Graz Economics Papers – GEP

GEP 2018–04

Higher Frequency Hedonic Property Price Indices: A State Space Approach

Robert J. Hill, Alicia N. Rambaldi, and Michael Scholz

February 2018

Department of Economics
Department of Public Economics
University of Graz

Higher Frequency Hedonic Property Price Indices: A State Space Approach

Robert J. Hill*

Department of Economics, University of Graz, Universitätsstr. 15/F4, 8010 Graz. Austria

Alicia N. Rambaldi†

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

Michael Scholz

Department of Economics, University of Graz, Universitätsstr. 15/F4, 8010 Graz. Austria

VERSION: 2 February 2018

Abstract

The hedonic imputation method estimates separate characteristic shadow prices for each period. These are used to construct matched samples, which are inserted into standard price index formulas. We implement two innovations to improve the method's effectiveness on housing data at higher frequencies. First, we use a time-varying parameter model in state-space form to increase the reliability of the estimated characteristic shadow prices. Second, we significantly reduce the number of parameters by replacing postcode dummies by a geospatial spline surface. Empirically, using a novel criterion, we show that in higher frequency comparisons our hedonic method outperforms competing alternatives. (***JEL.*** C33; C43; R31)

(***Keywords:*** Housing market; Hedonic imputation; State Space Model; Geospatial data; Spline; Quality adjustment; Matched sample)

*This project has benefited from funding from the Austrian National Bank (Jubiläumsfondsprojekt 14947). We thank Australian Property Monitors for supplying the data.

†Corresponding author: A.N.Rambaldi (+61(0)7 3365 6576), a.rambaldi@uq.edu.au

1 Introduction

Since the global financial crisis there is an increased awareness of the importance of the housing market to the broader economy. Hence there is a growing demand from central banks, governments, banks, real estate developers, and households for reliable and more timely house price indices, and for the development of tradable derivatives (Bokhari and Geltner, 2012). The increased availability of housing data and advances in computing power and econometric techniques is making it possible to deliver more timely indices to meet this demand.

Much progress has been made recently on computing higher frequency repeat-sales house price indices (Bokhari and Geltner, 2012; Bollerslev, Patton, and Wang, 2016; Bourassa and Hoesli, 2016). However, less progress has been made on computing higher frequency hedonic indices.[1] It is such higher frequency hedonic indices that are the focus of our attention here. More specifically, we focus on one specific type of hedonic method, referred to in the literature as the hedonic imputation method (Hill, 2013).

Hedonic methods estimate the price of a product (here housing) as a function of a vector of explanatory characteristics. The hedonic imputation method, first proposed by Court (1939) and further developed by Griliches (1961, 1971), reestimates the hedonic model each period. The hedonic model is then used to impute prices for matched samples, after which the overall price index can be computed using a standard price index formula. The method is more flexible and timely than other hedonic methods such as the time-dummy method in that the characteristic shadow prices are updated each period.[2] Unlike the time-dummy method, the hedonic imputation method also satisfies non-revisability (i.e. the indices once computed are never revised). Index users often find this useful. For example Eurostat (2016) advises European countries to use a non-revisable hedonic method when constructing their official house price indices.

The hedonic imputation method, however, becomes problematic at higher frequencies. For example, the hedonic model typically includes dummies to control for location (e.g. using postcodes) in addition to other hedonic characteristics of the dwelling.

[1] Quality-adjusted indices are typically computed using either hedonic or repeat-sales methods. The latter are more common in the US – the best known example being the S&P CoreLogic Case-Shiller indices. In Europe, hedonic methods are more widely used. For example the national statistical institutes (NSIs) of most member countries of the European Union now compute an official House Price Index (HPI) at a quarterly frequency using hedonic methods (Eurostat, 2016). One reason for this difference is that repeat-sales methods tend to work better when the frequency of transactions (i.e., turnover) is high as it is in the US. In Europe turnover is generally much lower. Elsewhere in the world, it is less clear which approach is preferred. CoreLogic for example computes both hedonic and repeat-sales indices for Australian cities.

[2] See Rosenthal (2006) for an example of an application of the time-dummy method to housing data.

At higher frequencies (e.g. weekly indices), even in large data sets there may not be enough price observations in each period to satisfactorily estimate the hedonic model. As a consequence computational and statistical problems occur (e.g. no observations for some postcodes, a loss in degrees of freedom, or an increased variance of estimated parameters).

Geltner and Ling (2006) describe the trade-off between statistical quality per period and the frequency of index reporting, holding constant the overall quantity and quality of raw valuation data and index construction methodology. They conclude that the usefulness of an index for research purposes clearly increases the greater the frequency of reporting, holding statistical quality (per period) constant (Bokhari and Geltner, 2012).

In this paper we implement two innovations to improve the effectiveness of the hedonic imputation method at higher frequencies. First, we use a time-varying parameter model in state-space form to increase the reliability of the estimated characteristic shadow prices in each period. Second, we replace postcode dummies by a geospatial spline surface. Replacing postcode dummies by values from the geospatial spline function at each location in the data set very significantly reduces the number of parameters that need to be estimated.

Our econometric modelling method is a two-step approach where in the first step we construct a location spline surface using the set of observed houses that have transacted at each period via a semi-parametric specification. This model provides both a spline measure of location for each house as well as a measure of spatial uncertainty. Both enter the second step as part of a state-space specification of a hedonic function with time-varying parameters (the temporal movements). This model is used to produce matched predictions of the sale price of each house in the sample across two time periods, providing a model based measure of price relatives which enter the computation of a superlative index formula (see Diewert 1976).

We are not aware of any prior research in econometrics that incorporates spatially and time-varying parameter features in the form we propose. Our two-stage approach provides a non-parametric estimate that varies over individual properties at any given time period in the first stage (to capture location of each individual property). These estimates (and their statistical uncertainty) enter the second stage, which is a time-varying parameter model written as state-space. The estimation of the second stage is by the Kalman filler and that is standard, although the state-space model itself has a number of non-standard features. Lastly, the predictions from the model needed to provide the correct matching as required by the Törnqvist price index formula (a superlative index based on the ratio of two predictions) are not from standard predictors. In summary, the model being estimated and its purpose, to provide a model based matching sample to construct a price index, are not a regular feature of either academic publications or standard practice by official or commercial providers of house price indices.

There are some similarities here with the literature on constructing monthly or weekly price indices for consumer goods using scanner data. In particular, Melser (2016) uses the hedonic imputation method to estimate superlative price indices, and impose non-revisability using a rolling window method. He then endogenizes the window length and allows each period in the window to be weighted differently. We impose non-revisability using an approach that mimics an endogenous rolling window. However, there are important differences as well. We focus on the construction of house price indices, not consumer price indices. Also, our approach embeds the hedonic model (which includes a geospatial spline surface) in a state-space model. The Kalman filter then optimally weights prior periods' information when constructing the estimates.[3]

We use a recently developed criterion proposed by Hill and Scholz (2017) to assess the performance of our method using data for Sydney (Australia) over the period 2003–2014. This criterion focuses on comparing the model's predicted price relatives of properties that have sold twice within the sample. The rationale behind the use of this measure is that imputed price relatives form the basic building blocks of the Törnqvist superlative price index, and thus it is the ability of the model to predict price changes over time rather than the price level of each property that matters. Based on this criterion, we find that our preferred index outperforms competing alternatives. Furthermore, we find that weekly indices are quite sensitive to the choice of method.

The remainder of this article is structured as follows. Section 2 provides an overview of the hedonic imputation method, the hedonic model and the methods used to estimate the generalized additive and the spatio-temporal components of the model. The criterion used to compare the performance of competing hedonic imputed indices is also considered here. Section 3 presents our data set, the empirical study and the results of our analysis. Section 4 concludes by summarizing our main findings.

[3]This is a feature of the Kalman filter as shown by Koopman and Harvey (2003).

2 Hedonic Imputation and Index Quality

2.1 Index Definition

Hedonic price indices for housing are typically constructed using one of the time-dummy, hedonic imputation, and average characteristic methods (Diewert, 2010; Hill, 2013; European Commission, Eurostat, OECD, and World Bank, 2013; Silver, 2016). All of them have in common that in a hedonic model the price of a product is regressed on a vector of characteristics (whose prices are not independently observed). The hedonic equation is a reduced form that is determined by the interaction of supply and demand. Hedonic models are used to construct quality-adjusted price indices in markets (such as computers) where the products available differ significantly from one period to the next. Housing is an extreme case in that every house is different.

Here we focus on the hedonic imputation method since it is more flexible than either the time-dummy or average characteristics methods (Silver and Heravi, 2007). The hedonic imputation method uses the predictions from a hedonic model to impute prices over a matched sample which can then be inserted into a standard price index formula. Let $x'_{t,h}$ be a vector of characteristics associated with property h sold in period t, and $\hat{p}_{t+1,h}(x'_{t,h})$ as the imputed price for that property had it sold in period $t+1$. The model used in this study to produce these predictions is presented in the next section. To obtain a hedonic imputed price index comparing periods t and $t+1$, we use a *Laspeyres*-type formula that focuses on the properties sold in the earlier period t, and a *Paasche*-type formula that focuses on the properties sold in the later period $t+1$. Our price indices are constructed by taking the geometric mean of the price relatives, giving equal weight to each house.[4] Taking a geometric mean of the Laspeyres and Paasche-type indices, we obtain a Törnqvist-type superlative index, that has the advantage that it treats both periods symmetrically and is consistent with a log-price hedonic model (Hill and Melser, 2008).

The indices presented below are all of the double imputation type.[5] This means that both prices in each price relative are imputed. For example, the double imputation

[4]This democratic weighting structure is in our opinion more appropriate in a housing context than weighting each house by its expenditure share. See Hill and Melser (2008), de Haan (2010), Rambaldi and Rao (2011) and Rambaldi and Fletcher (2014) for a discussion on alternative weighting schemes.

[5]Double imputation indices tend to be slightly more robust to omitted variables bias (Hill and Melser, 2008). We also calculated single imputation indices where only one price in each price relative is imputed. The results are virtually indistinguishable. Hence to save space we focus here only on double imputation indices.

Laspeyres (DIL), Paasche (DIP), and Törnqvist indices (DIT) are defined as follows:

$$P_{t,t+1}^{DIL} = \prod_{i=1}^{N_t} \left[\left(\frac{\hat{p}_{i,t+1}(x'_{i,t})}{\hat{p}_{i,t}(x'_{i,t})} \right)^{1/N_t} \right], \tag{1}$$

$$P_{t,t+1}^{DIP} = \prod_{i=1}^{N_{t+1}} \left[\left(\frac{\hat{p}_{i,t+1}(x'_{i,t+1})}{\hat{p}_{i,t}(x'_{i,t+1})} \right)^{1/N_{t+1}} \right], \tag{2}$$

$$P_{t,t+1}^{DIT} = \sqrt{P_{t,t+1}^{DIP} \times P_{t,t+1}^{DIL}}, \tag{3}$$

where $i = 1, \ldots, N_t$ indexes the dwellings sold in period t, and $i = 1, \ldots, N_{t+1}$ indexes the dwellings sold in period $t+1$. The overall price index is then constructed by chaining together these bilateral comparisons between adjacent periods. As is discussed in the next section, the predictions used to compute the bilateral indices must take into account the spatio-temporal nature of our modelling approach.

2.2 The Model

The objective of the hedonic model is to provide predictions of the prices of properties included in the Törnqvist index calculation. The econometric model combines elements from the work of Wikle and Cressie (1999)-WC and Rambaldi and Fletcher (2014). WC provide a temporally dynamic and spatially descriptive model and an efficient estimation algorithm designed to deal with a large scale spatio-temporal dataset. We adopt a similar modelling approach in that measurement error, location, property quality components, and a term that captures small scale spatial variability are incorporated. This term conceptually extends the spatio-temporal models proposed by Rambaldi and Fletcher (2014), where two parametric alternatives to model location are used. Following Hill and Scholz (2017), the model incorporates the measure of location obtained by estimating a geospatial spline surface within a semi-parametric framework using observed sales in each individual period. The periodwise estimation provides a required measure of spatial variability and identification of the parameters of the spatio-temporal model.

We denote the observed (log transformed) price by $y_{it} = \ln price_{it}$. The objective is to predict y^*_{it}, a smoother but unobservable (log) price of property i in period t, for i in any location and over all time periods t, regardless of when and where the data are observed.

We write this model as

$$y_{it} \quad = \quad y^*_{it} + \epsilon_{it}; \epsilon_{it} \sim N(0, \sigma^2_\epsilon). \tag{4}$$

The random process ϵ_{it} is not correlated across location or time and captures overall measurement error.

At a given time period, τ, N_τ properties are sold, and y^*_τ is given by

$$y^*_\tau = x^\dagger_\tau + v_\tau; v_\tau \sim N(0, V_\tau) \tag{5}$$

where, v_τ is a random error that does not have a temporally dynamic structure but might have some spatial structure and thus V_τ might not be diagonal. It is assumed that $E(v_{i\tau}\epsilon_{jt}) = 0$ for all $i, j = 1, \ldots, N$ and $-\infty \leq t \leq \infty$.

$x^\dagger_t$ is assumed to evolve according to three components, trend, property quality and location,

$$x^\dagger_{it} = \mu_t + \sum_{k=1}^{K} \beta_{k,t} z_{k,it} + \gamma_t g_{it}(z_{long}, z_{lat}) \tag{6}$$

where, μ_t is a trend component common to all i in period t and captures overall macroeconomic conditions that affect all locations in the market under study;

$z_{k,it}$ is the kth hedonic characteristic from a set of K providing information on the type/quality of the property (e.g., number of bedrooms, bathrooms, size of the lot).

Note that the vector $z_{k,t}$ contains the kth hedonic characteristic for the set of properties sold in period t, which is different to the set of properties sold in $t - s$ where $s \neq 0$, and thus these are not trending variables.

$g_{it}(z_{long}, z_{lat})$ is a measure of the location of property i on a continuous surface defined by all properties sold in period t. Thus, it is temporally uncorrelated as the set sold in period t is different from that sold in $t - 1$ or $t + 1$.[6]

$\beta_{k,t}$ and γ_t are time-varying parameters associated with the hedonic characteristics and location shadow prices, and are to be estimated.

$E(z_k v_t) = 0$, $E(z_k \epsilon_t) = 0$ for all $k = 1, \ldots, K$, $E(g_{it} v_{jt}) = 0$, $E(g_{it}\epsilon_{jt}) = 0$, for all i, j, t.

[6] The only overlap is that of properties that repeat sale. The hedonic imputation method treats these as independent sales. We use this feature of the approach to compute an index performance indicator as explained in section 2.5.

2.3 Identification and State Space Representation

Inspecting equation (6) it is clear that an identifying assumption is required to be able to estimate γ_t. We achieve this by obtaining an estimate of location, denoted by $\widehat{g}_{it}(z_{long}, z_{lat})$, from a semi-parametric model of the form in (7) estimated at each time period τ using only the transactions information of that period,

$$y_{i\tau} = \theta_{0\tau} + z'_{i\tau}\theta^{\dagger}_{\tau} + g_{i,\tau}(z_{long}, z_{lat}) + v_{i\tau} \tag{7}$$

where,

$\theta^{\dagger}_{\tau} = \{\theta_{1\tau}, \ldots, \theta_{K,\tau}\}'$

This model provides two key estimates (estimation is discussed in Section 2.5) to our spatio-temporal model, namely, the estimate of location, $\hat{g}_{i,t}(z_{lat}, z_{long})$ and an estimate of V_t.

With above definitions, the model can be written in familiar state-space representation

$$y_t = X_t\alpha_t + v_t + \epsilon_t; \epsilon_t \sim N(0, H) \tag{8}$$

$$\alpha_t = D\alpha_{t-1} + \eta_t; \eta_t \sim N(0, Q) \tag{9}$$

where,

X_t is $N_t \times (K+2)$ and with the ith row being $x'_{it} = \{1, z_{1,it}, \ldots, z_{K,it}, \hat{g}_{it}(z_{long}, z_{lat})\}$

y_t is the vector of log transformed observed prices of properties sold at t.

$H = \sigma^2_{\epsilon} I_{N_t}$

$\alpha_t = \{\mu_t, \beta_{1t}, \ldots, \beta_{K,t}, \gamma_t\}'$

$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & I_K & 0 \\ 0 & 0 & \rho \end{bmatrix}$; $0 \leq \rho \leq 1$; If $\rho < 1$ the estimate of γ_t is mean reverting. If $\rho = 1$, γ_t evolves as a random walk as do the other state parameters α_t.

$Q = \begin{bmatrix} \sigma^2_{\mu} & 0 & 0 \\ 0 & \sigma^2_{\beta} I_K & 0 \\ 0 & 0 & \sigma^2_g \end{bmatrix}$

The estimate $\widehat{g}_{it}(z_{long}, z_{lat})$ enters the spatio-temporal model as a generated regressor and the parameter γ_t, in (8) and (9), provides the flexibility for the vector of location spline estimates of properties sold in period t, $i = 1, \ldots, N_t$, to be shifted by temporal market information up to time t. The additional uncertainty induced by replacing $g_{it}(z_{long}, z_{lat})$ by an estimate is captured by v_t in (8). The combination of spatial and temporal information leads to two unconventional features of this model, compared to one in a standard time-series setting, with consequences for the form of the Kalman filter algorithm as well as the price prediction to be used for the computation of the Törnqvist price index. First the error has two components, ϵ_t, the conventional overall measurement error, and v_t capturing the spatial variability within each time period. This results in a Kalman gain, G_t, which is a function of the sum of the two covariances $(H + V_t)$ under the assumptions already stated. The second is that the value of the location spline for property i sold in period t will not be identical in value if property i is priced in a different time period. That is, a given property has fixed location coordinates and hedonic characteristics; however, its location spline value, unlike the size of the land, will differ between period t and period $t + 1$. We denote by $\hat{g}_{t(t)}(z_{long}, z_{lat})$ the vector of spline values for properties sold and priced in period t, and by $\hat{g}_{t(t-1)}(z_{long}, z_{lat})$ the vector of spline values for the set of properties sold in t when priced in $t - 1$. The implications for the form of the Kalman filter algorithm are presented next.

Using the innovation form of the filter, the state at time t is given by

$$\alpha_{t|t} = \alpha_{t|t-1} + G_t\{y_t - X_t^1 \alpha_{t|t-1}\} \tag{10}$$

where, the prediction step of the Kalman filter uses X_t^1 which is the X_t matrix with the $\hat{g}_{i,t(t)}(z_{long}, z_{lat})$ replaced by $\hat{g}_{i,t(t-1)}(z_{long}, z_{lat})$. This is necessary to obtain the conditional prediction error from a conditional prediction of the state. The mean square error matrix given information up to time period t is $P_{t|t} = P_{t|t-1} - G_t X_t P_{t|t-1}$,

Under the stated assumptions the Kalman gain is a function of two covariances, H and V_t,

$$G_t = P_{t|t-1} X_t' \{H + V_t + X_t P_{t|t-1} X_t'\}^{-1} \tag{11}$$

The updating equations are of the standard form, $\alpha_{t|t-1} = D\alpha_{t-1|t-1}$ and $P_{t|t-1} = DP_{t-1|t-1}D' + Q$.

Estimates of the state (10), $\hat{\alpha}_{t|t}$, are obtained by replacing H, Q, D, and V_t, by suitable estimates.

Estimating (7) predictions of (log) prices, denoted by $\hat{x}_\tau^\dagger$, and of the location spline $\hat{g}_\tau(z_{lat}, z_{long})$ can be obtained for each property. Residuals and estimates of the location spline for $j = -1, 0, 1$, $\hat{v}_{t(t+j)}$ and $\hat{g}_{t(t+j)}(z_{lat}, z_{long})$, respectively, are used to implement the predictions/imputations to construct the index (discussed in the next section). Details of the estimation of the semi-parametric model are provided in section 2.5.

2.4 Constructing the Predictions

The computation of the index, (3), depends crucially on the prediction of log price. The state vector at period t conditional on information up to time period t, the prediction of the log price for property h is given by the natural predictor plus a Goldberger's correction term (Goldberger, 1962) as follows,

$$\widehat{y^*_{t|t,h}} = x'_{t,h}\hat{\alpha}_{t|t} + c'_{vt,h}\Omega^{-1}e_t \tag{12}$$

where,

$\Omega = cov\{y_t, y_t\}$

$c'_{vt,h} = E(v_{ht}, v_t)$ is the row of V_t corresponding to property h and has elements $c_{v,hj} \equiv E\{v_{ht}v_{jt}\}$ which could be equal to zero for $h \neq j$.

$e_t = y_t - E(y_t)$

To show this, in addition to assumptions already stated, we assume v_{it} and y_t have a joint multivariate normal distribution. Taking the characteristics and location of properties as given, the predictor is derived as follows,

$$\begin{aligned}
\widehat{y^*_{i,t|t}} &= E\{y^*_{it}|y_t, y_{t-1}, \ldots, y_1\} \\
&= E\{X_{it}\alpha_t + v_{it}|y_t, y_{t-1}, \ldots, y_1\} \\
&= X_{it}E\{\alpha_t|y_t, y_{t-1}, \ldots, y_1\} + E\{v_{it}|y_t, y_{t-1}, \ldots, y_1\} \\
&= X_{it}\alpha_{t|t} + c'_{vt,h}\Omega^{-1}e_t
\end{aligned}$$

The last term is of this form since $E\{v_{it}y_{jt}\} = c_{v,ij}$; $c_{v,ij} \equiv E\{v_{it}v_{jt}\}$, and $c'_{v,it} = E\{v_{it}, v_t) = \{c_v(i, j_1), \ldots, c_v(i, j_{N_t})\}'$

In this study we implement this prediction by defining $\widehat{v_t} = y_t - \hat{x}^{\dagger}_t$, which are the residuals from estimating (7) and $e_t = y_t - \widehat{y_{t|t}}$, where $\widehat{y_{t|t}}$ is the state-space prediction of the (log) price of property h at time t, $\widehat{y_{t|t}} = X_t\hat{\alpha}_{t|t}$, which are the residuals from the estimated state-space model.

For the index calculation predictions and imputations are needed. The prediction of the price of property h sold in period $t = 1, \ldots, T$ is defined as

$$\hat{p}_{t,h}(z'_{t,h}, \hat{g}_{h,t(t)}) = \exp(\widehat{y^*_{t|t,h}}) \tag{13}$$

The imputation of the price of property h sold in period t for period $t-1$ is given by

$$\hat{p}_{t-1,h}(z'_{t,h}, \hat{g}_{h,t(t-1)}) = \exp({x^1_{t,h}}'\hat{\alpha}_{t-1|t-1} + c'_{v(t-1),h}\Omega^{-1}e_{t(t-1)}) \quad (14)$$

The crucial point is that the constructed location effect and parameters need to be matched with the correct period that is being imputed. In this case, $\hat{g}_{t(t-1),h}$ enters in ${x^1_{t,h}}'$, $c'_{v(t-1),h}$ and together with $\hat{\alpha}_{t-1|t-1}$ in $e_{t(t-1)}$.

2.5 Estimation

Step I - Location Spline and Spatial Uncertainty

This step provides estimates of $g_t(.)$ and V_t using spatial only information given by the sample of N_t transacted houses at each period. A semi-parametric hedonic model with the specification in (7) is implemented as a generalized additive model (GAM) – a flexible model class that generalizes linear models with a linear predictor combined with a sum of smooth functions of covariates. At any time period t, the model estimated is,

$$y_i = x'_i\theta^t + v_i \quad (15)$$

$$= z'_i\theta^{t,\dagger} + g(z_{long}, z_{lat})^t + v_i; \quad i = 1, \ldots, N_t \quad (16)$$

where the ' ^{t} ' indicates the parameters and spline are period specific.

The estimates of the spline surface, $\hat{g}_t(.)$, enter the spatio-temporal model's X_t matrix; while the vector of N_t predictions from this model, $\hat{x}^\dagger_t$, provides residuals, $\widehat{v_{it}} = y_{it} - \hat{x}^\dagger_{it}$, from which a sample estimate of V_t, and thus $c'_{vt,h}$, make operational the correction term in (12).

To estimate (16) the problem is to select the smooth functions and their degree of smoothness. Here, we use a penalized likelihood approach (see Wood 2006, and the references therein) based on a transformation and truncation of the basis that arises from the solution of the thin plate spline smoothing problem. This method is computationally efficient and avoids the problem of choosing the location of knots, known to be crucial for other basis functions.

For the selection of the smoothing parameter we refer to Wood (2011), who proposes a Laplace approximation to obtain an approximate restricted maximum likelihood (REML) estimate which is suitable for efficient direct optimization and computationally stable. The REML criterion requires that a Newton-Raphson approach is used in model fitting, rather than a Fisher scoring. The penalized likelihood maximization problem is solved by Penalized Iteratively Reweighted Least Squares (P-IRLS).

The semi-parametric model is estimated using the *mgcv* package of the statistical software R 3.4.3 (R Core Team 2017). The same basis dimension and sample size are used as in Hill and Scholz (2017).[7]

Step II - Estimation of the State-Space Model for Prediction/Imputation

This step produces estimates of the state-vector, α_t (and its mean squared prediction matrix, $P_t|t$) via the algorithm outlined in Section 2.3. The algorithm requires estimates of D, H, Q. These can be obtained by maximum likelihood estimation. Given y_t, $Z_t = \{z_{1t}, \ldots, z_{Kt}\}$, $\hat{g}_t(\cdot)$ and $\hat{V}_t$, the Kalman filter algorithm is run to evaluate the log-likelihood, $\ln L$, in predictive form[8]. The estimator's algorithm is a function of a prediction error, $\nu_{t|t-1} = y_t - X_t^1 \hat{\alpha}_{t|t-1}$, and the Kalman Gain (11), which is a function of $F_t = E(\nu_{t|t-1}\nu'_{t|t-1}) = H + \hat{V}_t + X_t P_{t|t-1} X'_t$.

The estimation of the model and computation of indices were coded by the authors.

2.6 Measuring the quality of the index

The constructed indices should be useful instruments for policymakers and market participants. A criterion is needed therefore to evaluate the quality of the proposed indices. An important distinction can be made here between the hedonic model and the resulting price index. What matters is the performance of the index. Hence performance criteria should focus on the Törnqvist index defined in (3), and not the within-period fit of the hedonic model itself. Guo, Zheng and Geltner (2014) and Jiang, Phillips and Yu (2015) take a similar view. Guo, Zheng and Geltner (2014) suggest criteria based on the autocorrelation and volatility of the index, and Jiang, Phillips and Yu (2015) create a testing sample which is used for out-of-sample evaluation of the model's fit.

[7] It is important that the sample size each period exceeds the basis dimension. Hill and Scholz select these values by comparing computing time and model fit as measured by the Akaike Information Criterion (AIC).

[8]

$$\ln L(\rho, \sigma^2_\epsilon, \sigma^2_\mu, \sigma^2_\beta, \sigma^2_\gamma; y_t, Y_{t-1}, Z_t, \hat{g}_t, \hat{g}_{t(t-1)}, V_t)$$
$$= -\frac{NT}{2} ln(2\pi) - \frac{1}{2}\sum_{t=d}^{T} ln|F_t| - \frac{1}{2}\sum_{t=d}^{T} \nu'_{t|t-1} F_t^{-1} \nu_{t|t-1}$$

where, $Y_{t-1} = y_{t-1}, y_{t-2} \ldots$. We use a standard Newton-Raphson algorithm to estimate $\hat{\sigma}^2_\epsilon$, $\hat{\sigma}^2_\mu$, $\hat{\sigma}^2_\beta$ and $\hat{\sigma}^2_\gamma$ within a grid search for ρ in the range (0.1 to 1); $N = \sum_{t=d}^{T} N_t$; d is sufficiently large to avoid the log-likelihood being dominated by the initial condition, $\alpha_0 \sim N(a_0, \Omega_0)$. In the empirical implementation we have 731 weeks and set $d = 105$ (this choice is explained in the empirical section). For details on estimation of state-space models see Harvey (1989) or Durbin and Koopman (2012).

We follow a more direct approach here that makes use of the underlying structure of our hedonic imputation price indices.

The Törnqvist index is the geometric mean of the Laspeyres and Paasche-type price index formulas (1) and (2). From inspection of (1) and (2) it can be seen that the building blocks of the Laspeyres-type index are the imputed price relatives $\hat{p}_{i,t+1}(x'_{i,t})/\hat{p}_{i,t}(x'_{i,t})$, while the building blocks of the Paasche-type index are the imputed price relatives $\hat{p}_{i,t+1}(x'_{i,t+1})/\hat{p}_{i,t}(x'_{i,t+1})$. Hence the performance of the index depends on the quality of these imputed price relatives. Following Hill and Scholz (2017), the key insight is that repeat sales price relatives can be used as a benchmark for evaluating the imputed price relatives. To ensure a large enough sample size, repeat-sales price relatives over any time horizon in our data set are compared to their imputed counterparts, and not just in adjacent periods.

More formally, suppose property i sells in both periods t and $t+k$. For this property therefore we have an observed repeat-sales price relative: $p_{i,t+k}/p_{i,t}$. The corresponding imputed price relative is $\hat{p}_{i,t+k}/\hat{p}_{i,t}$. The subsample of properties that have repeat-sale are indexed by $i = 1, \ldots, H_{RS}$. We can now define the ratio of imputed to actual price relative for house i as follows:

$$d_i = \frac{\hat{p}_{i,t+k}}{\hat{p}_{i,t}} \Big/ \frac{p_{i,t+k}}{p_{i,t}}. \tag{17}$$

Our quality measure is then the average squared error of the log price relatives of each hedonic method:

$$D = \left(\frac{1}{H_{RS}}\right) \sum_{i=1}^{H_{RS}} [\ln(d_i)]^2, \tag{18}$$

where the summation in (18) takes place across the whole repeat-sales sample. We prefer whichever hedonic imputation model generates the smallest value of D, on the grounds that the resulting Törnqvist index will be constructed from the most reliable imputed price relatives.

One potential problem with using repeat-sales as a benchmark is that a repeat-sales sample may have a "lemons bias", since starter homes sell more frequently as people upgrade as their wealth rises. This lemons bias has been documented by, amongst others, Clapp and Giaccotto (1992), Gatzlaff and Haurin (1997), and Shimizu, Nishimura and Watanabe (2010). The quality of the house between repeat sales may also decline due to depreciation or it could improve due to renovations and repairs.

We correct for any such bias by adjusting the repeat-sales price relatives $p_{t+k,h}/p_{t,h}$

as follows:

$$\left(\frac{p_{i,t+k}}{p_{i,t}}\right)^{adj} = \left[\left(\frac{P^{RS}_{t+k}}{P^{RS}_t}\right) \Big/ \left(\frac{P^{Hed}_{t+k}}{P^{Hed}_t}\right)\right] \left(\frac{p_{i,t+k}}{p_{i,t}}\right), \tag{19}$$

where P^{RS}_{t+k}/P^{RS}_t denotes the change in the standard Case-Shiller Home Price repeat-sales price index between periods t and $t + k$, while P^{Hed}_{t+k}/P^{Hed}_t is the change in a reference hedonic index, calculated using the Törnqvist formula in (3) over the same time interval.[9] Hence the ratios of actual to imputed price relatives are adjusted as follows:

$$d^{adj}_i = d_i \left[\left(\frac{P^{RS}_{t+k}}{P^{RS}_t}\right) \Big/ \left(\frac{P^{Hed}_{t+k}}{P^{Hed}_t}\right)\right]. \tag{20}$$

Bias corrected D coefficients, denoted by D^{adj} in Table 2, are then calculated as follows:

$$D^{adj} = \left(\frac{1}{H_{RS}}\right) \sum_{i=1}^{H_{RS}} [\ln(d^{adj}_i)]^2. \tag{21}$$

There remains the question of which set of hedonic price indices should be used to make the lemons bias correction when computing (19) and (20). As a robustness check we take each of three indices (details in the next section) in turn as the reference method when making the bias correction. Hence in Table 2 we present three alternative D^{adj} coefficients. In all cases the ranking of methods is the same. Hence our findings are robust to the treatment of lemons bias.

3 Empirical application

3.1 The data set

We use a data set obtained from Australian Property Monitors that consists of prices and characteristics of houses sold in Sydney (Australia) for the years 2001–2014. For each house we have the following characteristics: the actual sale price, time of sale, postcode, property type (i.e., detached or semi), number of bedrooms, number of bath-

[9] In both the D measures in (18) and the repeat sales index we exclude repeat sales where the house was renovated between sales. We attempt to identify such houses in two ways. First, we exclude repeat sales where one or more of the characteristics have changed between sales (for example a bathroom has been added). Second, we exclude repeat sales that occur within six months on the grounds that this suggests that the first purchase was by a professional renovator. Exclusion of repeat-sales within six months is standard practice in repeat-sales price indices such as the S&P CoreLogic Case-Shiller Home Price Index. Finally, for houses that sold more than twice during our sample period (2003-2014), we only include the two chronologically closest repeat sales (as long as these are more than six months apart). There are 83 258 repeat-sales houses in the full data set. As a result of the deletions explained above, the sample was reduced to 61 024 houses.

rooms, land area, exact address, longitude and latitude. (We exclude all townhouses from our analysis since the corresponding land area is for the whole strata and not for the individual townhouse itself.) Some summary statistics are provided in Table 1, and a plot of the number of sales per week is shown in Figure 1. As can be seen from Figure 1, the number of transactions falls very significantly each year during the summer holiday period from mid December to late January. Any method for computing weekly indices needs to be able to handle such seasonal fluctuations in transactions volume.

Table 1 and Figure 1 about here

For a robust analysis it was necessary to remove some outliers. This is because there is a concentration of data entry errors in the tails, caused for example by the inclusion of erroneous extra zeroes. These extreme observations can distort the results. The exclusion criteria we applied are also shown in Table 1. Complete data on all our hedonic characteristics are available for 433 202 observations. To simplify the computations we also merged the number of bathrooms and number of bedrooms to broader groups (one, two, and three or more bathrooms; one or two, three, four, five or more bedrooms). The quality of the data improves over time. In particular, missing characteristics are quite common in the first two years (i.e., 2001 and 2002). Thus we present the hedonic indices starting in 2003. Nevertheless, we use the full sample period to run the Kalman filter algorithm but compute the log-likelihood function (see section 2.4) with all weeks in the 2003-2014 period.

3.2 Comparing Parameter Estimates

In this section we report the parameter estimates of the spatio-temporal model. In addition we compare these with the estimates of shadow price parameters obtained by estimating the generalised additive model (16), $\hat{\theta}_t^{\dagger}$, which, following the standard practice in the price index literature, are obtained by estimation of the model period-by-period. These are displayed in Figures 2 and 3 with the $\hat{\theta}_{kt}^{\dagger}$ estimates labelled "GAM" in the figures' legends, and the corresponding estimate from the spatio-temporal model ($\hat{\alpha}_{k,t|t}$) labelled "SS+GAM". In addition the plots show an approximate 95% coverage of each element of the estimated state computed as two-standard errors from the estimated mean squared error matrix of the state.

The GAM estimates are much more volatile than their SS+GAM counterparts as expected. Estimating via a state-space representation provides a linking of the parameters overtime and reduces greatly the effect of the change in the composition of properties across periods. As transacted properties are not random samples of the market at each period, the effect of sales composition together with small samples in some periods can lead to this high volatility in the shadow price parameter estimates which should not be there in theory. We note that in a number of periods and across all the $\hat{\theta}^{\dagger}_{kt}$ there are estimates that fall outside the 95% bound of $\hat{\alpha}_{k,t|t}$. These can have potentially important implications for the index constructed using predictions from this model (presented in the next section).

Figures 2 and 3 about here

One additional parameter of the spatio-temporal representation is the process γ_t, which we modelled as possibly mean-reverting unlike the shadow price parameters of hedonic characteristics. This process provides the flexibility for the vector of location spline estimates of properties sold in period t, $i = 1, ..., N_t$, to be shifted by temporal market information up to time t, and is mean-reverting if the estimate of $|\rho| < 1$ (see equation (9). We found the estimate of ρ to be 0.4.[10]

3.3 Property price indices

We construct three hedonic price indices, a repeat sales index and a quality unadjusted index. A basic hedonic index is computed from the generalised additive hedonic model in (16) estimated separately each week. This index is referred to as **GAM**. The second hedonic is based on the spatio-temporal hedonic model presented in Section 2 and is referred to as **SS+GAM**. As discussed in Section 1, postcode dummies are often used to control for location. Thus, a simpler alternative to (16) is

$$y_t = \mu_t + Z_t\beta_t + D_t\pi_t + \varepsilon_t \tag{22}$$

where μ_t is local level trend, Z_t hedonic characteristics, D_t is a matrix of postcode dummies containing the location information and π_t is the vector of corresponding shadow prices for the postcodes.

Computing hedonic imputation price indices using period-by-period estimation with (22) is not feasible in a weekly context. It happens that for some postcodes we have no observations in some weeks causing both statistical and computational problems, especially in the hedonic prediction step. However, it can be estimated as a regression with time-varying parameters by setting it up as a state-space model. The index obtained

[10]We have not included a plot of $\hat{\gamma}_t$, but it is available from the authors.

from this model is referred to here as **SS+PC**. Imputed price relatives from the three models are inserted into the Törnqvist formula in (3) to generate the respective price index.

Figure 4 shows the three hedonic indices (chained), the repeat sales index calculated using the standard formula from Bailey, Muth, and Nourse (1963), and the quality unadjusted price index computed from the median of the prices of observed sales in each week. The median index is both a quality and location unadjusted index. It is extremely volatile, thus demonstrating the need for quality adjustment to generate an economically meaningful index. All indices except for SS+GAM lie below the median price index for most of the sample period. The GAM index appears to suffer from some chain drift. Prior to 2011 the index is closer to the median and the SS+GAM; however, it drifts down to the SS+PC and repeat-sales indices after 2011. Index drift may occur in the conventional approach to hedonic imputation when the market is thin as small samples and sales' composition in thin markets can affect the parameter estimates and lead to large changes in the price relatives. This is clearly the case in this instance as was discussed in Section 3.2. Chaining can then compound this drift. Rambaldi and Fletcher (2014) find chain drift occurs in monthly indices even when using a two-month rolling window to estimate the parameters of the model. The SS+PC and repeat-sales indices are uniformly below the median and virtually indistinguishable from each other.

The differences between the hedonic indices in Figure 4 are larger than one might expect to observe in hedonic indices computed at annual or quarterly frequency (Hill and Scholz, 2017). To illustrate this point we have estimated the three models and computed all the indices at a quarterly frequency. Figure 5 shows that at a quarterly frequency our hedonic indices approximate each other quite closely. Hence it can be seen that the choice of hedonic method is of greater importance when indices are computed at higher frequencies, such as weekly.

Figures 4 and 5 about here

3.4 Comparing the quality of the indices

The performance of our three indices according to the D and D^{adj} criteria is shown in Table 2.

Table 2 about here

All our criteria (D, D^{adj}_{GAM}, D^{adj}_{SS+GAM}, D^{adj}_{SS+PC}) generate the same ranking of hedonic methods. In all cases, the SS+GAM model performs best followed by GAM, with SS+PC performing worst.

Furthermore, the superior performance of SS+GAM is highly statistically significant. To show this we test whether D and D^{adj} are significantly different across different hedonic models. It is clear from equations (18) and (21) that these measures are averages. Thus, the null hypothesis is that the true difference between two means is zero ($H_0 : D_{M1} - D_{M2} = 0$) or ($H_0 : D^{adj}_{M1} - D^{adj}_{M2} = 0$), where $M1$ and $M2$ denote two of the hedonic models (e.g. GAM and SS+GAM). To decide on a suitable statistic to conduct the test we note that there is no dependence structure in the computed D measures. This is the case as we do not include repeat sales that are within six months of each other and for each pair the length between sales varies. In addition, any pair of repeat sales appears only once in the sample. Scatter plots of the $[\ln(d_i)]^2$ and $[\ln(d_i^{adj})]^2$ confirm there are no patterns.[11] Thus, based on the Central Limit Theorem (see for example pages 490-491 in Devore and Berk, 2012),

$$D_{M1} - D_{M2} \sim \mathcal{N}\left(0, \frac{s_1^2 + s_2^2}{H_{RS}}\right),$$

where s_j (j=1,2) is the sample standard deviation of the $D_j ({D_j}^{adj})$ for hedonic model j. The computed test-statistics and corresponding two-sided p-values of this exercise are shown in Table 3.

Table 3 about here

These results therefore show the importance of correctly modelling space and time in a unified framework which can account for all sources of error. The SS+GAM method generates the most accurate matched sample price relatives. The Törnqvist price index as defined in (3) is computed by taking a geometric mean of these price relatives.

4 Conclusion

This article has focused on the construction of weekly house price indices using the hedonic imputation method. The hedonic imputation method provides a flexible way of constructing quality-adjusted house price indices using a matching sample approach. We develop a state-space model that controls for location with a geospatial spline surface. Estimation of the model requires a modified form of the Kalman filter. The

[11]These scatter plots are available from the authors upon request.

geospatial spline surface replaces postcode dummies which are commonly used to control for location in the hedonic price index literature. The use of the spline achieves greater precision and a large reduction in the dimensionality of the spatio-temporal model. Imputed prices are obtained using a Goldberger's adjusted form of the predictor. These imputed prices provide a matched sample, thus allowing the price index to be computed using the superlative Törnqvist price index formula. Using a data set for Sydney, Australia, weekly hedonic indices are shown to be far more sensitive to the method of construction than indices computed at lower frequencies such as quarterly. Hence it is at these higher frequencies that the choice of hedonic method matters most. It is then shown, using a recently proposed criterion that our preferred method for computing weekly indices clearly outperforms alternative hedonic imputation methods.

References

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

Bokhari, S. and D. Geltner (2012). Estimating Real Estate Price Movements for High Frequency Tradable Indexes in a Scarce Data Environment. *Journal of Real Estate Finance and Economics* **45**(2), 522–543.

Bollerslev, T., A.J. Patton, and W. Wang (2016). Daily House Price Indices: Construction, Modeling, and Longer-Run Predictions. *Journal of Applied Econometrics* **31**, 1005–1025.

Bourassa, S.C. and M. Hoesli (2016). High Frequency House Price Indexes with Scarce Data. *Swiss Finance Institute Research Paper Series* **16-27**.

Clapp, J.M. and C. Giaccotto (1992). Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods. *Journal of Real Estate Finance and Economics* **5**(4), 357–374.

Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples, in *The Dynamics of Automobile Demand*. New York: The General Motors Corporation, 99–117.

de Haan, J. (2010). Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Re-Pricing Methods. *Journal of Economics and Statistics (Jahrbuecher für Nationalökonomie und Statistik)* **230**(6), 772–791.

Devore, J.L., and K.N. Berk (2012). *Modern Mathematical Statistics with Applications.* Springer, New York, 2nd edition.

Diewert, W. E. (1976). Exact and Superlative Index Numbers. *Journal of Econometrics* **4**, 115–145.

Diewert, W. E. (2010). Alternative Approaches to Measuring House Price Inflation. Discussion Paper 1010, Department of Economics, The University of British Columbia, Vancouver, Canada, V6T 1Z1, 2010.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods.* Oxford Statistical Science Series. Oxford University Press, 2nd edition.

European Commission, Eurostat, OECD, and World Bank (2013): Handbook on Residential Property Price indices (RPPIs). Eurostat: Luxembourg.

Eurostat (2016). Detailed Technical Manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices. Eurostat: Luxembourg.

Gatzlaff, D.H. and D.R. Haurin (1997). Sample Selection Bias and Repeat-Sales Index Estimates. *Journal of Real Estate Finance and Economics* **14**, 33–50.

Geltner, D. and D. Ling (2006). Considerations in the Design and Construction of Investment Real Estate Research Indices. *Journal of Real Estate Research* **28**(4), 411-444.

Griliches, Z. (1961). Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change, in *Government Price Statistics, Hearings Before the Subcommittee on Economic Statistics of the Joint Economic Committee*, 87th Congress.

Griliches, Z. (1971). Price Indexes and Quality Change, Cambridge MA: Harvard University Press., chap. Introduction: Hedonic Price Indexes Revisited.

Goldberger, A.S. (1962). Best Linear Unbiased Prediction in the Generalized Linear Regression Model. *Journal of the American Statistical Association* **57**, 298, 369 –375.

Guo, X., S. Zheng, D. Geltner, and H. Liu (2014). A new approach for constructing home price indices: The pseudo repeat sales model and its application in China. *Journal of Housing Economics* **25**, 20-38.

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

Hill, R.J. (2013). Hedonic Price Indexes for Housing: A Survey, Evaluation and Taxonomy. *Journal of Economic Surveys* **27**(5), 879–914.

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

Hill, R.J. and M. Scholz (2017). Incorporating Geospatial Data in House Price Indexes: A Hedonic Imputation Approach with Splines. *Review of Income and Wealth*, doi:10.1111/roiw.12303.

Jiang, L., P.C.B. Phillips, and J. Yu. (2015). New Methodology for Constructing Real Estate Price Indices Applied to the Singapore Residential Market. *Journal of Banking and Finance* **61**, S121–S131.

Koopman, S and A. Harvey. (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control*, **27**, 1317–1333.

Melser, D. (2016). Scanner Data Price Indexes Addressing Some Unresolved Issues. *Journal of Business & Economic Statistics*, forthcoming.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Rambaldi, A.N. and D.S.Pr. Rao (2011). Hedonic Predicted House Price Indices Using Time-Varying Hedonic Models with Spatial Autocorrelation. School of Economics Discussion Paper 432, School of Economics, University of Queensland.

Rambaldi, A.N. and C.S. Fletcher (2014). Hedonic Imputed Property Price Indexes: The Effects Of Econometric Modeling Choices.*Review of Income and Wealth* **60**, Supplement Issue, S423-S448.

Rosenthal, L. (2006). Efficiency and Seasonality in the UK Housing Market, 1991-2001.*Oxford Bulletin of Economics and Statistics* **68**(3), 289–317.

Shimizu, C., K.G. Nishimura and T. Watanabe (2010). Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures. *Journal of Economics and Statistics (Jahrbuecher für Nationalökonomie und Statistik)* **230**(6), 792–813.

Silver, M. (2016). How to Better Measure Hedonic Residential Property Price Indexes. International Monetary Fund Working Paper 16/213.

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.

Wikle C.K. and N. Cressie (1999). A Dimension-Reduced Approach to Space-Time Kalman Filtering. *Biometrika* **86**, 815–829.

Wood, S.N. (2006). *Generalized Additive Models: An introduction with R*, Chapman & Hall/CRC.

Wood, S.N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society B* **73**(1), 3–36.

Table 1: Summary of characteristics

	PRICE	AREA	LAT	LONG	FREQ.	BED	BATH
Minimum	56500	100.0	-34.20	150.6	1:	1348	190395
1st Quartile	420000	461.0	-33.93	150.9	2:	38578	174161
Median	610000	587.0	-33.84	151.0	3:	200428	57673
Mean	784041	626.1	-33.85	151.0	4:	147794	8835
3rd Quartile	900000	720.0	-33.76	151.2	5:	38734	1746
Maximum	3200000	4998.0	-33.40	151.3	6:	6320	392
Minimum Allowed	50000	100.0	-34.20	150.60		1.000	1.000
Maximum Allowed	4000000	5000.0	-33.40	151.35		6.000	6.000

Note: Price is measured in Australian dollars. Area is land area measured in square meters. The last two rows show the thresholds that were applied to delete outliers.

Table 2: Index quality based on D and D^{adj} criteria (2003-2014)

	D	D^{adj}_{GAM}	D^{adj}_{SS+GAM}	D^{adj}_{SS+PC}
GAM	0.0233	0.0272	0.0313	0.0230
SS+GAM	0.0102	0.0096	0.0099	0.0133
SS+PC	0.0246	0.0279	0.0320	0.0240

Note: GAM is based on periodwise estimation of the semiparametric model (16) with a geospatial spline; SS+GAM is the spatio-temporal model; SS+PC is the state space model applied to the semilog model in (22) with location effects captured using postcodes. D^{adj}_{GAM} refers to the adjusted D criteria with lemons bias corrected for using the GAM hedonic price index as the adjustment factor. Similarly, D^{adj}_{SS+GAM} and D^{adj}_{SS+PC} use the SS+GAM and SS+PC hedonic price indexers, respectively as the adjustment factors.

Table 3: p-values for the Null: $D_{M1} - D_{M2}(D^{adj}_{M1} - D^{adj}_{M2}) = 0$

	D	D^{adj}_{SS+PC}	D^{adj}_{SS+GAM}	D^{adj}_{GAM}
SS+PC vs. SS+GAM	0.0000	0.0000	0.0000	0.0000
SS+PC vs. GAM	0.0483	0.1004	0.2573	0.2732
GAM vs. SS+GAM	0.0000	0.0000	0.0000	0.0000

Note: These p-values imply that SS+GAM is highly significantly different from both SS+PC and GAM.

Figure 1: Number of Transactions per Week, 2001-2014

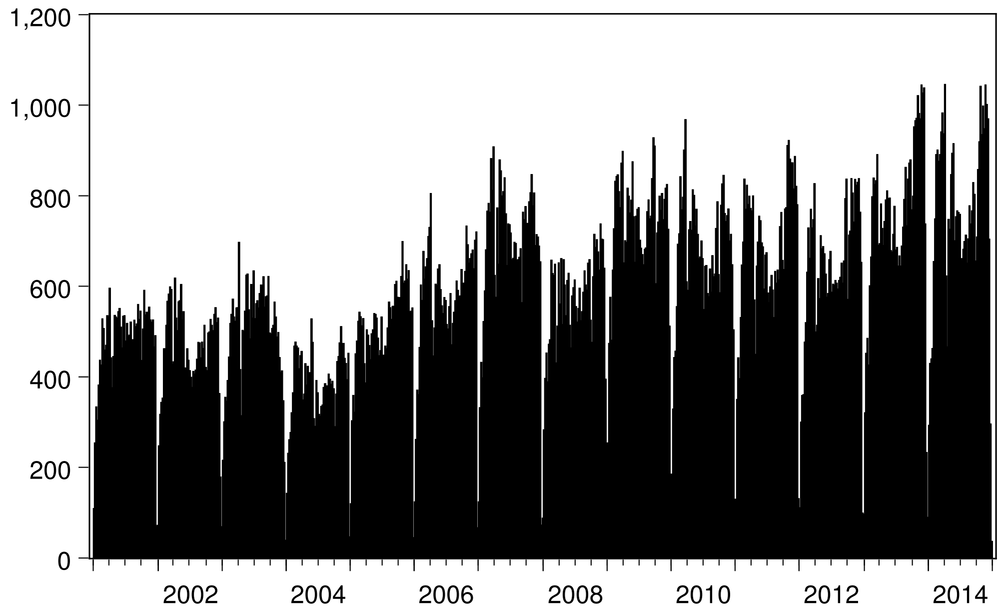


Figure 2: Parameter Plots (a)

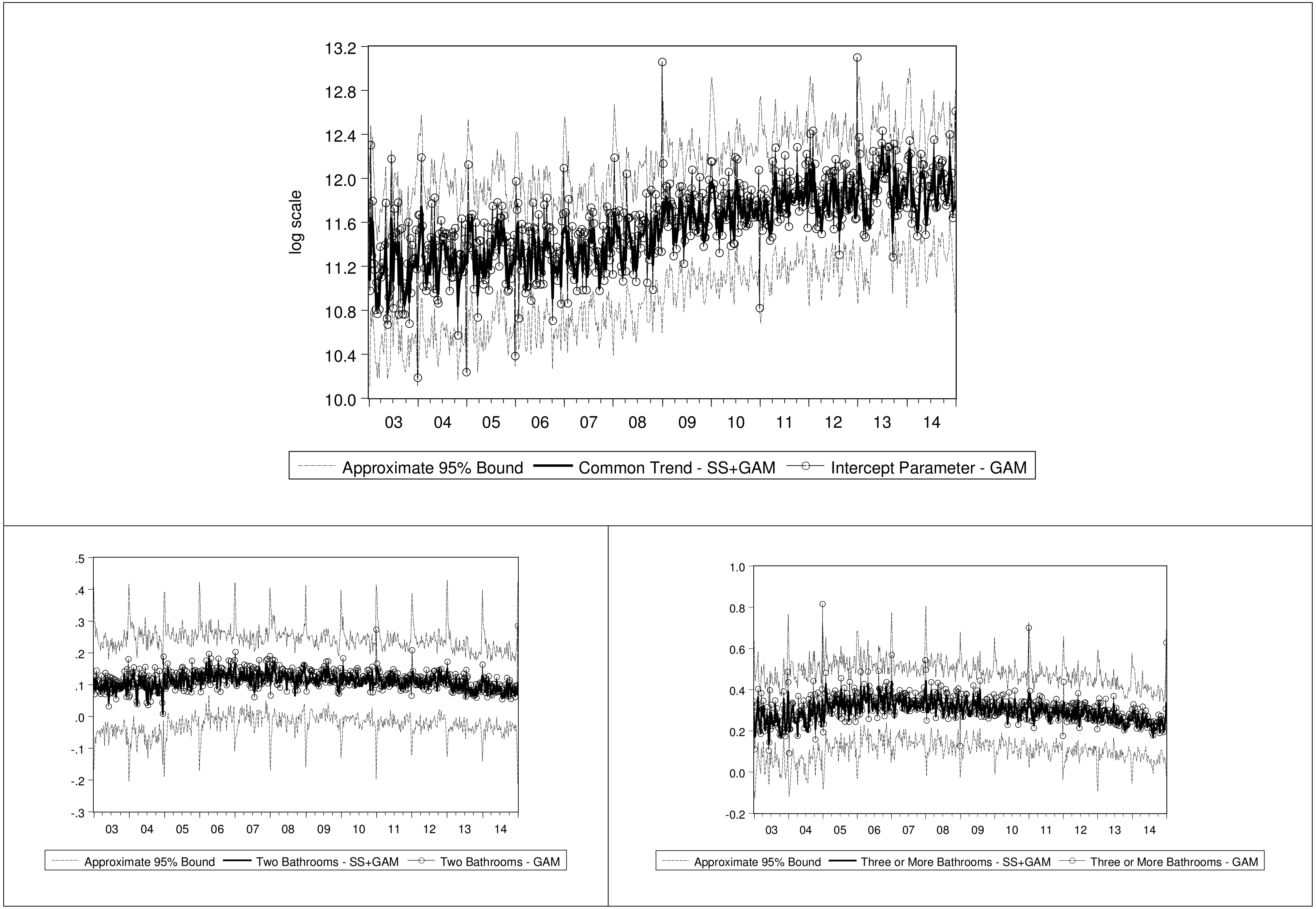


Figure 3: Parameter Plots (b)

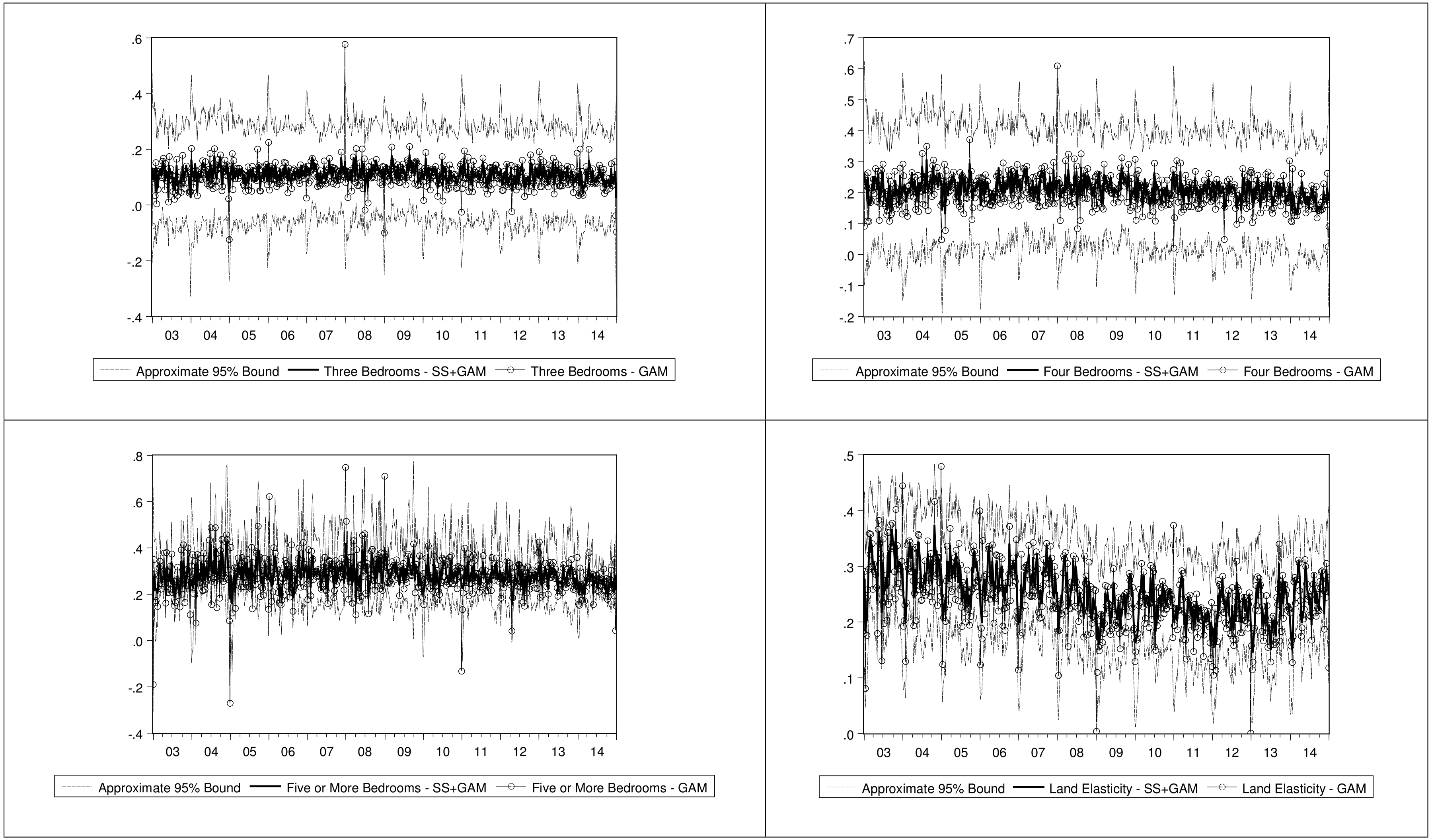
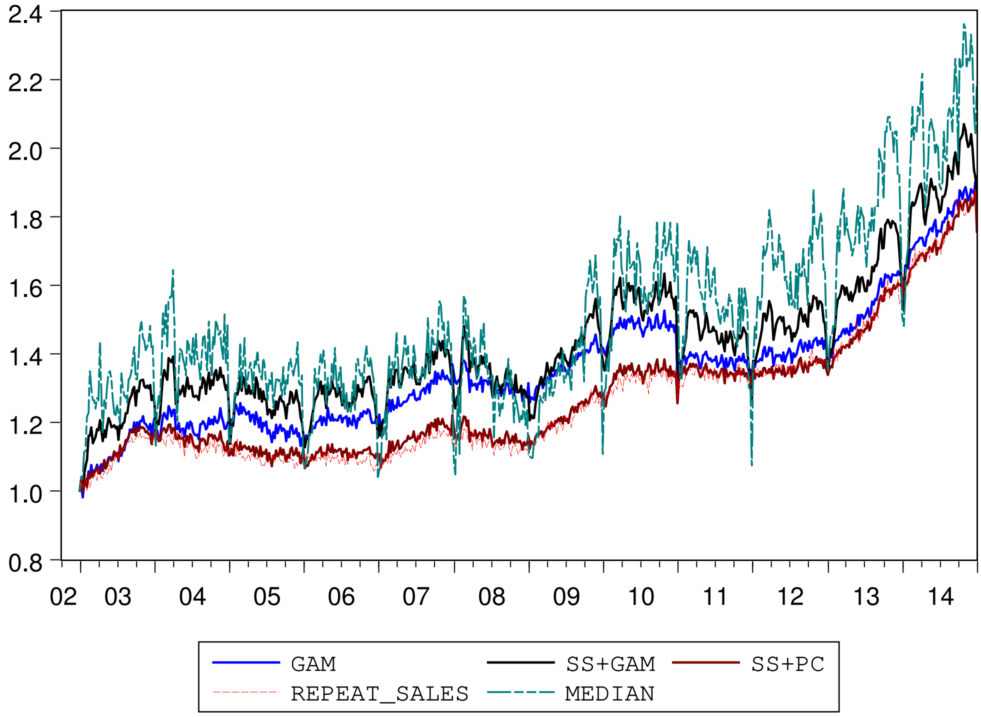
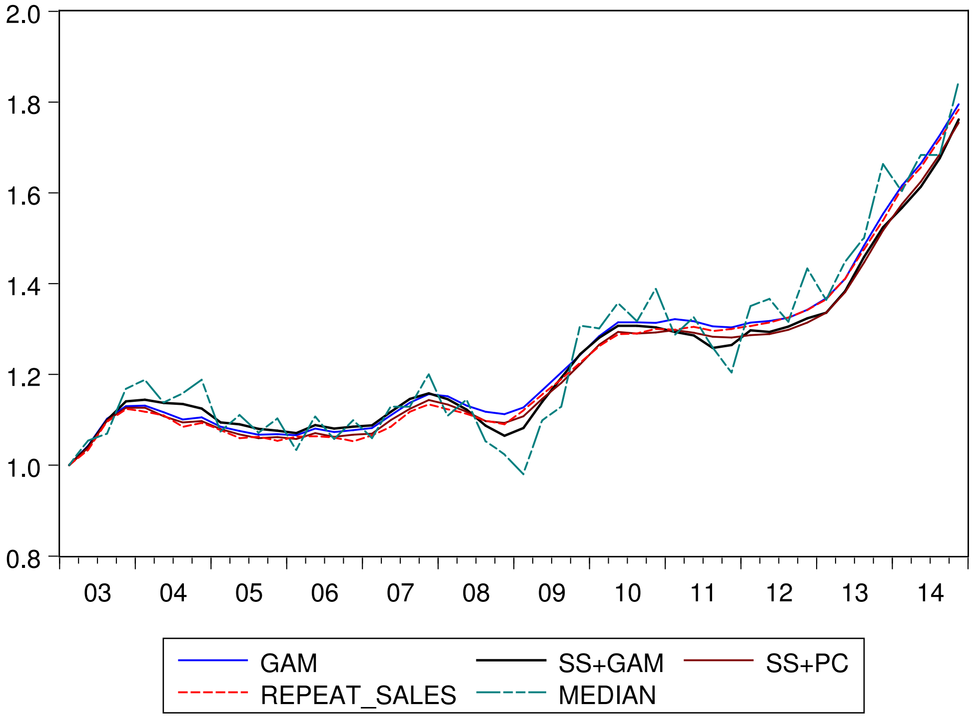


Figure 4: Weekly Property Price Indices from 2003 to 2014



Note: GAM is based on periodwise estimation of model (7); SS+PC is the state space model (22) with postcode dummies; SS+GAM is the spatio-temporal model; Repeat_Sales index is calculated using the Bailey, Muth, and Nourse (1963) formula; Median is the usual median index computed at a weekly frequency. Base: Week starting 30/12/2002 = 1

Figure 5: Quarterly Property Price Indices from 2003 to 2014



Note: GAM is based on periodwise estimation of model (7); SS+PC is the state space model (22) with postcode dummies; SS+GAM is the spatio-temporal model; Repeat_Sales index is calculated using the Bailey, Muth, and Nourse (1963) formula; Median is the usual median index computed at a quarterly frequency. Base: First Quarter 2003 = 1

Graz Economics Papers

For full list see:
http://ideas.repec.org/s/grz/wpaper.html
Address: Department of Economics, University of Graz,
Universitätsstraße 15/F4, A-8010 Graz

04–2018 **Robert J. Hill, Alicia N. Rambaldi, and Michael Scholz**: Higher Frequency Hedonic Property Price Indices: A State Space Approach

03–2018 **Reza Hajargasht, Robert J. Hill, D. S. Prasada Rao, and Sriram Shankar**: Spatial Chaining in International Comparisons of Prices and Real Incomes

02–2018 **Christoph Zwick**: On the origin of current account deficits in the Euro area periphery: A DSGE perspective

01–2018 **Michael Greinecker and Christopher Kah**: Pairwise stable matching in large economies

15–2017 **Florian Brugger and Jörn Kleinert**: The strong increase of Austrian government debt in the Kreisky era: Austro-Keynesianism or just stubborn forecast errors?

14–2017 **Jakob Mayer, Gabriel Bachner and Karl W. Steininger**: Macroeconomic implications of switching to process-emission-free iron and steel production in Europe

13–2017 **Andreas Darmann, Julia Grundner and Christian Klamler**: Consensus in the 2015 Provincial Parliament Election in Styria, Austria: Voting Rules,Outcomes, and the Condorcet Paradox

12–2017 **Robert J. Hill, Miriam Steurer and Sofie R. Waltl**: Owner Occupied Housing in the CPI and Its Impact On Monetary Policy During Housing Booms and Busts

11–2017 **Philipp Kohlgruber and Christoph Kuzmics**: The distribution of article quality and inefficiencies in the market for scientific journals

10–2017 **Maximilian Goedl**: The Sovereign-Bank Interaction in the Eurozone Crisis

09–2017 **Florian Herold and Christoph Kuzmics**: The evolution of taking roles

08–2017 **Evangelos V. Dioikitopoulos, Stephen J. Turnovsky and Ronald Wendner**: Dynamic Status Effects, Savings, and Income Inequality

07–2017 **Bo Chen, Yu Chen and David Rietzke**: Simple Contracts under Observable and Hidden Actions

06–2017 **Stefan Borsky, Andrea Leiter and Michael Paffermayr**: Product Quality and Sustainability: The Effect of International Environmental Agreements on Bilateral Trade

05–2017 **Yadira Mori Clement and Birgit Bednar-Friedl**: Do Clean Development Mechanism projects generate local employment? Testing for sectoral effects across Brazilian municipalities

04–2017 **Stefan Borsky, Alexej Parchomenko**: Identifying Phosphorus Hot Spots: A spatial analysis of the phosphorus balance as a result of manure application

03–2017 **Yu Chen, Yu Wang, Bonwoo Koo**: Open Source and Competition Strategy Under Network Effects

02–2017 **Florian Brugger**: The Effect of Foreign and Domestic Demand on U.S. Treasury Yields

01–2017 **Yu Chen**: Simplifiability of Contingency Plans with Applications to Optimal Mechanism Design

13–2016 **Jan de Haan, Rens Hendriks, Michael Scholz**: A Comparison of Weighted Time-Product Dummy and Time Dummy Hedonic Indexes

12–2016 **Katja Kalkschmied**: Complementary Institutional Elements and Economic Outcomes

11–2016 **Dieter Balkenborg, Josef Hofbauer, Christoph Kuzmics**: The refined best reply correspondence and backward induction

10–2016 **Karl W. Steininger, Wolf D. Grossmann, Iris Grossmann**: The economic value of dispatchable solar electricity: a Post-Paris evaluation

09–2016 **Sofie R. Waltl**: Estimating aggregate quantile-specific gross rental yields for residential housing in Sydney

08–2016 **Karl Farmer**: Financial integration and house price dynamics in equilibrium modeling of intra-EMU and global trade imbalances

07–2016 **Gabriel Bachner, Birgit Bednar-Friedl**: Counterbalancing the Effects of Climate Change Adaptation on Public Budgets: Factor Taxes, Transfers, or Foreign Lending?

06–2016 **Margareta Kreimer, Ricardo Mora**: Segregated Integration: Recent Trends in the Austrian Gender Division of Labor

05–2016 **Karl Farmer, Bogdan Mihaiescu**: Credit constraints and differential growth in equilibrium modeling of EMU and global trade imbalances

04–2016 **Daniel Eckert, Frederik Herzberg**: The birth of social choice theory from the spirit of mathematical logic: Arrow's theorem in the framework of model theory

03–2016 **Christoph Kuzmics, Jan-Henrik Steg**: On public good provision mechanisms with dominant strategies and balanced budget

02–2016 **Florian Gauer, Christoph Kuzmics**: Cognitive Empathy in Conflict Situations

01–2016 **Thomas Aronsson, Olof Johansson-Stenman, Ronald Wendner**: Redistribution Through Charity, and Optimal Taxation when People are Concerned with Social Status

09–2015 **Sofie R. Waltl**: Variation across price segments and locations: A comprehensive quantile regression analysis of the Sydney housing market

08–2015 **Karl Farmer, Bogdan Mihaiescu**: Pension systems and financial constraints in a three-country OLG model of intra-EMU and global trade imbalances

07–2015 **Javier López Prol, Karl W. Steininger**: Photovoltaic self-consumption regulation in Spain: profitability analysis and alternative regulation schemes.

06–2015 **Miriam Steurer, Caroline Bayr**: Quantifying Urban Sprawl using Land Use Data

05–2015 **Porfirio Guevara, Robert J. Hill, Michael Scholz**: Hedonic Indexes for Public and Private Housing in Costa Rica

04–2015 **Sofie R. Waltl**: A Hedonic House Price Index in Continuous Time

03–2015 **Ronald Wendner**: Do Positional Preferences for Wealth and Consumption Cause Inter-Temporal Distortions?

02–2015 **Maximilian Gödl, Christoph Zwick**: Stochastic Stability of Public Debt: The Case of Austria

01–2015 **Gabriel Bachner**: Land Transport Systems under Climate Change: A Macroeconomic Assessment of Adaptation Measures for the Case of Austria