Deriving a Rent-to-Price Ratio in Residential Markets:
A comparison of methods

DRAFT WORK. PLEASE DO NOT CITE

Andy Krause (Contact Author)

Department of Property

Faculty of Architecture, Building and Planning

University of Melbourne

andy.krause@unimelb.edu.au

&

Gideon Aschwanden

Department of Urban Analytics

Faculty of Architecture, Building and Planning

University of Melbourne

gideon.aschwanden@unimelb.edu.au

March 25, 2016

Abstract

In studies of residential real estate markets the relationship between rents and prices are an oft-used indicator of both investment potential as well as bubble identification. Most often, this relationships – referred to as rent-price ratio, price-rent ratio or rental yield – are constructed via simple metrics such as medians or means using spatially aggregated data. In such cases, inconsistencies between the types of properties that rent and sell as well as the constant quality differences between similar sales and rentals of a given property type may lead to biased measures. In this study we use a unique set of transaction-level observations - residential sales and rentals - in the Melbourne Metropolitan Area (Australia) to test for differences in the measured relationship between rents and prices using four different estimation techniques. More specifically we compare relatively simple spatial aggregation and index methods to a hedonic regression imputation method and a direct matching method. Our results show that, in general, the rent-price ratios (rental yields) suggested by the median and index methods are biased on the low side (up to 20%) when compared with direct matches of properties that have sold and rented (or vice versa) within a short time period. These biases have increased over the past five years suggesting that the mix of rental and sold units has changed during this period. Additionally, in a test to determine which method offers the best predictive accuracy of future sales prices and rental rates, the direct match method fares the best, followed closely by hedonic imputation.

Introduction

The relationship between home sale prices and rents is a fundamental metric in the analysis of housing markets. The specific method and related data used to estimate this relationship (the 'rent-price ratio') varies, with the choice of analytical method often tied closely to the type and spatial granularity of available data. In reviewing the extant literature and industry practices, we have identified four primary methods in use: 1) spatial median (or mean); 2) index; 3) hedonic imputation; and 4) direct match. The first two, spatial median and index, are aggregate methods that do not require property-level observations. Hedonic imputation and direct matching, on the other, are diffuse metrics that are able to offer estimates at a fine spatial scale. With this particularity, however, comes the requirement of transaction-level data.

Homes that sell and those that rent often differ in type, quality and location. As a result, the potential for the aggregate methods – patial median and index – to suffer from bias related to these differences has been long established in the literature (Himmelberg et al., 2005; Glaeser and Gyourko, 2007; Verbrugge, 2008; Garner and Verbrugge, 2009). Due to these perceived biases, more recent research has developed and utilized the diffuse methods – hedonic imputation or direct match methods(Smith and Smith, 2006; Hattapoglu and Hoxha, 2014; Bracke, 2015; Hill and Syed, 2016). What is not clear from this research, however, is the extent

to which, if any, the actual rent-price ratios estimated by aggregate methods differs from those derived by imputation or matching.

In this paper, we set out to compare results from the four available methods over a variety of geographic levels of observation and over a period where the housing market performance (appreciation rates) varied markedly. We also then compare the four methods for accuracy in predicting sales prices and rental values. Overall, our intent is three-fold: 1) To determine if the rent-price ratio estimates differ by method; 2) To determine if the geographic level of analysis and/or the overall market conditions explain the differences, if any, and 3) To gauge which method produces the most accurate predictions of prices and rent.

We conduct this analysis using a dataset of more than 710,000 sales and rental observations from the Melbourne (Australia) Metropolitan region over the January 2011 to December 2015 time period. Our findings indicate that there are large differences in rent-price ratio estimates between the four methods of analysis, with the spatial median and index methods offering the lowest estimates and the direct match the highest. The differences vary across the various dimensions of analysis noted above, but average around 20%. We also find that as the geographic level of aggregation shrinks, so to does the differences between the methods. Additionally, the differences between the aggregate and the diffuse methods have varied over time, with the absolute level of difference increasing markedly in the 2014 to 2015 period. Finally, our analysis of predictive accuracy shows the matching method to the be the most accurate, followed by the imputation method. The aggregate methods offer a reduced accuracy in predictive ability.

This paper continues as follows. Section two presents a review of existing methods and their uses, with a focus on both techiques as well as the end use of the metric. A discussion of the four methods tested in this paper along with an explanation of the data is found in the third section. Results of the method comparisons and the predictive accuracy tests follow. This paper concludes with a discussion of the results as well as suggestions for future research in this direction.

Literature Review

A common metric in a variety of housing analyses is the relationship between home prices and home rents. Depending on the discipline or use, it may be referred to (and calculated as) 'price-to-rent ratio' (Glaeser and Gyourko, 2007; Goetzmann et al., 2012; Himmelberg et al., 2005), 'rent-price ratio' (Bracke, 2015; Campbell et al., 2009), 'rental yield' (Fu and Ng, 2001) and 'dividend price ratio' (Hwang et al., 2006). Within this study we will use the term 'rent-price ratio.' Being expressed as such, the annual rental value is the numerator and the sale price is the denominator.

Rent-price ratios (RPRs) have multiple uses in the literature. One of the more common purposes is in the

testing of market fundamentals and/or equilibrium (Capozza and Seguin, 1996; Ayuso and Restoy, 2006; Ambrose et al., 2013; Sommer et al., 2013; Hill and Syed, 2016). When used for to this end, the rent-price ratio is often expressed as the dependent variable indicating market equilibrium or the lack thereof. Relatedly, RPRs are also employed in the identification of periods of mispricing or bubbles (Himmelberg et al., 2005; Smith and Smith, 2006; Brunnermeier and Julliard, 2008; Pavlidis et al., 2013). Here RPRs generally serve as independent or explanatory variables aimed at explaining bubble mechanisms, however, occasionally the RPR itself, or the change in it, is offered as evidence of mispricing in the market.

Studies on expected housing returns may also call on RPRs to help explain how expectations of housing return or future captial appreciation can impact market decisions (Gelain and Lansing, 2014; Engsted and Pedersen, 2015; Hattapoglu and Hoxha, 2014). Additionally, rent-price ratios are also used as independent or explanatory variables in studies explaining mortgage supply (Goetzmann et al., 2012), house prices levels (Hwang and Quigley, 2006), spatial arbitrage opportunities (Glaeser and Gyourko, 2007) and user-cost ratios (Verbrugge, 2008).

More recently, there has been increased interest in both describing the trends that explain movements in the rent-price ratios (Campbell et al., 2009) as well as improving the methods used to calculate the RPR estimates themselves (Hattapoglu and Hoxha, 2014; Bracke, 2015; Hill and Syed, 2016). In the vast majority of studies, RPRs are calculated by comparing a median or mean rent for a given area to a median or mean sale price for the same area or by comparing indexes of prices and rents over time. These both serve as 'aggregate' methods of calculation. While this method is computationally cheap and the data is often readily available for consumption, these two aggregate methds do suffer from a number of issues, including the the fact that the type, quality, condition and location of homes that sell are often vastly different from those that rent (Bracke, 2015; Hattapoglu and Hoxha, 2014).

To address this, and other issues discussed below, two additional methods have been used to create alternate estimates of rent-price ratios using property-level observations: 1) Hedonic Imputation; and 2) Direct Matching. The imputation method utilizes hedonic price models of rents and sales prices to create an imputed sales price for rental observations and an imputed rental amount for sales observations (Hattapoglu and Hoxha, 2014; Hill and Syed, 2016). The matching method pairs up sales and rental observations of the same property to determine the RPR (Hwang et al., 2006; Smith and Smith, 2006; Bracke, 2015).[1] More information on these methods are given in the 'Method' section that follows.

In review of the literature, rent-price ratios are used in a variety of ways in the sphere of housing economics. Traditionally, spatial aggregation or index methods of calculation dominated; however, recent literature suggests that these methods are biased and have propsed new, though data-heavy, techniques to improve

---

[1]Note that Smith and Smith's (2006) analysis actually used proximity matching instead of direct matching due to a limited number of direct matching pairs.

on the estimation of RPRs. Though these new methods have been presented and tested, no comparison to date shows the level of bias, if any, that exists when using the median method and whether or not additional clarifications such as bifurcation by property type or use of smaller geographic units can help negate issues in the median method. As a result, this paper seeks to compare the three methods used in the existing literture with a single dataset to determine the existance, and, potentially, the level of difference in the methods.

Method

Four primary methods are used to construct rent-price ratios: 1) Spatial Median; 2) Index; 3) Hedonic Imputation; and 4) Direct Matching. The first two can be considered aggregate metrics while the latter two represent diffuse measures of rent-price ratios. In discussing each method there there are four issues to consider. The first is data availability. The aggregate methods – spatial median and index – do not require property-specific observations and, therefore, are much more easily calculated. This issue is particulary important in studies that seek to compare RPRs across multiple metropolitan areas or countries.

In addition to data concerns, there are also three related estimation issues that frequently arise. The first is the issue of constant quality. If, for instance, a given geographic area sees a sudden increase of new detached homes for sale and there is no related change in the quality of rent units available then the resulting decrease (assuming the new homes are more expensive than existing homes) in the RPR is not driven by any change between the actualy relationship of rents and prices on a particular home, but rather by a change in the overall make up of homes in the area.

Next, is the issue of tenure mix location. In a given geographic area for which an RPR is being calculated if the rental properties are located in less desirable locations than the owner occupied properties then the RPR is not accurately measuring the the relationship between rent and price for an given home, as the differences in location values are biasing the figure. The smaller or more homogenous the geographic area being examined the less likely tenure mix location bias is to occur.

Finally, is the issue of omitted variables. If the information available on individual homes and transactions do not include variables that may differ between rental and sold properties then RPRs are likely to be biased. For example, many datasets do not include complete or accurate information on a property's construction quality and current condition. If, correspondingly, the quality and/or condition of homes that sell in the market are different from those that rent then RPRs can be biased. Quality and condition are two of the most commonly omitted variables in pricing model datasets, but others such as home age, number of bedrooms or bathrooms can also causes bias in RPR calculations.

The methods are presented in this order as each method corrects for one major issue not controlled for by the

prior method. Table 2 illustrates the differences in terms of data required and estimation issues controlled. Below, each is described as used in the comparative analysis that follow in the 'Results' section.

Table 1: Price-Rent Ratio Methods and Key Issues

Method	Type	Data	Quality	Mix/Loc.	Omit.
Spatial Median	Aggregate	✓			
Index	Aggregate	✓	✓		
Impute Regression	Diffuse		✓	✓	
Direct Match	Diffuse		✓	✓	✓

Spatial Median

Calculating the spatial median (or mean) method is straightword. For a given geographic area at time period t, divide the median (or mean) house rent, r, by the median (or mean) house price, p. The resulting dividend is the rent-price ratio. Due to the fact that economic observations such as home prices and rent are often right-skewed, the median is more commonly used than the mean.

$$RPR_t = \frac{median(R_t)}{median(P_t)} \tag{1}$$

A benefit of this method is the ease of calculation as well as the low data requirements since only aggregated values (median or means) at given geographic level are necessary. As most government or industry analysts publish aggregated median price and rent values, the median method can be quickly calculated for a given geographic area and, usually, over a long time period as historic data for aggregrate price, and to a less extent, rental trends are normally available. The ease of calculation comes with a considerable cost. The spatial median method does not control for constant quality, tenure mix location or omitted variable diffrences between houses that sell and those that rent.

Index

The index method for estimating rent-price ratios is, like the spatial median, an aggregate metric. Index calculations involve comparing a rent index value at at given time for a given geographic area to a price index value at the same time for the same area. Presuming that the actual rent-price ratio is known at the base time period of the index, then the RPR can be calculated by transforming the index values into nomimal rent and price series which can then be compared to generate an RPR. This is shown by:

$$RPR_t = \frac{BV_{r,0} * IV_{r,t}}{BV_{p,0} * IV_{p,t}} \tag{2}$$

where $BV_{r,0}$ is the nominal base (time $= 0$) rental value of the index, $BV_{p,0}$ is the nominal base (time $= 0$) price of the index, $IV_{r,t}$ is the rental index value in geographic area at time t, and $IV_{p,t}$ is the price index value at time t.

Hedonic Imputation

The hedonic imputation approach uses property-level transaction data to impute the likely rental amount for sales and likely sale price for rentals. More specifically, a hedonic price model is built using all sales observations and then used to predict the likely sale price of all rental observations at the time of the rental contract. Likewise, the same is done using all rental observations to predict the likely rental amount at the time of the sale. This process creates a sale price and rental amount for each observation in the dataset. For sales, the sale price is observed and the rental amount is imputed; for rentals the rental amount is observed and the sale price is imputed. The relationship between these two values represents the observation specific (as of the data of the observed transaction) rent-price ratio. These property specific rent-price ratio estimates are then aggregated at the desired geographic level and a measure of central tendency (usually a median) is taken.

For an observed sale, the rent-price ratio is expressed as:

$$RPR_{i,t} = \frac{Rpred_{i,t}}{SP_{i,t}} \tag{3}$$

where $RPR_{i,t}$ is the rent-price ratio for property i at time t (the time of the sale), $SP_{i,t}$ is the observed sale price at time t and $Rpred_{i,t}$ is the predicted or imputed rental value at time t expressed by the following standard hedonic price or rent equation:

$$Rpred_{i,t} = \alpha + \beta X_i + \epsilon_i \tag{4}$$

where X_i is the vector of independent variables for property i, β is a vector of estimated coefficients and ϵ_i is an error term.

Likewise, for an observed rental, the rent-price ratio is expressed as:

$$RPR_{i,t} = \frac{R_{i,t}}{SPpred_{i,t}} \tag{5}$$

where $RPR_{i,t}$ is the rent-price ratio for property i at time t (the time of the sale), $R_{i,t}$ is the observed rental value at time t and $SPpred_{i,t}$ is the predicted or imputed sale price at time t expressed by the following

7

standard hedonic price or rent equation:

$$SPpred_{i,t} = \alpha + \beta X_i + \epsilon_i \tag{6}$$

where X_i is the vector of independent variables for property i, β is a vector of estimated coefficients and ϵ_i is an error term.

The estimated rent-price ratios are then pooled at the geographic level desired and an overall rent-price ratio is estimated by taking a measure of central tendency for area j at time t.

The hedonic imputation method controls for both changes to constant quality of homes as the RPR is estimated at the property level. This method also controls for issues stemming from heterogenous tenure mix by location, again due to the fact that property level estimates are made. Omitted variables, especially those that are likely to differ between rental and owner occupied (sold) property do still remain an issue. For example, if all rental properties in an area are in poorer condition (an unobserved variable) then all imputed or predicted rental values are likely to be biased on the low side resulting in an RPR lower than actual.

Direct Matching

A final method used to calculate rent-price ratios is the direct matching approach. In this approach only properties which have both sold and rented in a given time period are considered in the calculations. As sales and rentals are rarely, if ever, simultaneous, the time between the two observations must be adjusted for. In the case of a sale that then rented sometime afterwards, the rental amount is adjusted back in time via the rent index to the time of the sale. Likewise, for a property which rented and then subsequently sold, the sale price is adjusted back in time to the date of rental using the sale price index. The adjustments are generally made through externally derived price and rent indices. In the analysis that follows we have adjusted each pair of rental-sale or sale-rental by the individual price and rent indices constructed at the level of geographical analysis. For example, in calculation done at the local government area (LGA) level, sales and rentals are adjusted with price and rent indices calculated at the LGA level. For suburb calculation we use suburb level adjustment indices, etc.

For any pair of sale-rental an RPR is calculated at both the time of the sale and at the rental. In other words, each matched pair yields two observations of RPR. All potential pairs of sales and rentals are utilized so a property that sold twice and rented three times, for instance, would be represented six times in the matched dataset, once for each pairing of a sale with a rental. This example would then yield 12 instance of RPR estimates, two for each pairing.

The rent-price ratio for the rental observation in a matched pair is represented by:

$$RPR_{i,t} = \frac{R_{i,t}}{SPadj_{i,t}} \tag{7}$$

where $RPR_{i,t}$ is the estimated rent-price ratio for property i at time t (when it was rented), $R_{i,t}$ is the value that it was rented at and $SPadj_{i,t}$ is the sale price adjusted by a market index to time t.

Similarly, the rent-price ratio for the sale observation in a matched pair is represented by:

$$RPR_{i,t} = \frac{Radj_{i,t}}{SP_{i,t}} \tag{8}$$

where $RPR_{i,t}$ is the estimated rent-price ratio for property i at time t (when it was sold), $Radj_{i,t}$ is the rental value adjusted by a market index to the time at which the property sold and $SP_{i,t}$ is the sale price at time t.

An obvious downfall to the matching approach is the greatly reduced set of data that is available for estimation. Similar to a repeat-sales estimator used in housing index creation, the matching method only uses properties for which more than one observations is recorded – in this case at least one sale and one rental. However, this reduction in sample size is balanced against the fact that direct, observed rental and sales prices for the same property are measured. As such, the direct match method does not suffer from any biases that may arise do to omitted variables on home type, quality or condition.[2]

Other Considerations

In addition to the choice of analytical method, two other considerations or dimensions of analysis must be considered when estimating rent-price ratios. The first is in regard to how to treat different property uses such as houses versus apartments/condominiums ('units' in Australia). A rent-price ratio analysis could merge the two or estimate the two separately. As the buyer and renters of houses likely differ from those of units, there is a good rationale for separating the two. Separation, however, is only possible if the data exists to do so. In the spatial median and index methods, this may or may not be available in all data sets. In the case of imputation and direct matching, any observation-specific dataset that provides the data necessary to conduct a hedonic pricing model and/or direct property matches should contain enough information to identify the correct property use of the observation.

In the analysis that follows we do have property use indicators through all of our data. Exploratory analylyis showed a large difference in the rent-price ratios as well as price and rent trends between houses and

---

[2]To the extent that homes have not been significantly remodeled or drastically depreciated between the two observations.

9

apartments/condominiums (units) in our dataset. As a result, we have conducted all analyses with the two property uses split. [3].

The second consideration to be made in a rent-price ratio analysis is the geographic level of aggregation. Much of the existing research presented in the literature review analyzes trends at fairly coarse geographic levels such as at the metropolitan region, or even the nation. As home vintages, property use (home vs. unit), condition, quality and value vary widely over space, there is likely to be markedly different estimates of RPRs depending on the size of the spatial aggregation used. Finer spatial aggregations are hypothesized to created more accurate yield estimates but also have more variability in the estimates themselves due to the decrease in sample size as the size of the geographic aggregration decreases.

To highlight the variability in rent-price ratio estimates over space, we have estimated quarterly ratios at a variety of successively finer levels of geographic aggregation. Beginning with the entire metropolitan region, we then calculated RPRs at the level of local government area (LGA), statistical local area 1 (SLA1), postcode and suburb, in decreasing order of size. Suburbs in Australia are often much smaller than the similarly named entities that exist in North America. Within Australia, local governance is carried out at the LGA level, with each LGA consisting of generally 6 to 12 different suburbs. Large LGAs located at the fringe of metropolitan areas often have more than 12 suburbs within their jurisdiction. Suburbs do not always fit perfectly within an LGA. Within the dataset described below we have 33 LGAs and 288 suburbs with enough data for computation. See Figure 1 for a location of the LGAs (largest units) and the suburbs (smallest units) in our analysis.

---

[3]Properties labeled as townhomes, villas, or terraces are treated as houses, while studios are treated as units. These use type represent approximately 8.5% of the cleaned observations)
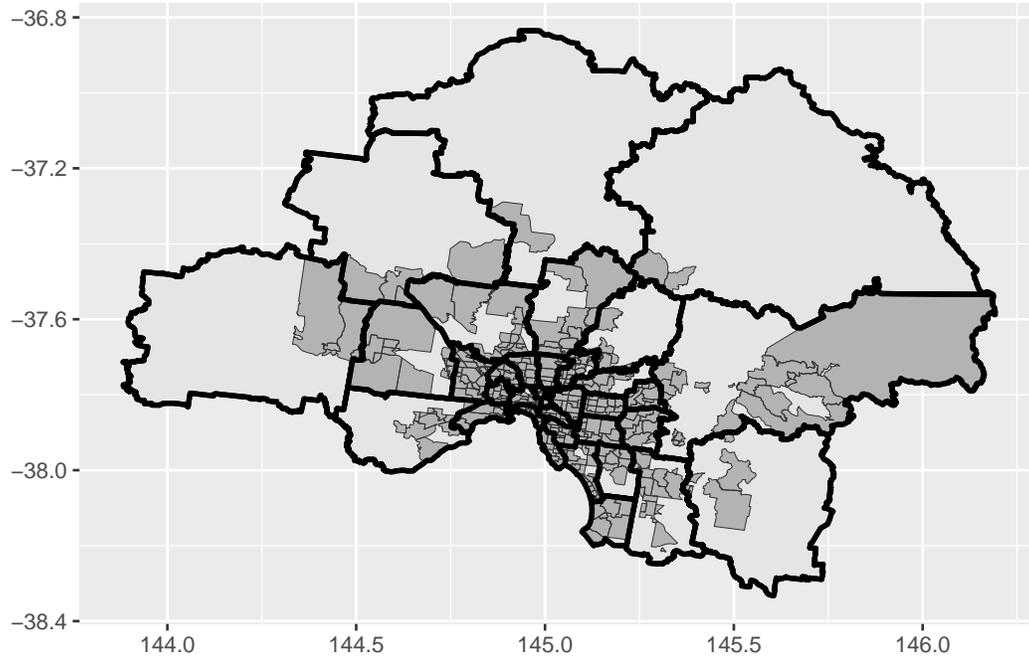
Figure 1: Map of Study Area LGAs and Suburbs

Data

To test the three models discussed above, we utilize property-level observations of home sales prices and rental lease amounts obtained from Australian Property Monitors (APM).[4] This data set includes all transactions of single family homes and units in the Melbourne Metropolitan region over the January 2011 to December 2015 time period. After removing incomplete and potentially suspect observations, more than 710,000 individual transactions remained – greater than 235,000 sales and 475,000 rentals. In addition to the transaction date and price, the data includes more than 15 individual physical attributes of the property along with the property address, latitude and longitude.

In the direct match method we match properties that have both sold and rented during our study period. Due to these constraints, the direct match method uses a much smaller set of data (140,000 vs. 910,000) that the median and imputation methods. However, this relatively large number of matched properties (1/6 of total) is due to the fact that in the Australian residential markets nearly all apartment units are owned by individuals and subsequently rented out on the private market. Additionally, the leasing of a single family home is more common than in North American markets.

---

[4]http://www.apm.com.au

Table 2: Data Overview

Data Type	Houses	Units	Total
Transaction Data			
Sales	191,908	44,806	236,714
Rentals	253,447	222,958	476,405
All	445,355	267,764	713,119
Matched Observations			
All	81,804	38,597	120,401

Comparitive Results

We begin by calculating and comparing the rent-price ratio for houses and units over the entire Melbourne metropolitan region using the four method described above. As Figure 2 shows, the spatial median and index method estimates for houses (left panel) are considerably lower than the those for the hedonic imputation which are in turn lower than the direct matching method. The relationship between the matching and imputation method stays relatively constant from 2011 to 2013, after which the rent-price ratio estimates for the hedonic imputation method drop off much more quickly than those of the matching. The gap between the spatial median and index methods slowly widens over time as well. Units see a similar story, with the direct matching method giving the largest RPR estimates and the spatial median and index methods the smallest. The differences between methods are similar on an absolute value scale to those found in the houses, however, theses differences remain relatively constant over time as opposed to the house market where the gap widens as time passes.

Overall, the housing market in Melbourne has appreciated markedly from 2013 through to the end of the study period (January 2015). The widening gap in the results over time suggests that the differences between the methods may be, at least partially, related to the performance of the housing market in general. Figures 3 and 4, illustrate the differences between rent-price ratio estimate for first three methods and the matching method plotted again overall market appreciation in the quarter.[5] For houses (Figure 3), we see that the differences between the spatial median (left panel) and index (middle panel) do show a positive relationship to market performance. The relationship between the differences in the imputation method and the direct matching method appear ambiguous. Similar trends (Figure 4) are shown when looking into the differences between methods when examining units in the market.

---

[5]The measure of quarterly appreciation is derived from a constant quality adjusted measure of median home value, smoothed for seasonality effects in the market.
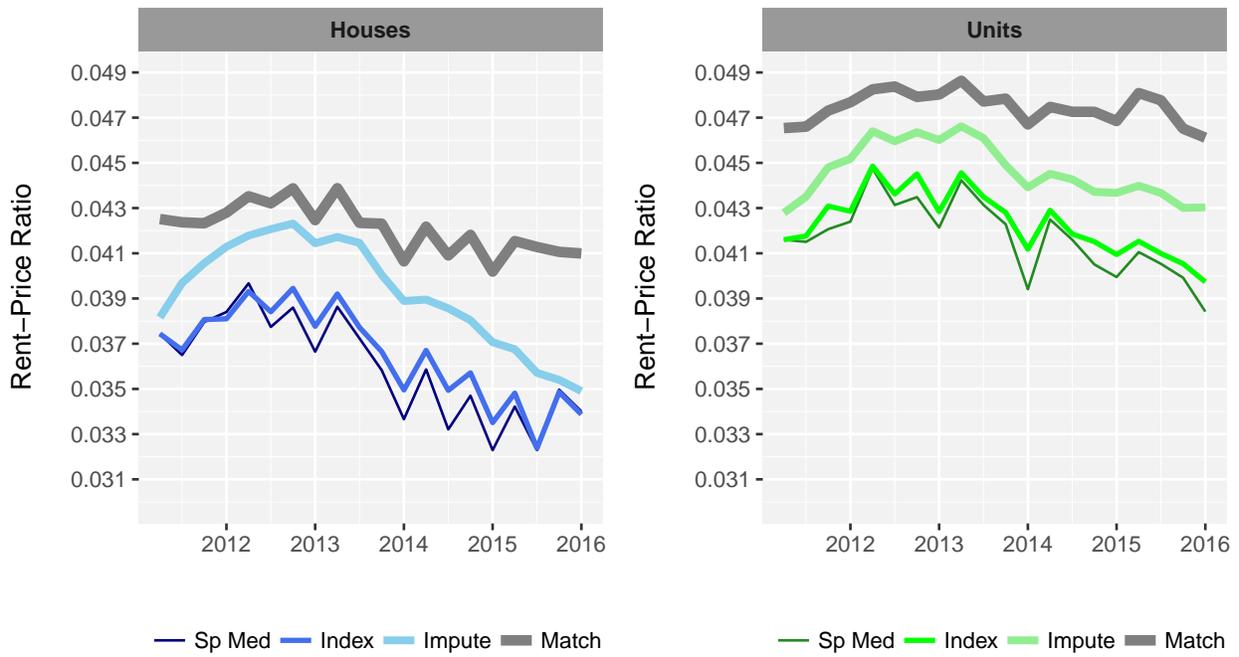
Figure 2: Method Comparison: All Metro Area
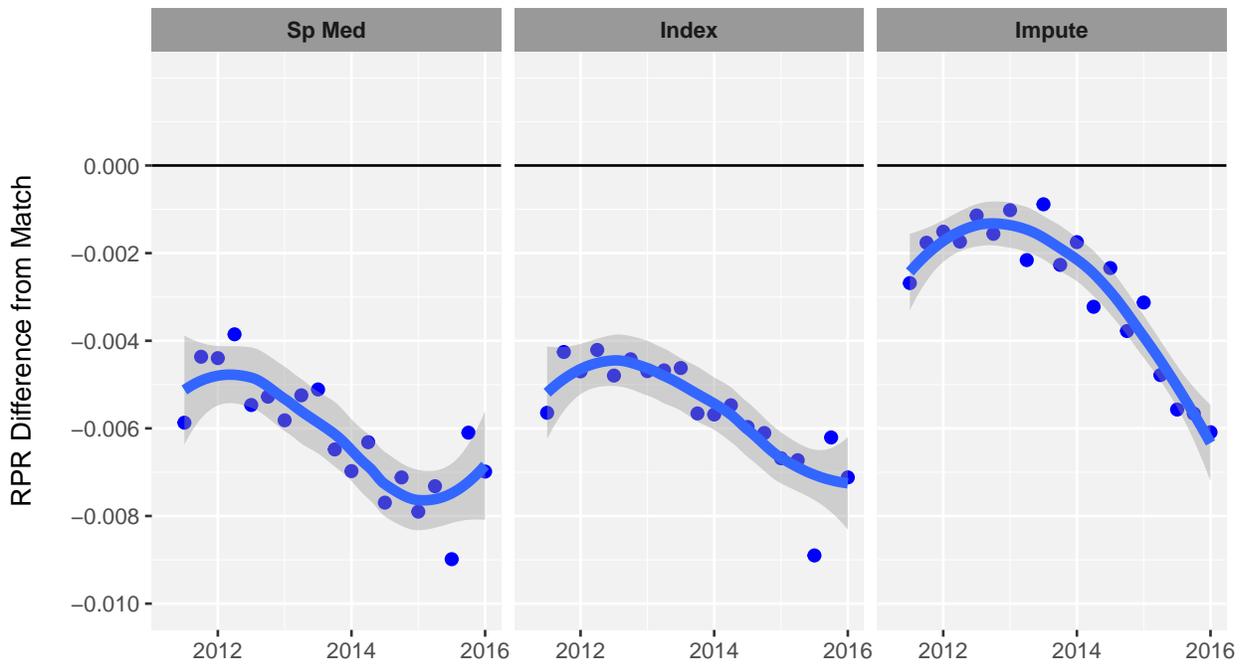
Figure 3: Method Differences: Metro Region - Houses
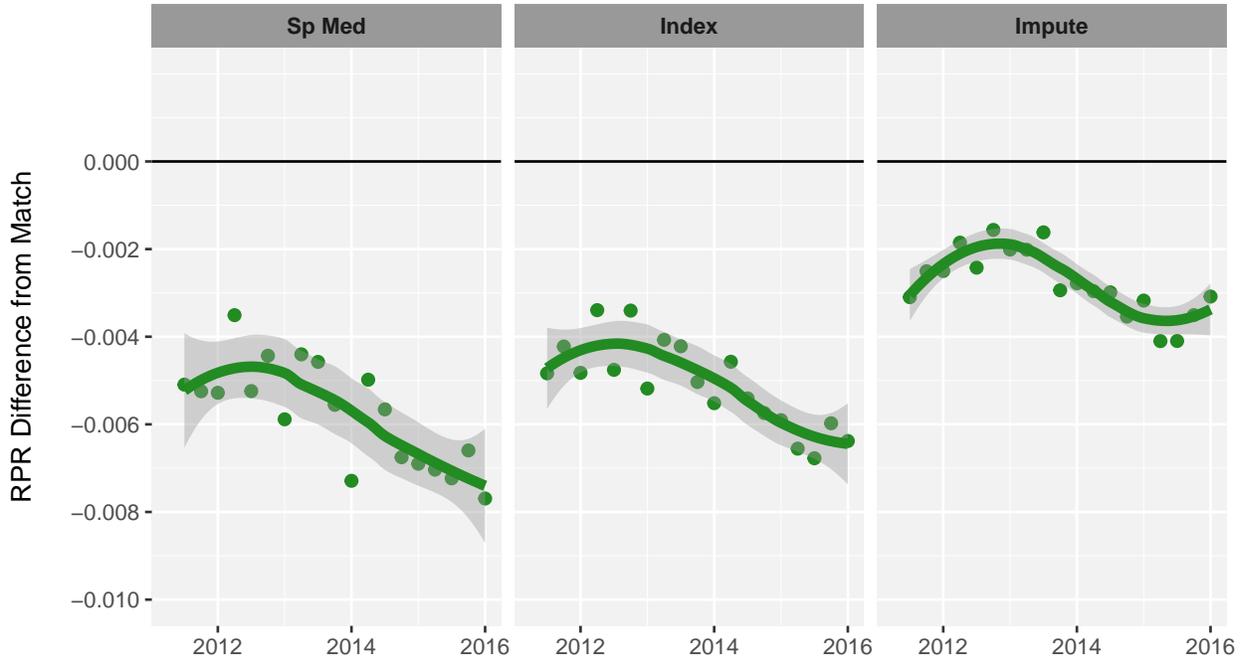
Figure 4: Method Differences: Metro Region - Units

Geographic Disaggregation

The above graphs show a comparison of metrics when calculated at the entire metropolitan region level. We have also estimated rent-price ratio at the LGA, SLA1, Post Code and Suburb level. Overall, as the geographic level of aggregations (LGA down to Suburb) decreased, the variation in trends within each method increased, but the differences between methods decreased. Figure 5 highlights the variation between RPR time trends of the individual suburbs across each methods (dark line shown metro level trend). Two additional things to note from this graph: 1) The hedonic imputation method offers a slightly smoother estimate over time since one of the observations in each pair is imputed; and 2) There are many less suburbs that contain enough unit transactions per period than there are eligible suburbs when looking into houses. This is due to the fact that most units in the Melbourne area are concentrated in the inner suburbs.

Next, we compare the differences between the match method and the other three methods across all five levels of geographic aggregation. For houses (Figure 6) we see that for the Spatial Median and Index methods each successively small geographic disaggration narrows the gap between these methods and the match method. As these methods are 'aggregate' method, each smaller geographic level of analysis eliminates some measure of contant quality conerns as well as issues regarding tenure mix location – rendering the results closer to those of the match method. The choice of level of geographic aggregation has no noticable impact on the differences between the imputation and match method. Additionally, we see that across all levels of

14
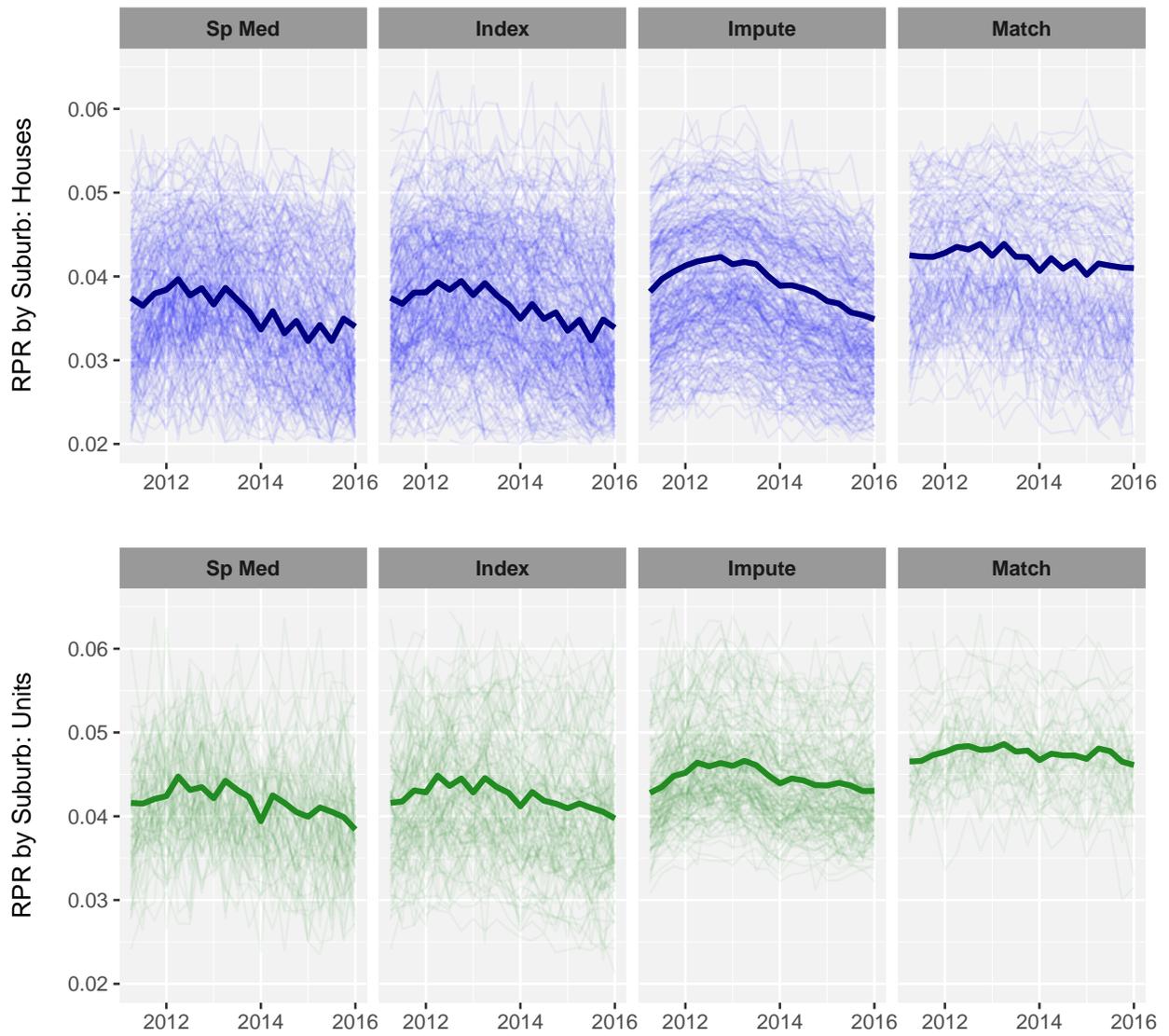
Figure 5: RPR Suburb Variation
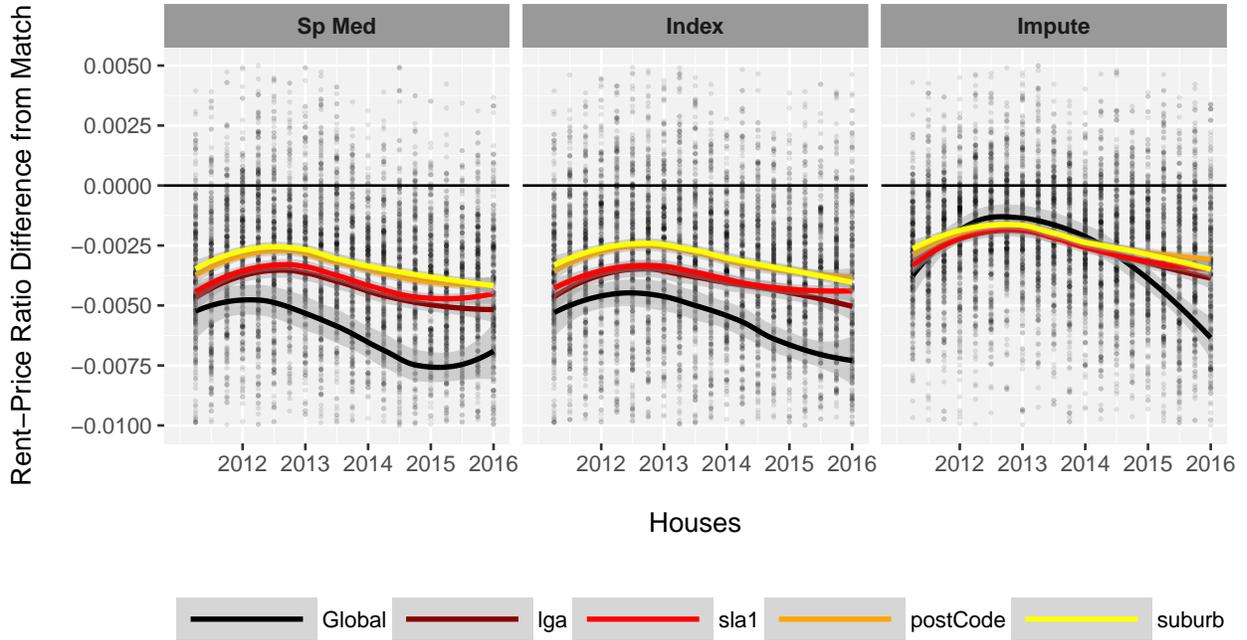
Figure 6: Method Differences: All Geographic Levels - Houses

geographic disaggregation that the differences between the match method and the other widens over the time period studied.

A close examination of figure 7 shows that, for units, differences between the match method and the other methods also decreases with increasing geographic disaggretion. The changes, however, are much less pronounced that those of houses. Much of this can be explained by two facts. First, most units are clustered in the inner suburbs and exhibit much less variation in location and prices than houses. The second reason is that units as a product are more homogenous and rely less on land values drive prices, factors which narrow the range of rents and prices by location, thereby providing less to be gained with performing analyses at small geographic levels.
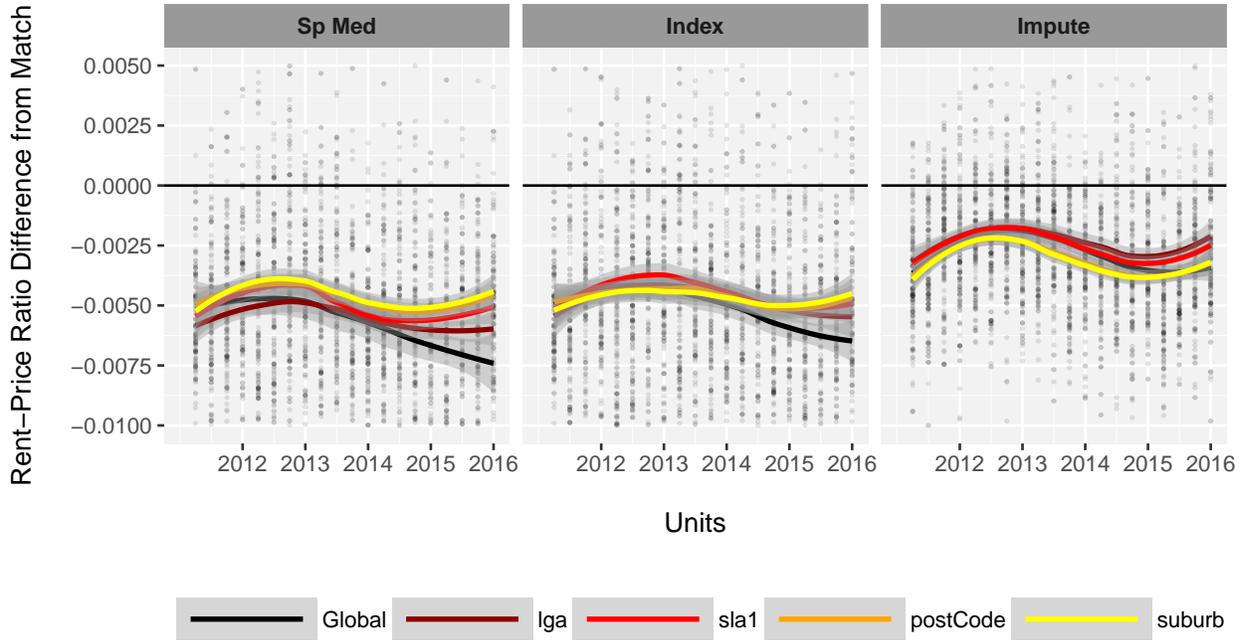
Figure 7: Method Differences: All Geographic Levels - Units

Predictive Results

It is clear from the comparative analysis that the four methods produce different price-rent ratio estimates. What the above analysis doesn't tell us is which analysis is more accurate. Considering the issues presented in Table 2 We would expect that the match method would be the most accurate as it controls for all of the potential concerns. However, it does rely on a vastly smaller sample size, a factor which could reduce predictive ability. Likewise we would expect analyses that are done at smaller geographic aggregation to be better than those completed at coarser ones. In this section we analyze the predictive ability of the rent-price ratios estimated by each method and compare them to determine if our hypotheses regarding the most accurate are correct, and if so, by how much.

To calculate the predictive accuracy we have used the rent-price ratio trends from each method to predict the future sale price or rental value of a matched observation. For example, if property A sold in Q1 2011 and then rented in Q3 2011, we will apply the RPR estimate from method X to determine a likely Q1 rental value. We then adjust this value for market changes based on the changes in the rental yield trend from method X to produce a prediction for the rental amount as of Q3 2011. The difference between the actual rental value in Q3 and the predicted value is the level of prediction error. Figure 8 shows the mean absolute prediction error for all four methods by use and across all five successively finer geographic levels of analysis.
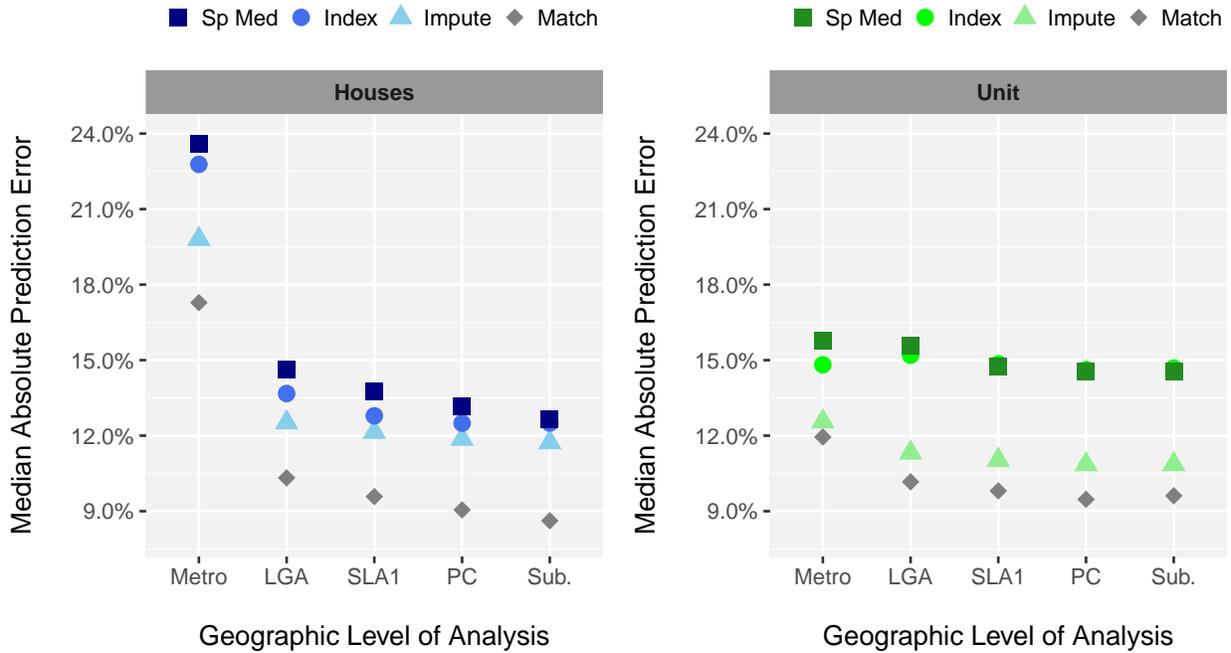
Figure 8: Predictive Accuracy

In looking across the geographic levels of analysis, we see that each step down to a smaller geographic level produces an improvement in predictive ability. Alternatively, this can be thought of as the accuracy of the rent-price ratio estimates. When analyzing house data, there is a large gain in accuracy when moving from a region wide analysis to one conducted at the local government area (LGA) level. Successive refinements in spatial scale see smaller and smaller gains in accuracy (moving left to right in the graph. For units, there is no large gain for moving from region-wide to LGA, rather a small gain exists at each step down the geographic scale.

Examining estimates from the four methods shows that in all situations the direct matching methods produces the most accurate estimates, followed by hedonic imputation. The spatial median and index methods systematically produce the least accurate measures of rent-price ratios. The gains in accuracy from the matching methods is greater for houses than for units and this degree of improvement actually increases at smaller geographic levels. The gain from the matching method over imputation is considerably smaller for units across all geographic levels.

Discussion

In reviewing previous research that employs some measure of relationship between home price and home rents, we find that most studies calculate this metric via a spatial median or index method. The ease of computation and the low data requirements for these aggregate methods are offset by potential biases due to differences between the types, location, condition and quality of the sets of homes that sell versus those that rent. Citing these issues, a number of studies use one of two more complex methods: the hedonic imputation method or the direct match method. In this study, we have set out to compare how results from the four methods differ using an identical database of sales and rentals. We have also examined these differences over a time period long enough to encounter distinct phases of a housing market cycle. Additionally, we look into difference between houses and units (apartments/condominiums) as well as the impacts of using finer geographic aggregations in estimating rental yields.

Across all analyses, the spatial median and index methods produce the lowest estimates of rent-price ratios, whereas the match method generates the RPRs. Differences between the match method and the other three methods average from between 0.003 and 0.01, which represents a difference of around 10% to 25% from the 0.036 to 0.04 rent-price ratios suggested by the spatial median method. This difference, however, is not contant over time. The differences between the other methods and the match method has increase over the study period, specifically since 2013. This finding suggests that even in studies where the trend of rent-price ratios is of more interest than the actual levels (such as those that are tracking housing bubbles) that the choice of method will influence the results.

Breaking the analysis down into smaller geographic aggregations has a moderate effect of lessening the difference in RPR estimates between the four methods. When computed at the smallest level, differences narrow to around 0.5% between spatial median and matching methods, however, in the case of houses (not units) these differences have increased over time. Comparisons between suburbs show very high variation with some areas showing rent-price ratios around 0.025 to 0.03 with others above 0.05. This high level of variation suggest that investors in this market should consider location when making decisions about rental yields. Additionally, analysts should attempt to employ the finest geographic aggegation possible if seeking to minimize bias.

Overall, the comparative analysis indicates that the matching method and the imputation method produce markedly different RPR estimates than the more common spatial median and index methods. These facts coupled with the predictive analysis that shows the matching method to be the most accurate (and the spatil median the least), future studies looking into the relationship between rents and prices should considered using the match and/or imputation method if observation level data is available.

It should also be noted that the purpose for which the RPRs are being created should also be considered

when choosing an estimation method. The hedonic imputation method will generate the likely rent-price ratio for all properties that are transacting in the market, either by sale or rent. On the other hand, the matching approach only measuring rent-price ratios for homes that are currently or have at least once during the period of study been rented. This distinction is important as the set of homes that sell are generally of higher quality than the set of homes that rent, especially when considering houses (as opposed to units). As a result, if users or researchers are interested in producing an estimate of imputed rents on owner occupied homes or in buliding user costs models then the hedonic imputation method is preferable. Conversely, if the intention of the analysis is to measure the actual rent-price ratio or rentaly yields on homes currently used as investment vehicles (leased out) then the matched approach is preferable. Futher research should be done to clarify this distinctions.

Finally, all code used to complete this analysis is available at https://github.com/andykrause/ausPropMrkt/prmc

References

Ambrose, B. W., Eichholtz, P., and Lindenthal, T. (2013). House prices and fundamentals: 355 years of evidence. *Journal of Money, Credit and Banking*, 45(2-3):477–491.

Ayuso, J. and Restoy, F. (2006). House prices and rents: An equilibrium asset pricing approach. *Journal of Empirical Finance*, 13(3):371–388.

Bracke, P. (2015). House Prices and Rents: Microevidence from a Matched Data Set in Central London. *Real Estate Economics*, 43(2):403–431.

Brunnermeier, M. K. and Julliard, C. (2008). Money illusion and housing frenzies. *Review of Financial Studies*, 21(1):135–180.

Campbell, S., Davis, M. A., Gallin, J., and Martin, R. (2009). What Moves Housing Markets: A Variance Decomposition of the Rent-Price Ratio. *Journal of Urban Economics*, 66(2):90–102.

Capozza, D. R. and Seguin, P. J. (1996). Expectations, efficiency, and euphoria in the housing market. *Regional Science and Urban Economics*, 26(3):369–386.

Engsted, T. and Pedersen, T. Q. (2015). Predicting returns and rent growth in the housing market using the rent-price ratio: Evidence from the OECD countries. *Journal of International Money and Finance*, 53:257–275.

Fu, Y. and Ng, L. K. (2001). Market Efficiency and Return Statistics: Evidence from Real Estate and Stock Markets Using a Present-Value Approach. *Real Estate Economics*, 29(2):227–250.

Garner, T. I. and Verbrugge, R. (2009). Reconciling user costs and rental equivalence: Evidence from the US consumer expenditure survey. *Journal of Housing Economics*, 18(3):172–192.

Gelain, P. and Lansing, K. J. (2014). House prices, expectations, and time-varying fundamentals. *Journal of Empirical Finance*, 29:3–25.

Glaeser, E. L. and Gyourko, J. (2007). Arbitrage in housing markets. Technical report, National Bureau of Economic Research.

Goetzmann, W. N., Peng, L., and Yen, J. (2012). The subprime crisis and house price appreciation. *The Journal of Real Estate Finance and Economics*, 44(1-2):36–66.

Hattapoglu, M. and Hoxha, I. (2014). The Dependency of Rent-to-Price Ratio on Appreciation Expectations: An Empirical Approach. *The Journal of Real Estate Finance and Economics*, 49(2):185–204.

Hill, R. J. and Syed, I. A. (2016). Hedonic pricerent ratios, user cost, and departures from equilibrium in the housing market. *Regional Science and Urban Economics*, (56):60–72.

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

Hwang, M. and Quigley, J. M. (2006). Economic fundamentals in local housing markets: evidence from US metropolitan regions. *Journal of Regional Science*, 46(3):425–453.

Hwang, M., Quigley, J. M., and Son, J.-y. (2006). The dividend pricing model: New evidence from the Korean housing market. *The Journal of Real Estate Finance and Economics*, 32(3):205–228.

Pavlidis, E., Yusupova, A., Paya, I., Peel, D., Martínez-García, E., Mack, A., and Grossman, V. (2013). Episodes of exuberance in housing markets: in search of the smoking gun. *The Journal of Real Estate Finance and Economics*, pages 1–31.

Smith, M. H. and Smith, G. (2006). Bubble, bubble, where's the housing bubble? *Brookings Papers on Economic Activity*, 2006(1):1–67.

Sommer, K., Sullivan, P., and Verbrugge, R. (2013). The equilibrium effect of fundamentals on house prices and rents. *Journal of Monetary Economics*, 60(7):854–870.

Verbrugge, R. (2008). The puzzling divergence of rents and user costs, 1980–2004. *Review of Income and Wealth*, 54(4):671–699.