Australian School of Business Research Paper No. 2012 ECON 45

# Hedonic Price-Rent Ratios, User Cost, and Departures from Equilibrium in the Housing Market

Robert J. Hill
Iqbal A. Syed

# Hedonic Price-Rent Ratios, User Cost, and Departures from Equilibrium in the Housing Market

**Robert J. Hill**[a] **and Iqbal A. Syed**[b,*]

[a]Department of Economics, University of Graz, Universitätsstrasse 15/F4, 8010 Graz, Austria: robert.hill@uni-graz.at

[b]School of Economics, University of New South Wales, Sydney 2052, Australia: i.syed@unsw.edu.au

October 15, 2012

**Abstract:** Departures of the housing market from equilibrium can be detected by comparing the actual price-rent ratio with the user cost of owner occupying. Empirical implementation of this idea, however, is problematic for two reasons. First, the price-rent ratio needs to be quality adjusted. Second, the expected capital gain – an important input into the user cost formula – is not directly observable. Using a large data set for Sydney-Australia, we show how these problems can be resolved using hedonic methods. Otherwise the user cost approach can generate highly misleading results.

**Keywords:** Real estate; Housing market; Hedonic model; Price-rent ratio; Rental yield; Quality adjustment; User cost; Capital gains

**JEL Classification Codes:** C43; E01; E31; R31

* Corresponding author.

We thank Australian Property Monitors and the NSW Department of Housing for providing the data and the Australian Research Council for providing financial assistance for the research (DP0667209 and LP0884095). Preliminary versions of this paper were presented at the Economic Measurement Group Workshops, December 2010 and 2011, Sydney; Joint Vienna Macroeconomics Seminar, Institute for Advanced Studies, December 2011, Vienna; 53rd Annual Conference of the New Zealand Association of Economists, June 2012, Palmerston North; Econometric Society Australasian Meeting, July 2012, Melbourne; 41st Australian Conference of Economists, July 2012, Melbourne; and 32nd General Conference of the International Association for Research in Income and Wealth, August 2012, Boston. We thank the participants of these conferences for making helpful comments. We also thank Erwin Diewert, Kevin Fox, Jan de Haan, Tran Van Hoa, Glenn Otto and Nigel Stapledon for their valuable comments.

# 1 Introduction

Housing markets seem to be particularly prone to booms and busts. Recent events have also shown how the housing market can impact on the rest of the economy, as a bust in the US housing market precipitated a global financial crisis. It is particularly important therefore that policy makers and other market participants can detect departures from equilibrium before they become too extreme.

One way of detecting such departures is to compare the user cost of owner-occupying with the cost of renting. In equilibrium households should be indifferent between owner-occupying and renting. When the user cost of owner-occupiers is higher (lower) than the cost of renting, then the price-rent ratio is too high (low). Departures from equilibrium therefore can be detected by comparing the actual median price-rent ratio with the price-rent ratio derived from the user-cost equilibrium condition.

Attempts to apply the user-cost equilibrium condition to actual housing markets encounter two serious problems. First, the price and rent used to compute the price-rent ratio in the user cost formula should refer to the same dwelling or to dwellings of the same quality. If there is a quality difference between the sold and rented dwellings this invalidates the user-cost equilibrium condition.

Differences clearly exist between owner-occupied and rented dwellings. Rental dwellings are concentrated more at the lower end of the housing distribution. For example, according to the American Housing Survey (2001), 82 percent of owner-occupied dwellings are detached single-family homes, while the corresponding figure for rental dwellings is only 23 percent (see also Gallin 2008 and Heston and Nakamura 2009). Also, Shilling, Sirmans and Dombrow (1991) show that the owner-occupied houses are better maintained than the rented houses. It follows therefore that the median owner-occupied dwelling is generally of better quality than the median rented dwelling. An implication of this is that a price-rent ratio obtained by dividing the median price by the median rent is likely to be higher than its quality-adjusted counterpart. This means that applications of the user-cost equilibrium condition based on median prices and rents (see for example Hatzvi and Otto 2008) will be biased towards finding that the price-rent ratio is above its equilibrium level.

Many applications of the user-cost equilibrium condition compare price and rent indexes rather than median prices and rents. Notable examples include Leamer (2002), Himmelberg, Mayer and Sinai (2005), Girouard, Kennedy, Noord and André (2006), Gallin (2008), Verbrugge (2008), Campbell, Davis, Gallin and Martin (2009), Duca, Muellbauer and Murphy (2011), and Hiebert and Sydow (2011). As noted by Smith and Smith (2006), such an approach is also problematic.

> [T]he dwellings included in price indexes do not match the dwellings in rent indexes, so that the resulting comparison is of apples to oranges. The ratio of a home sale price index to a rent index can rise because the prices of homes in desirable neighborhoods increased more than did the rents of apartment buildings in less desirable neighborhoods. Or perhaps the quality of the average home in the price index has increased relative to the quality of the average property in the rent index. In any case, gauging fundamental value requires actual rent and sale price data, not indexes with arbitrary scales. (p. 7)

The problem is that, even if the price and rent indexes are themselves quality adjusted, the derived price-rent ratios may not be (see section 5.5). Also the use of price and rent indexes only allows comparisons between the change in the price-rent ratio and the change in its corresponding equilibrium level. At any point in time, therefore, we cannot answer the most fundamental question which is whether the price-rent ratio is above or below its equilibrium level or whether it is moving towards or away from equilibrium.

The apples and oranges problem implied by using separate price and rent indexes can be seen clearly in some applications (all of which use US data). Leamer (2002) constructs price-rent ratios by dividing a median house price index by the rent of shelter index from the consumer price index (CPI) produced by the Bureau of Labor Statistics. Himmelberg et al. (2005) divide a repeat-sales price index calculated for single-family houses obtained from the Office of Federal Housing Enterprise Oversight (OFHEO) – now part of the Federal Housing Finance Agency (FHFA) – by an index of annual average rents of two-bedroom apartments obtained from REIS (a real estate consulting firm). Gallin (2008) and Campbell et al. (2009) use the same FHFA repeat-sales price index as Himmelberg et al., and the tenant rent index (part of the rent of shelter index) from the CPI. Duca et al. (2011) divide the FHFA repeat-sales index by the rental fixed dwelling index from the personal consumption expenditure (PCE) price index produced by the Bureau of Economic Analysis.

The best way to avoid quality mismatches is to use price-rent ratios matched at the level of individual dwellings. In general this is not possible since dwellings sell and rent only at irregular intervals, which typically do not coincide. Household surveys also cannot get information on the actual price and rent of the same dwelling, since a household is either an owner-occupier or a renter.

In this paper we show how price-rent ratios at the level of individual dwelling can be estimated using hedonic methods. Two complications arise when implementing this method. First, some of our dwellings are missing one or more characteristics (e.g., the land area is missing). We deal with this missing data problem by estimating multiple versions of our hedonic model each of which has a different mix of characteristics. Prices and rents for each dwelling are imputed from the hedonic model with exactly its mix of characteristics. Second, our hedonic models suffer from an omitted variables problem. We show how the subsample of dwellings that both sell and rent during our sample period provide a benchmark against which omitted variables bias can be measured.

Applying our method to about 730,000 price and rent observations for Sydney, Australia over the period 2001 to 2009 we find that the median sold dwelling is systematically about 18 percent better than the median rented dwelling. The difference for owner-occupied versus rented dwellings is even larger (i.e., 26 percent), since some sold dwellings are subsequently rented. About half of this quality difference is attributed to the observed variables in our model and the remainder to the omitted variables.

The matched price-rent ratios at the level of individual dwellings provide us with a rich micro level data set that allows us also to explore some features of the cross-sectional distribution of price-rent ratios. We find that the quality-adjusted price-rent ratio is systematically higher at the upper end of the market than at the lower end. We also find that the price-rent ratio converged during the last years of the boom (which ended in early 2004) and then diverged again once prices started falling. We suspect that this result may hold more generally for other housing markets that turn from boom to bust.

The second problem with the user-cost equilibrium condition is that the equilibrium price-rent ratio is very sensitive to the assumed expected real capital gain on housing (see Verbrugge 2008 and Diewert 2009). An alternative and perhaps more promising approach is to assume that the housing market is in equilibrium and then impute the expected real capital gain implied by the user cost formula. Such an approach requires price-rent ratios in levels rather than price and rent indexes. This perhaps explains why this approach has not received attention in the literature. We find in the case of Sydney the average expected real capital gain is 3.8 percent per year. This is at the upper limit of what is plausible, indicating that the price-rent ratio may in fact be unsustainably high in Sydney. However, we find that since its peak in 2003–2004, the market has corrected itself substantially.

Our methodology and results have applications that extend beyond the main issues we address here. In particular, some of our findings have implications for the measurement of GDP. Failure to account for the cross-section variation in the price-rent ratio may result in the flow of housing services (and hence GDP) being mismeasured. More generally, hedonic methods are used to quality adjust many items in the CPI and components of GDP (e.g., apparel, televisions, video equipment, computers, software, photocopiers, audio and video, household appliances, rent, education writing equipment in the US—see Wasshausen and Moulton 2006). The problems of missing data and omitted characteristics are applicable also to some of these data sets.

The remainder of this paper consists of six sections. Section 2 explains the user-cost equilibrium condition. Section 3 develops our hedonic approach for computing price-rent ratios at the level of individual dwellings. Section 4 describes our data set, and then explains our methods for correcting for missing characteristics and omitted variables bias. Our empirical results are presented in section 5. Section 6 uses the quality-adjusted price-rent ratios as inputs in the user cost formula to check for departures from equilibrium. Our conclusions are discussed in section 7.

# 2 Equilibrium in the Housing Market

The user cost of a durable good is the present value of buying it, using it for one period and then selling it (see Hicks 1946). In equilibrium this should equal the cost of renting the good for one period.[1] Following Himmelberg et al. (2005) and Girouard et al. (2006), the equilibrium condition can be written as follows:

$$R_t = u_t P_t, \tag{1}$$

[1] This assumes households are not credit constrained. In the presence of such constraints, user cost may be less than rent (see Duca, Muellbauer and Murphy 2011). We return to this point in section 5.4. Also, adjustments to equilibrium may be slow due to the high level of transaction costs.

where $R_t$ is the period $t$ rental price, $P_t$ the purchase price, $u_t P_t$ is user cost, and $u_t$ the per dollar user cost. In a housing context, per dollar user cost can be calculated as follows:

$$u_t = r_t + \omega_t + \delta_t + \gamma_t - g_t, \tag{2}$$

where $r$ denotes the risk-free interest rate, $\omega$ is the property tax rate, $\delta$ the depreciation rate for housing, $\gamma$ the risk premium of owning as opposed to renting, and $g$ the expected capital gain. That is, an owner occupier foregoes interest on the market value of the dwelling, incurs property taxes and depreciation, incurs risk (mainly due to the inherent uncertainty of future price and rent movements in the housing market) and benefits from any capital gains on the dwelling.[2] If $R_t > u_t P_t$, owner-occupying becomes more attractive and hence this should exert upward pressure on $P$ and downward pressure on $R$ until equilibrium is restored. The converse argument applies when $R_t < u_t P_t$.

Rearranging (1), we obtain that in equilibrium the price-rent ratio should equal the reciprocal of per dollar user cost (i.e., $P_t/R_t = 1/u_t$). If the actual price-rent ratio exceeds our estimate of the reciprocal of per dollar user cost it follows that the housing market is not in equilibrium.

Practical application of this approach to the housing market is not straightforward for two reasons. First, the equilibrium condition (1) implicitly assumes that $P_t$ and $R_t$ are calculated for properties of equivalent quality. Suppose instead that the price $P_t$ refers to dwelling $A$ while the rent $R_t$ refers to dwelling $B$ and that dwelling $A$ is of superior quality to dwelling $B$. In this case, when a household is indifferent between buying and owner-occupying $A$ or renting $B$, we should expect that $R_t < P_t u_t$ and hence that $P_t/R_t > 1/u_t$.

The ratio of the median dwelling price to the median rent, which is perhaps the easiest way to obtain an average measure of price-rent ratios, suffers from exactly this kind of quality mismatch. The median owner-occupied dwelling will tend to be of superior quality to the median rental dwelling. (This is certainly true for our data set.) By implication, observed price-rent ratios calculated from unmatched medians should be higher than matched price-rent ratios. An analysis of the housing market based on comparisons of price-rent ratios and per dollar user costs will therefore be subject to systematic bias. In the next two sections,

[2] In some countries, owner-occupiers get the benefit from the tax deduction on the mortgage interest payments (see Girouard et al. 2006 for a list of OECD countries providing such benefits). For these countries, $r_t$ should be adjusted to include the offsetting tax benefit. However, no such benefit is provided to the owner occupiers in Australia.

we develop a methodology for calculating quality-adjusted price-rent ratios that correct for this bias.

The second problem with this user cost approach is that the expected capital gain $g$ is not directly observable. $g$ can be separated into two components: the expected real capital gain and the expected rate of inflation. Of these, the expected real capital gain is more problematic. A standard approach is to assume that the expected real capital gain is extrapolated from the past performance of the housing market. Some insight into the speed at which expected capital gains can adjust is provided by Case and Shiller's (2006) surveys of individuals in US cities. For example, Shiller (2007) describes how the median expected capital gain in Los Angeles was 10 percent in 2003, 5 percent in 2006 and then 0 percent in 2007 (as house prices began to fall). This suggests that households may be extrapolating over relatively short time horizons (such as the average capital gain over the preceding two years), as witnessed by the quite rapid decline in expected capital gains in Los Angeles as boom turned to bust. By implication $g$ and hence the equilibrium price-rent ratio $1/u$ may fluctuate a lot over time, thus potentially seriously undermining the usefulness of this particular application of the user-cost approach (see Verbrugge 2008 and Diewert 2009).

An alternative and probably more fruitful way of using the user-cost concept assumes that the housing market is in equilibrium, and then imputes the implied expected capital gain. To the best of our knowledge, this approach – first suggested by Diewert (1983) – has not been applied to data. This is perhaps because it requires price-rent ratios in levels which are difficult to obtain. If the imputed expected capital gain is deemed unrealistically high (low), then by implication we can conclude that the current price-rent ratio is too high (low). More specifically, rearranging the user cost formula in (2) and imposing the equilibrium condition in (1) yields the following:

$$g_t = r_t + \omega_t + \delta_t + \gamma_t - \frac{R_t}{P_t}. \tag{3}$$

Setting $R_t/P_t$ equal to the reciprocal of the median quality-adjusted price-rent ratio in period $t$ and inserting estimates of $r_t$, $\omega_t$, $\delta_t$ and $\gamma_t$, we obtain an estimate of $g_t$. An implausibly high $g$ indicates that the price-rent ratio is unsustainably high. We apply this approach to our Sydney data in section 6.2.

# 3 A Hedonic Approach to Constructing Quality-Adjusted Price-Rent Ratios

## 3.1 The hedonic imputation method

The hedonic method dates back at least to Waugh (1928) and Court (1939). It was, however, only after Griliches (1961) that hedonic methods started to receive serious attention (see Schultze and Mackie 2002 and Triplett 2006). The conceptual basis of the approach was laid down by Lancaster (1966) and Rosen (1974). The hedonic model is a reduced form equation which regresses the price of a product on a vector of characteristics (whose prices are not independently observed).

The hedonic approach can be implemented in different ways (see Triplett 2006 and Hill 2012 for surveys of the literature). However, in our context the most appropriate method is the hedonic imputation method where a separate hedonic model is estimated for each comparison period typically using a semilog functional form.[3]

$$y_t = X_t\beta_t + u_t, \tag{4}$$

where $y_t$ is an $H_t \times 1$ vector with elements $y_h = \ln p_h$ (where $H_t$ denotes the number of dwellings sold in period $t$), $X_t$ is an $H_t \times C$ matrix of characteristics (some of which may be dummy variables), $\beta_t$ is a $C \times 1$ vector of characteristic shadow prices, and $u_t$ is an $H_t \times 1$ vector of random errors. Examples of characteristics include the number of bedrooms, number of bathrooms, land area, and postcode.

Once the hedonic model has been estimated separately for each period, the prices of dwellings sold in one period can be imputed from the hedonic model of another period. For example, let $\hat{p}_{th}(x_{sh})$ denote the estimated price in period $t$ of a dwelling $h$ sold in period $s$. This price is imputed by substituting the characteristics of dwelling $h$ into the estimated hedonic model of period $t$ as follows:

$$\hat{p}_{th}(x_{sh}) = \exp(\sum_{c=1}^{C} \hat{\beta}_{ct} x_{csh}),$$

where $c = 1, \ldots, C$ indexes the set of characteristics included in the hedonic model. A Laspeyres-type hedonic index that compares periods $s$ and $t$ using the dwellings sold in

[3]Alternative functional forms, such as linear or Box-Cox transformations, are sometimes also considered. See Diewert (2003) and Malpezzi (2003) for a discussion of some of the advantages of semilog in a hedonic context and Diewert, Heravi and Silver (2009) for advantages of the hedonic imputation method.

period $s$ can now be constructed in one of two ways:

$$\begin{aligned}
\text{L1}: \quad P_{st}^{L1} &= \sum_{h=1}^{H_s} w_{sh} \left[\hat{p}_{th}(x_{sh})/p_{sh}\right] = \sum_{h=1}^{H_s} \hat{p}_{th}(x_{sh}) \Bigg/ \sum_{h=1}^{H_s} p_{sh} \\
\text{L2}: \quad P_{st}^{L2} &= \sum_{h=1}^{H_s} \hat{w}_{sh} \left[\hat{p}_{th}(x_{sh})/\hat{p}_{sh}(x_{sh})\right] = \sum_{h=1}^{H_s} \hat{p}_{th}(x_{sh}) \Bigg/ \sum_{h=1}^{H_s} \hat{p}_{sh}(x_{sh}) \, , \qquad (5)
\end{aligned}$$

where $w_{sh}$ and $\hat{w}_{sh}$ denote actual and imputed expenditure shares calculated as follows:

$$w_{sh} = p_{sh}/ \sum_{m=1}^{H_s} p_{sm}, \quad \hat{w}_{sh} = \hat{p}_{sh}(x_{sh})/ \sum_{m=1}^{H_s} \hat{p}_{sm}(x_{sm}).$$

In an analogous manner corresponding Paasche-type hedonic indexes that compare periods $s$ and $t$ using the dwellings sold in period $t$ can be constructed:

$$\begin{aligned}
\text{P1}: \quad P_{st}^{P1} &= \left\{ \sum_{h=1}^{H_t} w_{th} \left[p_{th}/\hat{p}_{sh}(x_{th})\right]^{-1} \right\}^{-1} = \sum_{h=1}^{H_t} p_{th} \Bigg/ \sum_{h=1}^{H_t} \hat{p}_{sh}(x_{th}) \\
\text{P2}: \quad P_{st}^{P2} &= \left\{ \sum_{h=1}^{H_t} \hat{w}_{th} \left[\hat{p}_{th}(x_{th})/\hat{p}_{sh}(x_{th})\right]^{-1} \right\}^{-1} = \sum_{h=1}^{H_t} \hat{p}_{th}(x_{th}) \Bigg/ \sum_{h=1}^{H_t} \hat{p}_{sh}(x_{th}) \, . \qquad (6)
\end{aligned}$$

A Fisher-type hedonic index, that treats periods $s$ and $t$ symmetrically, is obtained by taking the geometric mean of Laspeyres and Paasche:

$$\text{F1}: \quad P_{st}^{F1} = \sqrt{P_{st}^{L1} \times P_{st}^{L1}} = \sqrt{\frac{\sum_{h=1}^{H_s} \hat{p}_{th}(x_{sh})}{\sum_{h=1}^{H_s} p_{sh}} \times \frac{\sum_{h=1}^{H_t} p_{th}}{\sum_{h=1}^{H_t} \hat{p}_{sh}(x_{th})}}; \qquad (7)$$

$$\text{F2}: \quad P_{st}^{F2} = \sqrt{P_{st}^{L2} \times P_{st}^{L2}} = \sqrt{\frac{\sum_{h=1}^{H_s} \hat{p}_{th}(x_{sh})}{\sum_{h=1}^{H_s} \hat{p}_{sh}(x_{sh})} \times \frac{\sum_{h=1}^{H_t} \hat{p}_{th}(x_{th})}{\sum_{h=1}^{H_t} \hat{p}_{sh}(x_{th})}}. \qquad (8)$$

In the hedonic literature L1, P1 and F1 are referred to as single imputation price indexes, and L2, P2 and F2 as double imputation price indexes (see Triplett 2006 and Hill and Melser 2008). No clear consensus has emerged in the literature as to which approach is better. Single imputation uses less imputations and therefore is preferred by statistical agencies (see de Haan 2004). Double imputation may reduce omitted variables bias (see Hill and Melser 2008). We find that for our data set both the F1 and F2 price indexes and the F1 and F2 rent indexes are almost indistinguishable.

## 3.2 Hedonic price-rent ratios for individual dwellings

Here we apply the logic of the hedonic imputation method in a new context. Our objective is to compute a matched price-rent ratio for each individual dwelling. We achieve this by first estimating separate price and rent hedonic models. A price for each rented dwelling can then be imputed from the hedonic price model, and a rent for each sold dwelling imputed from the hedonic rent model. In this way a price-rent ratio can be calculated for each rented dwelling and each sold dwelling. A feature of this approach is that the hedonic price and rent models need to be defined on the same set of characteristics.

Papers that have implemented some of these steps include Arévalo and Ruiz-Castillo (2006), Kurz and Hoffmann (2009), Crone, Nakamura and Voith (2009) and Davis, Lehnert and Martin (2008). Among these papers, only Davis et al. consider estimation of price-rent ratios. Using US Census survey data, they impute rents for individual dwellings from a hedonic model. These rents are then matched with price estimates for these same dwellings obtained directly from the survey. The rent-price ratio is then averaged and interpolated from one Census benchmark to the next.

The methodological scope of our paper is broader than Davis et al. in that (as noted above) it estimates both price and rent hedonic models and then uses them to impute a rent for each dwelling sold and a price for each dwelling rented. Price-rent ratios at the level of individual dwellings are hence calculated using a double-imputation approach. More importantly, we develop extensions of our basic method to account for missing characteristics and omitted variables (see sections 4.2 and 4.3). Likewise, our empirical focus is broader in that we consider both the cross-section of price-rent ratios and the evolution of the average over time, and then insert the average into the user cost formula to detect departures from equilibrium.

Our starting point is the hedonic price equation, which is assumed to take the following form:

$$y_{Pt} = X_{Pt}\beta_{Pt} + u_{Pt}, \tag{9}$$

where $y_{Pt}$ is the vector of log prices of the dwellings sold in period $t$, and $X_{Pt}$ is the corresponding matrix of sold dwelling characteristics and $u_{Pt}$ is the random error term with zero mean and a constant variance.[4] Similarly, the hedonic rent equation is as follows:

$$y_{Rt} = X_{Rt}\beta_{Rt} + u_{Rt}, \tag{10}$$

where $y_{Rt}$ is the vector of log rents of the dwellings rented in period $t$, and $X_{Rt}$ is the corresponding matrix of rented dwelling characteristics. A rent for each dwelling $h$ sold in

[4]Spatial dependence in our model is captured through the inclusion of postcode dummies. With data on individual dwelling longitudes and latitudes, the spatial dependence could be modeled more rigorously using a spatial autoregressive model with autoregressive errors (see for example Badinger and Egger 2011).

period $t$ is imputed from (10) as follows:

$$\ln \hat{r}_{th} = \sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Pthc}, \tag{11}$$

where $c = 1, \ldots, C$ indexes the list of characteristics over which the price and rent hedonic models are defined. Similarly, a price for each dwelling $j$ rented in period $t$ is imputed from (9) as follows:

$$\ln \hat{p}_{tj} = \sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Rtjc}. \tag{12}$$

We can also use the hedonic rent equation to impute a rent for a dwelling $j$ actually rented in period $t$:

$$\ln \hat{r}_{tj} = \sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Rtjc}, \tag{13}$$

and the hedonic price equation to impute a price for a dwelling $h$ actually sold in period $t$:

$$\ln \hat{p}_{th} = \sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Pthc}. \tag{14}$$

Exponentiating, it follows that:[5]

$$\hat{r}_{th}(x_{Pth}) = \exp\left(\sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Pthc}\right),$$

$$\hat{p}_{tj}(x_{Rtj}) = \exp\left(\sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Rtjc}\right),$$

$$\hat{r}_{tj}(x_{Rtj}) = \exp\left(\sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Rtjc}\right),$$

$$\hat{p}_{tj}(x_{Pth}) = \exp\left(\sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Pthc}\right).$$

The distinction between single and double imputation arises again in the calculation of our hedonic price-rent ratios. A single imputation price-rent ratio $P/R(sold)^{SI}_{th}$ for a dwelling $h$ sold in period $t$ divides the actual price at which dwelling $h$ is sold by its imputed rent in period $t$ obtained from (11):

$$P/R(sold)^{SI}_{th} = \frac{p_{th}}{\hat{r}_{th}(x_{Pth})} = \frac{p_{th}}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Pthc}\right)}.$$

A corresponding double imputation price-rent ratio $P/R(sold)^{DI}_{th}$ divides the imputed price

[5] Strictly speaking, $\hat{r}$ and $\hat{p}$ are biased estimates of $r$ and $p$ since by exponentiating we are taking a nonlinear transformation of a random variable. Possible corrections have been proposed by Goldberger (1968), Kennedy (1981) and Giles (1982). From our experience, however, these corrections are small enough that they can be ignored.

for dwelling $h$ obtained from (14) by its imputed rent obtained from (11):

$$P/R(sold)_{th}^{DI} = \frac{\hat{p}_{th}(x_{Pth})}{\hat{r}_{th}(x_{Pth})} = \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Rthc}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Pthc}\right)}. \tag{15}$$

We can likewise generate two alternative matched price-rent ratios for each dwelling $j$ rented in period $t$. A single imputation price-rent ratio $P/R(rented)_{tj}^{SI}$ divides the imputed price for dwelling $j$ obtained from (12) by its actual rent:

$$P/R(rented)_{tj}^{SI} = \frac{\hat{p}_{tj}(x_{Ptj})}{r_{tj}} = \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Rtjc}\right)}{r_{tj}}.$$

Finally, a double imputation price-rent ratio $P/R(rented)_{tj}^{DI}$ divides the imputed price for dwelling $j$ obtained from (12) by its imputed rent obtained from (13):

$$P/R(rented)_{tj}^{DI} = \frac{\hat{p}_{tj}(x_{Rtj})}{\hat{r}_{tj}(x_{Rtj})} = \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Rtjc}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Rtjc}\right)}. \tag{16}$$

Empirically, we find that on average our double imputation price-rent ratios are 3.4 percent lower than their corresponding single-imputation counterparts (see Table 10). The choice between single and double imputation methods does not affect the general thrust of our results in section 5. The subsequent analysis focuses on the double imputation price-rent ratios.[6]

## 3.3 Median and quartile matched price-rent ratios

Let $Med[P/R(sold)^{DI}]$ denote the median price-rent ratio derived from the double-imputation price-rent distribution defined on the dwellings actually sold, while $Med[P/R(rented)^{DI}]$ denotes the corresponding median price-rent ratio defined on the dwellings actually rented. An overall median is obtained by averaging these two population specific medians as follows:

$$Med[P/R^{DI}] = \sqrt{Med[P/R(sold)^{DI}] \times Med[P/R(rented)^{DI}]}. \tag{17}$$

An alternative approach is to first pool the price-rent distributions defined on sold and rented dwellings and then calculate the median.

$$Med[P/R_{pooled}^{DI}] = Med[P/R(sold)^{DI}, P/R(rented)^{DI}]$$

[6] A robustness check is also conducted where the hedonic models in (9) and (10) are specified on price and rent levels instead of the log of prices and rents. In this case, (15) is replaced by $P/R(sold)_{th}^{DI} = \sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Rthc} / \sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Pthc}$ and (16) by $P/R(rented)_{tj}^{DI} = \sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Rtjc} / \sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Rtjc}$. The price-rent ratios obtained from the level models are on average about the same as those obtained from the log models (see Table 10).

Intuitively, we prefer the former approach (i.e. averaging rather than pooling) in (17) since it gives equal weight to the price and rent data sets. Empirically we find that the averaged and pooled medians are very close. A similar approach can be applied to any other quantile of the price-rent distribution. In particular, we compute lower and upper quartiles LQ and UQ as follows:

$$LQ[P/R^{DI}] = \sqrt{LQ[P/R(sold)^{DI}] \times LQ[P/R(rented)^{DI}]}, \quad (18)$$

$$UQ[P/R^{DI}] = \sqrt{UQ[P/R(sold)^{DI}] \times UQ[P/R(rented)^{DI}]}. \quad (19)$$

# 4 Data Sets and Empirical Strategy

## 4.1 The hedonic price and rental data sets

The data sets used in this paper are for Australia's largest city, Sydney, over the period 2001 to 2009. These are assembled from three sources. The data pertain to separate houses, where each house is built on its own piece of land. The data set on actual transaction prices is obtained from Australian Property Monitors (APM) and consists of a total of 395,110 observations over the 2001 to 2009 period.[7] The characteristics included in the data set are the transaction price, exact date of sale, land area, number of bedrooms, number of bathrooms, exact address and a postcode identifier. The rental data set is obtained by combining rental data from the New South Wales (NSW) Department of Housing (of which we have 331,877 observations) with data from APM (of which we have 89,495 observations that are not also in the NSW Housing data set). In total, therefore, we have 421,372 rental observations.[8]

A problem with the data sets is that there are many observations for which one or more characteristics are missing, even after filling in some missing values through the matching of addresses across the three data sets. In particular, all the characteristics are available for 61.67 percent of the price data and for 45.60 percent of the rental data (see Table 1). For

[7] APM provides real estate related research service and data for the Australian market. See http://apm.com.au in order obtain access to their data sets.

[8] While the recorded rents in the NSW Housing data refer to new rental contracts, the rents in the APM data refer to rents as advertised in the media. However, we find that there is virtually no difference between the actual and advertised rents when we test their mean difference on the houses which appeared in both data sets in a given quarter.

the remainder, at least one of the three characteristics of land area, number of bedrooms and number of bathrooms is missing. We explain in section 4.2 how we deal with this problem.

**Insert Table 1 Here**

With regard to the nature of the missing data, there are reasons to believe that values are randomly missing in the sense that a particular missing value is not related to the value itself. The original sources are government agencies (except for the APM rental data). The NSW Valuer-General's Office updates record of holding of properties and assesses the value of land which is used to determine property tax. The NSW Rental Bond Authority stores bonds (deposits provided by the tenants) and acts as an intermediary in the case of disputes between landlords and tenants. In each case the physical characteristics information are not so important. The key information for these agencies are addresses, prices, rents, bond amount, contract date, name of owners and renters). It does not benefit any party to withhold characteristic information (such as the number of bedrooms). Characteristic information can go missing both at the submission and data entry stage. APM, however, has supplemented the data with characteristic information obtained from other sources (such as newspapers, online housing databases and real estate offices). Similarly, APM's rental data are obtained from advertisements posted on online websites and at real-estate offices. While this process of supplementing the existing databases could in theory cause the missing characteristics to no longer be random, there is no particular reason to expect such an occurrence.[9]

The data sets are expected to provide a comprehensive picture of the purchase and rental markets in Sydney. It is mandatory for the parties to inform the State Valuer-General in the case of any change in the ownership of a property. The Rental Bond Authority obtains the information on new rental contract when the renter deposits the amount of bond with the agency. The authority does not charge any party for their service. While it is not mandatory, most new contracts are recorded with the Rental Bond Authority. Many of the contracts not lodged with the Authority are captured in the APM rental data.

Before proceeding with the estimation of our hedonic models, we removed some extreme observations (most of which are data-entry errors). The houses whose recorded prices, rents and land areas are located in the extreme 1.0 percent in both tails of the distributions are

[9] The percentage of missing characteristics has decreased (i.e. the quality of the data has improved) over the years in the sample. However, this pattern should not affect our results because we conducted our regression analysis separately for each year of the data.

deleted. In addition, houses with the number of bedrooms greater than 6 and the number of bathrooms greater than 5 are also deleted (these correspond to the 99.68 and 99.95 percentiles in the price data and 99.99 and 99.99 in the rent data). We had to undertake some further deletions in order to implement our hedonic approach since it requires that both price and rent models are specified on the same set of characteristics. For example, if the hedonic price model includes dwellings in a particular postcode, then the rental model must include dwellings rented in the same postcode.[10] In total, the deletion of extreme observations and the deletions due to the matching requirement led to the exclusion of 10.6 percent of observations from the total number of price and rental observations. This leaves us with 371,604 observations in the price data and 358,381 observations in the rental data (see Table 1 for detailed descriptions of the data sets).

Our expectation is that owner-occupied (and hence sold) dwellings on average are of better quality than rented dwellings.[11] This hypothesis is supported by the figures in Tables 1 and 2. From Table 1 it can be seen that the mean number of bedrooms and bathrooms and mean land area are all higher for sold dwellings than for rental dwellings. Table 2 compares the bedroom, bathroom, land area and locational distributions of the price and rental data. Of particular interest in Table 2 are the locational distributions. These were constructed by ranking the postcodes from cheapest to most expensive in terms of their median prices and median rents, and then allocating the postcodes to decile groups (i.e., the first decile is the cheapest and the tenth is the most expensive). From Table 2, it is clear that the rented

[10] If we had used a larger geographical area, such as 'local government area' instead of postcodes, we would have needed to delete fewer observations. However, using a larger area worsens the quality of the matches when adjusting for quality difference between sold and rented dwellings. We could also have tried matching at a lower level of aggregation, such as suburbs, where the suburbs typically cover smaller geographical regions than postcodes. The choice of postcode as the location-specific hedonic characteristic, however, is a natural one, partly because the presence of postcode is universal in addresses and also because postcodes are not prone to mismatches due to name abbreviations (which happens in the case of suburbs). With further improvement in data quality, matching at suburb level may in future become more feasible.

[11] After a house is sold it can be either occupied by the new owner or rented. ABS (2010) reports that the home-ownership rate in Australia remained stable at around 70 percent over the period 1971-2006. This indicates that 70 percent of the houses sold in each year can be expected to be occupied by the new owner. The home-ownership rate in Australia is similar to that of other countries including Canada, New Zealand, the European Union (EU) and the US (see AFTF 2007, Eurostat 2011, and Sinai and Souleles 2005).

dwellings are concentrated relatively more in the cheaper postcodes.[12]

**Insert Table 2 Here**

While these results support the hypothesis that sold dwellings are of better quality than rented dwellings, the quality differences are not that large. When imputing prices for rented dwellings from the price equation and rents for sold dwellings from the rent equation, the mean values of the characteristics corresponding to the predicted dwellings are quite close to the mean values of the characteristics that enter in the corresponding hedonic equations (see Table 1). These factors combined with our large sample size indicate that our approach of imputing prices for rented dwellings and rents for sold dwellings is viable.

## 4.2 Imputing prices and rents for dwellings with missing characteristics

Dwellings with missing characteristics are a common problem in housing data sets.[13] Instead of deleting these observations, we develop an alternative way of dealing with this issue that may be also applicable in other contexts (such as when estimating price and rent indexes and equivalent rent for owner-occupied houses) and to other data sets (such as electronics data used to construct quality-adjusted price indexes).

Our solution entails estimating a number of different versions of our basic hedonic price and rent equations. This allows the price and rent for each dwelling to be imputed from a hedonic equation that is tailored to its particular mix of available characteristics.

More specifically, focusing on the the case of the hedonic price equation, we estimate the following eight hedonic models (HM1,. . . ,HM8) for each year in our data set:

(HM1): *ln price = f(quarter dummy, land area, squared land area, num bedrooms, num bathrooms, postcode, land area & bedroom inter., land area & bathroom inter.)*

(HM2): *ln price = f(quarter dummy, num bedrooms, num bathrooms, postcode)*

(HM3): *ln price = f(quarter dummy, land area, squared land area, num bathrooms, postcode, land area & bathroom inter.)*

(HM4): *ln price = f(quarter dummy, land area, squared land area, num bedrooms, postcode, land area & bedroom inter.)*

[12] A more detailed description of the data set is provided in the supplementary material to this paper.

[13] For example, Crone, Nakamura and Voith (2009) mention that they experience this problem with the American Housing Survey (AHS) data set that they use.

(HM5): *ln price = f(quarter dummy, num bathrooms, postcode)*

(HM6): *ln price = f(quarter dummy, num bedrooms, postcode)*

(HM7): *ln price = f(quarter dummy, land area, postcode)*

(HM8): *ln price = f(quarter dummy, postcode)*

Each of these eight models is estimated using all the available dwellings that have at least these characteristics. For example, a dwelling for which land area, number of bedrooms and number of bathrooms are all available is included in all eight regressions. A dwelling that is missing the land area is included only in HM2, HM5, HM6, and HM8. A dwelling that is missing land area and number of bathrooms is included only in HM6 and HM8, etc.

The imputed price for each dwelling that is entered into (15) and (16), however, is only taken from the equation that exactly matches its list of available characteristics. This means that a dwelling for which all characteristics are available will have its price imputed from HM1. A dwelling that is missing only land area will have its price imputed from HM2. A dwelling missing land area and number of bathrooms will have its price imputed from HM6, etc.

The imputed rents are obtained in an analogous manner from 8 versions of the hedonic rent equation. If we had only estimated the HM1 model, then the price-rent ratios of a large number of dwellings could not have been calculated. Estimating multiple versions of our hedonic model allows us to calculate the price-rent ratio of every dwelling in the data sets.

## 4.3 Correcting for omitted variables bias

Omitted variables are a problem in all our hedonic models, even in HM1. The omitted variables may be physical (e.g., the quality of the structure, its energy efficiency, the general ambience, floor space, sunlight, the availability of parking, and the convenience of the floor plan), or locational (e.g., street noise, air quality and the availability of public transport links). The impact of some locational characteristics can sometimes be captured by locational dummies, as long as the geographical zones are sufficiently small. This should be the case for the postcode dummies used here (there are about 213 postcodes in Sydney). Many studies (such as Arévalo and Ruiz-Castillo 2006, Davis et al. 2008, Crone et al. 2009, and Kurz and Hoffmann 2009), however, use locational dummies defined on only a few divisions of a large metropolitan area. These zones are probably too big and heterogeneous to effectively capture locational effects.

Omitted variables may cause bias in our quality-adjusted price-rent ratios if the sold dwellings tend to perform better on the omitted variables than the rented dwellings. If so, our quality-adjusted price-rent ratios will be too high since they will fail to fully adjust for quality differences. We show later that this is exactly what seems to be happening in our data.

The required omitted variables bias correction will differ for each of our eight models. For example, it is bigger for HM8 than for HM1, since HM8 includes less explanatory variables.

The first step in correcting for omitted variables bias is to obtain reference quality-adjusted price-rent ratios that are free of omitted variables bias. This can be done by collecting dwellings that are both sold and rented over our sample period. We use a house price index and rent index to extrapolate forwards and backwards prices and rents on the same dwelling in different quarters. For example, suppose dwelling $h$ sells in period $s$ at the price $p_{sh}$ and is rented in period $t$ at the rate $r_{th}$. An address-matched price-rent ratio for this dwelling in period $s$ can be calculated by extrapolating the rental rate back to period $s$ using a rental index $R_{st}$ as follows:

$$P/R_{sh}^{AM} = \frac{p_{sh} \times R_{st}}{r_{th}}, \tag{20}$$

or by extrapolating the selling price forward to period $t$ using a price index $P_{st}$ as follows:

$$P/R_{th}^{AM} = \frac{p_{sh} \times P_{st}}{r_{th}}. \tag{21}$$

We now pool all the price-rent ratios derived using (20) and (21), and take the median for each period $t$:

$$P/R_{t}^{AM} = Med_{h=1,\ldots,H_t}[P/R_{th}^{AM}], \tag{22}$$

where $h = 1, \ldots, H_t$ indexes all the address-matched price-rent ratios in period $t$ in our data set. $P/R_{t}^{AM}$ should by construction be free of omitted variables bias.[14]

[14] For dwellings with multiple prices and rents in our sample, we select the chronologically closest price and rent observations to construct our address-matched price-rent ratio. Alternatively, we could consider each price-rent pair. For example, 12 address-matched price-rent ratios can be constructed from (20) and (21) for a dwelling that sold three times and rented twice in our data set. Our concern with this approach is that dwellings with multiple prices and rents may exert too much influence on (22). We try both approaches and find that they generate similar price-rent ratios (see Table 10). For dwellings that sell and rent in the same period, we count these price-rent ratios twice. Hence we have exactly two address-matched price-rent ratios for each dwelling that both sold and rented.

The price and rent indexes in (20) and (21) are calculated using the repeat-rent and repeat-sales index formulas. We use Calhoun's (1996) method, which corrects for heteroscedasticity by giving greater weight to repeats that are chronologically closer together (see also Hill, Melser and Syed 2009). Here we prefer repeat the rents/sales method over a hedonic method for computing the rent and price indexes since the former should be free of omitted variables bias. As a robustness check though we also estimate address-matched price-rent ratios using the price and rent indexes obtained from the double imputation hedonic (F2) method. The two approaches generate similar price-rent ratios (see Table 10).

With our methodology in place for constructing quality-adjusted price-rent ratios that are free of omitted variables bias, we can now compute bias correction factors for models HM1,...,HM8. We consider first the omitted variables bias of our HM8 model (which can be estimated over the largest data set since it contains the least characteristics). We calculate this as follows:

$$\lambda_{t,HM8} = \frac{HM8m(AMs_t)}{AMm(AMs_t)}, \tag{23}$$

where HM8m(AMs$_t$) denotes the median price-rent ratio obtained from (17) using the hedonic model HM8 applied to the address-matched sample (AMs) in period $t$. More precisely, we estimate the HM8 model over the HM8 price and rent data sets and then pick out the imputed price-rent ratios for dwellings in the address-matched sample (AMs). The median is then calculated only over the imputed price-rent ratios in the address-matched sample

The median in the denominator of (23) [i.e., AMm(AMs$_t$)] is calculated over the same address-matched sample as the median in the numerator. The difference now is that the imputed price-rent ratios are calculated by extrapolation from (20) and (21) using price and rent indexes rather than by using the HM8 hedonic model. The price and rent indexes themselves are calculated using the HM8 sample (i.e., the same sample that is used for the regressions in the numerator). Given it is derived from the sample of address-matched price-rent ratios, the median AMm(AMs$_t$) should be free of omitted variables bias.

The samples used to calculate the numerator and denominator of (23) are matched in two senses. First, prices and rents are imputed using the HM8 sample (using the HM8 hedonic model in the numerator and price and rent indexes in the denominator). Second, both medians HM8m(AMs$_t$) and AMm(AMs$_t$) are then calculated only over the address-matched samples. Any systematic deviation of $\lambda_{t,HM8}$ from 1 can therefore be attributable to omitted variables bias in the HM8m(AMs$_t$) median price-rent ratio. In our empirical results we find in every year that $\lambda_{t,HM8} > 1$, indicating that omitted variables bias is causing the price-rent ratios obtained from the HM8 model to be systematically too high.[15]

The omitted variables bias for each of our other models HMj (where $j = 1, \ldots, 7$) relative to HM8 is calculated as follows:

$$\lambda_{t,HMj|HM8} = \frac{HMjm(HMjs_t)}{HM8m(HMjs_t)}. \tag{24}$$

That is, we compare the median price-rent ratio obtained from HMj estimated over the HMj sample with the median price-rent ratio obtained from HM8 estimated over the HMj sample. We use HM8 as our reference hedonic model since it does not include any of land area, number of bedrooms, or number of bathrooms as characteristics. We can therefore impute the price-rent ratios of all dwellings in the AMs sample including those which do not provide information on physical characteristics. Furthermore, HM8 models are estimated over the whole data and hence using it as the reference hedonic method maximizes the sample size over which the omitted variables bias of each hedonic method is measured.

Given that the median imputed price-rent ratios $HMjm(HMjs_t)$ and $HM8m(HMjs_t)$ in (24) are calculated over the same sample of dwellings (i.e., the HMj sample), any systematic deviation of $\lambda_{t,HMj|HM8}$ from 1 can be attributed to omitted variables bias. While both $HMjm(HMjs_t)$ and $HM8m(HMjs_t)$ will be affected by omitted variables bias, our expectation is that the bias will be bigger for $HM8m(HMjs_t)$ than for $HMjm(HMjs_t)$ (for $j = 1, \ldots, 7$). The other models include more characteristics than HM8. Given our hypothesis that sold dwelling perform better than rental dwellings on these characteristics, it follows that $\lambda_{t,HMj|HM8}$ should be systematically less than 1. Our empirical results confirm this finding.[16]

Our estimate of the overall omitted variables bias of HMj is then given by:

$$\lambda_{t,HMj} = \lambda_{t,HM8} \times \lambda_{t,HMj|HM8}. \tag{25}$$

That is, first we calculate the omitted variables bias of HM8 (i.e., $\lambda_{t,HM8}$), and then we calculate the omitted variables bias of model HMj relative to that of HM8 (i.e., $\lambda_{t,HMj|HM8}$). The overall omitted variables bias of model HMj is then obtained by multiplying $\lambda_{t,HM8}$ by $\lambda_{t,HMj|HM8}$.

[15] As a robustness check, we also compute the denominator in equation (23) using hedonic imputation price and rent indexes. The results are quite similar (see Table 10).

[16] As a robustness check we also try using HM1 as the reference sample, where $\lambda_{t,HMj|HM8} = HMjm(HM1s)/HM8m(HM1s)$. Again this has little impact on the results (see Table 10).

Our expectation is that $\lambda_{t,HMj} < \lambda_{t,HM8}$ for $j = 1, \ldots, 7$ since as already noted each of these other models has less omitted variables. Applying the same logic we should also expect that:

$$\lambda_{t,HM1} < \lambda_{t,HM2} < \lambda_{t,HM5} < \lambda_{t,HM8};$$

$$\lambda_{t,HM1} < \lambda_{t,HM2} < \lambda_{t,HM6} < \lambda_{t,HM8};$$

$$\lambda_{t,HM1} < \lambda_{t,HM3} < \lambda_{t,HM5} < \lambda_{t,HM8};$$

$$\lambda_{t,HM1} < \lambda_{t,HM3} < \lambda_{t,HM7} < \lambda_{t,HM8};$$

$$\lambda_{t,HM1} < \lambda_{t,HM4} < \lambda_{t,HM6} < \lambda_{t,HM8};$$

$$\lambda_{t,HM1} < \lambda_{t,HM4} < \lambda_{t,HM7} < \lambda_{t,HM8}. \tag{26}$$

For example, taking the first of these inequalities, we have that HM2 is obtained by deleting land area from HM1. HM5 is then obtained from HM2 by deleting number of bedrooms. Finally, HM8 is obtained by deleting number of bathrooms.

We therefore adjust the price-rent ratio of a dwelling $h$ sold in period $t$ with the HMj mix of characteristics for omitted variables bias by dividing it by $\lambda_{t,HMj}$ as follows:

$$\begin{aligned} P/R(sold)^{adj}_{th,HMj} &= \frac{P/R(sold)_{th,HMj}}{\lambda_{t,HMj}} = \frac{P/R(sold)_{th,HMj}}{\lambda_{t,HMj|HM8} \times \lambda_{t,HM8}} \\ &= P/R(sold)_{th,HMj} \times \left(\frac{AMm(AMs_t)}{HM8m(AMs_t)}\right) \times \left(\frac{HM8m(HMjs_t)}{HMjm(HMjs_t)}\right). \end{aligned} \tag{27}$$

Similarly, a dwelling $j$ with the HMj mix of characteristics rented in period $t$ is adjusted for omitted variables bias as follows:

$$\begin{aligned} P/R(rented)^{adj}_{tj,HMj} &= \frac{P/R(rented)_{tj,HMj}}{\lambda_{t,HMj}} = \frac{P/R(rented)_{tj,HMj}}{\lambda_{t,HMj|HM8} \times \lambda_{t,HM8}} \\ &= P/R(rented)_{tj,HMj} \times \left(\frac{AMm(AMs_t)}{HM8m(AMs_t)}\right) \times \left(\frac{HM8m(HMjs_t)}{HMjm(HMjs_t)}\right). \end{aligned} \tag{28}$$

# 5 Empirical Results

## 5.1 The estimated hedonic models

We estimate our eight versions of the price and rent hedonic models, HM1–HM8, separately for each of the 9 years in the data set (altogether 144 regressions are run). Focussing on the HM1 model first, which is our most general model, Table 3 provides the average results of some key statistics for the 9 yearly regressions, separately for the prices and rents. The average adjusted R-squares for the price and rent models are 78.4 and 79.3 percent, respectively.[17] The postcode dummies explain 54.9 and 48.1 percent of the variations in the price and rent regressions, respectively. The next largest contribution is the group of physical characteristics, contributing 9.6 and 12.7 percent to the price and rent variations, respectively. The regression results also show that the percentage of significant coefficients is high, their economic significance is plausible and the directions implied by the estimated coefficients accord with our prior expectations. With some small variations in the exact numbers, these results generally hold separately for each of the 9 yearly regressions. Given this performance, our hedonic approach is expected to control for a large portion of the quality difference between sold and rented houses.[18]

**Insert Table 3 Here**

With regard to the regression results of the HM2–HM8 models, the explanatory power of these models falls as less characteristics are included (as expected), with the smallest model, HM8, explaining 63.9 and 62.8 per cent of the variation in prices and rents, respectively (see Table 4). Around 94.0 per cent of the signs of the estimated coefficients remain the same as the corresponding coefficients of the HM1 model. The premiums to an additional bedroom or bathroom or more land area are in most cases in HM2–HM7 higher than those found in the HM1 model. This is expected because the estimated coefficients in the HM2–HM7 models include a positive effect of the omitted characteristics. In summary, we find the performance of the HM2–HM8 models is stable across years and is as expected in relation to the HM1 model.[19]

**Insert Table 4 Here**

[17]The lowest adjusted R-square is 72.1 percent for the price model and 76.5 percent for the rent model. Our adjusted R-squares for the rent models are much larger than those reported by Arévalo and Ruiz-Castillo (2006), Crone et al. (2009) and Kurz and Hoffmann (2009).

[18]The models do not include interactions between number of bedrooms and number of bathrooms, since the inclusion of interactions between pairs of discrete variables would create problems when calculating our quality-adjusted price-rent ratios. Our hedonic approach requires that both the price and rent models are specified on the same set of characteristics. If a particular combination of characteristics, say 3 bedrooms and 2 bathrooms, is explicitly included in the hedonic models in the form of a dummy variable, then our approach requires that this combination is observed in both the sold and rental data. In many cases, the matching of characteristics at such a level of detail is not observed.

[19]See the supplementary material for more details.

## 5.2 Adjustments for omitted variables bias in our hedonic models

Our distributions of quality-adjusted price-rent ratios, from which medians and quartiles can then be calculated, are obtained by bringing together the price-rent ratios from our 8 models (HM1, HM2, . . . , HM8). However, as is explained in section 4.3, a different omitted variables adjustment is made to the imputed price-rent ratios of each model, prior to their pooling into a single data set.

A point of reference is provided by address-matched price-rent ratios, which directly control for quality differences. We have 42,153 dwellings in our data set for which we observe both prices and rents. We have a total of 49,388 selling prices for these dwellings (13.3 percent of the sold data) and 71,566 rents (20.0 percent of the rented data), respectively.[20] As shown in (20) and (21), the matching of time periods is attained by extrapolating the prices and rents over time (both backwards and forwards) using price and rent indexes.[21] The number of houses which are sold and rented more than once within the sample period are 38,612 and 81,017, respectively (corresponding to 81,568 price and 217,575 rent observations).

Table 5-column 2 provides estimates of the omitted variables bias of the price-rent ratios derived from the HM8 model [i.e., $\lambda_{t,HM8}$ derived from (23)]. Conforming to our expectations, we find that for every year $\lambda_{t,HM8} > 1$. The average $\lambda_{t,HM8}$ for 9 years is 1.115, implying that HM8 models fail to fully adjust for the quality difference between the sold and rented dwellings and, as a result, the price-rent ratios obtained from the HM8 model are on average 11.5 percent higher than those obtained from the address-matched model. Table 5 also provides estimates (see columns 3-9) of $\lambda_{t,HMj|HM8}$ in (24) for $j = 1, \ldots, 7$. Conforming to our expectations, these estimates are less than 1 (with only a few exceptions for individual years). This provides strong support for our hypothesis that sold dwellings perform better than the rented dwellings on the omitted variables. A model with more explanatory variables has less omitted variables and hence on average lower price-rent ratios.

The overall omitted variables bias $\lambda_{t,HMj}$ of model HMj is obtained by multiplying

[20] The number of observations is greater than the number of dwellings because of repeat-sales and repeat-rents. Only around 1500 of these matched houses were sold and rented in the same quarter.

[21] The average time span over which prices and rents are extrapolated is 2 and a quarter years, with 90 percent of the extrapolation done for less than 6 years (the larger the time span the less reliable is the extrapolation). Harding, Rosenthal and Sirmans (2007) report that the median time between two sales was 5 years for US data.

$\lambda_{t,HM8}$ by $\lambda_{t,HMj|HM8}$, as shown in (25). The estimates of $\lambda_{t,HMj}$ are broadly consistent with the inequalities in (26). While there are some slight inconsistencies for individual years, the average results for each model correspond exactly with (26).

**Insert Table 5 Here**

### 5.3 Quality-adjustment bias in price-rent ratios

Raw and quality adjusted price-rent ratios for the lower quartile, median and upper quartile for each year in our data set are shown in Table 6. As expected, the raw price-rent ratios are systematically larger than their quality adjusted counterparts, thus indicating that on average owner-occupied dwellings are of higher quality than rented dwellings. The raw price-rent ratio on average is 20.8 percent larger for the lower quartile, 18.4 percent larger for the median and 12.8 percent larger for the upper quartile. This suggests that dwellings with smaller price-rent ratios are more affected by quality adjustment bias.

In summary, sold dwellings on average are of 18.4 percent better quality than rented dwellings, and hence failure to quality adjust, will cause the median price-rent ratio to be too large on average by 18.4 percent.

If we had not adjusted for the omitted variables bias, the estimated quality difference would have been only 8.7 percent (see Table 10-column 1 for the estimates of the median price-rent ratios under this scenario). The difference is large, around 10 percent, and, therefore, has implications for studies that estimate equivalent rents or flow of services of owner-occupied houses (for example, see Arévalo and Ruiz-Castillo 2006, Crone et al. 2009 and Kurz and Hoffmann 2009).

**Insert Table 6 Here**

It is noticeable that the magnitude of this bias decreases significantly towards the end of our sample. One possible explanation for this finding is a fall in the average quality of dwellings sold during the financial crisis (which admittedly did not affect Australia as much as many other OECD countries), perhaps due to an increase in the number of distressed sales.

The average quality difference between owner-occupied and rented dwellings will be even higher than the difference between sold and rented dwellings, since some fraction of sold dwellings are subsequently rented. Suppose 70 percent of the dwellings sold in our data set are owner-occupied and 30 percent are rented (as is the case on average in Australia). Given that sold dwellings on average are of 18.4 percent better quality than rented dwellings, an estimate of the average quality difference between owner-occupied and rented dwellings is $(118.4 - 0.3 \times 100)/0.7 = 126.3$. In other words owner-occupied dwellings are on average of 26.3 percent better quality than rented dwellings.

## 5.4 Cross-sectional variation of price-rent ratios

We observe that the price-rent ratios increase steadily as we move from the lower to the upper end of the market. By pooling data across years, and by regressing the log of the estimated quality-adjusted price-rent ratio for sold dwellings against the log of prices, we find that the price-rent ratio increases by 0.21 percent for each percent increase in prices, and, similarly, by regressing the log of the quality-adjusted price-rent ratio for rented dwellings against the log of rents, we find that the price-rent ratio increases by 0.11 percent for each percent increase in rents (see Table 7). This pattern of a rising price-rent ratio can also be discerned as we move from cheaper to more expensive postcodes.

**Insert Table 7 Here**

The same pattern can be seen from a different perspective. Ordering all dwellings sold and rented each year from cheapest to most expensive, we then compute a quality-adjusted price-rent ratio for the lower quartile, median and upper quartile sold dwellings and likewise for the lower quartile, median and upper quartile rented dwellings (see Table 8). The quality-adjusted price-rent ratio in Table 8 is lowest for the first quartile, followed by the median, and is highest for the upper quartile. The results are very similar for the sold and rented dwellings. We find that the difference in the price-rent ratio between the third and first quartiles is 17.6 percent when the postcodes are ordered by median price, and 14.9 percent when ordered by median rent.

**Insert Table 8 Here**

This trend has been previously noted by Heston and Nakamura (2009) and Aten, Figueroa and Martin (2011). Heston and Nakamura for example find that the price-rent ratio rises by more than 50 percent as the price of a dwelling increases from \$50,000 to \$500,000. Their study uses survey data on four regions in the US (Alaska, the Caribbean, the Pacific and Washington D.C.) for 1990, where the rent and price of an owner-occupied dwelling is self-estimated by owners.[22]

[22] Heston and Nakamura further report that this cross-sectional variation in the price-rent ratio is not widely

We consider five explanations for this finding. A rising price-rent ratio implies that the rental yield (i.e., the reciprocal of the price-rent ratio) is lower at the top end. According to the user cost equilibrium condition, per dollar user cost should equal rental yield. If rental yield is lower at the top end this therefore implies that either per dollar user cost must be likewise lower at the top end or that the equilibrium condition does not hold for all segments of the market. The first two explanations below focus on the latter scenario, while the last three explanations focus on the former.

Following Diewert (2009), the first explanation argues that the owner of an expensive dwelling may wish to rent for temporary purposes or to rent to someone reliable who will maintain the property properly. The rent is offered at a discount in order to attract either lower income households or higher income households who would otherwise prefer to owner-occupy. To the extent this is true, it follows that rental yield will be lower than per dollar user cost at the top end of the market.

Second, households – particularly those with low incomes or wealth – may be credit constrained. They may prefer to buy than rent, but cannot get a large enough mortgage to do so. This will cause rental yield to be higher than per dollar user cost at the low end of the market. As is explained later, these first two explanations have implications for the measurement of GDP.

The third explanation is that the observed pattern for rental yield may be due to the depreciation rate being lower at the top end of the market. Structures depreciate while land does not, and the value of the land relative to the value of the structure is typically higher at the top-end of the market (see Diewert 2009, and Diewert, de Haan and Hendriks 2012). Himmelberg et al. (2005) make a similar argument in the context of comparisons of price-rent ratios across cities. They argue that the high value of land in San Francisco and New York should act to lower the per dollar user cost in these cities.

In order to see whether the cross-sectional difference in price-rent ratios can be explained by the difference in depreciation costs, we decompose house prices into structural and land components using a hedonic model. Other papers which have previously followed a similar approach include Bostic, Longhofer and Readfearn (2007), Diewert et al. (2012) and Rambaldi, McAllister, Collins and Fletcher (2011). In our context, we regress separately for each postcode and quarter the price of sold dwellings on the bedroom dummies, bathroom dum-

known outside official statistical circles.

mies, land area and squared land area. The regressions have no intercept term. This setup ensures that value is divided between the structure (represented by number of bedrooms and bathrooms) and land (i.e., none is attributed to a constant, quarters or postcodes). The squared land area is included to capture the non-linearity of land's contribution to the value of the property. The share of land in the price of each house is obtained by dividing the sum of the contribution of the land and the squared land area by the predicted price of that house. The remaining share is attributed to the structural components of each house.[23]

We pool our results across all postcodes and then regress separately for each year the log price-rent ratios on an intercept and the estimated structure shares in total value. The slope coefficients shown in Table 8 are negative for each year. Our findings therefore support the claim that part of the cross-section variation in price-rent ratios can be explained by variation in the depreciation rate.

The fourth explanation is that the risk premium is lower at the top end of the market, thus lowering the per dollar user cost. Some evidence to support this hypothesis is provided in Figure 1(a), where it is shown that prices at the low end of the market rose more at the end of the boom in Sydney and then fell more as boom turned to bust.

Fifth, the expected capital gain may be higher at the top end of the market (which again acts to lower the per dollar user cost). While this is unlikely to always be the case, the rise in inequality observed in many countries since the 1970s may be supporting such expectations over a sustained period.

## 5.5 Movement of ratios of price and rent indexes

Figure 1 shows the price and rent indexes obtained from the median, hedonic and repeats methodologies for the whole, lower and upper end of the market. The lower and upper end are defined here as the bottom and top 40 percentiles of postcodes, respectively, where these

[23] One important characteristic missing in this context is the age of the house, which Diewert et al. (2012) and Rambaldi et al. (2011) include in their models. Two houses may have the same floor space and land area, but because of depreciation, the value of the structure may be lower for the older houses. We find that for many postcodes the estimated hedonic coefficients are quite unstable across quarters and years. Diewert et al. (2012) report the same problem. They deal with it by imposing monotonicity restrictions on the hedonic parameters. Furthermore, for some postcodes we find either the contribution of the structures or land is negative. We delete these postcodes before estimating the elasticity of price-rent ratio and price to changes in the structural shares of dwellings reported in Table 8.

postcodes are ordered from the cheapest to the most expensive, separately for each year.

**Insert Figure 1 Here**

The three methodologies reveal some common themes. It can be seen that prices rose faster at the lower end during the boom (ending in 2004), after which they fell again. By contrast, at the upper end prices continued to rise after 2004. The movements of rents at the upper and lower end are more synchronized. Both rose throughout the sample, although at a faster rate at the lower end of the market.

Figure 1 highlights the potential distortions that can arise from using price and rent indexes to measure changes in the price-rent ratio, as is done frequently in the literature (see the discussion in section 1). For example, dividing a hedonic price index by a repeat-rent index, or a hedonic price index defined on the upper end of the market by a hedonic rent index defined on the lower end, may generate a distorted price-rent ratio series.

Figure 1 also sheds light on the convergence/divergence trend of the cross-section price-rent ratios. The variance of the log of the price-rent ratio in each quarter is graphed in Figure 2.[24] The dispersion of the price-rent ratios is U-shaped, with the minimum dispersion being observed in 2004. This corresponds to the peak of the boom in Sydney house prices. The U-shape in Figure 2 (i.e., $\sigma$ convergence followed by later divergence) can be attributed to the fact that the 1993-2004 housing boom in Sydney started at the top-end of the market (triggered by strong income growth at the top end and the scarcity of dwellings in prime locations) and then gradually rippled down to the low end. As a result, towards the end of the boom, the prices of lower quality dwellings rose faster than those at the top end thus causing price convergence. Price rises at the low end however were probably driven more by momentum than genuine scarcity. Also, buyers at the low end tended to have higher loan-to-value ratios. Hence when the boom ended prices fell at the low end, triggered partly by distressed sales, thus generating the subsequent price divergence. Meanwhile, the standard deviation of rents over this period was relatively stable. Combining these strands, it follows that price-rent ratios, like prices, first converged and then diverged, generating the U-shaped curve in Figure 2.

**Insert Figure 2 Here**

[24]The variance of the log of the price-rent ratio is scaled up by a factor of ten to fit in the Figure with the variances of price and rent.

## 5.6 Some implications for the measurement of GDP

Housing services are an important component of GDP. For example, imputed rent of owner-occupied housing accounts for about 8-9 percent of GDP in the US while tenant rent accounts for about 2-3 percent (see Grist 2010). Our findings have highlighted some of the difficulties that can arise when measuring the flow of housing services. In particular, imputing rent for owner-occupied housing by matching characteristics of rented and owner-occupied dwellings could, in the absence of an omitted variables adjustment, cause the service flow from owner-occupied housing to be significantly underestimated. In our case, we find a 13 percent quality difference between rented and owner-occupied dwellings matched on land area, number of bedrooms, number of bathrooms and postcode. Assuming owner-occupied housing's share in GDP is 8.5 percent, this translates into a downward bias in GDP of over 1 percent (i.e., $0.13 \times 0.085$). With less exact matching of characteristics the bias will be even larger.

Mismatches between per dollar user cost and rental yield may also have implications for GDP. According to the user cost condition, in equilibrium, per dollar user cost should equal the rental yield. When they are not equal, it is not clear which out of user cost and rent (actual or imputed) should be used to measure the flow of housing services. According to Diewert, Nakamura and Nakamura (2009), the value of housing services equals the maximum of rent and user cost (i.e., the opportunity cost). In this case, our results imply that the current methodology that equates the service flow to rent may overstate GDP during booms (where the price-rent ratio is typically above its equilibrium level and hence per dollar user cost exceeds the rental yield), and understate GDP during busts (where the reverse is true).

Over and above these variations over the business cycle, housing services may be over-estimated at the low end (where because of credit constraints rental yield may exceed per dollar user cost) and underestimated at the top end of the market (where because of a lack of high income renters per dollar user cost may exceed rental yield). The overall effect on the level of GDP of these cross-section variations therefore is ambiguous. However, the expansion of the subprime market in the US and other countries in the decade leading up to 2008 may have narrowed the gap between rental yield and per dollar user cost at the low end of the market by allowing more low wealth households to switch from renting to owner-occupying. An implication of the Diewert-Nakamura-Nakamura approach is that failure to account for this trend will impart an upward bias to the growth rate of GDP. The direction of the bias should have reversed since 2008 with the contraction of the subprime market.

# 6 Detection of Departures from Equilibrium

## 6.1 Equilibrium versus actual price-rent ratios

Our user cost equation in (2) contains the following variables:[25]

$r$ – the risk-free interest rate;

$\omega$ – the land tax rate;

$\delta$ – the depreciation rate for housing;

$\gamma$ – the risk premium of owning as opposed to renting;

$g$ – the expected nominal capital gain.

The values we use for these variables are shown in Table 9.

$r$ is the 10-year interest rate on Australian government bonds.[26] The value is updated each quarter. Our use of the 10-year rate rather than a 1-year rate can be justified as follows:

> Looking at the opportunity costs of owning a house from the viewpoint of an owner occupier, the relevant time horizon ... is the expected time the owner expects to use the dwelling before reselling it. This time horizon is typically some number between six and twelve years. (Diewert 2009, p. 494).

The 10-year government bond rate remained reasonably stable over the 2001-9 period, ranging between a minimum value of 4.1 percent in 2009 and a maximum value of 6.1 percent in 2008. (Source: Reserve Bank of Australia)

$\omega = 1.0$ percent. This is an estimate for an average land tax over the 2001-2009 period. (Source: Office of State Revenue, New South Wales, Australia)

[25]Mortgage interest payments are not tax deductible for owner-occupiers in Australia.

[26]Alternatively, we could have used the mortgage interest rate. Whether this is appropriate depends on the loan-to-value ratio of purchasers. The relevant interest rate for a purchaser with a 100 percent loan-to-value ratio is the mortgage interest rate $r^M$, while for a purchaser with a 0 percent loan-to-value ratio it is the risk-free 1-year rate $r^{rf}$. According to Green and Wachter (2005, Table 2), the average loan-to-value ratio in Australia is 63 percent. Assuming this figure remains constant, we could calculate $r$ as follows: $r^* = 0.37 \times r^{rf} + 0.63 \times r^M$. Interestingly, over our sample, $r^*$ and the 10-year interest rate are quite similar. On average $r^*$ is 0.09 percentage points higher. It follows that the choice between using the 10-year government bond rate and $r^*$ has virtually no impact on our results.

$\delta = 2.5$ percent. This is the gross depreciation rate estimated by Harding, Rosenthal and Sirmans (2007) using American Housing Survey data over the period 1983 to 2001. This is also the rate used by Himmelberg et al. (2005).

$\gamma = 2.0$ percent. This is the risk premium estimated by Flavin and Yamashita (2002) and used by Himmelberg et al. (2005).[27]

It should be noted that Girouard et al. (2006) fix $\delta + \gamma$ at 4 percent for the 18 OECD countries (including Australia) they studied over the 1990-2004 period. Verbrugge (2008) fixes $\omega + \delta + \gamma$ at 7 percent for the US over the 1980-2004 period. The values of our parameters lie in between these estimates.[28]

$g$ is the expected nominal capital gain which consists of the sum of the expected real capital gain and expected inflation.[29] The expected real capital gain in year $t$ is assumed to equal the moving average of real capital gain over the preceding $x$ years. We consider two different values of $x$ (i.e., 10 and 20 years).[30] More precisely, the expected capital gain in year $t$ is calculated as follows:

$$\text{Expected real capital gain}_t = \left( \frac{EHPI_t / CPI_t}{EHPI_{t-x} / CPI_{t-x}} \right)^{1/x}.$$

Here $EHPI_t$ is the level of the Established House Price Index and $CPI_t$ is the level of the consumer price index for Sydney in year $t$. Both the EHPI and CPI are computed by the

[27] The high price volatility relative to rent volatility in housing, implies a higher level of risk associated with home purchases. Han (2010) reports that the home price risk is one of the most significant risks that homeowners in the U.S. face. In our data set, we find that the standard deviation of the log of prices is higher than the standard deviation of the log of rents in every year and on average 20 percent higher over the whole sample period.

[28] The higher the values of these parameters, the more likely it is that the price-rent will be found to be above its equilibrium level.

[29] We need to separate expected real capital gains from inflation due to the change over time in the inflation environment in Australia. From 1981-1990 the average inflation rate in Sydney (as measured by the consumer price index) was 8 percent. By contrast, it was 2.6 percent from 1990-2000, and 3.1 percent from 2000-2009. The expected nominal capital gain on housing in the 1980s (most of which is inflation) must therefore have been much higher than in the 1990s and 2000s.

[30] Girouard et al. (2006) estimate that housing cycles in a sample of English-speaking countries (including Australia) last on average about 18 years. Hence extrapolation over 20 years may provide quite a good approximation of the long-run underlying trend.

Australian Bureau of Statistics (ABS).[31,32] For $x$=20 years, the expected annual real capital gain ranges from a peak of 5.0 percent in 2004 to a low of 1.8 percent in 2009 (see Table 9).

The expected rate of inflation is assumed to be 3 percent. This is very close to the average rate of inflation over the 2001-9 period which equalled 3.07 percent. It is also the upper bound on the Reserve Bank of Australia's inflation target (which is 2-3 percent).

**Insert Table 9 Here**

Inserting these values into (2) yields the values shown in Table 9 for the equilibrium price-rent ratio $1/u_t$ each year. The assumed time horizon of past performance over which expected capital gains are calculated plays a pivotal role. When the time horizon is 20 years, the equilibrium price-rent ratio ranges between 17.1 and 31.3. When it is 10 year, the range is much larger from 18.6 to 62.6. If the time horizon is reduced to 5 years, then in some years the equilibrium price-rent ratio is not even defined since the expected capital gain is large enough to make the user cost become negative.

The extreme volatility of per dollar user cost when expected capital gains are extrapolated from past performance over short time horizons has been noted previously by Verbrugge (2008) and Diewert (2007, 2009). Diewert (2009), citing evidence on the length of housing booms and busts from Girouard et al. (2006), argues that a longer time horizon (between 10 and 20 years) is more plausible in terms of how market participants form their expectations.

> [I]t is possible that landlords may have some idea of the long-run average rate of property inflation for the type of property that they manage, and this long-run average annual rate of price appreciation could be inserted into the user cost formula. (Diewert 2009, p. 494)

In the case of Sydney at least, to prevent excessive volatility in the equilibrium price-to-rent ratio, a 20-year horizon is preferable to a 10-year horizon.

[31] The Established House Price Index (EHPI) only goes back to 1986. To obtain prices back to 1981 (for the case where $x$=20), the EHPI was spliced together with an index calculated by Abelson and Chung (2005). See Stapledon (2007) for a discussion of why the Abelson and Chung series is probably the best available option for extending the EHPI back before 1986. In addition, the methodology underlying the EHPI changed slightly in 2005. Hence to obtain our full series, it was also necessary to splice together the pre and post 2005 EHPI series.

[32] The EHPI is computed using the stratified-median approach, which may fail to fully adjust for quality changes over time. Given the EHPI is probably the most widely followed house price index for Sydney, it nevertheless is a useful benchmark for describing expectations of capital gains.

Equilibrium price-rent ratios are compared with actual quality-adjusted and unadjusted price-rent ratios in Table 9. The quality-adjusted price-rent ratio is above its equilibrium level (when $x$=20) in 2001, 2002 and 2009, and below its equilibrium level in 2003, 2004, 2005, 2006, 2007 and 2008. By contrast, the quality-unadjusted price-rent ratio is above its equilibrium level for every year in the sample.

## 6.2 Imputed expected capital gains assuming the housing market is in equilibrium

An alternative way of using the user cost formula suggested by Diewert (1983), provided one has data on actual price-rent ratios in levels, is to assume that the housing market is in equilibrium. The implied expected capital gain can then be imputed using the method described in Section 2. If the resulting implied expected capital gain is unrealistically high (low) then it follows that the price-rent ratio is too high (low).

Substituting the values for $r_t$, $\omega$, $\delta$, and $\gamma$ from Table 9 and the quality-adjusted median price-rent ratios from Table 6 into (3), yields the expected capital gains series shown in Table 9. On average, the expected annual capital gain is 6.83 percent over the sample period. Assuming an expected inflation rate of 3 percent (the average for our sample is 3.07 percent), this implies an average expected real capital gain of 3.83 percent per year. If we had failed to quality-adjust then the impled average expected real capital gain would have been 4.42 percent.

Is this figure (i.e., 3.8 percent) realistic? Gyourko, Mayer and Sinai (2006) find that the average annual real capital gain for the 50 US cities in their sample over the period 1950 to 2000 was 1.7 percent, with the highest result of 3.5 percent being observed for San Francisco. There are in fact a number of similarities between San Francisco and Sydney, ranging from desirable coastal locations and scarcity of land to population growth. In this sense San Francisco is perhaps not a bad benchmark for Sydney. Nevertheless, these figures suggest that an expected real capital gain of 3.83 per cent is at the upper limit (if not beyond that) of what can be believed as realistic.

By comparison, based on the Established House Price Index (EHPI), the average real capital gain in Sydney per year over the following periods was: Dec 1980-Dec 2009: 2.77 percent; Dec 1989-Dec 2009: 2.97 percent; Dec 2000-Dec 2009: 4.42 percent; Dec 2004-Dec

2009: -0.02 percent. While the performance of the Sydney housing market over the period of our data set (i.e., 2001-2009) has exceeded the implied expected capital gain obtained from the user cost formula, over the last five years the real capital gain has been negative. Given the current state of the largest economies of the world it is hard to believe that Sydney can sustain a real capital gain of 3.83 percent per year.[33] In the final year of our sample (i.e., 2009), however, the implied real expected capital gain falls to 1.96 percent (see Table 9), due to the 2 percentage point fall in the 10-year interest rate.

Our two approaches to using the user cost formula hence lead us to somewhat contradictory conclusions. Our actual quality-adjusted price-rent ratios are quite close to their equilibrium levels throughout our sample period. By contrast, the expected real capital gain implied by market equilibrium is high enough to suggest that the price-rent ratio – at least until 2009 – may have been above its equilibrium level. This apparent contradiction arises from the very strong real capital gains experienced over the 1992-2004 period, which act to raise the equilibrium price-rent ratio in Table 9 to an unusually high level (even when $x = 20$).

In our opinion the most useful way of using the user cost formula is to compute the expected real capital gain implied by market equilibrium. From this perspective, our conclusion is that while the price-rent ratio was unsustainably high for most of our sample, this is no longer true in 2009. The market has gone through a gradual correction process since 2004. This can be attributed to the combination of stable or falling prices since 2004 accompanied by a steady rise in rents leading to a gradual fall in the price-rent ratio, and a fall in the 10-year interest rate.

**Insert Table 10 Here**

# 7 Conclusion

We find that the ratio of median price to median rent overstates the quality-adjusted price-rent ratio by 18 percent in Sydney. The quality difference between the owner-occupied and rented medians is even larger (i.e., 26 percent). Quality mismatches between the dwellings in price and rent indexes likewise distort movements in the price-rent ratio when the movements are estimated from their ratios.

[33] World Output grew by 5.3 percent in 2010 and 3.9 percent in 2011, and is projected to grow by 3.5 percent in 2012 and 3.9 percent in 2013 (see IMF 2012).

We use a hedonic approach to quality-adjust price-rent ratios. As part of this process an omitted variables bias correction must be made to account for the fact that sold dwellings perform better than rented dwellings on the omitted characteristics. This omitted variables correction contributes about half of the overall adjustment.

Our results have implications for the measurement of GDP. Our findings on omitted variables bias demonstrate the difficulties that can arise when imputing rent for owner occupied housing, which in the US for example accounts for 8-9 percent of GDP. Also, we show how housing services (and hence GDP) may be mismeasured when rent (actual or imputed) does not equal user cost. This can happen when the housing market is out of equilibrium, and at the low and top end of the market even in equilibrium. Given the large share of housing services in GDP, the impact on GDP of these factors could be quite large.

Finally, using our quality-adjusted price-rent ratios we find that the price-rent ratio in Sydney was above its equilibrium level for most of our sample period. However, it went through a correction process from its peak in 2004 and by the end of our sample in 2009 had returned to equilibrium.

# References


Abelson, P. and D. Chung (2005), "The Real Story of Housing Prices in Australia from 1970 to 2003," *Australian Economic Review* 38(3), 265-81

ABS (2010), Measures of Australia's Progress, 2010, Cat. No. 1370.0, Australian Bureau of Statistics, Canberra.

AFTF (2007), "Challenges and Directions for Australia's urban and Regional Future: Identifying the Key Issues," The Australia Futures Task Force (AFTF), Australian Research Council Network in Spatially Integrated Social Sciences. Available at: `www.siss.edu.au/siss/documents/AFTF_Task_Force_Report.pdf`

American Housing Survey (2001). Washington DC: US department of Commerce.

Arévalo, R. and J. Ruiz-Castillo (2006), "On the Imputation of Rental Prices to Owner-Occupied Housing," *Journal of the European Economic Association* 4(4), 830-861.

Aten, B., E. Figueroa and T. Martin (2011), "How Can the American Community Survey (ACS) Be Used to Improve the Imputation of Owner-Occupied Rent Expenditures? Available at: `www.bea.gov/papers/pdf/WP_ACS_OORE_020112.pdf`

Badinger, H. and P. Egger (2011), "Estimation of Higher-Order Spatial Autoregressive Cross-Section Models with Heteroscedastic Disturbances," *Papers in Regional Science* 90(1), 213-235.

Bostic, R. W., S. D. Longhofer and C. L. Redfearn (2007),"Land Leverage: Decomposing Home Price Dynamics," *Real Estate Economics* 35(2), 183-208.

Calhoun, C. A. (1996), OFHEO House Price Indexes: HPI Technical Description, Office of Federal Housing Enterprise Oversight, Washington, DC.

Campbell, S. D., M. A. Davis, J. Gallin and R. F. Martin (2009), "What Moves Housing Markets: A Variance Decomposition of the Rent-Price Ratio," *Journal of Urban Economics* 66(2), 90102.

Case, K. E., and R. J. Shiller (2006), "Home Buyer Survey Results 1988-2006," Unpublished Paper, Yale University.

Crone, T. M., L. I. Nakamura and R. P. Voith (2009), "Hedonic Estimates of the Cost of Housing Services: Rental and Owner Occupied Units," in Diewert, W. E., B. M. Balk, D. Fixler, K. J. Fox and A. O. Nakamura (eds.), *Price and Productivity Measurement: Volume 1 - Housing*. Trafford Press, Chapter 4, 51-68.

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

Davis, M. A., A. Lehnert and R. F. Martin (2008), "The Rent-Price Ratio for the Aggregate Stock of Owner-Occupied Housing," *Review of Income and Wealth* 54(2), 279-284.

de Haan, J. (2004),"Hedonic Regression: The Time Dummy Index As a Special Case of the Imputation Törnqvist Index," Paper Presented at the 8th Ottawa Group Meeting, August 23–25, 2004, Helsinki.

Diewert, W. E. (1983), "The Theory of the Cost of Living Index and the Measurement of Welfare Change," pp. 163-233 in Price Level Measurement, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada (reprinted as pp. 79-147 in Price Level Measurement, W.E. Diewert (ed.), Amsterdam: North-Holland, 1990.)

Diewert, W. E. (2003), "Hedonic Regressions: A Review of Some Unresolved Issues," Paper Presented at the Seventh Meeting of the Ottawa Group, Paris, 27–29 May.

Diewert, W. E. (2007), "The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions," Discussion Paper 07-01, Department of Economics, University of British Columbia, Vancouver.

Diewert, W. E. (2009), "Durable and Owner-Occupied Housing in a Consumer Price Index," in Diewert W. E., J. S. Greenless and C. R. Hulten (eds.), *Price Index Concepts and Measurements*. NBER: University of Chicago Press, Chapter 12, 445-500.

Diewert, W. E., J. de Haan and R. Hendriks (2012), "Hedonic Regressions and the Decomposition of a House Price Index into Land and Structure Components," *Econometric Reviews*, forthcoming.

Diewert, W. E., S. Heravi and M. Silver (2009),"Hedonic Imputation versus Time Dummy Indexes," in Diewert W. E., J. S. Greenless and C. R. Hulten (eds.), *Price Index Concepts and Measurements*. NBER: University of Chicago Press, Chapter 4, 161-196.

Diewert, W. E., A. O. Nakamura and L. I. Nakamura (2009),"The Housing Bubble and a New Approach to Accounting for Housing in a CPI," *Journal of Housing Economics* 18, 156-171.

Duca, J. V., J. Muellbauer and A. Murphy (2011), House Prices and Credit Constraints: Making Sense of the US Experience, *The Economic Journal* 121, 533-551.

Eurostat (2011), Housing Statistics. Available at: `http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Housing_statistics`

Flavin, M. and T. Yamashita (2002), "Owner-Occupied Housing and the Composition of the Household Portfolio," *American Economic Review* 92(1), 345-362.

Gallin, J. (2008), "The Long-Run Relationship Between House Prices and Rents," *Real Estate Economics* 36(4), 635-658.

Giles, D. E. A. (1982),"The interpretation of dummy variables in semilogarithmic equations," *Economics Letters*, 10, 7779.

Girouard, N., M. Kennedy, P. van den Noord, and C. André (2006), "Recent House Price Developments: The Role of Fundamentals," OECD Economics Department Working Papers 475. Paris: OECD.

Goldberger, A. S. (1968), "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica* 35, 464-472.

Green, R. K. and S. M. Wachter (2005), "The American Mortgage in Historical and International Context," *Journal of Economic Perspectives* 19(4), 93-114.

Griliches, Z. (1961), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in *The Price Statistics of the Federal Government*, G. Stigler (chairman). Washington D.C.: Government Printing Office.

Grist P. (2010), "Housing and GDP," HousingEconomics.com

Gyourko, J., C. Mayer and T. Sinai (2006), "Superstar Cities," NBER Working Paper 12355.

Han, L. (2010),"The Effects of Price Risk on Housing Demand: Empirical Evidence from U.S. Markets," *Review of Financial Studies*, 23 (11), 3889–3928.

Harding, J. P., S. S. Rosenthal, and C. F. Sirmans (2007), "Depreciation of Housing Capital, Maintenance, and House Price Inflation: Estimates from a Repeat Sales Model," *Journal of Urban Economics* 61(2), 193-217.

Hatzvi, E. and G. Otto (2008),"Prices, Rents and Rational Speculative Bubbles in the Sydney Housing Market," *Economic Record* 84 (267), 405-420.

Heston, A. and A. O. Nakamura (2009),"Reported Prices and Rents of Housing: Reflections of Costs, Amenities or Both?" in Diewert, W. E., B. M. Balk, D. Fixler, K. J. Fox and A. O. Nakamura (eds.), *Price and Productivity Measurement: Volume 1 - Housing*. Trafford Press, Chapter 7, 117-124.

Hicks, J. R. (1946), *Value and Capital*, Second Edition. Oxford: Clarendon Press.

Hiebert, P. and M. Sydow (2011), "What Drives Returns to Euro Area Housing? Evidence from a Dynamic DividendDiscount Model," *Journal of Urban Economics* 70, 8898.

Hill, R. J. (2012), "Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy," *Journal of Economic Surveys*, forthcoming.

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

Hill, R. J., D. Melser and I. Syed (2009), "Measuring a Boom and Bust: The Sydney Housing Market 2001-2006," *Journal of Housing Economics* 18(3), 193-205.

Himmelberg, C., C. Mayer and T. Sinai (2005), "Assessing High House Prices: Bubbles, Fundamentals and Misperceptions," *Journal of Economic Perspectives* 19(4), 67-92.

IMF (2012),"World Economic Outlook Update," released on July 16, 2012, International Monetary Fund, Washington D.C.

Kennedy, P. E. (1981), "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations." *American Economic Review* 71 (4), 801.

Kurz, C. and J. Hoffmann (2009), "A Rental Equivalence Index for Owner-Occupied Housing in West Germany," in Diewert, W. E., B. M. Balk, D. Fixler, K. J. Fox and A. O. Nakamura (eds.), *Price and Productivity Measurement: Volume 1 - Housing*. Trafford Press, Chapter 5, 69-86.

Lancaster, K. J. (1966), A New Approach to Consumer Theory, *Journal of Political Economy* 74, 13257.

Leamer, Edward E. (2002),“Bubble Trouble? Your Home Has a P/E Ratio Too,” UCLA Anderson Forecast (June).

Malpezzi, S. (2003), “Hedonic Pricing Models: A Selective and Applied Review,” in A. OSullivan and K. Gibb (eds.) *Housing Economics: Essays in Honor of Duncan Maclennan*, 67-89. Blackwell: Malder MA.

Rambaldi, A. N., R. R. J. McAllister, K. Collins and C. S. Fletcher (2011),“An Unobserved Components Approach to Separating Land from Structure in Property Prices: A Case Study for the City of Brisbane,” School of Economics Discussion Paper No. 428, University of Queensland, Australia.

Rosen, S. (1974), “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,” *Journal of Political Economy* 82(1), 34-55.

Schultze, C. S., and C. Mackie (2002), At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes, in *Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of- Living Indexes.* Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC, National Academy Press.

Shiller, R. J. (2007), “Understanding Recent Trends in House Prices and Home Ownership,” Cowles Foundation Discussion Paper No. 1630.

Shilling, J. D., C. F. Sirmans and J. Dombrow (1991),“Measuring Depreciation in Single-family Rental and Owner-occupied Housing,” *Journal of Housing Economics* 1, 368-383.

Sinai, T. and N. Souleles (2005), “Owner-Occupied Housing as a Hedge Against Rent Risk,” *Quarterly Journal of Economics*, 120, 763-89.

Smith, M. H. and G. Smith (2006), “Bubble, Bubble, Where’s the Housing Bubble,” *Brooking Papers on Economic Activity* 92(1), 1-67.

Stapledon, N. (2007), Long Term Housing Prices in Australia and Some Economic Perspectives, PhD Thesis, School of Economics, University of New South Wales, Sydney, Australia.

Triplett, J. E. (2006), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, Directorate for Science, Technology and Industry, Organisation for Economic Co-operation and Development, Paris.

Verbrugge, R. (2008), “The Puzzling Divergence of Rents and User Costs, 1980-2004,” *Review of Income and Wealth* 54(4), 671-699.

Wasshausen, D. and B. R. Moulton (2006), "The Role of Hedonic Methods in Measuring Real GDP in the United States," Paper Presented at 31st CEIES Eurostat Seminar *Are We Measuring Productivity Correctly?*, Rome, 12-13 October.

Waugh, F. V. (1928), "Quality Factors Influencing Vegetable Prices," *Journal of Farm Economics* 10, April, 185-196.

Table 1: Data Description

| Statistics | Price Data | Rental Data |
|---|---|---|
| No. of observations | 371,604 | 358,381 |
| Period of coverage (years) | 9 (2001–09) | 9 (2001–09) |
| Median price or annual rent (AU $) | 495,000.00 | 18,250.00 |
| Median land area (square meters) | 592.00 | 584.00 |
| Mean land area (square meters) | 684.39 (568.54) | 640.86 (390.11) |
| Median no. of bedrooms | 3 | 3 |
| Mean no. of bedrooms | 3.31 (0.84) | 3.17 (0.82) |
| Modal no. of bedrooms | 3 | 3 |
| Median no. of bathrooms | 2 | 1 |
| Mean no. of bathrooms | 1.68 (0.74) | 1.45 (0.64) |
| Modal no. of bathrooms | 1 | 1 |
| Percentage of observations having the following characteristics: | | |
| Land area, no. of bedrooms and no. of bathrooms | 61.67 | 45.60 |
| No. of bedrooms and no. of bathrooms | 62.43 | 49.33 |
| Land area and no. of bathrooms | 61.67 | 45.60 |
| Land area and no. of bedrooms | 73.50 | 47.97 |
| No. of bathrooms | 62.43 | 49.33 |
| No. of bedrooms | 74.46 | 99.95 |
| Land area | 98.43 | 47.99 |

Note: The figures in the parentheses are the estimated standard errors.

Table 2: Distributions of Characteristics in the Price and Rental Data (in %)

| Characteristics | Data | Counts | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | Total | | | |
| Bedrooms | Price | 0.62 | 13.30 | 49.40 | 29.22 | 6.48 | 0.98 | 100.00 | | | |
| | Rent | 1.52 | 15.48 | 53.22 | 23.79 | 5.88 | 0.11 | 100.00 | | | |
| Bathrooms | Price | 46.74 | 40.07 | 11.61 | 1.40 | 0.18 | n.a. | 100.00 | | | |
| | Rent | 61.74 | 31.54 | 6.20 | 0.48 | 0.03 | n.a. | 100.00 | | | |
| Characteristics | Data | Deciles | | | | | | | | | |
| | | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th |
| Land Area* | Price | 9.10 | 10.01 | 9.99 | 10.95 | 9.50 | 9.97 | 10.55 | 9.91 | 10.19 | 9.83 |
| | Rent | 11.93 | 10.17 | 10.15 | 9.61 | 9.44 | 10.03 | 9.39 | 9.67 | 9.54 | 10.07 |
| Postcodes | Price | 17.62 | 14.91 | 11.79 | 8.30 | 9.66 | 8.28 | 8.97 | 8.09 | 7.33 | 5.04 |
| (by Price)† | Rent | 20.40 | 15.73 | 12.35 | 7.16 | 8.98 | 8.09 | 7.97 | 7.65 | 7.15 | 4.52 |
| Postcodes | Price | 18.56 | 15.46 | 9.31 | 7.60 | 10.96 | 9.40 | 9.22 | 7.91 | 6.46 | 5.11 |
| (by Rent)§ | Rent | 19.98 | 16.97 | 10.07 | 7.27 | 10.55 | 8.58 | 8.46 | 7.69 | 5.92 | 4.50 |

*The price and rental data are pooled before dividing them into deciles in terms of land area. Therefore, each decile corresponds to the same land area in both data sets.
†Houses are ordered from the cheapest to the most expensive in terms of price.
§Houses are ordered from the cheapest to the most expensive in terms of rent.

Table 3: HM1 Regression Results

| Statistics | | Price Models | Rent Models |
|---|---|---|---|
| No. of observations | | 25462 (12273) | 18157 (16235) |
| No. of parameters | | 204 (17) | 204 (17) |
| Adjusted $R^2$ (%) | | 78.35 (3.29) | 79.33 (1.95) |
| Location attributes: | Joint contribution (%) | 54.9 (2.89) | 48.08 (1.94) |
| | % of significant coefficients | 93.26 (2.10) | 85.77 (7.53) |
| Temporal attributes: | Joint contribution (%) | 0.32 (0.33) | 0.13 (0.11) |
| | % of significant coefficients | 96.30 (11.11) | 55.56 (52.70) |
| Physical attributes: | Joint contribution (%) | 9.60 (0.93) | 12.74 (3.76) |
| | % of significant coefficients | 84.13 (8.58) | 68.25 (8.07) |

Note: The numbers are the mean results obtained from the 9 yearly regressions. The numbers in parentheses are the standard deviations of the 9 yearly regressions, indicating how stable or dispersed the statistics are across yearly regressions. The joint contribution is calculated by taking the difference in the adjusted $R^2$ between the unrestricted and restricted models. The statistical tests are conducted at the 5% significance level.

Table 4: HM2–HM8 Regression Results

| Statistics | Data | HM2 | HM3 | HM4 | HM5 | HM6 | HM7 | HM8 |
|---|---|---|---|---|---|---|---|---|
| Adjusted $R^2$ (in %) | Price | 75.92 | 76.67 | 74.90 | 74.19 | 72.36 | 67.06 | 63.87 |
| | | (4.55) | (2.71) | (3.65) | (3.86) | (5.09) | (2.87) | (4.26) |
| | Rent | 78.54 | 74.44 | 75.18 | 73.97 | 72.50 | 65.85 | 62.83 |
| | | (2.15) | (1.10) | (1.96) | (0.95) | (2.40) | (1.74) | (0.95) |
| % of coefficients having the same sign as HM1 | Price | 92.29 | 98.76 | 99.20 | 92.26 | 91.15 | 97.32 | 90.33 |
| | | (5.43) | (1.17) | (6.64) | (5.18) | (6.06) | (1.81) | (4.34) |
| | Rent | 97.44 | 95.80 | 97.04 | 95.52 | 92.48 | 91.37 | 87.16 |
| | | (2.15) | (1.49) | (1.99) | (1.89) | (4.35) | (2.96) | (3.06) |

Notes: The numbers are the mean results obtained from the 9 yearly regressions. The numbers in parentheses are the standard deviations of the 9 yearly regressions.

Table 5: Omitted Variables Adjustment Factors: $\lambda_{t,HM8}$ and $\lambda_{t,HMj|HM8}$

| Year | $\lambda_{t,HM8}$ | $\lambda_{t,HMj\|HM8}$, $j = 1, \ldots, 7$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | HM1 | HM2 | HM3 | HM4 | HM5 | HM6 | HM7 |
| 2001 | 1.160 | 0.943 | 0.946 | 0.944 | 0.981 | 0.944 | 0.981 | 0.989 |
| 2002 | 1.169 | 0.948 | 0.951 | 0.947 | 0.980 | 0.948 | 0.987 | 0.993 |
| 2003 | 1.120 | 0.943 | 0.945 | 0.944 | 0.977 | 0.948 | 0.981 | 0.993 |
| 2004 | 1.115 | 0.939 | 0.953 | 0.939 | 0.975 | 0.953 | 0.987 | 0.992 |
| 2005 | 1.103 | 0.962 | 0.967 | 0.960 | 0.987 | 0.962 | 0.992 | 0.999 |
| 2006 | 1.121 | 0.960 | 0.963 | 0.962 | 0.987 | 0.963 | 0.991 | 0.995 |
| 2007 | 1.117 | 0.954 | 0.949 | 0.966 | 0.975 | 0.960 | 0.974 | 1.002 |
| 2008 | 1.084 | 0.956 | 0.956 | 0.966 | 0.973 | 0.964 | 0.976 | 1.005 |
| 2009 | 1.052 | 0.965 | 0.960 | 0.972 | 0.980 | 0.968 | 0.984 | 1.004 |
| Average | 1.115 | 0.952 | 0.954 | 0.955 | 0.979 | 0.957 | 0.984 | 0.997 |

Note: Overall adjustment factor: $\lambda_{t,HMj} = \lambda_{t,HM8} \times \lambda_{t,HMj|HM8}$. For example, $\lambda_{2001,HM1} = 1.160 \times 0.943 = 1.094$.

Table 6: Actual and Quality-Adjusted Price-Rent Ratios and Quality Bias

| Year | Actual Price-Rent | | | Quality-Adjusted Price-Rent | | | Quality Bias (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Lower Quartile | Median | Upper Quartile | Lower Quartile | Median | Upper Quartile | Lower Quartile | Median | Upper Quartile |
| 2001 | 22.51 | 25.63 | 27.68 | 18.37 | 20.48 | 23.17 | 22.58 | 25.15 | 19.45 |
| 2002 | 27.51 | 30.42 | 32.41 | 21.43 | 24.29 | 27.48 | 28.38 | 25.24 | 17.92 |
| 2003 | 31.69 | 34.72 | 36.39 | 25.43 | 28.59 | 32.12 | 24.60 | 21.45 | 13.28 |
| 2004 | 32.76 | 35.48 | 37.45 | 26.91 | 29.78 | 32.68 | 21.74 | 19.13 | 14.43 |
| 2005 | 30.69 | 33.41 | 34.93 | 24.82 | 27.09 | 29.76 | 23.62 | 23.32 | 17.37 |
| 2006 | 29.53 | 32.06 | 34.08 | 23.13 | 25.44 | 28.26 | 27.68 | 26.01 | 20.61 |
| 2007 | 25.22 | 27.45 | 30.17 | 21.23 | 23.60 | 26.91 | 18.78 | 16.33 | 12.09 |
| 2008 | 21.77 | 23.01 | 25.28 | 19.16 | 21.40 | 24.61 | 13.63 | 7.52 | 2.74 |
| 2009 | 20.63 | 21.85 | 24.04 | 19.41 | 21.62 | 24.69 | 6.29 | 1.10 | -2.63 |
| Average | 26.92 | 29.34 | 31.38 | 22.21 | 24.70 | 27.74 | 20.81 | 18.36 | 12.81 |

Table 7: Cross-Sectional Variation of Price-Rent Ratios

| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | Pooled |
|---|---|---|---|---|---|---|---|---|---|---|
| Elasticity of price-rent ratio to changes in price or rent:* | | | | | | | | | | |
| $\Delta$ in price | 0.213 | 0.200 | 0.181 | 0.140 | 0.134 | 0.145 | 0.188 | 0.211 | 0.213 | 0.206 |
| | (0.001) | (0.001) | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| $\Delta$ in rent | 0.201 | 0.171 | 0.100 | 0.077 | 0.106 | 0.144 | 0.187 | 0.209 | 0.225 | 0.111 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) |
| Elasticity of price-rent ratio or price to changes in the share of structural components:† | | | | | | | | | | |
| Elasticity of price-rent | -0.238 | -0.234 | -0.177 | -0.195 | -0.194 | -0.136 | -0.245 | -0.219 | -0.225 | -0.295 |
| | (0.010) | (0.010) | (0.008) | (0.009) | (0.005) | (0.005) | (0.004) | (0.005) | (0.005) | (0.002) |
| Elasticity of price | -0.729 | -0.603 | -0.502 | -0.579 | -0.473 | -0.467 | -0.805 | -0.603 | -0.694 | -0.676 |
| | (0.031) | (0.031) | (0.023) | (0.032) | (0.018) | (0.018) | (0.013) | (0.017) | (0.016) | (0.006) |

Notes: *The figures are the estimated slope coefficients obtained when the log of the quality adjusted price-rent ratios are regressed on an intercept and the log of price or the log of rent. The figures in the brackets are the estimated standard errors.

†These figures are the estimated slope coefficients obtained when the log of price-rent ratios or prices are regressed on an intercept and the structural shares of dwellings.

Table 8: Quality-Adjusted Price-Rent Ratios for Different Market Segments

**Houses are ordered from cheapest to most expensive**

| Year | Price-Rent from Price Data | | | Price-Rent From Rent Data | | |
|---|---|---|---|---|---|---|
| | Lower Quartile | Median | Upper Quartile | Lower Quartile | Median | Upper Quartile |
| 2001 | 19.28 | 21.89 | 22.63 | 18.49 | 20.09 | 21.82 |
| 2002 | 21.17 | 24.79 | 26.98 | 21.72 | 24.20 | 25.96 |
| 2003 | 27.39 | 29.17 | 31.47 | 26.27 | 28.91 | 30.16 |
| 2004 | 28.61 | 30.11 | 30.91 | 28.35 | 29.99 | 30.14 |
| 2005 | 26.52 | 27.07 | 27.35 | 26.14 | 27.36 | 28.19 |
| 2006 | 23.88 | 26.02 | 29.40 | 23.67 | 24.93 | 25.84 |
| 2007 | 21.57 | 25.65 | 25.54 | 21.24 | 23.01 | 25.66 |
| 2008 | 19.70 | 21.07 | 24.08 | 19.47 | 21.26 | 23.58 |
| 2009 | 19.82 | 21.36 | 26.13 | 19.36 | 21.27 | 23.81 |
| Average | 23.10 | 25.24 | 27.17 | 22.75 | 24.56 | 26.13 |

Table 9: Departure of Market from Equilibrium

| Year | $r_t$ | $g_t$ ($x$=10) | $g_t$ ($x$=20) | Equilib. $P_t/R_t$ ($x$=10) | Equilib. $P_t/R_t$ ($x$=20) | Actual Adjusted $P_t/R_t$ | Actual Unadj. $P_t/R_t$ | Implied $g_t$ Adjusted $P_t/R_t$ | Implied $g_t$ Unadj. $P_t/R_t$ |
|---|---|---|---|---|---|---|---|---|---|
| 2001 | 0.054 | 0.025 | 0.020 | 18.64 | 17.05 | 20.48 | 25.63 | 0.060 | 0.070 |
| 2002 | 0.059 | 0.038 | 0.031 | 22.09 | 19.13 | 24.29 | 30.42 | 0.073 | 0.081 |
| 2003 | 0.053 | 0.054 | 0.044 | 42.84 | 30.07 | 28.59 | 34.72 | 0.073 | 0.079 |
| 2004 | 0.057 | 0.066 | 0.050 | 62.55 | 31.26 | 29.78 | 35.48 | 0.078 | 0.084 |
| 2005 | 0.054 | 0.059 | 0.047 | 51.56 | 31.30 | 27.09 | 33.41 | 0.072 | 0.079 |
| 2006 | 0.052 | 0.056 | 0.043 | 46.48 | 29.66 | 25.44 | 32.06 | 0.068 | 0.076 |
| 2007 | 0.059 | 0.053 | 0.045 | 32.80 | 25.75 | 23.60 | 27.45 | 0.072 | 0.078 |
| 2008 | 0.061 | 0.048 | 0.042 | 26.55 | 22.61 | 21.40 | 23.01 | 0.069 | 0.073 |
| 2009 | 0.041 | 0.034 | 0.018 | 31.21 | 21.07 | 21.62 | 21.85 | 0.050 | 0.050 |
| Average | 0.054 | 0.048 | 0.038 | 37.19 | 25.32 | 24.72 | 29.34 | 0.068 | 0.074 |

Notes: $\omega_t = 0.01$, $\delta_t = 0.025$, $\gamma_t = 0.02$ and $\pi_t^e = 0.03$ for the entire period.

Table 10: Robustness Checks

| Year | Median Price-rent Ratios Obtained from Alternative Methods | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 2001 | 23.35 | 21.10 | 20.76 | 21.35 | 20.69 | 20.08 | 20.66 |
| 2002 | 28.19 | 25.47 | 24.55 | 24.92 | 24.53 | 23.98 | 24.29 |
| 2003 | 31.81 | 29.68 | 28.82 | 28.96 | 29.04 | 28.66 | 28.66 |
| 2004 | 32.11 | 30.38 | 29.90 | 29.96 | 30.05 | 29.53 | 29.96 |
| 2005 | 29.48 | 28.08 | 26.98 | 27.28 | 27.43 | 26.37 | 27.11 |
| 2006 | 27.76 | 26.21 | 25.36 | 25.63 | 25.51 | 24.84 | 25.44 |
| 2007 | 25.31 | 24.56 | 23.61 | 23.63 | 23.41 | 23.35 | 23.57 |
| 2008 | 22.47 | 22.06 | 21.54 | 21.22 | 21.27 | 21.11 | 21.42 |
| 2009 | 22.46 | 22.44 | 21.76 | 21.17 | 21.51 | 21.34 | 21.64 |
| Average | 27.00 | 25.55 | 24.81 | 24.90 | 24.83 | 24.36 | 24.75 |

Notes: (1) Obtained from running HM1 models on observations which have all three physical characteristics, without correcting for omitted variables bias. (2) Single imputation price-rent ratios—$P/R(sold)^{SI}_{th}$ and $P/R(rented)^{SI}_{tj}$ (see section 3.2)—are estimated, instead of double imputation price-rent ratios. (3) Hedonic models in equations 9 and 10 are specified on price and rent levels instead of log of prices and rents. (4) Double imputation hedonic price and rent indexes are used in order to estimate $P_{st}$ and $R_{st}$ in equations 21 and 20, respectively (the main results use repeat-sales and -rents indexes). (5) For dwellings with multiple prices and rents, we consider each price-rent pair in order to obtain $P/R^{AM}_t$ (in equation 22) where the main results use only the chronologically closest price and rent observations. (6) The regressions required to obtain $\lambda_{t,HM8}$ (equation 23) are estimated using the sample HM8s\AMs (instead of using HM8s which includes AMs) and then the price-rent ratios of the AMs dwellings are imputed from the regression results. (7) The correction factor $\lambda_{t,HMj|HM8}$ (equation 24) is obtained using the HM1s, i.e. $\lambda_{t,HMj|HM8} = HMjm(HM1s)/HM8m(HM1s)$.

Figure 1: Price and Rent Indexes of Lower and Upper Ends of the Market

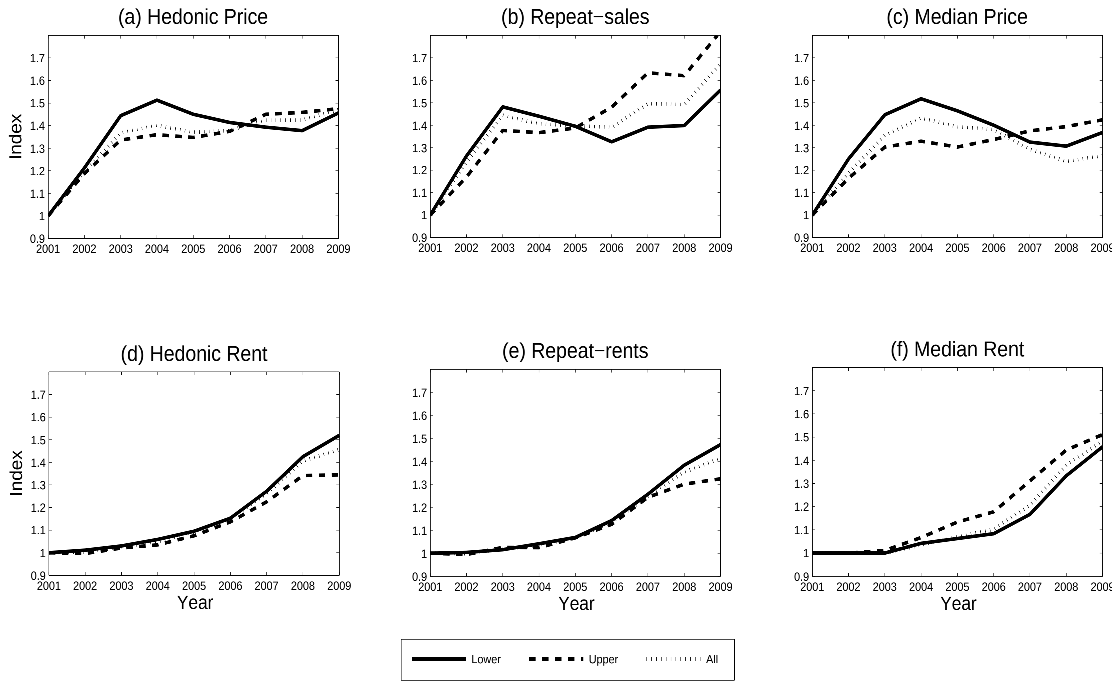


Figure 2: $\sigma$-Convergence