Session 7A: Role of the Housing Market in the Economy Time: Friday, August 10, 2012 AM

> Paper Prepared for the 32nd General Conference of The International Association for Research in Income and Wealth

## Boston, USA, August 5-11, 2012

## The Effect of Alternative Estimators of the Hedonic Model on the Revisions of Hedonic Imputed Residential Property Price Indexes

Alicia N. Rambaldi and Cameron Fletcher

For additional information please contact:

Name: Alicia N. Rambaldi Affiliation: University of Queensland, Australia

Email Address: a.rambaldi@uq.edu.au

This paper is posted on the following website: http://www.iariw.org

# The Effect of Alternative Estimators of the Hedonic Model on the Revisions of Hedonic Imputed Residential Property Price Indexes<sup>\*</sup>

Alicia N. Rambaldi<sup>1</sup>, Cameron Fletcher<sup>2</sup>

<sup>1</sup>School of Economics. The University of Queensland. St Lucia, QLD 4072. Australia. Email: a.rambaldi@uq.edu.au

<sup>2</sup> CSIRO Ecosystem Sciences, PO Box 780, Atherton, QLD 4883, Australia

## version: 15 July, 2012

Parallel Session 7A: The Role of the Housing Market in the Economy. Organiser: Robert Hill, University of Graz, Austria.

#### Abstract

The main objective of the paper is to study the effects that choices made in modeling, estimation and prediction of the transaction price have on the computed hedonic imputed price index. The paper first presents a general representation of the hedonic model that serves as a general framework to allow a discussion of alternative estimators and predictors. These lead to a number of alternatives that vary in the conditioning set used, the estimated time-path of shadow price parameters, and the restrictions imposed on the variance-covariance of the model (such us the presence and time-invariance of spatial correlation). It is argued that the theoretically consistent model is one where the parameters vary over time following a defined stochastic process estimated using the Kalman filter as this combination results in minimum revisions to the index while incorporating information of transactions in previous periods. Using data from a town in the state of Queensland, Australia, preliminary results are presented using two alternative hedonic imputed indexes and two estimators that do not lead to revisions in the previous values of the index es as new sales are observed.

<sup>\*</sup>The authors acknowledge funding from the Australian Research Council (DP120102124). The data from Fletcher et al. (2011) which were collected, cleaned and checked with funding from the Department of Climate Change and Energy Efficiency, CSIRO Climate Adaptation Flagship, and the 2011/12 UQ Summer Scholarship Program funded by the School of Economics and The University of Queensland. Many thanks to Ryan McAllister for insightful discussion and data contributions, and Thi Thuy (Kelly) Dung Trinh and Kian Nam Loke for their work on the dataset.

## 1 Introduction

Hedonic imputed indexes, considered superior to the traditional time-dummy index (Silver and Heravi (2007), de Haan and Diewert (2011)), are computed using predictions from hedonic models. Hedonic regressions have a theoretical foundation along the lines discussed in Diewert (2003) and Hill and Melser (2008) which imply appropriately derived econometric models should have hedonic coefficients that evolve over time. Recent work by Rambaldi and Rao (2011) has studied the predictive power of a rolling-window (of two periods) approach compared to that of a time-varying hedonic model where the evolution of the parameters obeys a flexible but specified stochastic process and found there are substantial gains in prediction ability in using the latter. The estimator of the hedonic parameters used by Rambaldi and Rao (2011) was a Kalman Smoother (KS). One drawback of this estimator is that the time series of estimates of the hedonic parameters changes when a new time period is added to the sample and as a consequence the resulting set of imputed prices also changes. The revised imputed prices would then result in the need for continuous revisions of the housing price index computed from these estimates. A second draw back is in relation to the conditioning set used to construct the prediction. This issue was first raised by Pace et al. (2000) in the context of a pooled GLS estimator of the hedonic regression. The conditioning set used to construct a price prediction in a pooled model is the complete sample without any information on the ordering of the transactions. Although the KS uses the ordering information, it still results in estimates that lead to in-sample predictions (except for those for the last time period in the sample) based on both past and *future transactions*. Pace et al. (2000) proposed to solve this problem through the use of a GLS estimator with a strictly lower triangular specification of the covariance which ensures the information ordering is preserved. However, this is still a fixed parameter estimator of the shadow price parameters of the hedonic function. A comparison of the Pace et al. spatial-temporal estimator to a Kalman filter (KF) estimator of a time-varying parameter model was studied by Svetchnikova et al. (2008) who found the time-varying framework provided superior prediction ability. The KF is an estimator of the conditional (on information up to the current period) mean of the state vector which in this case is defined as a subset of the unknown parameters of the hedonic model. The main objective of the paper is to study the effects that choices made in modeling, estimation and prediction of the imputed price, have on the computed price index. To this end, the paper first presents a general representation of the hedonic model that serves as a nesting framework and allows a discussion of alternative estimators and predictors. The aim is to identify the possible sources of revisions of the index that arise from the choice of model, estimators of

the parameters and predictors of the transaction price. The sources identified include: the conditioning set; the stability of shadow prices parameters; and parameters of the variance-covariance structure of the model. The loss of precision in the imputation process and the effect on the computed hedonic imputed price indexes is studied under a number of scenarios. These include the polar cases of continuous revision vs no-revision as well intermediate and alternative scenarios. To illustrate the study uses data from a town in the state of Queensland, Australia.

The paper is organized as follows. Section 2 briefly presents the form of the price indexes used in this study in order to introduce the role of the prediction of the property price in the index. Section 3 presents a general form of the hedonic model as well as several special cases which are commonly used in the literature. The methods of estimation of the parameters of each model and the resulting form of the predictor are discussed in each case. Through this discussion the possible reasons that would lead to index revision are identified in each case. Section 4 is used to illustrate with actual data and Section 5 concludes.

## 2 Hedonic Imputed Residential Property Price Indexes Used in This Study

Hill and Melser (2008) and Rambaldi and Rao (2011) discuss of a range of index number formulae that are based on different sets of weighting systems and on different sets of imputed prices. In this paper the general recommendations from these works are taken and two Tornqvist indexes are used, a democratic and a plutocratic type. These are weighted indexes. When weights represent the relative value of each of the houses included in the sample, this is a plutocratic index, while the use of the number of transactions in each time period produces a democratic index. The plutocratic index is computed using *actual shares* based on actual prices as defined in (1), and *imputed (or predicted)* prices in both the base and current periods. Let  $P_t^h$  represent the sale price of house h in period t. Further, let  $w_t^h$  be the value share of the house h defined as:

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

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.

The plutocratic Tornqvist index is defined as follows:

$$T_{s,t}^P = \sqrt{GL_{s,t}^p \times GP_{s,t}^p} \tag{2}$$

where  $GL_{s,t}^p$  and  $GP_{s,t}^p$  are the plutocratic geometric Laspyeres and geometric Paasche indexes which are defined as:

$$GL_{s,t}^{P} = \prod_{h=1}^{N_{s}} \left[ \frac{\hat{P}_{t}^{h}(x_{s}^{h})}{\hat{P}_{s}^{h}(x_{s}^{h})} \right]^{w_{s}^{h}}$$
(3)

$$GP_{s,t}^{P} = \prod_{h=1}^{N_{t}} \left[ \frac{\hat{P}_{t}^{h}(x_{t}^{h})}{\hat{P}_{s}^{h}(x_{t}^{h})} \right]^{w_{t}^{h}}$$
(4)

where,

 $\hat{P}_i^h(x_j^h)$  for i, j = s, t is a prediction or imputation of the price of house h with vector of hedonic characteristics  $x_j^h$  at periods s, t.

These indexes are "plutocratic" and are influenced by houses with large price tags. Despite this, it measures the changes in the housing stock values that can be attributed exclusively to price changes, and therefore provides useful information (see Hill and Melser (2008) for extensive discussion of these and other types of indexes for residential housing).

The Tornqvist democratic index consistent with the use of a log-price hedonic model is defined as:

$$T_{s,t}^{D} = \sqrt{\left[\prod_{h=1}^{N_{s}} \left(\left[\frac{\hat{P}_{t}^{h}(x_{s}^{h})}{\hat{P}_{s}^{h}(x_{s}^{h})}\right]^{\frac{1}{N_{s}}}\right)\right] \left[\prod_{h=1}^{N_{t}} \left(\left[\frac{\hat{P}_{t}^{h}(x_{t}^{h})}{\hat{P}_{s}^{h}(x_{t}^{h})}\right]^{\frac{1}{N_{t}}}\right)\right]}$$
(5)

$$= \sqrt{GL_{s,t}^D GP_{s,t}^D} \tag{6}$$

where  $N_t$  and  $N_s$  are respectively the number of houses sold in periods s and t.

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<sup>1</sup>. When the geometric Tornqvist index

 $<sup>^{1}</sup>$ Analogously if the prices are normally distributed then a simple arithmetic mean would be used in the place of a geometric mean.

is computed, we explicitly recognize the unequal numbers of houses sold in the two periods and define a geometric mean of the geometric Laspyeres and Paasche indexes (see equation 5 and 6).<sup>2</sup> Rambaldi and Rao (2011) argue that 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 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 driven by the main objective behind the housing price index construction.

Whichever form the imputed index takes, the commonality amongst these indexes is that they depend on a prediction,  $\hat{P}_i^h(x_j^h)$ . Silver and Heravi (2007) and Hill and Melser (2008) discuss the importance of computing this prediction using estimates of the parameters which vary over time and regions (see equation (9) and related discussion in Hill and Melser (2008)). Section 3 discusses alternative modeling frameworks available to construct the prediction (imputation) required to construct these indexes.

## 3 Hedonic Models of Property Prices

For the purpose of discussion it will be convenient to set a general framework that can accommodate a number of alternative models and estimators that are candidates for computing  $\hat{P}_i^h(x_j^h)$ . The general model is that proposed by Rambaldi and Rao (2011) where all the parameters are allowed to vary following a given stochastic process and errors are assumed to be spatially correlated to account for omitted hedonic characteristics that are likely to be captured by location.

$$y_t = \mu_t + X_t \beta_t + \epsilon_t \qquad \epsilon_t \sim NID(0, H_t) \tag{7}$$

$$\epsilon_t = \rho W_t \epsilon_t + \sigma_u u_t \qquad u_t \sim NID\left(0, \sigma_u^2 I_{N_t}\right) \tag{8}$$

$$\mu_t = \mu_{t-1} + \sigma_\mu \xi_t \qquad \xi_t \sim NID(0, \sigma_\mu^2) \tag{9}$$

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

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

 $<sup>^{2}</sup>$ It is possible to consider a more sophisticated approach after stratifying the sample into different regions and by the type of houses.

where,

 $y_t - N_t \times 1$  vector of observations of the dependent variable, typically the log of sale price  $(P_t)$ ,  $y_t = \ln P_t$ for the  $N_t$  transactions observed in period t;

 $\mu_t$  – local level which captures overall macroeconomic trends

 $\beta_t{-}K\times 1$  vector of unknown slope (shadow prices) parameters;

 $X_t - N_t \times K$  matrix of independent variables, house and land attributes, which will typically include measures of spatial characteristics (such as distances to transport, schools, etc);

 $u_t - N_t \times 1$  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);

 $\epsilon_t$ - $N_t \times 1$  vector of correlated errors with covariance  $H_t$ ;

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

The spatial weights matrix, W, has the following characteristics

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

#### That is,

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

We note here that the parameter  $\rho$  is assumed to be constant over time and the form of  $H_t$  can easily be shown to be

$$H_t = \sigma_u^2 \left( I_{N_t} - \rho W_t \right)^{-1} \left( I_{N_t} - \rho W_t \right)^{-1}$$
(11)

since  $\epsilon_t \sim N(0, H_t)$ , and  $\epsilon_t = (I_{N_t} - \rho W_t)^{-1} u_t$  using (8).

To aid the discussion the system in (7)-(9) is written in the following form (12) and (13)

$$y_t = Z_t \alpha_t + \epsilon_t \tag{12}$$

$$\alpha_t = \alpha_{t-1} + \eta_t \tag{13}$$

where,  

$$Z_{t} = \begin{bmatrix} \mathbf{1} & X_{t} \end{bmatrix}$$

$$\alpha_{t} = \begin{bmatrix} \mu_{t} \\ \beta_{t} \end{bmatrix}$$

$$\eta_{t} = \begin{bmatrix} \xi_{t} \\ \zeta_{t} \end{bmatrix} \text{ with } Q_{t} = E(\eta_{t}\eta'_{t}), Q_{t} \sim N(0, \begin{bmatrix} \sigma_{\mu}^{2} & 0 \\ 0 & \sigma_{\beta}^{2}I_{K} \end{bmatrix}).$$

where, **1** is an  $N_t \times 1$  vector of ones. Several models are special cases of this general framework.

## 3.1 Special Cases of the General Model

## 3.1.1 Models with determinist time-varying intercept and fixed slopes

The conventional time dummy model (TD) is given by setting  $\sigma_{\mu}^2 = \sigma_{\beta}^2 = 0$  in (8) and (10) and replacing the local level (9) with a fixed time effect,  $\sum_{t=1}^{T} \delta_t D_t$ , where  $D_t$  is a time dummy with value equal to one at time period t and zero otherwise and  $\delta_t$  is the intercept parameter for period t. This model would typically be used to estimate the hedonic model with annual data. If the errors are assumed spatially correlated, the model is,

$$y_t = \delta_t + X_t \beta + \epsilon_t \qquad \epsilon_t \sim NID(0, H_t) \tag{14}$$

$$\epsilon_t = \rho W_t \epsilon_t + u_t \qquad u_t \sim NID\left(0, \sigma_u^2 I_{N_t}\right) \tag{15}$$

$$\beta_t = \beta_{t-1} \tag{16}$$

It is easily seen that if there is no spatial correlation,  $\rho = 0$  in (15), the error term  $\epsilon_t = u_t$  and  $H_t = \sigma_u^2 I_N$ . However, once the errors are assumed to follow (15), two alternatives of the specification of (14)-(16) can be found in the literature. Their difference resides in how the spatio-temporal dimensions of transactions are accounted for.

To show these it is convenient to write the model for the  $N = \sum_{t=1}^{T} N_t$  observations in the sample

$$y = \sum_{t=1}^{T} \delta_t D_t + X\beta + \varepsilon \tag{17}$$

$$y = Z\alpha + \varepsilon \tag{18}$$

$$\varepsilon = \rho W \varepsilon + u \tag{19}$$

In this case the definitions of the matrices are,  $\ensuremath{\case$ 

$$Z = \begin{bmatrix} D_1 & 0 & 0 & 0 & X_1 \\ 0 & D_2 & 0 & 0 & X_2 \\ 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & D_T & X_T \end{bmatrix}_{N \times (T+K)}$$
$$\alpha = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_T \\ \beta \end{bmatrix}$$
$$\eta_t = 0 \text{ with } Q_t = 0$$

Only space relationship among transactions accounted for The random error structure is  $\epsilon = (I_N - \rho W)^{-1} u$ , and W is defined as above; however, the t dimension is not taken into account. The matrix is defined over the N transactions in the sample and does not take into account the timing of the transaction, only the distance between neighbors. The resulting variance-covariance is of the form

$$H = \sigma_u^2 \left( I_N - \rho W \right)^{-1} \left( I_N - \rho W \right)^{-1'}$$
(20)

Space-Time relationships among transactions accounted for Pace et al. (2000)'s contribution is to propose an alternative form of the W matrix in the model (17), such that the variance-covariance takes into account the spatio-temporal relationship of the transactions effectively acting as a space-time filter. In their framework the space-time ordering enters H through W. The following definitions are used

$$\hat{H} = \sigma_u^2 \left( I_N - \rho \hat{W} \right)^{-1} \left( I_N - \rho \hat{W} \right)^{-1\prime}$$
(21)

where,

 $\hat{W} = \hat{\varphi}_S S + \hat{\varphi}_{\Upsilon} \Upsilon + \hat{\varphi}_{ST} S \Upsilon + \hat{\varphi}_{\Upsilon S} \Upsilon S.$ 

The  $\Upsilon$  and S are lower triangular matrices with zeros in the diagonals that specify the temporal  $(\Upsilon)$  and

spatial (S) relations among previous observations and the  $\varphi's$  are parameters that need to be estimated. The resulting  $\hat{W}$  is also lower triangular with zeros in the main diagonal. While this specification addresses the conditioning set problem by imposing a space-time ordering, it still assumes shadow price parameters are time invariant.

#### 3.1.2 Models with non-systematic time-varying intercept and slopes

Given the currently available data on multiple transactions, it is usually the case that even if the t index denotes a month,  $N_t$  will be sufficiently large to run a regression (or a SEM model) leading to a time series of estimates of both the intercept and the slope parameters. This is the approach advocated in Hill and Melser (2008).

To compare to the previous alternatives, note that in the framework (7)-(9), this approach can be represented by re-writing equations (10) and (9) as follows:

$$egin{array}{rcl} eta_t & 
eq & eta_{t-1} \ \mu_t & 
eq & \mu_{t-1} \end{array}$$

and the representations in (12) and (13) become

$$y_t = Z_t \alpha_t + \epsilon_t \tag{22}$$

$$\alpha_t \neq \alpha_{t-1} \tag{23}$$

This is the polar case from a pooled model and it indirectly implies  $\rho_t \neq \rho_{t-1}$  and  $\sigma_{ut}^2 \neq \sigma_{u,t-1}^2$ . Here the parameters of the hedonic model are not fixed over time like in (14), (15), (16); however, no assumption is made on the memory of the shadow price parameters regarding previous values. This approach effectively ignores the correlation across time of the consumer's marginal willingness to pay for a particular characteristic (i.e. extra bathroom, a swimming pool, closeness to a park or public transport). Both from a theoretical and empirical perspective, it is difficult to justify ignoring this correlation completely.

The consequences of these alternative modeling choices for estimation and prediction are presented in the next subsection.

## 3.2 Estimators, Predictors and Price Indexes of Property Prices

In this section the estimation and construction of a prediction of the price are discussed within the context of the above modeling framework.

#### 3.2.1 Estimators and Price Indexes of Models with Fixed Slope Parameters

When the shadow price parameters are assumed fixed, (18)-(19), the estimation of  $\alpha$  will be by least squares if there is no spatial error. When the error structure is assumed to have spatial correlation, the maximum likelihood estimator is an "estimated" (or feasible) generalized least squares with form:

$$\hat{\alpha}_{MLE} = (Z'\tilde{H}^{-1}Z)^{-1}Z'\tilde{H}^{-1}y$$

where the definition of  $\tilde{H}$  will vary depending on whether only spatial dimension (20) or both spatiotemporal relationships  $(21)^3$  are taken into account.

The conventional price index from this model is not a hedonic imputed index, it is the time-dummy hedonic index (TDH) given by

$$Index_t = exp(\hat{\delta}_t) \qquad t = 1, .., T$$

where,

 $\hat{\delta}_t$  is the estimate of the intercept for time period t

A hedonic imputed price index (HI) can also be computed by computing a prediction of the price (24) for i, j = s, t and using the indexes in (2) and (5):

$$\hat{P}_{i}^{h}(x_{j}^{h}) = \exp\left[\left(\begin{array}{cc} 1 & x_{j}^{h} \end{array}\right)\hat{\alpha}_{MLE}\right]$$

$$(24)$$

$$= \exp\left[ \left( \begin{array}{cc} 1 & x_j^h \end{array} \right) \left( \begin{array}{c} \hat{\delta}_i \\ \hat{\beta} \end{array} \right) \right]$$
(25)

where,

i, j = s, t in (2) and (5)

 $<sup>^{3}</sup>$ Pace et al. (2000) develop a computationally efficient maximum likelihood routine. The reader is referred to their paper for details

 $x_j^h$  is the vector of characteristics for property h in time period j, evaluated at the parameter values of time period i.

In this case both the TDH and HI will be subjected to revisions once new transactions are available. This is well recognized in the literature and thus it is one of the reasons why the price index literature is advocating the use of time-varying parameters. These are considered in the next section.

#### 3.2.2 Estimators and Price Indexes in Models with Time-Varying Parameters.

The polar case to (17) is the approach presented in Section 3.1.2. In this case the model is assumed to have time-varying parameters; however, it is also assumed that shadow price parameters at time t and time t + 1 are not statistically related. In this case the estimator is a GLS (in (26)) and the definition of  $H_t$  could vary from an identity (in which case the estimator is OLS) or a spatial covariance of the form in (11) where neighboring transactions for period t are used to construct the  $W_t$  matrix.

$$\hat{\alpha}_t = (Z'_t H_t^{-1} Z_t)^{-1} Z'_t H_t^{-1} y_t \tag{26}$$

with variance-covariance,

$$\Omega_t^{GLS} = (Z_t' H_t^{-1} Z_t)^{-1}$$

As the estimates are computed for each t = 1, ..., T, there is no issue with revisions of the index as past transactions do not enter the estimation. While this is appealing from the point of view of index revision, there is a likelihood that the parameter estimates could be unduly influenced by the seasonality of sales (if this is estimated monthly or even quarterly) as well as other influential observations such as sales of very expensive property in a particular time period.

The predictions required to compute the HI index are of the form

$$\hat{P}_{i}^{h}(x_{j}^{h}) = \exp\left[\left(\begin{array}{cc} 1 & x_{j}^{h} \end{array}\right)\hat{\alpha}_{i}\right]$$

$$\left[\left(\begin{array}{cc} 1 & x_{j}^{h} \end{array}\right)\left(\begin{array}{c} \hat{\delta}_{i} \end{array}\right)\right]$$

$$(27)$$

$$= \exp\left[ \left( \begin{array}{cc} 1 & x_j^h \end{array} \right) \left( \begin{array}{c} \delta_i \\ \hat{\beta}_i \end{array} \right) \right]$$
(28)

for i, j = s, t in (2) and (5)

An alternative to computing the model for each time period is to use a rolling overlapping window. This has been advocated by Triplett (2004) and the prediction performance studied by Rambaldi and Rao (2011); however, it will not be considered in this study.

An alternative time-varying parameter structure to this polar case is given by estimating the general framework in (7) - (9) using an estimator based on the appropriate conditioning set for the purpose, i.e. to obtain a prediction. The proposed estimator is the Kalman filter. The main advantages of this approach are that estimates of shadow prices incorporate past transactions (in a consistent space-time ordering) and under some assumptions (detailed shortly) will not result in revisions of past estimates of  $\alpha_t$ .

To show how this estimator works it will be convenient to incorporate equation (8) into (7)

$$y_{t} - \begin{bmatrix} i & X_{t} \end{bmatrix} \alpha_{t} = \rho W_{t} \left( y_{t} - \begin{bmatrix} \mathbf{1} & X_{t} \end{bmatrix} \begin{bmatrix} \mu_{t} \\ \beta_{t} \end{bmatrix} \right) + u_{t}$$
$$(I_{N_{t}} - \rho W_{t}) y_{t} = (I_{N_{t}} - \rho W_{t}) \begin{bmatrix} \mathbf{1} & X_{t} \end{bmatrix} \begin{bmatrix} \mu_{t} \\ \beta_{t} \end{bmatrix} + u_{t}$$
$$\tilde{y}_{t} = \tilde{Z}_{t} \alpha_{t} + u_{t}$$
(29)

where,

$$\begin{aligned} \widetilde{y}_t &= (I_{N_t} - \rho W_t) y_t, \\ \widetilde{Z}_t &= (I_{N_t} - \rho W_t) Z_t, Z_t = \begin{bmatrix} \mathbf{1} & X_t \end{bmatrix} \\ u_t &\sim NID \left( 0, \sigma_u^2 I_{N_t} \right) \\ H^u &= \sigma_u^2 I_{N_t} \end{aligned}$$

and  $\alpha_t$  follows the process in (13) which is repeated here for convenience,

$$\alpha_t = \alpha_{t-1} + \eta_t \tag{30}$$

where,  

$$\begin{aligned} \alpha_t &= \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} \\ \eta_t &= \begin{bmatrix} \xi_t \\ \zeta_t \end{bmatrix} \text{ with } Q = E(\eta_t \eta'_t), \ \eta_t \sim N(0, \begin{bmatrix} \sigma_{\mu}^2 & 0 \\ 0 & \sigma_{\beta}^2 I_K \end{bmatrix}) \end{aligned}$$

In this specification, the Kalman filter estimate of  $\alpha_t$ , denoted  $a_t$ , is given by<sup>4</sup>

$$a_t = a_{t-1} + G_t \nu_t, \quad \text{for } t = 1, ..., T$$
 (31)

with covariance matrix,  $\Omega_t$ , given by

$$\Omega_t = \Omega_{t|t-1} - M_t F_t^{-1} M_t', \quad \text{for } t = 1, ..., T$$
(32)

where,

 $u_t = \tilde{y}_t - \tilde{Z}_t a_{t-1}$  is the prediction error using the parameter estimates at t - 1,  $\nu_t \sim (0, F_t)$   $F_t = \tilde{Z}_t \Omega_{t|t-1} \tilde{Z}'_t + H^u$  is the variance-covariance of the prediction error,  $\nu_t$ .  $G_t = M_t F_t^{-1}$  is known as the Kalman gain. Captures the information gain from t - 1 to t.  $M_t = \Omega_{t|t-1} \tilde{Z}_t$   $\Omega_{t|t-1} = \Omega_{t-1} + Q$  $\alpha_0 \sim N(a_0, \Omega_0)$  is the initial distribution

The estimate of the macro-level trend  $(\mu_t)$  and shadow prices  $(\beta_t)$  will change from t - 1 to t if the prediction error at time t is not zero, and the change will be proportional to the information gained,  $G_t$ , between t - 1 and t.

It is clear from equation (31) that this estimator has two desirable characteristics. The first is that the estimates of the macro-level trend and shadow price parameters are functions of past information on sale prices, hedonic characteristics and the spatial distribution. The second is that the addition of a new time period does not result in a revision of the previous periods macro-level trend, { $\mu_{t-1}, \mu_{t-2}, ...$ }, or shadow price parameters { $\beta_{t-1}, \beta_{t-2}, ...$ }, unless estimates of the parameters  $\psi = {\rho, \sigma_u^2, \sigma_\mu^2, \sigma_\beta^2}$  change. The parameters  $\psi$  define  $H^u$ ,  $Q_t$ ,  $\tilde{y}_t$  and  $\tilde{Z}_t$  and therefore are required for the Kalman filtering algorithm in (31) and (32) to be feasible and provide the estimate of  $\alpha_t$  and its Mean Squared Error Matrix ( $\Omega_t$ ). The estimation of these parameters can be achieved using maximum likelihood for a sample defined by N, the transactions over T time periods, and defining the likelihood function as

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

which will provide

$$\hat{\psi} = \operatorname{argmax}_{\psi} \ln L(y_t | \psi)$$

<sup>&</sup>lt;sup>4</sup>For derivations and more details See Harvey (1989) or Durbin and Koopman (2001)

leading to

$$\hat{a}_t = \hat{a}_{t-1} + \hat{G}_t \hat{\nu}_t \tag{33}$$

with covariance matrix,  $\hat{\Omega}_t$ , given by

$$\hat{\Omega}_t = \hat{\Omega}_{t|t-1} - \hat{M}_t \hat{F}_t^{-1} \hat{M}_t' \tag{34}$$

where, the  $\hat{i}$  indicates that  $\hat{\psi}$  has been used instead of  $\psi$ .

Although theoretically  $\psi$  does not change with time or space dimensions, the estimates will change depending on the sample N used. The empirical question is how often should  $\psi$  be re-estimated.

Using these estimates the prediction for the purpose of imputation and construction of the price index is given by

$$= \exp\left[ \left( \begin{array}{cc} 1 & x_j^h \end{array} \right) \left( \begin{array}{c} \mu_i \\ \hat{\beta}_i \end{array} \right) \right]$$
(36)

for i, j = s, t in (2) and (5)

Before moving to the next section, the Kalman smoother estimator is presented to show that this estimator leads to revisions of the estimated  $\alpha_t$  and  $\Omega_t$ , even if  $\psi$  were to be fixed and known. Further, as the purpose of the Kalman smoother is to produce estimates that are conditioned on T, this results in the undesirable property that estimates are functions of future transactions.

The fixed-interval smoothing used by (Rambaldi and Rao (2011)) is used here to demonstrate how the estimates are computed

$$\hat{a}_{t|T} = \hat{a}_t + \Omega_t^* (\hat{a}_{t+1|T} - \hat{a}_t), \quad \text{for } t = T - 1, ..., 1$$
(37)

and

$$\hat{\Omega}_{t|T} = \hat{\Omega}_t + \Omega_t^* (\hat{\Omega}_{t+1|T} - \hat{\Omega}_{t+1|t}) \Omega_t^{*'}, \quad \text{for } t = T - 1, ..., 1$$
(38)

where,

$$\hat{a}_{T|T} = \hat{a}_{T}$$
$$\hat{\Omega}_{T|T} = \hat{\Omega}_{T}$$
$$\Omega_{t}^{*} = \hat{\Omega}_{t}\hat{\Omega}_{t+1|t}^{-1}$$

The purpose of the smoother is to obtain an estimate of  $E(\alpha_t|y_1, ..., y_T, Z_1, ..., Z_T)$ , that is, all estimates are conditional on the complete sample over T time periods. These equations show that to obtain these estimates, the estimates obtained using the Kalman filter, (31) - (32) or (33) -34), are now revised to be made conditional on the complete sample, making clear that smoothed estimates of the macro-level trend and shadow price parameters are a function of all observed transactions over the period t = 1, ..., T. A consequence of this is that even if  $\psi$  were to remain fixed, a new set of transactions in T + 1 will lead to a complete revision of the predictions/imputations and therefore the index.

The predictions using this estimator will be given by

$$\hat{P}_{i}^{h}(x_{j}^{h}) = \exp\left[\left(\begin{array}{cc} 1 & x_{j}^{h} \end{array}\right) \hat{a}_{i|T}\right]$$

$$\left[\left(\begin{array}{cc} & & \\ \end{array}\right) \left(\begin{array}{c} \hat{\mu}_{i|T} \end{array}\right)\right]$$

$$(39)$$

$$= \exp\left[ \left( \begin{array}{cc} 1 & x_j^h \end{array} \right) \left( \begin{array}{c} \mu_{i|T} \\ \hat{\beta}_{i|T} \end{array} \right) \right]$$
(40)

for i, j = s, t in (2) and (5)

Using (27) as the benchmark, some preliminary results are presented in the next section where several scenarios are studied with (35) to evaluate the effects on the estimated  $\alpha_t$  resulting from changing the sample period of estimation of  $\psi$  as well as the consequences of using spatial correlation in the errors when regressors identifying location features of the propety are not available.

## 4 Empirical Illustration

To illustrate the concepts, this section presents *preliminary* empirical results using a data set from a town in the state of Queensland, Australia. The comparison uses the two estimators presented in Section 3.2.2, (26) and (33), as they share the desirable properties of not leading to revisions of previous values of the index when a new time period is observed and allow the shadow price parameters to vary over time.

Two versions of the hedonic model are used, the first assumes errors are not to be spatially correlated  $(\rho = 0)$ ; however, a number of regressors are including in the model measuring locational characteristics.

Further discussion is presented in Section 4.2.

## 4.1 Data

The data used were compiled, cleaned and checked with funding from the Department of Climate Change and Energy Efficiency, CSIRO Climate Adaptation Flagship, and were used to produce some of the results presented in Fletcher et al. (2011). These data are sourced primarily from one of Australia's leading providers of real estate sales transaction data (RPData) and a number of spatial hedonic characteristics, such as distances to landmarks (more details in Table 1), were added through GIS analysis. Further cleaning were performed as part of the 2011/12 UQ Summer Scholarship Program funded by the School of Economics and The University of Queensland. The dataset consists of individual transactions of family dwelling residential property (i.e. units, townhouses and terraces are not included) for the period May 1991 to September 2010. Sales of vacant land as well as land and structure are included. The number of transactions per month is presented in Figure 1.

## [Figure 1 here]

The area in this study is in the south east corner of the state of Queensland (SEQ) and it is a coastal area. The market in the SEQ area went through a boom period between 2001 and 2005 and this is reflected in the number of transactions per month for that period. The market was more stable between 2005 and 2008. The 2008 GFC is noticeable in that the number of transactions drops to levels similar to the early 1990s and has remained volatile since.

The hedonic characteristics used in this study with descriptive statistics are presented in Table 1.

## [Table 1 here]

## 4.2 Estimation and Comparison of Predictions

In all cases the models are estimated as log-linear in all regressors except for lot size which is also logtransformed. Monthly predictions are produced in all cases. When errors are assumed to be spatially correlated (Section 4.2.2), all regressors measuring distances to landmarks are dropped from the model. This is done to study the options of using a smaller model (dropping ten regressors) and using the spatial error structure to capture the location features that are now not explicitly controlled for in the model.

#### 4.2.1 Models without Spatially Correlated Errors

When errors are assumed to be non-spatially correlated, the covariance matrix  $H_t$ , (20) reduces to  $\sigma_u^2 I$ . Therefore estimator, (26), reduces to least squares. The Kalman filter estimator in this case is given by

$$a_t = a_{t-1} + G_t \nu_t, \quad \text{for } t = 1, ..., T$$
 (41)

where,

$$\begin{split} \nu_t &= \tilde{y}_t - \tilde{Z}_t a_{t-1} , \ \nu_t \sim (0, F_t) \\ G_t &= M_t F_t^{-1} \\ F_t &= \tilde{Z}_t \Omega_{t|t-1} \tilde{Z}_t' + \sigma_u^2 I \\ M_t &= \Omega_{t|t-1} \tilde{Z}_t \\ \Omega_{t|t-1} &= \Omega_{t-1} + Q \\ Q &= \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 I_K \end{bmatrix} \\ \alpha_0 &= 0_{K+1} , \ \Omega_0 &= 10^6 I_{K+1} \text{ is the initial distribution} \end{split}$$

For the OLS estimates two consecutive months of data are used to estimate  $\alpha_t$ , t - 1 and t. That is, for a given month the data of the current and previous months are used. The reason for this approach is that for some months the number of observations was not large enough for the model to produce reliable estimates. Even using this approach there are some instances when estimates could not be obtained.

To obtain the Kalman filter estimates of  $\alpha_t$ , estimates of  $\sigma_u^2$ ,  $\sigma_\mu^2$  and  $\sigma_\beta^2$  are need as they define the matrices Q and H. The sample from 1991:5 - 1999:12 (first 104 months of data) was used initially to produce these estimates and these were then used to estimate  $\alpha_t$  and the predictions for the complete sample. The estimates are  $\hat{\sigma}_u^2 = 0.001$ ,  $\hat{\sigma}_\mu^2 = 1$  and  $\hat{\sigma}_\beta^2 = 0.5$ . The model was re-estimated using the sample 1991:5-2005:12 and 1991:5-2010:9 (full sample) and the estimates of these three parameters did not change. This is an important result as these parameters are the only possible source of revisions for Kalman filter estimates of  $\alpha_t$ .

In order to evaluate and compare the performance of these two approaches the two sets of estimates are used to produce predictions of the price for each house in the sample,  $\hat{P}_t^h(x_t^h)$ , using prediction equations (27) and (35). For each month the following prediction error measure is computed using the least squares and Kalman filter estimates respectively,

$$\%Deviation_t = \left(\frac{1}{N_t} \sum_{i=1}^{N_t} \frac{|\hat{P}_t^h(x_t^h) - P_t^h(x_t^h)|}{P_t^h(x_t^h)}\right) \times 100$$
(42)

Figure 2 presents the results for the sample. For ease of visual presentation the sample is presented in four panels, 1991:6-1995:12, 1996:1-1999:12, 2000:1-2004:12, and 2005:1-2010:9. As a further summary measure the average prediction error for each of these sub-periods are presented in Table 2. Unlike the OLS estimates, the Kalman filter estimates are functions of the previous period estimates. This is expected to diminish the effect of individual outlier observations especially once the filter has run for a few months. This is supported by the results where the %*Deviation* measure shows a consistently superior prediction performance by the Kalman filter based estimates.

#### [Figure 2 and Table 2 here]

#### 4.2.2 Models with Spatially Correlated Errors

The use of a spatially correlated error is justified on the basis that it would capture omitted hedonic information. This is potentially important for the computation of price indexes in a larger scale where the index must be computed for many regions. It might be desirable to use models with a minimum number of hedonic characteristics and use the spatial error structure to capture the location characteristics as this only requires information on the latitude and longitude of each property, and this is part of the information routinely provided by RPData. To test this possibility a much smaller model with a reduced number of regressors is used. The model does not include any of the distance measures, all other regressors in Table 1 remain in the model. The estimator when the model is assumed to have non-systematic time-varying parameters is (26) and the Matlab *seem.m* routine developed by Pace and Lesage is used to obtain the estimates. The Kalman filter with the matrix  $H_t$  defined as in (11) is used to obtain the estimates (33). The spatial error adds one more parameter to the estimation,  $\rho$ . The estimate using the state-space model is 0.5. The estimate of the spatial parameter when using (26) with data from months t - 1 and t as the previous case changes across time periods with both positive and negative estimates of  $\rho$  observed.

Figure 3 and Table 3 present the computed measure of prediction error (42). Two interesting points are noteworthy. First, the results are similar to those from the non-spatial model in that the Kalman filter based estimates produce superior predictions. Second, comparing the prediction error between Tables 2 and 3, it would appear that the trade-off of several regressors for the spatial error structure is suitable as the decrease in prediction ability is not large.

[Figure 3 and Table 3 here]

#### 4.3 Computed Price Indexes and Discussion of Results

Figure 4 presents computed HI using the estimates from the non-spatial model (as reported in 4.2.1). Four indexes are reported,  $TP\_SS$  and  $TP\_OLS$  are the indexes computed using (2) and the estimates from the Kalman filter and the least squares, respectively, and,  $TD\_SS$  and  $TD\_OLS$  are the indexes computed using (5), respectively. The presentation has been divided into the four sub-samples previous used (1991:6-1995:12, 1996:1-1999:12, 2000:1-2004:12, and 2005:1-2010:9). As reported previously the OLS estimator could not be computed in a few cases, the first is 2008:9. As a consequence the indexes based on OLS are reported until that date for these preliminary results (these will be updated once the source of the problem has been located).

## [Figure 4 here]

The plot for the period 1996:1-1999:12 show divergent trends between the indexes based on OLS and Kalman filter estimates. The " SS" indexes indicate prices remained largely unchanged over this period (using base 1996:1=1, 1999:12 they are 0.999) while the "\_OLS" indicate a downward trend (using base 1996:1=1, 1999:12, TP is 0.785 and TD is 0.736). This is a period where the number of observed transactions per month is smallest (see Figure 1). Although two consecutive months are merged to obtain the estimates by OLS, this would appear not to be sufficient to mitigate outlier obsevarions and it the likely reason why the predictions of prices differ the most between the two approaches, and thus the result is that the indexes also deviate from each other. Another noticeable period when they deviate is the period 2000:1-2004:12 (based 2000:1=1). Around March 2003, "TD\_OLS" indicates the prices are not growing as fast as the other indexes indicate, and "TP OLS" shows a slow down by July 2003. All indexes are at similar value by November 2003 but start to show a different trend in 2004, where the "SS" indexes indicate prices are stable, while "\_OLS" indexes show a decreasing trend. The overall consequence of these divergent patterns can be seen if the indexes are based on June 1991, Figure 5, where the " OLS" show an overall increase in prices of two and half times by September 2008, while the " SS" based indexes indicate the price increase has been around four times. A median price index is also shown as reference and although much more volatily it would indicate the "\_SS" indexes are on the correct path. It is also noticeable that the TP\_SS and TD\_SS indexes deviate from each other from around June 2003, and although the trend remains similar, the consequence is that the actual index values differ in this period in

the order of 8-10%. All indexes (except for the median) show the effect of the GFC clearly as they show a sudden drop in March 2008 from their highest value in the first two months of 2008.

### [Figure 5 here]

These results highlight some of the differences that can arise from alternative econometric estimation of the same model as well as the weighting that is used to compute the index. In addition, the median index highlights some of the well known issues related to the use of an index that does not control for the composition of sales.

## 5 Conclusions

This paper extends the current literature which has advocated the use of hedonic models with timevarying parameters to compute hedonic price indexes. However, it shows that the use of models that do not impose some systematic memory in the evolution of the parameters can lead to estimates being dominated by influential observations which in turn lead to potentially highly misleading property price indexes. Although the use of a state-space formulation estimated with a Kalman smoother had been previously proposed for this purpose, the drawback is that the Kalman smoother produces a revision of the previous estimates of the parameters (and therefore index) once a new time period of data is added. Thus, this paper uses the Kalman filter to estimate the state-space formulation of the hedonic model. In this case there are only a small number of time-invariant covariance parameters that needed to be estimated with a reasonable amount of data; however, it is shown that the estimates of these parameters using the first nine years of data do not change from those obtained using fifteen or twenty years of data. Thus, once these parameters are estimated, the estimates of the macro-trend (intercept) and shadow prices are obtained using the Kalman filter and thus there is no revision in their values as new time periods are added.

The empirical results are preliminary but show a substantial bias in the computed index when the model is estimated using only the current and immediately previous month of data. This bias arises because the shadow prices parameters are likely to be influenced by instances when there are sales that are unusual (such as a very expensive property or a property sold at a much lower value than would be expected). Shadow prices are likely to be slow moving and thus there is a large benefit in using an estimator which includes the memory of previous values instead of re-estimating these every one or two periods. The results also indicate that the weights used in constructing the index might play a non-trivial

role. Using value shares, a plutocratic type index, compared to weights that are proportional to the number of transactions observed in the base and comparison period appear to lead to a difference of around 8% to 10% over some periods.

The paper has not compared the indexes resulting from methods that lead to revisions (continuos or frequent), but this will be added to the next version of the paper.

## References

- de Haan, J. and Diewert, E. W. (2011). Handbook on residential property price indexes. Technical report,B. Balk (Statistics Netherlands) project coordinator, de Haan and Diewert (eds). EuroStat.
- Diewert, E. W. (2003). Hedonic Regressions: A consumer Theory Approach, in Scanner Data and Price Indexes, NBER Conference on Research in Income and Wealth. Studies in Income and Wealth., volume 64, chapter 10, page 318:348. The University of Chicago Press.
- Durbin, J. and Koopman, S. (2001). Time Series Analysis by State Space Methods. Oxford Statistical Science Series. Oxford University Press.
- Fletcher, C. S., McAllister, R. R. J., Rambaldi, A. N., and Collins, K. (2011). The economics of adaptation to protect appreciating assets from coastal inundation. Final Report F110609, CSIRO Climate Adaptation Flagship.
- Harvey, A. C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge.
- Hill, R. J. and Melser, D. (2008). Hedonic imputation and the price index problem: An application to housing. *Economic Inquiry*, 46(4):593:609.
- Pace, R. K., Barry, R., Gilley, O. W., and Sirmans., C. F. (2000). A method for spatial-temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16:229:246.
- Rambaldi, A. N. and Rao, D. P. (2011). Hedonic predicted house price indices using time-varying hedonic models with spatial autocorrelation. School of Economics Discussion Paper 432, School of Economics, The University of Queensland. Australia.
- Silver, M. and Heravi, S. (2007). The difference between hedonic imputation indexes and time dummy hedonic indexes. *Journal of Business & Economic Statistics*, (25):239:246.

- Svetchnikova, D., Rambadi, A. N., and Strachan, R. (2008). A comparison of methods for spatial-temporal forecasting with an application to real estate prices. In *Econometric Society Australasian Meeting* (*ESAM08*), page 1:34, 9-11 July, Wellington, NZ.
- 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,.



Figure 1: Number of Single Transactions Per Month

| Table 1: Description of the Dataset |                      |                      |                      |  |  |
|-------------------------------------|----------------------|----------------------|----------------------|--|--|
| Variable                            | Median               | Min                  | Max                  |  |  |
| Sale Price                          | \$164,000            | \$5,250              | \$1,250,000          |  |  |
| LAND CHARACTERISTICS                |                      |                      |                      |  |  |
| Lot Size                            | $637 \text{ m}^2$    | $261 \text{ m}^2$    | $10610 \text{ m}^2$  |  |  |
| Distance to Waterway                | 251.25  m            | 5 m                  | 863.54 m             |  |  |
| Distance to Park                    | 110.11 m             | $7.07 \mathrm{~m}$   | 983.92 m             |  |  |
| Distance to Boat Ramp               | $1600.28~\mathrm{m}$ | $55.90~\mathrm{m}$   | 6294.16 m            |  |  |
| Distance to School                  | $322.50~\mathrm{m}$  | 10 m                 | $6547.09 {\rm m}$    |  |  |
| Distance to Shops                   | $399.55~\mathrm{m}$  | 10 m                 | $4796.77 {\rm m}$    |  |  |
| Distance to Industry                | $530.50~\mathrm{m}$  | $5 \mathrm{m}$       | 2708.44 m            |  |  |
| Distance to Bus Stop                | 226.11 m             | 15m                  | 4349 m               |  |  |
| Distance to Rail Station            | $6606.74~\mathrm{m}$ | $2453.59~\mathrm{m}$ | $13583.90 {\rm m}$   |  |  |
| Distance to Pubs/Clubs              | $1244.08 {\rm m}$    | 14.14 m              | 5336.24  m           |  |  |
| Distance to Hospital                | $2273.79~\mathrm{m}$ | 10 m                 | $13255~\mathrm{m}$   |  |  |
| Min-Height Parcel                   | $7.54 \mathrm{~m}$   | 0 m                  | $26.66~\mathrm{m}$   |  |  |
| STRUCTURE CHARACTERISTICS           |                      |                      |                      |  |  |
| Structure Footprint                 | $174.38 \text{ m}^2$ | 0 m                  | $920.90 \text{ m}^2$ |  |  |
| Bedrooms                            | 3                    | 0                    | 8                    |  |  |
| Bathrooms                           | 1                    | 0                    | 4                    |  |  |
| Carspaces                           | 1                    | 0                    | 12                   |  |  |
| Max-Height                          | $6.36 \mathrm{~m}$   | 0 m                  | 27.02 m              |  |  |
| Age                                 | 10  ys               | 0 ys                 | 86 ys                |  |  |
| Number of Transactions              | 13321                |                      |                      |  |  |
| Number of Vacant Land Sales         | 3326                 |                      |                      |  |  |

Table 1. Description of the Dataset



Figure 2: Average Monthly Per cent Prediction Error

|                | % Prediction Error Kalman Filter Estimates | % Prediction Error OLS Estimates |  |
|----------------|--|----------------------------------|--|
| 1991:6-1995:12 | 9.84                                       | 13.26                            |  |
| 1996:1-1999:12 | 7.75                                       | 14.22                            |  |
| 2000:1-2004:12 | 12.66                                      | 14.52                            |  |
| 2005:1-2010:9  | 8.53                                       | 9.62                             |  |

Table 2: Average % Prediction Error for the sub-samples

Table 3: Average % Prediction Error for the sub-samples -Spatial Model

|                | % Prediction Error Kalman Filter Estimates | % Prediction Error OLS Estimates |
|----------------|--|----------------------------------|
| 1991:6-1995:12 | 10.16                                      | 13.70                            |
| 1996:1-1999:12 | 8.09                                       | 15.14                            |
| 2000:1-2004:12 | 13.08                                      | 14.77                            |
| 2005:1-2010:9  | 8.89                                       | 9.85                             |



Figure 3: Average Monthly Per cent Prediction Error - Spatial Model



Figure 4: Hedonic Imputed Indexes - Non-Spatial Model



Figure 5: Hedonic Imputed Price Indexes - base 1991:6