ECONOMIC RECORD

THE ECONOMIC SOCIETY OF AUSTRALIA

THE ECONOMIC RECORD, 2009

House Price Measurement: The Hybrid Hedonic Repeat-Sales Method

CALLUM JONES

Economics Program, Business School, University of Western Australia, Perth WA, Australia

This article supplements James Hansen's (2009) recent article on the measurement of house prices in Australia. It provides a specific outline of the hybrid hedonic repeat-sales measure, a model which combines the hedonic and repeat-sales approaches examined in Hansen.

I Introduction

James Hansen (2009) recently applied regression measures of house prices to Sydney, Melbourne and Brisbane house price data. Hansen constructed price indexes from the hedonic and repeat-sales regression methods and evaluated these against a median, and a measure developed by Prasad and Richards (2006) controlling for the composition of broad types of houses in a sales sample (the mix-adjusted approach).

In his study, Hansen (2009) identifies the hybrid hedonic repeat-sales (hybrid) approach as a third regression measure of house prices. Introduced by Case and Quigley (1991), the hybrid approach, as its name suggests, combines the ideal features of the hedonic and repeat-sales measures. Although it was identified by Hansen, in that article a hybrid house price index was not constructed for comparison with the other two regression approaches and the mix-adjusted measure. This short note therefore contributes to the recently flourishing Australian literature on house price measurement by extending the examination in Hansen to the hybrid measure. The exposition is supported with an example.

JEL classification: C33

Correspondence: Callum Jones, University of Western Australia, 35 Stirling Highway, Crawley, Perth WA 6009, Australia. Email: jonescallum@gmail.com

II The Hybrid Approach

The building block for the hybrid model is the time-dummy hedonic regression, used in Hansen (2009):

$$p_h^t = \alpha + \boldsymbol{\beta}'\mathbf{x}_h^t + \boldsymbol{\gamma}'\mathbf{d}_h^t + \varepsilon_h^t, \tag{1}$$

where p_h^t is the price of house h selling in time t, $\boldsymbol{\beta} = \{\beta_k\}$ is a $K \times 1$ vector of implicit prices of the $\mathbf{x}_h^t = \{x_{k,h}^t\}$ vector of characteristic observations, γ is the $T \times 1$ vector of price changes and $\mathbf{d}_h^t$ is the $T \times 1$ vector of time-dependent dummy variables equal to 1 in the time period the house is sold and 0 otherwise.

In almost all detailed house price datasets are houses that sell once across the examined time period, and houses that sell two or more times across the time period – the repeat sales. Inherent in the repeat-sales data is useful information; assuming the house has not changed in characteristics and quality, the change in that house's observed price should reflect the pure price change of that house. Regressing those pure price changes can give a measure of general house price movements. This regression forms the repeat-sales methodology, which was also examined in Hansen (2009).

The key to understanding the contribution of the hybrid approach is to realise repeat-sales

doi: 10.1111/j.1475-4932.2009.00596.x

observations are not independent of each other because they involve the same property, and to separate out the hedonic regressions for single-sales $i = 1,\ldots,I$ and repeat-sales $j = 1,\ldots,J$:

$$p_i^t = \alpha + \boldsymbol{\beta}'\mathbf{x}_i^t + \boldsymbol{\gamma}'\mathbf{d}_i^t + \eta_i + e_i^t, \qquad (2a)$$

$$p_j^t = \alpha + \boldsymbol{\beta}'\mathbf{x}_j^t + \boldsymbol{\gamma}'\mathbf{d}_j^t + \eta_j + e_j^t, \qquad (2b)$$

$$p_j^\tau = \alpha + \boldsymbol{\beta}'\mathbf{x}_j^\tau + \boldsymbol{\gamma}'\mathbf{d}_j^\tau + \eta_j + e_j^\tau. \qquad (2c)$$

Note here the error term for house $h = \{i, j\}$ is decomposed into a time-independent specification error η_h and a white noise process e_h^t; in the hedonic regression of Equation (1) the error was expressed as ε_h^t. The hybrid contribution comes from differencing Equation (2b) from Equation (2c) to yield the following system of equations:

$$p_i^t = \alpha + \boldsymbol{\beta}'\mathbf{x}_i^t + \boldsymbol{\gamma}'\mathbf{d}_i^t + \eta_i + e_i^t, \qquad (3a)$$

$$p_j^t = \alpha + \boldsymbol{\beta}'\mathbf{x}_j^t + \boldsymbol{\gamma}'\mathbf{d}_j^t + \eta_j + e_j^t, \qquad (3b)$$

$$p_j^\tau - p_j^t = \boldsymbol{\gamma}'\boldsymbol{\delta}_j^\tau + e_j^\tau - e_j^t, \qquad (3c)$$

where $\boldsymbol{\delta}_j^\tau$ is a $T \times 1$ vector of time-dependent variables equal to 1 for the second of the pair of sales for house j, -1 for the first sale of the pair and 0 otherwise. Here, the error structure is updated as the unobserved specification errors are removed using repeat-sales observations (Quigley, 1995). This extracts the most information available in the data.

Assuming, for house $h = \{i, j\}, \mathrm{E}(e_h^t) = 0$, $\mathrm{var}(e_h^t) = \sigma_e^2, \mathrm{cov}(e_i^t, e_j^t) = \mathrm{cov}(e_h^t, e_h^\tau) = 0$ (white noise) and $\mathrm{E}(\eta_h) = 0, \mathrm{var}(\eta_h) = \sigma_\eta^2$ and $\mathrm{cov}(e_h^t, \eta_j) = 0$, the $(I + 2J) \times (I + 2J)$ covariance matrix of the error structure is:

$$\Omega = \begin{bmatrix} (\sigma_\eta^2 + \sigma_e^2)\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\sigma_\eta^2 + \sigma_e^2)\mathbf{I} & -\sigma_e^2\mathbf{I} \\ \mathbf{0} & -\sigma_e^2\mathbf{I} & 2\sigma_e^2\mathbf{I} \end{bmatrix}. \qquad (4)$$

Note that with the error assumptions we can write $\sigma_\varepsilon^2 = \sigma_\eta^2 + \sigma_e^2$, where σ_ε^2 is the variance of the error of regression (1). Conducting feasible generalised least squares on equation system (3a), (3b) and (3c) with an estimate of the covariance matrix (4) illustrates a hybrid framework.

To emphasise the value of the hybrid approach, it is worth noting the following.

Proposition 1. *In a dataset with a mix of properties that sell once or twice, if no mis-specifications are made in the hedonic equation, the hybrid estimation is equivalent to the hedonic estimation.*

Proof. *If no mis-specifications are made in the hedonic equation $\sigma_\eta^2 = 0$, the hybrid method is a non-singular transformation of the hedonic measure, making the estimators the same.*

Thus, because most hedonic studies introduce errors in hedonic specification, the hybrid method is a useful way to improve the precision with which time-dummy price indexes are calculated.

(i) Estimation Procedure

An estimation procedure for the hybrid model is outlined in this section. It was assumed that mis-specification errors are uncorrelated with the included variables. This opens up an interesting comparison between the hybrid model and the random effects model. Consistent estimators of components of the covariance matrix are based on the analysis of the random effects model by Greene (2003).

To derive $\hat{\sigma}_e^2$, a repeat-sales regression using repeat-sales properties for the time period analysed is conducted. This regression is equivalent to estimating Equation (3c) separately. Let the $n = 1, \ldots, N_{\mathrm{RS}}$ residuals of this regression be $\hat{\xi}_n$. From the error term of Equation (3c), $\sigma_e^2 = 1/2\mathrm{var}(\xi_n)$. Using the estimated residuals, an unbiased and consistent estimator with the degrees of freedom adjustment is (applying Greene, 2003):

$$\hat{\sigma}_e^2 = \frac{1}{2}\left(\frac{1}{N_{\mathrm{RS}} - T}\right)\sum_{n=1}^{N_{\mathrm{RS}}} \hat{\xi}_n^2. \qquad (5)$$

It can be shown that the estimate of $\hat{\sigma}_e^2$ is consistent. To estimate $\sigma_\varepsilon^2 = \sigma_\eta^2 + \sigma_e^2$, apply theorem 10.8 in Greene (2003) stating that the ordinary least squares estimators are consistent asymptotically, the ordinary least squares residuals $\hat{\varepsilon}_n$ of a hedonic regression (1) on all $n = 1,\ldots,N$ observations, single and repeated, may be used:

$$\hat{\sigma}_\varepsilon^2 = \left(\frac{1}{N - K - T - 1}\right)\sum_{n=1}^{N} \hat{\varepsilon}_n^2. \qquad (6)$$

Using $\hat{\sigma}_e^2$ and $\hat{\sigma}_\varepsilon^2$, an estimate of covariance matrix (4) is formed, and feasible generalised least squares estimators of the hedonic and time-dummy variables are estimated.

III Illustration: Mandurah House Prices

Here, a brief comparison of the hybrid approach with the other two regression methods

TABLE 1
Mandurah Price Index Statistics, 1994–2007

Measure (1)	Total % growth (2)	CI width (3)
Mean	343.6	—
Median	324.2	—
Repeat-sales	300.3	12.810
Hedonic	287.8	14.819
Hybrid	284.9	11.385

Note: CI, 95% confidence interval.

is conducted. It suffices to say here that 22,958 house sales observations for the southern Western Australian city of Mandurah across 1994–2007 were filtered and separated into single and repeat-sales pairs with hedonic characteristics; more information and tests using datasets of Geraldton and Perth are available on request.

Table 1 gives statistics on five price indexes calculated for Mandurah across 1994–2007. The basic criterion for precision used is the width of the confidence interval around the time-dummy coefficient estimated for each regression.

The price index estimated using the hybrid approach is the most accurately measured, according to the confidence band around the price index (column (3)). This result was expected in light of the exposition of the hybrid technique detailed in Section II.

IV Conclusion

This brief note provides a relevant extension to Hansen's (2009) paper on measuring house prices in Australia. It contributes to the Australian literature on house price measures by detailing a relatively underused regression approach, the hybrid hedonic repeat-sales approach. The hybrid measure is a conceptually appealing approach, and one which, with little additional effort to the regression methods examined in Hansen, yields gain in the precision with which a house price index is calculable. This conclusion is supported with an empirical illustration using West Australian data.

ACKNOWLEDGEMENTS

Generous financial support through the Brian Gray Scholarship Program organised by the Australian Prudential Regulation Authority is acknowledged, as is a C. A. Vargovic Bursary and write-up scholarship from the UWA Business School. The author thanks numerous others for their help, including, but not limited to, Prof. Ken Clements of the UWA Business School, Dr Katrina Ellis of the Australian Prudential Regulation Authority, Mr Stewart Darby of the Real Estate Institute of Western Australia, and Dr James Fogarty and Giri Parameswaran for their helpful comments. Any mistakes are the sole responsibility of the author.

REFERENCES

Case, K.E. and Quigley, J.M. (1991), 'The Dynamics of Real Estate Prices', *Review of Economics and Statistics*, **73**, 50–8.

Greene, W.H. (2003), *Econometric Analysis, international* 5th edn. Prentice Hall, New Jersey.

Hansen, J. (2009), 'Australian House Prices: A Comparison of Hedonic and Repeat-Sales Measures'. *Economic Record*, **85**, 132–45.

Prasad, N. and Richards, A. (2006), 'Measuring Housing Price Growth: Using Stratification to Improve Median-based Measures', Research Discussion Paper 2006-04, Reserve Bank of Australia.

Quigley, J. (1995), 'A Simple Hybrid Model for Estimating Real Estate Price Indexes'. *Journal of Housing Economics*, **4**, 1–12.