Hedonic Price-Rent Ratios for Housing: Implications for the Detection of Departures from Equilibrium

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

June 6, 2012

Preliminary Draft

In equilibrium the quality-adjusted price-rent ratio for housing should equal its user cost. Actual median price-rent ratios may be misleading since purchased dwellings on average tend to be of better quality than rented dwellings. Combining house sales and rents data for Sydney, Australia over the period 2001 to 2009 we construct a data set consisting of in excess of 900,000 observations. We then use an innovative hedonic approach to impute a rent for each dwelling sold and a purchase price for each dwelling rented, thus allowing us to compute price-rent ratios at the level of individual dwellings. Using these price-rent ratios, which by construction are quality adjusted, we find that the actual median price-rent ratio is systematically about 8 percent larger than its quality-adjusted counterpart. We also find that for most of our sample the quality-adjusted median price-rent ratio exceeds its equilibrium level derived from the user cost formula. The equilibrium price-rent ratio is itself highly sensitive to the assumed rate of expected capital gains. Our estimate of 21 for the equilibrium price-rent ratio is obtained using the average real capital gain during our sample of 3.4 percent per year. This is high by historical standards, thus suggesting that our equilibrium price-rent ratio may also be too high. An alternative approach is to assume that the housing market is in equilibrium and then use the user-cost formula to impute the expected capital gain. Using this approach we generate an imputed expected real capital gain of about 4.5 percent per year, which is even more implausible. This again indicates that, for at least most of our sample, the price-rent ratio in Sydney was at an unsustainable level. (**JEL.** C43, E01, E31, R31)

Keywords: Real estate; Housing market; Hedonic model; Price-rent ratio; User cost; Missing data

[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

1 Introduction

Housing markets seem to be particularly prone to booms and busts. Recent events have also shown how developments in 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 observe departures from equilibrium in the housing market.

One way of addressing this issue is through comparisons of the price-rent ratio with the reciprocal of the user cost of housing. In equilibrium these terms should be equal. If the price-rent ratio is greater than the reciprocal of user cost, then renting should be relatively more attractive thus implying that the price-rent ratio is too high.[1] Conversely, if the price-rent ratio is lower then buying is more attractive than renting and hence the price-rent ratio is too low.

Empirical implementation of this idea, however, is hampered by the fact that actual price-rent ratios are typically calculated as the ratio of median house price to median rent. The problem with comparing medians is that there is likely to be a quality differential between the median dwelling sold and the median dwelling rented. In particular, it is likely that the median dwelling sold will in most cases be of better quality than the median dwelling rented. The equilibrium condition, by contrast, implicitly assumes that the stated price and rent apply to dwellings of equivalent quality. If in fact the median price refers to a better quality dwelling than does the median rent then a comparison of price-rent ratios with user cost will be biased in favor of finding that the price-rent ratio is above its equilibrium level.

We have two main objectives in this paper. First, we show how quality-adjusted price-rent ratios can be constructed by applying hedonic methods at the level of individual dwellings. Our hedonic approach entails imputing a rental price for each dwelling

[1] If the price-rent ratio is above its equilibrium level, this does not necessarily imply that house prices are too high. Alternatively, both house prices and rents could be too low, but rents are even further than house prices below their equilibrium level.

actually sold in a given year, while simultaneously imputing a sale price for each dwelling rented in that year. In this way we are able to obtain a matched price-rent ratio for every dwelling either sold or rented in a given year. Comparison of these two distributions of price-rent ratios (i.e., one derived from dwellings sold and the other from dwellings rented) also provides an indication of the plausibility of our underlying assumptions. Using in excess of 900,000 price and rent observations for Sydney, Australia over the period 2001-2009 we show that on average the median price-rent ratio is about 8 percent larger than its quality-adjusted counterpart. The difference is even larger (i.e., 9 percent) for the lower quartile, but lower (i.e., 3 percent) for the upper quartile of the price-rent distribution. We also show that for both dwellings sold and rented the price-rent ratio is higher for more expensive dwellings. It follows therefore that the quality-adjustment bias resulting from comparing matched percentiles of the price and rent distributions is more pronounced at the lower end of the market.

Our second objective is to use the user-cost equilibrium condition to check for departures from equilibrium in the Sydney housing market. One problem with the user cost formula is that one of its key components is the expected capital gain, which cannot be directly observed. An estimate can be obtained either by extrapolating from past trends, or implicitly by assuming that the price-rent ratio is in equilibrium. Using the first approach we find that the price-rent ratio in Sydney was above its equilibrium level for most of our sample period, although by June 2008 this gap was largely eliminated. Using the second approach, which assumes the market is in equilibrium, we find that the implied real expected capital gain of 4.5 percent per year is implausibly large, thus again indicating that for at least most of our sample the price-rent ratio is too high.

Our approach of imputing the expected capital gain from the user cost formula also allows us to explore how the expected capital gain differs at the upper and lower ends of the market. We find that the expected capital gain is about half a percentage point higher at the upper quartile than at the lower quartile (when houses are ordered from cheapest to most expensive.

2 Price-Rent Ratios, User Cost, and Equilibrium in the Housing Market

The housing market is in equilibrium when the expected annual cost of owner-occupying equals the annual cost of renting. Following Himmelberg, Mayer and Sinai (2005) the equilibrium condition can be written as follows:

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

where R_t is the period t rental price, P_t the purchase price and u_t the per dollar user cost.[2] Abstracting from tax deductibility of mortgage interest payments by owner occupiers (which is not possible in most countries), per dollar user cost (henceforth 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, ω is the property tax rate, δ the depreciation rate for housing, γ 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. 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 $R_t < u_t P_t$. Transaction costs might slow the adjustment process but should not affect the equilibrium itself. Rent controls may prevent adjustment to equilibrium. In our data set this is not an issue since there is no rent control in Australia.

Rearranging (1), we obtain that in equilibrium the price-rent ratio should equal the reciprocal of user cost (i.e., $P_t/R_t = 1/u_t$). If the actual price-rent ratio exceeds

[2] The rental cost here excludes running costs. Since running costs must be paid by both owner occupiers and renters, they drop out of (1).

our estimate of the reciprocal of 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 $P_t/R_t > 1/u_t$.

It seems likely that publicly available price-rent ratios, which are typically calculated as the ratio of the median dwelling price to the median rent, suffer from exactly this kind of quality mismatch. The median owner-occupied dwelling will tend to be of superior quality to the median rental dwelling. 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 user costs will therefore be subject to systematic bias. In the next two sections, 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. We illustrate the extent of this problem in section 5.

An alternative and probably more fruitful way of using the user-cost concept is to assume the housing market is in equilibrium, and then impute the implied expected capital gain. If this 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, ω_t, δ_t and γ_t, we obtain an estimate of g_t. We apply this approach to our Sydney data in section 5.

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). Other early contributors include Court (1939) and Stone (1954). It was, however, only after Griliches (1961, 1971) 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). A hedonic model regresses the price of a product on a vector of characteristics (whose prices are not independently observed). The hedonic equation is a reduced form equation that is determined by the interaction of supply and demand.

Hedonic methods have been widely used for constructing quality-adjusted price indexes. Three main approaches have been used in the literature. Following the terminology used in Triplett (2006) and Hill (2012), we refer to these as the time-dummy, imputation and characteristics index methods. The time-dummy method estimates a hedonic model for the whole data set that includes time dummy fixed effects. The price index for each period is then obtained directly from these time dummies. The hedonic imputation and characteristic index methods by contrast both estimate a separate hedonic model for each time period. The imputation method then imputes a price for each dwelling in each period from that period's hedonic model, after which the price index can be calculated using a standard price index formula. The characteristics index imputes the price of the same average dwelling in each period again using that period's hedonic model. The estimated price of the average dwelling, in this case, is the price index.[3]

In this paper we focus exclusively on the second approach, i.e., imputation methods. Our main reason for preferring the imputations approach is that it can be easily adapted to deal with the problem of observations in our data set that are missing some characteristics. We return to this issue later.

Imputation methods make use of standard price index formulas. In a housing context, this requires the price of each dwelling in the comparison to be available in both periods being compared. Given that dwellings typically sell only at infrequent and irregular intervals, to make this approach operational it is necessary to impute at least some of the prices. For example, suppose we are trying to measure the change in house prices from 2008 to 2009. We could consider all the dwellings that sold in 2008 and impute prices for them in 2009. Conversely, we could consider all dwelling sold in 2009 and impute prices for them in 2008. The former is a Laspeyres-type price index and the latter a Paasche-type index.

An imputations method obtains these imputed prices from the hedonic model,

[3]This brief description brushes over a number of subtleties of each approach. See Hill (2012).

which is estimated separately for each period typically using a semilog functional form:[4]

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

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), β_t is a $C \times 1$ vector of characteristic shadow prices, and u_t is an $H_t \times 1$ vector of random errors.

The first column in X consists of ones, and hence the first element of β is an intercept term. Examples of characteristics include the number of bedrooms, number of bathrooms, land area, and postcode or some other locational identifier. It is possible also to include functions of characteristics (such as land size squared), and interaction terms between characteristics. For example, one might want to interact bedrooms and land area, bathrooms and land area, or bedrooms and bathrooms. Focusing specifically on the last of these, the inclusion of bedroom-bathroom interaction terms could be justified by the fact that the value of an extra bathroom may depend on how many bedrooms there are.

Once the hedonic model has been estimated separately for each year, it is now possible to use it to impute prices for individual dwellings. 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 can now be constructed in one of two ways:

$$\text{L1}: \quad P_{st}^{L1} = \sum_{h=1}^{H_s} w_{sh} \left[\hat{p}_{th}(x_{sh})/p_{sh}\right] = \sum_{h=1}^{H_s} \hat{p}_{th}(x_{sh}) \Big/ \sum_{h=1}^{H_s} p_{sh}$$

[4]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.

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

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

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

In an analogous manner corresponding Paasche-type hedonic indexes can be constructed:

$$\text{P1}: \quad P_{st}^{P1} = \left\{ \sum_{h=1}^{H_t} w_{th} \left[p_{th}/\hat{p}_{sh}(x_{th}) \right]^{-1} \right\}^{-1} = \sum_{h=1}^{H_t} p_{th} \Bigg/ \sum_{h=1}^{H_t} \hat{p}_{sh}(x_{th})$$

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

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).[5] No clear consensus has emerged in the literature as to which approach is better. Single imputation uses less imputations. Double imputation may reduce omitted variables bias (see Silver and Heravi 2001 and Hill and Melser 2008).

As is explained later we use hedonic price indexes in the construction of our quality-adjusted price-rent ratios. We find that for our data set F1 and F2 price indexes are almost indistinguishable. So in this context the choice between single and double imputation is of little consequence.

[5] We have simplified matters here by not considering the case of repeat-sales. In any comparison, there are likely to be a small number of dwellings that sell in both periods. These repeat sales could be used by both the single and double imputation methods. One reason for not doing so particularly in a double imputation setting is that the dwelling may have been renovated (e.g., an extra bathroom added) between sales. Also, a different version of the double imputation method is obtained if the expenditure shares w_{sh} in L2 in (5) and w_{th} in P2 in (6) are not imputed.

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. An important feature of this approach is that the hedonic price and rent models need to be defined on the same set of characteristics.

Some of these steps have been implemented previously by other authors. Arévalo and Ruiz-Castillo (2006) estimate a hedonic model using rental data and then use it to impute rents for owner-occupied dwellings in Spain. They then compare the imputed rents from the hedonic model with corresponding self-imputed rents obtained from household budget surveys. Similarly, Kurz and Hoffman (2009) estimate a hedonic model using rental data and another using self-imputed rents of owner-occupiers in Germany obtained from a household survey. Both papers find that the two approaches generate similar results and hence that the hedonic rent model is a viable approach for imputing rents for owner-occupied housing. None of these papers consider price-rent ratios. Crone, Nakamura and Voith (2009) estimate hedonic models for prices and rents. However, rather than trying to impute from one to another or compute price-rent ratios, they focus on trying to decompose price changes into a capitalization rate and a housing service component. The paper that is closest to ours in its approach is probably Davis, Lehnert and Martin (2008). Using US Census data, they estimate a hedonic model for rental data. Prices for owner-occupied housing are then imputed from this model. Finally, they compare the imputed rents with Census estimates of market value of owner-occupied dwellings to obtain rent-price ratios for owner-occupied housing.

The hedonic price equation 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.

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 period t is imputed from (10) as follows:

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

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:

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

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

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

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

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

Exponentiating, it follows that:[6]

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

[6]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.

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

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

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

The distinction between single and double imputation arises again in the calculation of our hedonic price-rent ratios. A single imputation price-rent ratio $P/R(sold)_{th}^{SI}$ 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):

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

A corresponding double imputation price-rent ratio $P/R(sold)_{th}^{DI}$ divides the imputed price for dwelling h obtained from (14) by its imputed rent obtained from (11):

$$P/R(sold)_{th}^{DI} = \frac{\hat{p}_{th}(x_{Pth})}{\hat{r}_{th}(x_{Pth})} = \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{Ptc} x_{Rthc}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{Rtc} x_{Pthc}\right)}. \tag{16}$$

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):

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

Empirically, we find that on average our double imputation price-rent ratios are 2.5 percent lower than their corresponding single-imputation counterparts. While the choice between single and double imputation is important, it turns out that it does not affect the general thrust of our results in section 5.

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:

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

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

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

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:

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

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

4 Empirical Strategy and Data Sets

4.1 The hedonic price and rental data sets

The data set used in this paper is for Australia's largest city, Sydney, over the period 2001 to 2009. It is 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 for individual dwellings in Sydney is obtained from Australian Property Monitors

(APM).[7] It consists of a total of 395,110 observations over the 2001 to 2009 period. 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 341,877 observations) with data from APM (of which we have 99,495 observations that are not also in the NSW Housing data set). In total, therefore, we have 441,372 rental observations. An important difference between the two rental data sets is that while the recorded rents in the NSW Housing data refer to the new rental contracts, the recorded 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.[8]

Apart from the fact that combining the two rental data sets gives us more observations to work with, the data sets also complement each other in terms of available characteristics information. The characteristics in the APM rental data set are identical to those in the sales data set. However, the NSW Housing data set has only the following characteristics: transaction price, exact date of sale, number of bedrooms, exact address and a postcode identifier, i.e., it is missing the number of bathrooms and land area. By matching the addresses in the NSW Housing data set with those in the APM price and rental data sets it was possible to obtain the missing characteristics for some observations. This process reduced the number of observations missing the characteristics of land area, the number of bedrooms and the number of bathrooms in the combined rental data by 29.68, 47.73 and 31.28 percent, respectively. The same exercise conducted for the price data reduced the number of observations missing land area, number of bedrooms and number of bathrooms by 14.69, 12.23 and 6.73 percent,

[7] APM provides real estate related research service and data for the Australian market (see http://apm.com.au).

[8] We checked this by calculating the mean difference of rents of the 41,853 houses which were recorded in the same quarter and whose addresses were matched perfectly between the two data sets. The mean of the pairwise difference between the log of NSW Housing and APM rents is -0.0032, and the median and mode are 0 (with 76% of observations having no difference).

respectively. Even after conducting this filling-in exercise, there are many observations for which one or more characteristics are missing. The exact figures are given in Table 1. In particular, all the characteristics are available for 62.2 percent of the price data and for 38.1 percent of the rental data. For the remainder, at least one of the three characteristics of land area, number of bedrooms and number of bathrooms is missing. We explain in the next section how we deal with this problem.

Insert Table 1 Here

Before proceeding with the estimation of our hedonic models and carrying out the above filling-in exercise, we removed some extreme observations. The main reason for removing these observations is the presence of data-entry errors, which are concentrated in the tails of the price, rent and characteristic distributions. The following extreme observations were deleted:

(1) houses with the number of bedrooms greater than 6 and bathrooms greater than 5 (these correspond to the 99.68 and 99.95 percentiles in the price data and 99.99 and 99.99 in the rent data);

(2) houses with land areas lower than 1.0 percentile and greater than 99.0 percentiles (after the deletion, the land area of the remaining observations ranges between 94 and 7609 square meters in the price data and 84 and 5891 square meters in the rental data);

(3) observations with missing prices (3178 observations) and rents (5088 observations);

(4) houses with prices and rents lower than the 1.0 percentile and greater than the 99.0 percentile (this leads to prices ranging between \$117,500 and \$3,300,000 and annual rents ranging between \$6,779 and \$86,036).

All these deletions taken together led to the exclusion of around 5 percent of the observations in our data set.[9]

We had to undertake some further deletions in order to implement our hedonic

[9]It should be noted that the removal of extreme observations was absolutely necessary. Otherwise we would be contaminating our results by including, say, a \$400,000 house which was recorded as \$4,000,000 or a \$10,000 rent recorded as \$1,000. The deletions are expected to minimize the number of observations with such errors.

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 houses in a particular postcode, then the rental model must include houses rented in the same postcode. Since we have applied our hedonic approach separately for each of the 9 years in the data set, this matching of characteristics between the price and rental data is done separately for each year. This matching reduces the number of observations in the price data by a small percentage (0.8%), but reduces the rental data by a large percentage (13.67%). The reason for such a large reduction in the rental data is that from 2001–2004 while the rental data included 210 postcodes, the price data included only 190 postcodes. Therefore the additional postcodes in the rental data had to be deleted. In the later years both data sets included 210 postcodes of data (there are 213 postcodes in the Sydney Metropolitan Area). 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 will tend to worsen the quality of the matches when adjusting for quality difference between sold and rented dwellings.[10]

In total, the deletion of extreme observations and the deletions due to the matching requirement led to the exclusion of 12.3% of observations from the combined total number of price and rental data observations. This leaves us with 371,652 observations in the price data and 362,108 observations in the rental data. A brief description of the data sets is provided in Table 1. The median price is $495,000.00 and the median annual rent is $16,685.71, giving a median price-rent ratio for the whole data set of 29.67.

There is a close link between the sold and owner-occupied dwellings. After a house is sold and, therefore, appears in the sold data, it can be either occupied by

[10]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 post-code 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.

the new owner or rented. ABS (2010) reports that the home-ownership rate, i.e. the percentage of households living in their own houses, in Australia remained stable at around 70 percent over the period 1971–2006 (see also AHURI 2010 and Yates 2000). 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). For example, Eurostat (2011) reports that 73.6 percent of the population in the 27 EU countries lived in owner-occupied dwellings in 2009, and Sinai and Souleles (2005) report that, according to the 2000 Decennial Census, 68 percent of the US households own the house they live in. In our data sets, we find that only 6 percent of the houses in the sold data appeared in the rental data within a year of the sale of the house, indicating that 94 percent of the houses were occupied by the owner. However, it should be noted that our matching exercise may identify only a portion of the total matched houses.[11]

Our expectation is that owner-occupied (and hence sold) dwellings on average are of better quality than rented dwellings. This hypothesis is confirmed by 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 dwellings are concentrated relatively more in the

[11]The matching of addresses between data sets runs into many practical problems because of the lack of uniformity across data sets including the use of abbreviations, spelling errors and missing parts of the address (such as whether it is street (or ST), avenue (AVE), road (RD), etc.). Throughout the paper, our matching process followed strict guidelines, a match was counted as a match only if every element matched perfectly.

cheaper postcodes.

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. Given these similarities and our large sample size, our imputations of prices for rented dwellings and rents for sold dwellings should achieve an acceptable level of accuracy.

4.2 Imputing prices and rents for dwellings with missing characteristics

The problem of missing characteristics can be dealt with by 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.)

(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 HM1. The omitted variables may take two forms. Omitted variables of the physical variety may include 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. Omitted variables of the locational variety include street noise, air quality and the availability of public transport links. The impact of some but not all of these locational characteristics may be captured by the postcode dummies.

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.

It follows that the omitted variables bias will be more severe in HM8 than in HM1, since HM1 includes land area, number of bedrooms and number of bathrooms as explanatory variables, while HM8 does not. In other words, the required omitted variables bias correction will differ for each of our eight models.

The first step in correcting for omitted variables bias is to obtain 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:

$$P/R_{sh}^{AM} = \frac{p_{sh} \times R_{st}}{r_{th}}. \qquad (20)$$

Similarly, an address-matched price-rent ratio for this dwelling in period t can be calculated by extrapolating the selling price forward to period t using a price index P_{st} as follows:

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

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

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

where $h = 1, \ldots, H_t$ indexes all the address-matched price-rent ratios in period t in our data set. 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.[12] 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.

Our address-matched price-rent ratios P/R_t^{AM} should by construction be free of omitted variables bias. There remains the issue of how the rent index R_t and price index P_t should be calculated. We consider two ways of doing this. The first is to use the double imputations hedonic formula $F2$ in (8). One concern with using a hedonic index is that it might indirectly reintroduce omitted variables bias. The second approach we consider is to compute P_t using the repeat-sales method and R_t using the repeat-rents method. Both the repeat-sales and repeat-rent indexes are calculated using Calhoun's (1996) method, which attempts to correct for heteroscedasticity by giving greater weight to repeats that are chronologically closer together (see also Hill, Melser and Syed 2009).

Repeat-sales (and repeat-rent) indexes however also have their disadvantages. In particular, all dwellings that sell (or rent) only once are deleted, dwelling may change in quality between sales (e.g., due to renovations or depreciation), and repeat-sales may not be representative of the broader sample. On this last point, Clapp and Giaccotto (1992), Gatzlaff and Haurin (1997), and Meese and Wallace (1997) all find evidence of a 'lemons' bias, since starter homes sell more frequently as people upgrade their dwelling as their wealth rises. Nevertheless, despite these shortcomings, repeat-sales and repeat-rent indexes should not be affected by the type of omitted variables bias that concerns us here. In our empirical analysis later in the paper we find that the address-matched median price-rent ratios derived using the double imputation hedonic (F2) and repeat-sales approaches are quite similar, and hence this decision is not so

[12] 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 virtually identical median address-matched price-rent ratios. So this decision does not have any bearing on our results.

important.

With our methodology in place for constructing quality-adjusted price-rent 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. 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. Due to its larger sample size here we actually 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 sample of dwellings as the median in the numerator [i.e., HM8m(AMs$_t$)]. The difference now is that the imputed price-rent ratios in the denominator are calculated by extrapolation from (21) and (20) using price and rent indexes rather than from HM8. To increase their reliability, these indexes are calculated over the full HM8 sample.

The median AMm(AMs$_t$) should be free of omitted variables bias since it is constructed from address-matched dwellings. Given that HM8m(AMs$_t$) and AMm(AMs$_t$) are calculated over the same sample of dwellings [i.e., the address-matched sample (AMs)], any sample selection bias in the numerator and denominator of (23) should be more or less offsetting. Any systematic deviation of $\lambda_{t,HM8}$ from 1 can therefore be attributed 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. We therefore adjust for omitted variables bias the price-rent ratio of a dwelling h sold in period t with the HM8 mix of characteristics by dividing it by $\lambda_{t,HM8}$ as follows:

$$P/R(sold)^{adj}_{th,HM8} = \frac{P/R(sold)_{th,HM8}}{\lambda_{t,HM8}} = P/R(sold)_{th,HM8} \times \left(\frac{AMm(AMs_t)}{HM8m(AMs_t)}\right).$$

Similarly, a dwelling j with the HM8 mix of characteristics rented in period t is adjusted for omitted variables bias as follows:

$$P/R(rented)^{adj}_{tj,HM8} = \frac{P/R(rented)_{tj,HM8}}{\lambda_{t,HM8}} = P/R(rented)_{tj,HM8} \times \left(\frac{AMm(AMs_t)}{HM8m(AMs_t)}\right).$$

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. HM8 can be estimated over any of our samples HM1,...,HM8 since it does not include any of land area, number of bedrooms, or number of bathrooms as characteristics.

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 $j = 1, \ldots, 7$). This is because HM8 is a special case of each of these other models. In other words, the other models all include more characteristics than HM8. Given our hypothesis that rental dwellings perform worse on these characteristics, it follows that $\lambda_{t,HMj|HM8}$ should be systematically less than 1. Our empirical results confirm this finding.

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 $\lambda_{t,HMj|HM8}$.

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. In fact we can go further and say that we expect to find that

$$
\begin{aligned}
&\lambda_{t,HM1} < \lambda_{t,HM2} < \lambda_{t,HM5} < \lambda_{t,HM8};\\
&\lambda_{t,HM1} < \lambda_{t,HM2} < \lambda_{t,HM6} < \lambda_{t,HM8};\\
&\lambda_{t,HM1} < \lambda_{t,HM3} < \lambda_{t,HM5} < \lambda_{t,HM8};\\
&\lambda_{t,HM1} < \lambda_{t,HM3} < \lambda_{t,HM7} < \lambda_{t,HM8};\\
&\lambda_{t,HM1} < \lambda_{t,HM4} < \lambda_{t,HM6} < \lambda_{t,HM8};\\
&\lambda_{t,HM1} < \lambda_{t,HM4} < \lambda_{t,HM7} < \lambda_{t,HM8}.
\end{aligned} \tag{26}
$$

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 for omitted variables bias the price-rent ratio of a dwelling h sold in period t with the HMj mix of characteristics 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). The specifications of the HM1–HM8 models are provided in the beginning of section 4.2. The characteristics in the regressions fall into three groups: temporal (3 quarterly dummies with the first quarter in the year as the base category), physical (dummies corresponding to the number of bedrooms and bathrooms, and the lot size) and location-specific (postcode dummies) characteristics.[13] With the exception of HM8, which does not include any physical characteristics, all other models include the three groups of characteristics. The exact number of postcodes included in the hedonic regressions varies across years, with the yearly average being 197 postcodes (there are 213 postcodes in the Sydney Metropolitan Area). Below, we summarize some key results obtained from the HM1–HM8 models.

Insert Table 3 Here

Focussing on the HM1 model first, which is our benchmark 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 77.4 and 73.9 per cent, respectively. The joint contributions of location in the regressions are, as expected, the largest, the postcode dummies explain 54.4 and 52.5 percent of the variations in the price and rent regressions, respectively.[14] The next largest contribution is the group of physical characteristics, contributing 9.4 and 9.0 percent to the price and rent variations, respectively. The quarter dummies add only 0.3 and 0.1 per cent to

[13]The HM1 model for 2001 is run on 209 variables. These include: an intercept, 3 quarterly dummies, 14 physical characteristics including their interaction terms and 191 postcode dummies (see Table 3 for the total number of covariates in the HM1 regressions).

[14]The marginal contribution is the difference in the adjusted R-squared obtained from the unrestricted and restricted model. The restricted model in this case does not include the postcode dummies.

prices and rents, respectively. With some small variations in the exact numbers, these results generally hold separately for each of the 9 yearly regressions.

The models include interaction terms between number of bedrooms and land area and between number of bathrooms and land area (HM1, HM3 and HM4 models). However, the joint contributions of these interaction terms are quite small, 0.25 and 0.27 percent respectively. 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.[15]

We now turn our attention to the individual significance of the estimated coefficients in the HM1 regressions. Land area, while statistically significant in 7 out of the 9 price equations, is significant in only 1 out of the 9 rent equations (all statistical significance are evaluated at the 5 per cent level unless stated otherwise). With the exception of one case in the price equation, all significant coefficients have a positive sign, in accordance with our prior expectation. The squared land area is significant in 9 price and 3 rent equations, and 8 and 2 of these significant coefficients have negative signs, respectively. The regressions include 3 bedroom dummies denoting 2, 3 and, 4 and above bedroom houses, with the 1 bedroom house as the base. Almost all the bedroom coefficients are significant in both the price and rent equations. In all cases, the difference between two consecutive bedroom coefficients, say the 3 bedroom coefficient minus the 2 bedroom coefficient, is positive, indicating that an additional bedroom contributes positively to the price and rent. The results with regard to the coefficients denoting 2, 3 and 4 bathroom dummies, with 1 bathroom as the base, are similar. The estimated bathroom coefficients are statistically significant in all, except for one, case.

[15]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.

The signs of the estimated coefficients and the signs of the difference between the two consecutive estimated coefficients (such as, 2 bathroom minus 1 bathroom coefficient) are positive in all except for 2 cases, one of which is not statistically significant. The quarter dummies are significant for 96 and 67 percent of the estimated coefficients in the price and rent equations, respectively, indicating the appropriateness of including quarterly dummies in the yearly hedonic regressions.[16]

The magnitude of the estimated coefficients on the physical characteristics shed light on the relative importance of these characteristics for sold and rented dwellings. However, in our hedonic regressions these magnitudes are not directly interpretable because of the interaction terms. Therefore, we evaluate the mean changes in price and rent due to a change in a particular characteristic for fixed values of other characteristics. These fixed values are 3 bedrooms and 1 bathroom (their modal values in both the data sets) and a land area of 750 square meters (the mean value of all the dwellings in the two data sets combined). Table 3 shows the results. For example, the table shows that if the number of bedrooms is increased from 2 to 3, the price increases by 9.7 percent and the rent increases by 12.79 percent for a house with 1 bathroom and a land area of 750 square meters. These results, with regard to the number of bedrooms and bathrooms, provide evidence of increasing returns to additional bedrooms and bathrooms for sold dwellings, but decreasing returns for rented dwellings. While land area is important to owners, adding 2.71 percent to price for every 100 square meters, it is not important to the renters (the land area coefficients are statistically or/and economically insignificant in the rent equations). The results indicate that the price-rent ratio for a large house is expected to be higher than that for a small house, providing us with an early indication that the deviation of the housing market from its equilibrium may differ for different segments of the market, such as between the upper and lower end of the market.

So far we have focused on the results for our benchmark, HM1, model. The adjusted R-squares show that the regressions performed quite well given the small number

[16]Note that though the joint contribution of the quarter dummies are small, the F-test shows that they are in most cases jointly significant.

of characteristics included. The performance of the price models is slightly better than the rent models, as indicated by the R-squares. The regressions are also stable across years, as indicated by the overall performance and the joint contributions of the groups of characteristics. 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. Given this performance, our hedonic approach is expected to control for a large portion of the quality difference between sold and rented houses. However, with the omission of some characteristics, a significant portion of the quality difference between the sold and rented houses may be unaccounted for. Methods for correcting for this were discussed in Section 4.2. Before turning to this issue, we briefly discuss the results obtained for the HM2–HM8 models.

The average adjusted R-squares of the 9 yearly regressions of these models are provided in Table 4. As expected, the explanatory power of these models falls as less characteristics are included, with the smallest model, HM8, explaining 63 and 61 per cent of the variation in prices and rents, respectively. More than 95 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 and more land area are in most cases higher than those found in the HM1 model. This is expected because the estimated coefficients in the HM2–HM7 model 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.

Insert Table 4 Here

The estimated price-rent ratios obtained from the HM1–HM8 models can be directly compared by estimating these 8 models on the dwellings for which all 3 physical characteristics (i.e., number of bedrooms, number of bathrooms, and land area) are available. This process produces 8 different price-rent ratios for each of these dwelling. The difference between them can be attributed to the difference in the model specifications. We compare the correlation between the estimated price-rent ratios derived from HM1 and each of models HM2 through to HM8. The results are shown in Table 5.

The correlation coefficients are close to 1, the closest is between HM1 and HM2, 0.997, obtained from the sold data using the single imputation method. As expected, the correlation coefficients decline as the overlap of characteristics falls. Nevertheless, the high correlations indicate that the price-rent ratios obtained from the models with less characteristics (e.g., HM8) still provide good approximations to the price-rent ratios obtained from the HM1 model.

Insert Table 5 Here

While the correlations show the strength of the linear relationship between the 8 sets of price-rent ratios, they do not reveal whether there are any scale differences between the price-rent ratios. In fact, as discussed in Section 4.2, the price-rent ratios obtained from a smaller model are expected to be on average higher than those obtained from a larger model. This is because the smaller models, which include less characteristics, are unable to adjust for quality differences as much as the larger models. In fact, these scale differences, shown in Table 5, can be interpreted as measuring the relative ability of models to adjust for quality differences. More specifically, they can be interpreted as average measures over the whole data set of $\lambda_{HM2}/\lambda_{HM1}$, $\lambda_{HM3}/\lambda_{HM1}$,. . . ,$\lambda_{HM8}/\lambda_{HM1}$, where λ_{HM8} is defined in (23) and λ_{HMj} for $j = 1, \ldots, 7$ is defined in (25). These adjustment factors are consistent with the list of inequalities in (26) from Section 4.2.[17] For example, in Table 5 we have that the double imputation coefficient for sold and rented dwellings for model HM8 both equal 0.951. This means that a median price-rent ratio initially adjusted for location (as is the case in HM8), gets scaled down by 5.2 percent (i.e., 100(1/0.951)-1) when it is further adjusted for differences between sold and rented dwellings in the number of bedrooms, number of bathrooms and land area.

[17] A slight modification is that HM1 is used here as the reference model rather than HM8. The choice of reference model should not affect these inequalities, as is indeed the case here.

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). The price-rent ratio for each dwelling is imputed from the model that has exactly its mix of characteristics. For example, HM6 is used for a model missing both the bathroom count and land area. However, as is explained in Section 4.2, 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 48,446 dwellings in our data set for which we observe both prices and rents. We have a total of 56,144 selling prices for these dwellings (14.22% of the sold data) and 85,649 rents (17.12% of the rented data), respectively (see Table 6).[18] However, dwellings are typically not sold and rented in the same quarter, and hence they are not matched with regard to the time period. The matching of time periods is attained by extrapolating the prices and rents over time (both backwards and forwards) using price and rent indexes as shown in (20) and (21). The average time span over which prices and rents are extrapolated is 2 and a quarter years, with 90% of the extrapolation done for less than 6 years (the larger the time span the less reliable is the extrapolation).

Insert Table 6 Here

Price and rent indexes can be calculated using a repeat sales and repeat rents approach. The number of houses which are sold and rented more than once within the sample period are 19,239 and 47,609, respectively (corresponding to 40,936 price and 119,182 rent observations (again see Table 6). These dwellings are used to estimate the repeat-sales and repeat-rents indexes (see Section 3.1). The resulting indexes are

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

shown in Table 7. The repeat-sales index shows that house prices in Sydney rose steadily between 2001 and 2003 (by 64 percent), remained almost flat between 2004 and 2006, and then rose slightly towards the end of the sample period. The repeat-rents index shows that the rents remained nearly the same between 2001 and 2003 (in contrast to house prices), then grew steadily until the end of the sample period. Over the whole sample period of 2001-9 house prices and rents grew by around 90 and 40 per cent, respectively.

Insert Table 7 Here

We also estimate price and rent indexes using the F2 double imputation hedonic method defined in (8), with the price and rent for each dwelling imputed from the model HMj with exactly its mix of characteristics. [19] Table 7 shows that while the repeat-sales index exhibits a larger price increase than its hedonic counterpart, the repeat-rent index exhibits a slower price increases than its hedonic counterpart. The difference between the repeat and hedonic indexes may be attributed to a combination of factors, including that the two methodologies are applied on a different set of observations, the potential sample selection bias arising from using only the repeat observations in the repeat regressions and the potential omitted variable bias in the hedonic regressions.[20]

We pool the address-matched price-rent ratios derived from (20) and (21), and then compute the median for each year. This gives us the denominator $AMm(AMs_t)$ in (23). Dividing $HM8m(AMs_t)$ – the median price-rent ratio obtained by imputing prices and rents for the address-matched data set from the HM8 model – by $AMm(AMs_t)$ we obtain an estimate of the omitted variables bias of the price-rent ratios derived from HM8.

Conforming to our expectations, we find that for every year $\lambda_{t,HM8} > 1$, irrespective

[19] The difference between the results obtained using the single and double imputation methods are very small.

[20] We do not attempt here to decompose these differences into their constituent parts, but simply use these indexes to estimate our address-matched price-rent ratios. For a discussion of some of the empirical implications of the choice of method see Clapham et al. (2006), Bourassa et al. (2006) and Hill, Melser and Syed (2009).

of whether the repeat sales/rents or hedonic approach is used to construct the address-matched price rent ratios. The omitted variables adjustment factors for models HM1-HM7 are calculated relative to the adjustment factor of HM8 using (25). The adjustment factors for all eight models for each year are given in Table 8 based on both the repeat sales/rents and hedonic approaches. The results in Table 8 are broadly consistent with the inequalities in (26). While there are some slight inconsistencies for individual years, the average results for each model for both the repeat sales/rents and hedonic approaches correspond exactly with (26).

Insert Table 8 Here

5.3 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 price-rent ratios 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 price-rent ratios against the log of rents, we find that the price-rent ratio increases by 0.10 percent for each percent increase in rents (see Table 9 for the yearly results and corresponding standard deviations). Also, Figure 1 plots the postcode deciles (ordered by median price in panel (a) and by median rent in panel (b)) against postcode median price-rent ratios separately for each year in our sample. This pattern of a rising price-rent ratio for more expensive postcodes can be clearly discerned every year in Figure 1.[21] The difference in the price between the 9th and the 1st deciles is

[21] A similar pattern emerges if we estimate the following equation: $log(price - rentratios) = f(bedroomdummies, bathroomdummies, logoflotsize, postcodedummies)$. Pooling the sold and rental dwellings and the dwellings across years, we find the estimated coefficient on the log of land area to be xx indicating that a one percent increase in land area leads to a xx percent increase in the estimated price-rent ratio (the estimated coefficient is positive in all cases when the same regression is run separately for the data sets and years). The estimated coefficients of bedrooms and bathrooms are generally supportive in this regard, though there are some cases whether the estimated coefficients are insignificant or negative.

43.52 percent when the postcodes are ordered by median price, and 36.39 percent when ordered by median rent.

Insert Table 9 Here

Insert Figure 1 Here

This finding may have implications for our understanding of departures of the housing market from equilibrium. The user cost, and hence the equilibrium price-rent ratio, should in principle be independent of whether a postcode is expensive or cheap. Hence the presence of systematic differences across expensive and cheap postcodes in the price-rent ratio seems to imply a cross-sectional violation of efficiency irrespective of what is happening at the aggregate level. Furthermore, it implies that when the price-rent ratio is above equilibrium for the cheapest postcodes then it will be true for all postcodes. Similarly, if the price-rent ratio is below equilibrium for the most expensive postcodes then this will be true for all postcodes. However, it is also possible that the the price-rent ratio is below equilibrium for cheap postcodes and above equilibrium for expensive postcodes.

Include discussion here of cross-section variation in quality adjustment bias?

5.4 Quality-adjustment bias in price-rent ratios

Raw and quality adjusted price-rent ratios for the lower quartile, median and upper quartile (as measured from the raw price-rent ratios) for each year in our data set are shown in Table 10. The raw price-rent ratios do not make any quality adjustment. They are calculated by simply dividing the median price by the median rent (or upper/lower quartile price by upper/lower quartile rent). As noted earlier, the median dwelling sold tends to be of better quality than the median dwelling rented.

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 20.2 percent larger for the lower quartile, 17.8 percent larger for the median and 12.1 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 17.8 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 17.8 percent.

Insert Table 10 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.

5.5 Quality differences between owner-occupied and rented dwellings

In total, 70 percent of the dwellings sold in our data set are owner-occupied and 30 percent are rented. Given that sold dwellings on average are of 17.8 percent better quality than rented dwellings, an estimate of the average quality difference Z between owner-occupied and rented dwellings is obtained as follows: $Z = (117.8 - 0.3 \times 100)/0.7 = 125.4$. In other words owner-occupied dwellings are on average of 25.4 percent better quality than rented dwellings.

5.6 Equilibrium versus actual price-rent ratios

Our user cost equation in (2) contains the following variables:[22]

r – the risk-free interest rate;

ω – the land tax rate;

δ – the depreciation rate for housing;

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

γ – 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 11. r is a rolling 10-year interest rate. It remained reasonably stable over the 2001-9 period, equalling 5.41 percent in 2001,and 4.09 percent in 2009. (Source: Reserve Bank of Australia)

ω = 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)

δ = 2.5 percent. This is the depreciation rate assumed by Himmelberg, Mayer and Sinai (2005).

γ = 2.0 percent. This is the risk premium assumed by Himmelberg, Mayer and Sinai (2005).

g is the expected nominal capital gain which consists of the sum of the expected real capital gain and expected inflation. The expected real capital gain in year t is assumed to equal the moving average real capital gain over the preceding x years. We consider two different values of x. These are 10 and 25 years respectively. For example, when x is set to 25, the expected real capital gain in 2001 is set equal to the actual real capital gain over the interval 1976-2001. More precisely, the real annual capital gain over the interval (s, t), where s and t are x years apart, is calculated as follows:

$$\text{Real capital gain}(s,t) \; = \left(\frac{EHPI_t/CPI_t}{EHPI_s/CPI_s} \right)^{1/x} .$$

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 Australian Bureau of Statistics (ABS).[23,24] For x=25 years, the implied expected

[23]The Established House Price Index (EHPI) only goes back to 1986. To obtain prices back to 1976 (for the case where x=25), the EHPI was spliced together with an index calculated by Abelson and Chung (2004). See Stapledon (2007) for a discussion of why the Abelson and Chung series is probably the best available option for extending the EHPI back to the 1970s. 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.

[24]The EHPI is computed using the stratified-median approach, which may fail to fully adjust for

annual real capital gain ranges from a peak of 4.5 percent in 2004 to a low of 2.8 percent in 2001 (see Table 11).

The expected rate of inflation is assumed to be 3 percent (which is very close to the average rate of inflation over the period 2001-9 which equalled 3.07 percent).

Insert Table 11 Here

Inserting these values into (2) yields the values shown in Table 11 for u_t and 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 25 years, the equilibrium price-rent ratio ranges between 18.3 and 29.7. When it is 10 year, the range is much larger from 17.8 to 63.0. 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 sensitivity of the equilibrium price-rent ratio to the way expected capital gains are calculated undermines the usefulness of this approach. From an economic perspective, it is not very useful to have the equilibrium price-rent ratio fluctuating wildly over time. Hence, when expected capital gains are calculated from past performance, a long time horizon is needed.

Our median quality-adjusted price-rent ratios are also shown in the last column of Table 11. The quality-adjusted price-rent ratio is above its equilibrium level (focusing on the x=25 case) in every year except 2009. On average, the quality-adjusted price-rent ratio is 8.9 percent above its equilibrium level. By comparison, the quality-unadjusted price rent ratio (taken from Table 10) is on average 29.1 percent above its equilibrium level.

---

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.

5.7 Imputed expected capital gains assuming the housing market is in equilibrium

A better way of using the user cost formula may be to assume the housing market is in equilibrium, and then impute the implied expected capital gain 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, ω, δ, and γ from Table 11 and the quality-adjusted median price-rent ratios from Table 10 into (3), yields the expected capital gains series shown in Table 12. The average expected annual capital gain is 6.83 percent. 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.

Insert Table 12 Here

Is this figure 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 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 1975-Dec 2009: 3.13 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. More generally, in spite of the continuing boom in the mining sector and immigration of high income earners it is hard to believe that Sydney can sustain a real capital gain of 3.83 percent per year. Hence we incline towards the conclusion that the price-rent ratio is above its equilibrium level.

Table 12 also shows the expected capital gain implied by using the unadjusted median price-rent ratio in the user cost formula. Failure to quality adjust results in an implied average expected real capital gain of 4.42 percent (again assuming the expected rate of inflation is 3 percent). This is almost a percentage point above Gyourko, Mayer and Sinai's upper limit for 50 US cities. Failure to quality-adjust therefore leads to a conclusion that the Sydney price-rent is significantly further above its equilibrium level than it actually is.

A similar exercise can be undertaken for segments of the housing market. Suppose we order 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 dwellings sold and likewise for the lower quartile, median and upper quartile dwellings rented. These results are shown in Table 13.

Insert Table 13 Here

The results for the sold and rented data, while differing slightly, exhibit the same pattern. The quality adjusted price-rent ratio is lower for cheaper dwellings than for more expensive dwellings. This in turn implies that households expect higher capital gains at the upper end of the market than they do at the lower end.

6 Conclusion

Failure to quality-adjust median price-rent ratios may cause the housing market to appear to be further from its equilibrium level than it actually is. We estimate the quality-adjustment bias to be about 8 percent for Sydney. Even after quality adjusting we still find that the price-rent ratio in Sydney was above its equilibrium level from 2001 to 2007, although in 2008 and 2009 this was no longer the case. These types of comparisons, however, are inherently problematic since the equilibrium price-rent level, which is derived from the user cost formula, depends critically on the expected capital gain which is not directly observable. A more promising approach, therefore, may be to assume the housing market is in equilibrium and then derive the expected capital gain implicitly. Using this approach we find that the expected capital gain in Sydney is implausibly high. While this may not preclude the market being in equilibrium given the prevailing expectations, it suggests that these expectations are unrealistic and hence not sustainable.

References

Abelson P. and D. Chung (2004), Housing Prices in Australia: 1970 to 2003, Macquarie University Economics Research Paper, No 9/2004, September, Macquarie University.

ABS (2010), Measures of Australia's Progress, 2010, Cat. No. 1370.0. Available at: `http://www.abs.gov.au/ausstats/abs@.nsf/2f762f95845417aeca25706c00834efa/8cba7b65667c8d59ca25779e001c481d!OpenDocument`

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`

AHURI (2010), "Australia's Changing Patterns of Home Ownership," Australian Housing and Urban Research Institute, Research and Policy Bulletin, Issue 133, December 2010. Available at: `www.ahuri.edu.au/publications/download/rap_issue_133`

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.

Bourassa S. C., M. Hoesli and J. Sun (2006),"A Simple Alternative House Price Index Method", *Journal of Housing Economics* 15(1), 80-97.

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

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

Clapham, E., P. Englund, J. M. Quigley and C. L. Redfearn (2006),"Revising the Past and Settling the Score: Index Revision for House Price Derivatives", *Real Estate Economics* 34(2), 275-302.

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

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), "Direct and Indirect Time Dummy Approaches to Hedonic Price Measurement," *Journal of Economic and Social Measurement* 29(4), 427-443.

Diewert W. E. (2003), "Hedonic Regressions: A Review of Some Unresolved Issues," Mimeo.

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

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

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

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.

Griliches Z. (1971), "Introduction: Hedonic Price Indexes Revisited," in Z. Griliches (ed), *Price Indexes and Quality Change*, 3-15. Cambridge MA: Harvard University Press.

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

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), forthcoming.

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

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

Kurz C. and J. Hoffman (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.

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.

Meese R. A. and N. E. Wallace (1997), "The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches," *Journal of Real Estate Finance and Economics* 14, 51-73.

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.

Silver M. and S. Heravi (2001), "Quality Adjustment, Sample Rotation and CPI Practice: An Experiment," presented at the Sixth Meeting of the International Working Group on Price Indices, Canberra, Australia, April 2-6.

Sinai T. and N. Souleles (2005), 'Owner-Occupied Housing as a Hedge Against Rent Risk', Quarterly Journal of Economics, 120, 763-89.

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.

Stone R. (1954), *The Measurement of Consumer Behaviour and Expenditure in the United Kingdom, 1920 1938*, Vol.1, Studies in the National Income and Expenditure of the United Kingdom (assisted by D. R. Rowe, W. J. Corlett, R. Hurstfield, and M. Potter). Cambridge: Cambridge University Press.

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.

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

Yates, J. (2000), "Is Australia's Home-Ownership Rate Really Stable? An Examination of Change between 1975 and 1994," *Urban Studies* 37(2), 319-342.

Figure 1: Quality-Adjusted Price-Rent Ratios by Deciles of Postcodes (Postcodes are ordered from cheapest to most expensive in terms of median price or rent)

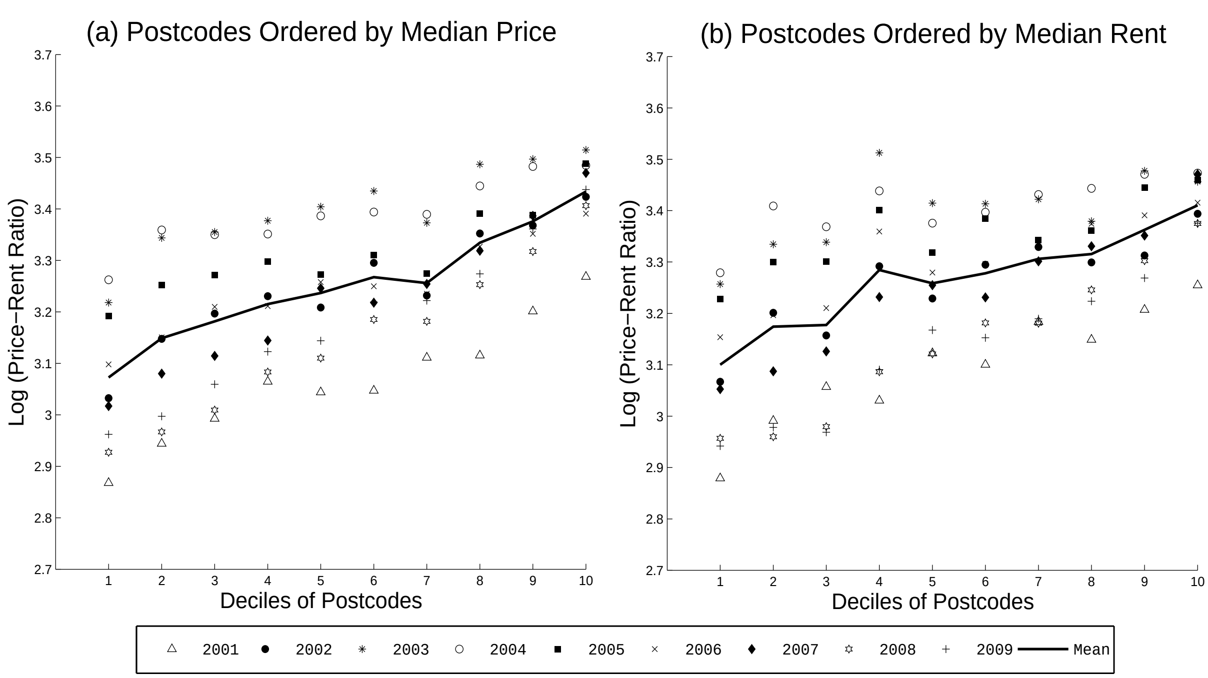

Table 1: Data Description

Statistics	Price Data	Rental Data
Median price or annual rent ($)	495,000.00	16,685.71
Median land area (square meters)	592.00	589.00
Mean land area (square meters)	683.87 (568.67)	640.43 (383.45)
Median num of beds	3	3
Mean num of beds	3.31 (0.84)	3.17 (0.82)
Modal num of beds	3	3
Median num of baths	2	1
Mean num of baths	1.68 (0.74)	1.45 (0.63)
Modal num of baths	1	1
Percentage of observations having the following characteristics:		
Land area, num beds and num baths	62.16	38.05
Num beds and num baths	62.95	41.12
Land area and num baths	62.16	38.05
Land area and num beds	73.72	39.80
Num baths	62.95	41.12
Num beds	74.72	91.36
Land area	98.26	39.83

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

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

Counts	Bedrooms		Bathrooms	
	Price Data	Rent Data	Price Data	Rent Data
1	0.63	1.46	46.68	62.32
2	13.31	15.41	40.11	31.22
3	49.40	53.43	11.63	5.97
4	29.21	23.75	1.40	0.46
5	6.47	5.85	0.18	0.03
6	0.98	0.10	n.a	n.a
Total	100.00	100.00	100.00	100.00

Deciles	Land Area		Postcodes (by Price)*		Postcodes (by Rent)**	
	Price Data	Rent Data	Price Data	Rent Data	Price Data	Rent Data
1st	9.34	11.31	17.56	24.44	16.68	25.93
2nd	10.10	9.71	14.83	16.46	12.68	13.26
3rd	10.04	10.34	11.83	12.19	13.95	14.98
4th	10.62	9.77	8.32	6.85	7.62	5.80
5th	9.49	9.69	9.62	8.15	10.64	8.92
6th	9.85	10.25	8.38	8.08	9.30	7.66
7th	10.25	9.44	8.48	6.97	8.63	7.11
8th	10.07	9.95	7.63	6.53	8.32	6.86
9th	10.30	9.42	8.28	6.41	7.04	5.56
10th	9.93	10.12	5.08	3.90	5.14	3.93
Total	100.00	100.00	100.00	100.00	100.00	100.00

* 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	27795	19294
	(16004)	(18756)
No. of parameters	215	215
	(32)	(32)
Adjusted R^2	0.774	0.739
	(0.035)	(0.035)
Location attributes:		
Joint contribution	0.544	0.525
	(0.035)	(0.035)
% of significant coefficients	0.923	0.808
	(0.018)	(0.096)
Temporal attributes:		
Joint contribution	0.003	0.001
	(0.003)	(0.001)
% of significant coefficients	0.963	0.667
	(0.110)	(0.441)
Physical attributes, including interactive attributes:		
Joint contribution	0.094	0.090
	(0.010)	(0.045)
% of significant coefficients	0.744	0.641
	(0.085)	(0.098)
Physical attributes, only interactive attributes:		
Joint contribution	0.002	0.003
	(0.002)	(0.001)
% of significant coefficients	0.543	0.481
	(0.196)	(0.136)
Contribution of an additional unit of physical attribute:		
Bedrooms: 2 to 3 (Bathroom: 1, Lot size: $750m^2$)	0.097	0.128
	(0.052)	(0.031)
Bedrooms: 3 to 4 (Bathroom: 1, Lot size: $750m^2$)	0.109	0.063
	(0.022)	(0.031)
Bathrooms: 1 to 2 (Bedroom: 3, Lot size: $750m^2$)	0.098	0.094
	(0.009)	(0.033)
Bathrooms: 2 to 3 (Bedroom: 3, Lot size: $750m^2$)	0.142	0.072
	(0.022)	(0.074)
Lot size: by $100m^2$ (Bedroom: 3, Bathroom: 1)	0.027	-0.001
	(0.021)	(0.002)

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.

Table 4: HM2–HM8 Regression Results

Statistics	HM2	HM3	HM4	HM5	HM6	HM7	HM8
Price Models							
Adjusted R^2	0.754	0.759	0.741	0.736	0.717	0.659	0.630
	(0.047)	(0.028)	(0.039)	(0.039)	(0.053)	(0.032)	(0.036)
% of coefficients having	0.932	0.983	0.988	0.924	0.918	0.963	0.903
the same sign as HM1	(0.051)	(0.010)	(0.012)	(0.051)	(0.059)	(0.019)	(0.047)
Contribution of an additional unit of physical attribute:							
Bedrooms: 2 to 3	0.130	n.a	n.a.	0.137	0.169	n.a	n.a.
(Bathroom: 1, Lot size: $750m^2$)	(0.058)	(n.a.)	(n.a.)	(0.028)	(0.034)	(n.a.)	(n.a.)
Bathrooms: 1 to 2	n.a.	0.182	n.a.	0.100	n.a.	0.198	n.a.
(Bedroom: 3, Lot size: $750m^2$)	(n.a.)	(0.031)	(n.a.)	(0.015)	(n.a.)	(0.016)	(n.a.)
Lot size: by $100m^2$	0.026	0.030	0.030	n.a.	n.a	n.a	n.a.
(Bedroom: 3, Bathroom: 1)	(0.020)	(0.026)	(0.024)	(n.a.)	(n.a)	(n.a.)	(n.a.)
Rent Models							
Adjusted R^2	0.735	0.705	0.712	0.701	0.697	0.649	0.610
	(0.038)	(0.022)	(0.021)	(0.021)	(0.028)	(0.011)	(0.007)
% of coefficients having	0.990	0.964	0.970	0.973	0.958	0.938	0.921
the same sign as HM1	(0.023)	(0.019)	(0.020)	(0.016)	(0.044)	(0.025)	(0.036)
Contribution of an additional unit of physical attribute:							
Bedrooms: 2 to 3	0.150	n.a	n.a.	0.130	0.166	n.a.	n.a.
(Bathroom: 1, Lot size: $750m^2$)	(0.040)	(n.a.)	(n.a.)	(0.030)	(0.029)	(n.a.)	(n.a.)
Bathrooms: 1 to 2	n.a	0.184	n.a.	0.092	n.a.	0.182	n.a.
(Bedroom: 3, Lot size: $750m^2$)	(n.a.)	(0.061)	(n.a.)	(0.031)	(n.a.)	(0.057)	(n.a.)
Lot size: by $100m^2$	-0.000	-0.001	0.002	n.a.	n.a.	n.a	n.a.
(Bedroom: 3, Bathroom: 1)	(0.001)	(0.002)	(0.003)	(n.a.)	(n.a.)	(n.a.)	(n.a.)

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.

Table 5: Comparisons of the Price-rent Ratios Between HM1 and HM2–HM8 Models

Data	Imputation Method	HM2	HM3	HM4	HM5	HM6	HM7	HM8
Correlation coefficients (HM1 versus another model):								
Sold Houses	Single imputation	0.971	0.974	0.910	0.997	0.969	0.973	0.908
	Double imputation	0.984	0.964	0.943	0.903	0.887	0.887	0.871
Rented Houses	Single imputation	0.976	0.958	0.901	0.932	0.903	0.895	0.832
	Double imputation	0.988	0.960	0.940	0.844	0.829	0.824	0.809
Ratio of median price-rent ratios (HM1 over another model):								
Sold Houses	Single imputation	0.999	0.992	0.984	0.990	0.984	0.953	0.951
	Double imputation	0.994	0.991	0.979	0.982	0.974	0.956	0.949
Rented Houses	Single imputation	0.998	0.992	0.986	0.989	0.985	0.953	0.951
	Double imputation	0.995	0.993	0.986	0.986	0.983	0.957	0.952

Note: Sold and rented houses include observations where all three physical characteristics are available. These houses consist of 62.16 and 38.05 percent of total sold and rented observations, respectively.

Table 6: Summary Description of the Matched, Repeat-sales and Repeat-rent Observations

Statistics	Matched Sales and Rents	Repeat Sales	Repeat Rents
No. of observations	141,793	40,936	119,182
No. of dwellings	48,446	19,239	47,609
No. of postcodes	245	242	248
Median price	462,437	525,000	n.a.
Mean price	585,320	663,705	n.a.
	(401,788)	(448,411)	(n.a.)
Median rent	18,250	n.a.	16,686
Mean rent	21,582	n.a.	19,835
	(12,630)	(n.a.)	(11,198)
Mean lot size	670	622	713
	(626)	(484)	(743)
Modal bedroom	3	3	3
Modal bathroom	1	1	1
Mean period of mis-match (in years)	2.27		

Table 7: Median and Quality-Adjusted Price and Rental Indexes

Year	Price Indexes			Rental Indexes		
	Median	Repeat-sales	Hedonic	Median	Repeat-rents	Hedonic
Dec-2000	1.00	1.00	1.00	1.00	1.00	1.00
Dec-2001	1.16	1.14	1.14	0.97	0.99	0.99
Dec-2002	1.36	1.41	1.36	0.95	0.99	1.01
Dec-2003	1.57	1.64	1.54	0.97	1.01	1.04
Dec-2004	1.54	1.60	1.52	1.00	1.03	1.06
Dec-2005	1.51	1.59	1.48	1.07	1.06	1.13
Dec-2006	1.51	1.58	1.49	1.10	1.12	1.21
Dec-2007	1.44	1.70	1.60	1.22	1.24	1.34
Dec-2008	1.26	1.70	1.53	1.37	1.34	1.47
Dec-2009	1.56	1.90	1.70	1.47	1.40	1.52

Table 8: Omitted Variables Adjustment Factors λ_{HMj}

(a) Address-matched benchmarks constructed using repeat sales/repeat rents indexes

	HM1	HM2	HM3	HM4	HM5	HM6	HM7	HM8
2001	1.080	1.083	1.080	1.123	1.081	1.123	1.133	1.145
2002	1.089	1.092	1.088	1.126	1.089	1.134	1.140	1.149
2003	1.051	1.052	1.052	1.088	1.056	1.093	1.106	1.114
2004	1.046	1.061	1.045	1.086	1.061	1.099	1.105	1.113
2005	1.056	1.061	1.054	1.083	1.056	1.089	1.097	1.097
2006	1.069	1.073	1.070	1.099	1.073	1.103	1.108	1.113
2007	1.048	1.041	1.060	1.070	1.054	1.069	1.100	1.098
2008	1.027	1.026	1.037	1.045	1.035	1.048	1.079	1.074
2009	0.990	0.984	0.996	1.004	0.992	1.008	1.030	1.025
Average	1.051	1.053	1.054	1.080	1.055	1.085	1.100	1.103

(b) Address-matched benchmarks constructed using hedonic DI indexes

	HM1	HM2	HM3	HM4	HM5	HM6	HM7	HM8
2001	1.053	1.056	1.053	1.095	1.053	1.095	1.104	1.116
2002	1.083	1.087	1.082	1.121	1.084	1.128	1.134	1.142
2003	1.057	1.059	1.059	1.095	1.063	1.101	1.114	1.122
2004	1.055	1.070	1.054	1.096	1.070	1.109	1.114	1.123
2005	1.064	1.070	1.062	1.091	1.064	1.097	1.105	1.106
2006	1.054	1.058	1.056	1.084	1.058	1.087	1.093	1.098
2007	1.038	1.032	1.050	1.060	1.044	1.059	1.089	1.087
2008	1.030	1.029	1.041	1.048	1.039	1.052	1.082	1.077
2009	1.005	1.000	1.012	1.021	1.008	1.025	1.046	1.041
Average	1.049	1.051	1.053	1.079	1.054	1.084	1.098	1.101

Note: From equation 25, $\lambda_{t,HMj} = \lambda_{t,HM8} \times \lambda_{t,HMj|HM8}$. Since $\lambda_{t,HM8|HM8} = 1$, $\lambda_{t,HMj}$ equals $\lambda_{t,HM8}$ of equation 23. $\lambda_{t,HMj|HM8}$ of equation 24 can be obtained by dividing the HMj column by the $HM8$ column.

Table 9: Percentage Change in Price-Rent Ratios due to 1% Change in Price or Rent

Δ in Price or Rent	2001	2002	2003	2004	2005	2006	2007	2008	2009	Pooled
Price	0.213	0.201	0.179	0.136	0.132	0.145	0.188	0.209	0.211	0.211
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Rent	0.200	0.170	0.100	0.076	0.106	0.143	0.187	0.210	0.224	0.102
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)

Note: The figures are the estimated coefficients when the log of the quality adjusted price-rent ratios are regressed on the log of price or the log or rent. The figures in the brackets are the estimated standard errors.

Table 10: 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	18.61	20.75	23.52	20.95	23.47	17.68
2002	27.51	30.42	32.41	21.88	24.77	28.12	25.73	22.84	15.24
2003	31.69	34.72	36.39	25.76	29.01	32.63	23.01	19.67	11.51
2004	32.76	35.48	37.45	26.50	29.33	32.25	23.65	20.97	16.12
2005	30.69	33.41	34.93	24.91	27.15	29.87	23.18	23.07	16.96
2006	29.53	32.06	34.08	23.26	25.54	28.41	26.98	25.56	19.96
2007	25.22	27.45	30.17	21.37	23.74	27.09	18.00	15.62	11.36
2008	21.77	23.01	25.28	19.16	21.38	24.58	13.65	7.65	2.87
2009	20.63	21.85	24.04	19.42	21.61	24.68	6.22	1.11	-2.59
Average	26.92	29.34	31.38	22.32	24.81	27.90	20.16	17.77	12.13

Table 11: Equilibrium Price-Rent Ratios Derived from User Cost Formula

	r_t	w_t	δ_t	γ_t	π_t^e	g_t (x=10)	g_t (x=25)	P_t/R_t (x=10)	P_t/R_t (x=25)	P_t/R_t Actual
2001	0.054	0.01	0.025	0.02	0.03	0.025	0.028	18.64	19.46	20.75
2002	0.059	0.01	0.025	0.02	0.03	0.038	0.036	22.09	21.00	24.77
2003	0.053	0.01	0.025	0.02	0.03	0.054	0.042	42.84	27.94	29.01
2004	0.057	0.01	0.025	0.02	0.03	0.066	0.045	62.55	26.78	29.33
2005	0.054	0.01	0.025	0.02	0.03	0.059	0.035	51.56	23.13	27.15
2006	0.052	0.01	0.025	0.02	0.03	0.056	0.029	46.48	20.76	25.54
2007	0.059	0.01	0.025	0.02	0.03	0.053	0.031	32.80	19.05	23.74
2008	0.061	0.01	0.025	0.02	0.03	0.048	0.037	26.55	20.48	21.38
2009	0.041	0.01	0.025	0.02	0.03	0.034	0.032	31.21	29.71	21.61

Table 12: Implied Expected Capital Gains Obtained from the User Cost Formula

	Using Quality Adjusted P_t/R_t	Using Actual P_t/R_t
2001	6.09	7.00
2002	7.32	8.07
2003	7.33	7.89
2004	7.79	8.38
2005	7.17	7.86
2006	6.79	7.58
2007	7.17	7.73
2008	6.90	7.23
2009	4.96	5.01
Average	6.83	7.42

Table 13: 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	19.37	22.30	22.82	19.02	20.57	22.38
2002	21.65	25.01	26.49	22.44	24.73	26.72
2003	27.76	29.31	31.86	26.75	29.54	30.81
2004	28.16	29.56	30.10	28.02	29.70	29.81
2005	26.64	27.10	27.43	26.21	27.41	28.30
2006	23.70	25.95	29.36	23.95	25.17	26.20
2007	21.71	25.98	25.78	21.51	23.31	26.02
2008	20.14	21.10	24.01	19.43	21.22	23.49
2009	19.79	21.10	25.81	19.45	21.43	23.96
Average	23.21	25.27	27.07	22.98	24.79	26.44