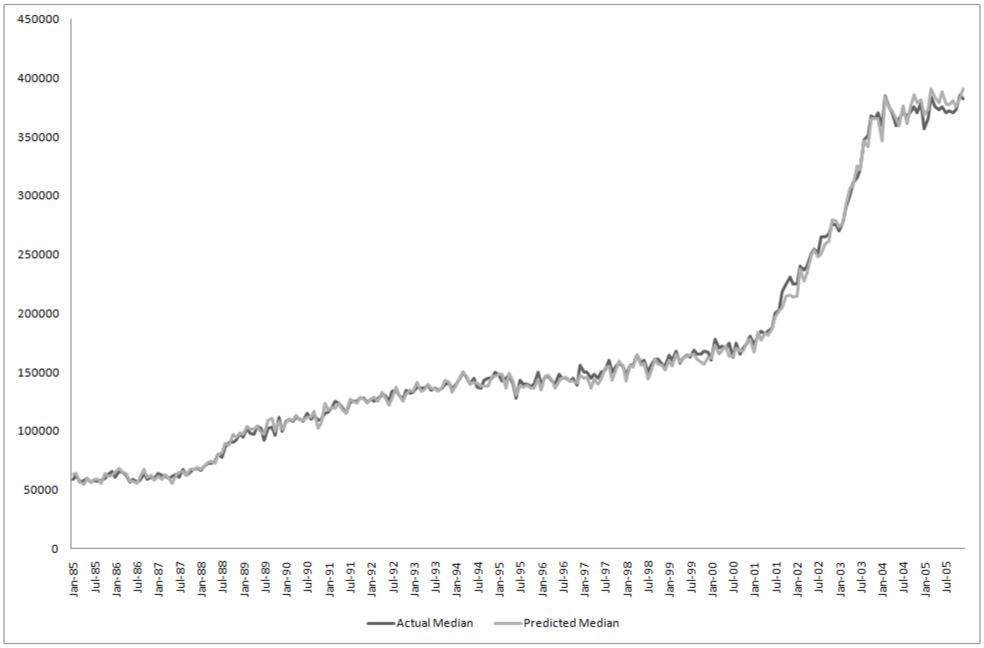


Figure 6: Prediction of median sale prices over the sample period

Figure 6 provides an interesting profile of prices of houses 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 July 1988, a surge in house prices is evident over the two year period until July 1989. Over the next decade from January 1990 until January 2001 there has been a steady increase in median prices from just above 100,000 dollars to 175,000 dollars. There has been a sharp rise in house prices from January/July 2001 until January 2004 where the prices had more than doubled. It is possible that structural breaks have occurred over this period; however, we have not formally tested for this possibility.

5 Hedonic Imputed Price Index Numbers for Housing

In this paper we report several sets of hedonic imputed price index numbers for housing. Hill and Melser (2008)-HM provide an extensive discussion of a range of index number formulae that are based on different sets of weighting systems and using different sets of imputed prices. The general conclusion by HM 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. We basically follow these recommendations.

As a deviation from the general practice in this area, we construct price index numbers with *plutocratic* *and democratic weights.* Plutocratic weights reflect the prices of different houses and higher priced houses are accorded higher weights in the index construction. These are essentially value shares of different houses sold at a given point of time. The use of plutocratic weights along with a Laspeyres type index (as in equation 22) measures the price change by comparing the total value of the housing stock in the base period and current period using hedonic imputations[8]. Similarly the Paasche index compares the housing stock of the current period at the base and current period prices. However, the geometric indices like the Tornqvist indices cannot be interpreted along the same lines. All the indices discussed in HM are essentially plutocratic indices.

In contrast, the democratic weighting system gives the same weight for 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 houses sold in any given area is taken as a random sample and therefore the price observations are assumed to have the same variance. However, when the geometric Tornqvist index is computed, we explicitly recognise the unequal numbers of houses sold in the two years and define a weighted geometric mean of the geometric Laspeyres and Paasche indices (see equation 27).[9] 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 house prices. Given that the expenditure weights used in hedonic imputed price indices 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 really depends upon the main objective behind the housing price index construction.

5.1 Index Number Formulae Used

Let P_t^h represent the price of house h in period t. Further, let w_t^h be the value share of the house h defined as:

$$w_t^h = \frac{P_t^h}{\sum\limits_{n=1}^{N_t} P_t^n} \tag{20}$$

where,

[8]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.

[9]It is possible to consider a more sophisticated approach after stratifying the sample into different regions and by the type of houses. However, we are yet to implement the stratified sampling approach.

P_t^h is the actual price. Typically in our case t refers to a particular month as we are making use of monthly sales data. Construction of annual indices is considered in Section 5.2.

We define the following types of indices used in the study.

PLUTOCRATIC INDICES: These indices are weighted indices where weights represent the relative value of each of the houses included in the sample. In the paper we use the Fisher and Tornqvist variants of this index. These indices are computed using:

(i) *actual shares* based on actual prices as defined in (20); and

(ii) imputed prices in both the base and current periods.

Hence, the Hedonic Imputed price indices for period t, with period s as the base, used in our study are defined as follows.

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

$$F_{s,t}^P = \sqrt{L_{s,t}^P P_{s,t}^P} \tag{21}$$

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

$$L_{s,t}^p = \sum_{h=1}^{N_s} w_s^h \left(\frac{\hat{P}_t^h(x_s^h)}{\hat{P}_s^h(x_s^h)} \right) \tag{22}$$

$$P_{s,t}^p = \left[\sum_{h=1}^{N_t} w_t^h \left(\frac{\hat{P}_s^h(x_t^h)}{\hat{P}_t^h(x_t^h)} \right) \right]^{-1} \tag{23}$$

with the value shares defined as in (20).

We note here that if the shares are based on predicted prices then the Laspeyres and Paasche indices defined in (22) and (23) simply turn out to be *ratios of the value of the stock of houses* in periods t and s respectively evaluated using the hedonic price models in these periods. Thus the index in (21) simply measures the change in the value of the housing stock due to changes in prices as reflected in the hedonic model of prices.

The **Tornqvist indices** are defined similar to equations (21), (22) and (23). Following HM, we define these indices as follows:

$$T^P_{s,t} = \sqrt{GL^p_{s,t} \times GP^p_{s,t}} \tag{24}$$

where $GL^p_{s,t}$ and $GP^p_{s,t}$ are the plutocratic geometric Laspyeres and geometric Paasche indices which are defined as:

$$GL^P_{s,t} = \prod_{h=1}^{N_s} \left[\frac{\hat{P}^h_t}{\hat{P}^h_s}\right]^{w^h_s} \tag{25}$$

$$GP^P_{s,t} = \prod_{h=1}^{N_t} \left[\frac{\hat{P}^h_t}{\hat{P}^h_s}\right]^{w^h_t} \tag{26}$$

These indices are "plutocratic" and are influenced by houses with large price tags. Despite this, the Fisher and Tornqvist indices in (21) and (24) measure the changes in the housing stock values that can be attributable to price changes, and therefore provide useful information.

We now deviate from the HM approach and define democratic indices which are statistically based measures of price changes.

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

$$T^D_{s,t} = \sqrt{GL^D_{s,t}\, GP^D_{s,t}} = \sqrt{\left[\prod_{h=1}^{N_s}\left(\left[\frac{\hat{P}^h_t(x^h_s)}{\hat{P}^h_s(x^h_s)}\right]^{\frac{1}{N_s}}\right)\right]\left[\prod_{h=1}^{N_t}\left(\left[\frac{\hat{P}^h_t(x^h_t)}{\hat{P}^h_s(x^h_t)}\right]^{\frac{1}{N_t}}\right)\right]} \tag{27}$$

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 geometric mean is consistent with a general log-normal distribution of price relatives.

5.2 Annual Price Indices

In the presence of seasonality we consider the problem of construction of chained indices. 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, a chained index will not normally return to unity. Hence, the use of chained indices for noisy monthly or quarterly series is not recommended.

In view of the drift caused by chaining in the presence of oscillations and in view of the presence of seasonality in the sales of houses and the types of houses sold it may be better if we compute month-on-month housing price indices and combine them to yield a year-on-year index. The following two methods from chapter 22 of the ECE-ILO Manual on the Consumer Price Index (ILO, 2006) are considered. In the ensuing empirical work the simple method due to Yule is employed.

5.2.1 Yule (1921)'s method (page 8, Chapter 22, ILO, 2006)

Step 1: 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 indices with plutocratic and democratic weights.

Step 2: The year-on-year index is then computed as a simple unweighted geometric mean of the month-on-month index

5.2.2 Stone (1965)'s index (pp. 15-16, Chapter 22, ILO, 2006)

Step 1: Compute the year-over-year monthly indices using standard index number formulae.

Step 2: Compute the year-on-year annual indices as follows:

$$L_{t,t+12} = \sum_{m=1}^{12} \sigma_m^t L_{t,t+12,m} \tag{28}$$

$$P_{t,t+12} = \sum_{m=1}^{12} \sigma_m^{t+1} P_{t,t+12,m} \tag{29}$$

$$F_{t,t+12} = \sqrt{L_{t,t+12} P_{t,t+12}} \tag{30}$$

where,

$$\sigma_m^s = \frac{\sum_{h \in H_m^s} p_h^{s,m} \cdot q_h^{s,m}}{\sum_{m=1}^{12} \sum_{h \in H_m^s} p_h^{s,m} \cdot q_h^{s,m}} \quad \text{with} s = t, t+12$$

are the value shares of houses sold in different months.

Step 3: Step 2 provides plutocratic indices. Weights in the Laspeyres and Paasche indices can be replaced by the number of houses sold in different months.

Step 4: We can use geometric versions of these formulae leading to Tornq$\ddot{v}$ist indices.

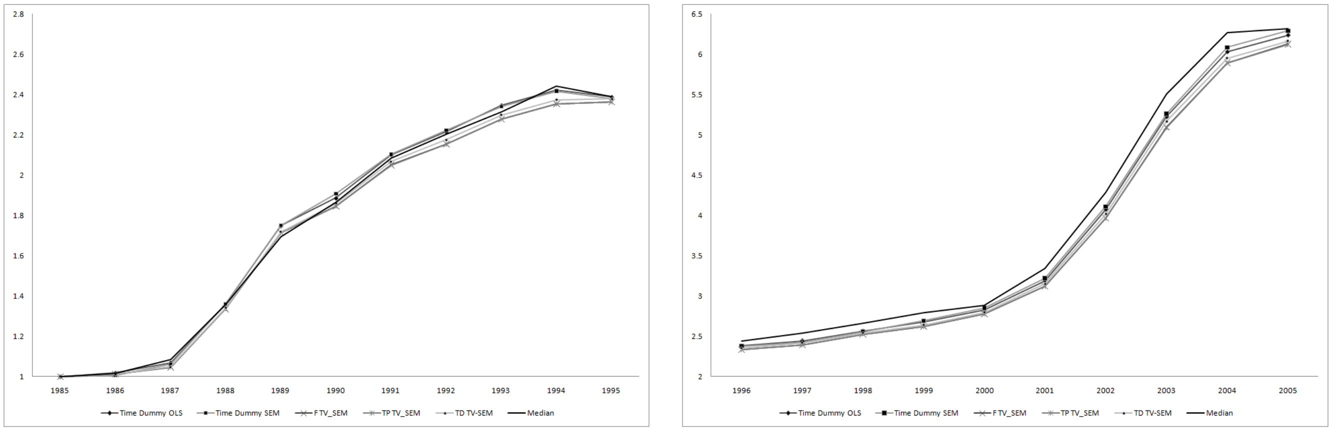
5.2.3 Estimates of Annual Housing Price Indices

In this section we present annual housing price indices which provide a measure of changes in housing prices from one year to the next starting from 1985 computed using the method described in section 5.2.1. We present indices based on the application of the *time dummy hedonic* (TDH) model as well as its extension that accounts for the presence of spatial correlation of errors, TDH_SEM, generated through locational characteristics. Even though the *rolling window* (RW_SEM) method is quite popular in the hedonic price index literature, we found the RW_SEM method to be the least performing model in terms of its predictive power within the sample. As hedonic price indices depend upon imputed prices of houses sold in different time periods, the use of RW method could introduce serious biases. As the main focus of the paper is on *time-varying* (TV) *parameter hedon*ic models we present several indices based on the TV_SEM which accounts for spatially correlated errors. We note here that the annual price indices from the TDH and TDH_SEM models are automatically given by the estimates of the parameters of the dummy variables contained therein, there is no need to use any specific index number formula. In contrast, when the RW, TV and TV_SEM models are used we need to decide whether a Fisher or Tornqvist index number formula is used and whether we compute these two indices using plutocratic or democratic weights.

In Figure 7, we present chained annual housing price index numbers from different models and computed using different index number formulae.[10] We also present chained annual index computed using the median price observed in each year. The median price index provides a frame of reference. All the indices are computed using 1985 as the base year. - As the Fisher and Tornqvist indices are both superlative [11]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 TDH indices are above the median index for a large portion of the period while the HI indices are consistently below the median based index. For the 1996-2005 period, all the chained indices are significantly below the median-based chain index of housing prices indicating an upward bias in the median price indices normally reported in the popular press.

[10] Annual housing price indices are first computed for each year relative to the previous period as an average of year-over-year monthly indices. These indices are then chained from year to year to yield price indices with a fixed base.

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



1985-1995 (1985=1) 1996-2005 (1985=1)

Figure 7: Annual Indices - Chained

5.3 Monthly Chained Indices from TV_SEM Model

From the annual price indices we now turn to chained prices indices constructed using month-to-month price indices. here we compute the price indexes from one month to the next and cumulate the changes over time through chaining. In this section we mainly focus on the plutocratic and democratic weighted price indices computed using the Fisher and Tornqvist index numbers. The Median housing price indices are also presented. Based on the results reported in Figure 7, we do not present price indices based on time dummy methods which implicitly assume constancy of parameters of the hedonic model. As the *time-varying parameter model with spatial errors* (TV_SEM) has the best predictive power within the sample period, we present only results based on the TV_SEM model.

Figure 8 presents chained monthly indices with January, 1985 as the base computed using the plutocratic Fisher and Tornqvist indices and the democratic weighted Tornqvist indices. The median housing price index is also presented. By the end of the study period, there has been a significant difference between the median and the hedonic price index numbers and of the magnitude of 20 to 30 percent higher when median is used relative to the Fisher and Tornqvist indices. We also note that the democratic weighted Tornqvist index is uniformly higher than the plutocratic weighted index but the percentage difference is much smaller compared to the median based price index.

As with the annual chained price indices, we note the presence of three episodes of acceleration in the housing prices. However, there is evidence of seasonal fluctuations in the indices but we do not notice any major drift in the indices. In order to facilitate visual examination of the differences, we split the period into two periods, 1985 to 1995 and from 2001 to 2005 and present the indices in Figures 9 and

10 separately for the two periods. We found these periods to represent periods of accelerated increases in prices. From Figure 9 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 effect of the mix of houses sold in different periods. However, all the indices are much more closely aligned during the period 2001 to 2005 and there are no appreciable differences between the median and hedonic price indices. This is in sharp contrast to the significant deviations between these two sets of indices observed during 1985 to 1995 period. The close alignment in the indices observed during 2001 to 2005 could be due to the fact that the during housing price boom price increases were uniform across all types of houses sold which in turn implies that the mix of houses sold will not significantly affect the housing price indices. This is an aspect that requires further analysis.

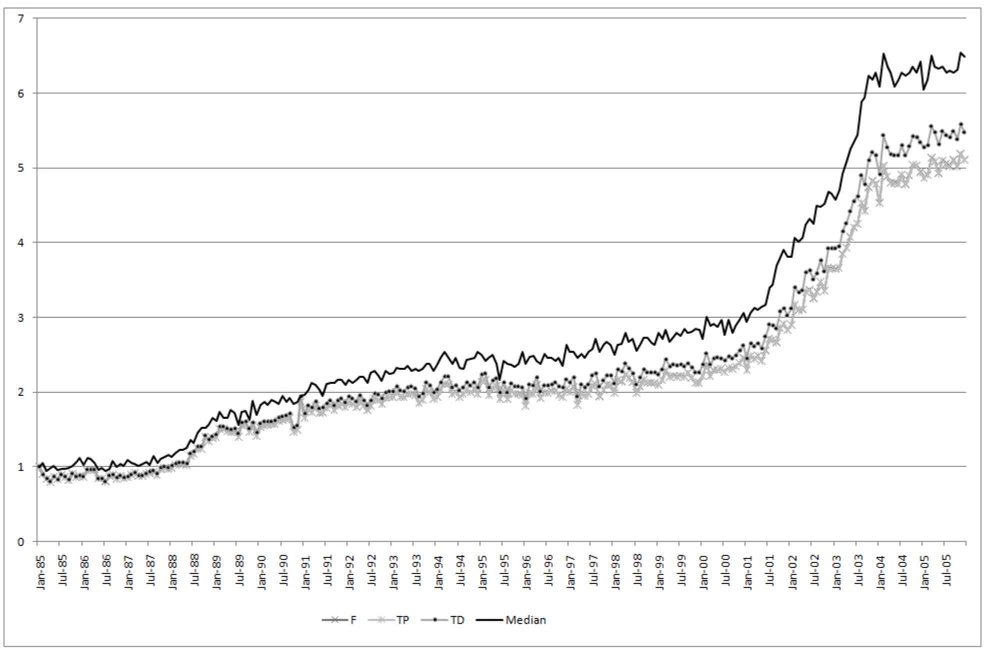


Figure 8: Chained Monthly Indices. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 1985:1 to 2005:12

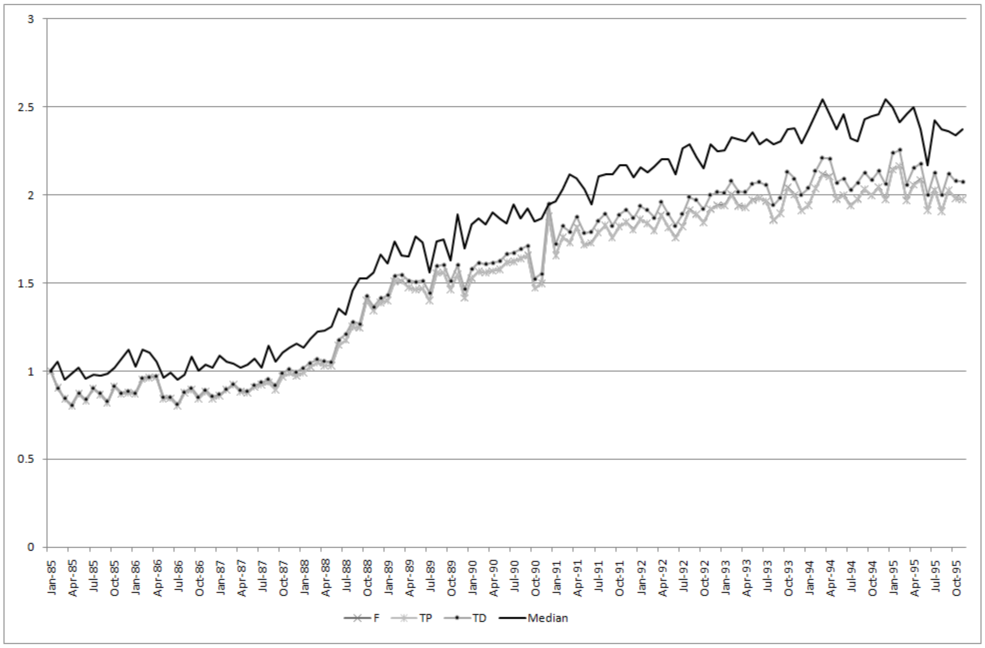


Figure 9: Chained Monthly Indices. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 1985:1 to 1995:12

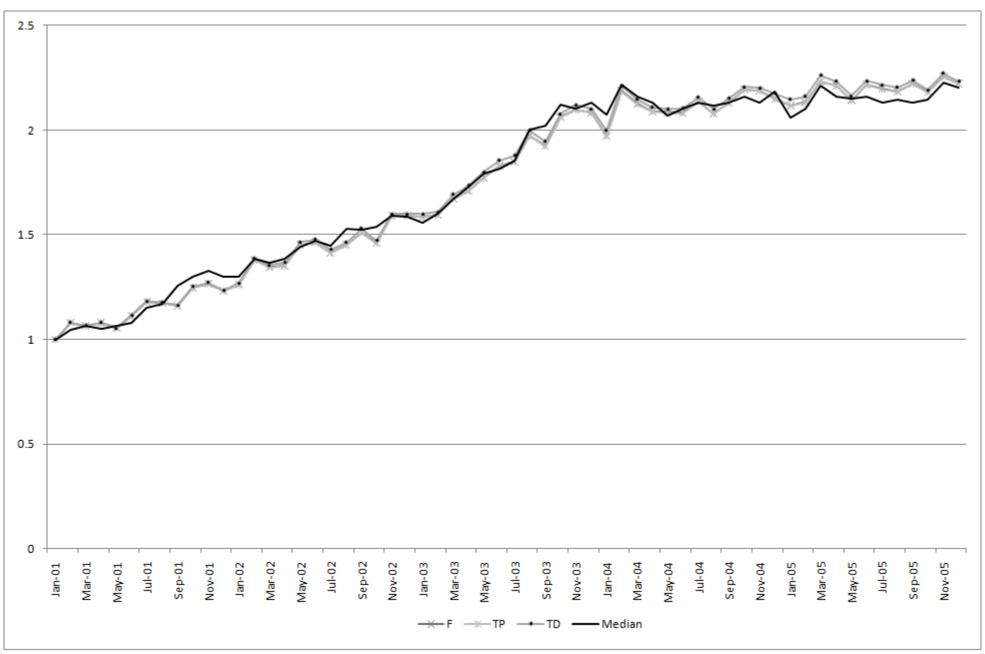


Figure 10: Chained Monthly Indices. F:Fisher Plutocratic, TP: Tornqvist Plutocratic, TD: Tornqvist Democratic for the period 2000:1 to 2005:12

6 Conclusions

The paper has dealt with several important issues relating to hedonic modelling 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 highlights the need to model the time-varying nature of the hedonic coefficients and also the importance of making optimal use of the information on the influence of locational characteristics available in the form of spatially correlated errors. A related 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 housing price index numbers. We also focus on the influence of the weights, plutocratic versus democratic weights, and on the chained annual indices. Finally, we examine the month-to-month housing price indices based on time-varying hedonic regression models to examine the general trends in the index series. 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 predictive power of the time-varying hedonic model with spatially correlated errors and we also show that the worst performing model is the rolling window approach commonly recommended in the hedonic price index literature. We also find that the use of the time dummy approach is likely to mask important underlying movements and features of the hedonic price index numbers. It is not clear if this applies mainly to our Brisbane sample but it is likely that this is an intrinsic feature of the time-dummy hedonic price index numbers. In general, the median price index provides an upper-bound and clearly well above the price indices from all the other approaches with the possible exception during periods of rapid and uniform price increases. 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 indices during the period 2001-2005. This result is particularly interesting as the housing market experienced a price boom during this period. We attribute this feature to the possibility that house price increases were uniform across different types of dwellings and in different locations. Trends in the chained annual price indices as well as chained monthly price indices 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.

Acknowledgement

The authors would like to acknowledge expert research assistance by Linh T. Huynh.

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.

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.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics*, 4, 114-145.

Diewert, W.E. (2001), "Hedonic Regressions: A consumer Theory Approach", Discussion Paper 01-12, Department of Economics, University of British Columbia, Vancouver, Canada.

Diewert, W. E., Jan de Haan and R. Hendriks (2011), "Hedonic Regressions and the Decomposition of a House Price index into Land and Structure Components ," Discussion Paper 11-02, Department of Economics, The University of British Columbia, Vancouver, Canada, V6T 1Z1.

Durbin, J. and S. J. Koopman (2001), Time Series Analysis by State Space Methods, Oxford Statistical Science Series. Oxford University Press Inc.

Hansen, J. (2009), "Australian House Prices: A Comparison of Hedonic and Repeat-Sales Measures," *Economic Record*, 85:269, pp. 132-145.

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

Hill, R.C., J. R. Knight, and C. F. Sirman (1997), "Estimating Capital Asset Price Indices," *The Review of Economics and Statistics,* 79:2, pp. 226-233.

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.

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.

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, p. 418-424.

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

Syed,I., R.J. Hill and D. Melser (2008), "Flexible Spatial and Temporal Hedonic Price Indices for Housing in the Presence of Missing Data," School of Economics Discussion Paper: 2008/14. Australian School of Business. University of New South Wales.

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 Indices and Quality Adjustments in Price Indices: Special Application to Information Technology Products", OECD Science, Technology and Industry Working Papers, 2004/9, OECD Publishing. doi:10.1787/643587187107

Statistics Netherlands and EuroStat (2011), *Handbook on Residential Property Price Indices*, v3.0 (draft version available for comments). Statistics Netherlands and EuroStat. `http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/documents/Tab/Tab/RPPI_Handbook%20_combined_V3_0.pdf`