Regional Science and Urban Economics 68 (2018) 204–225

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regsciurbeco

Estimating quantile-specific rental yields for residential housing in Sydney

Sofie R. Waltl

Luxembourg Institute of Socio-Economic Research, Maison des Sciences Humaines, 11, Porte des Sciences, 4366 Esch-sur-Alzette/Belval, Luxembourg

ARTICLE INFO

JEL:
R31
C21
C43

Keywords:
Rental yield
Price-to-rent ratios
Hedonic models
Quantile regression
Housing markets

ABSTRACT

Rental yields are widely used by investors, central bankers, researchers, and policy makers to assess and detect disorders in housing markets. This paper proposes a framework to measure rental yields across the distribution, time and space thus providing a comprehensive picture of housing markets. The two-step procedure based on hedonic quantile regression and propensity score matching is designed to fundamentally control for differences in house characteristics. The methodology is applied to micro-data on house transactions and asking rents in Sydney, Australia, between 2004 and 2014. The paper finds large temporal and spatial variation in rental yields, decreasing yields when moving from the low end of the distribution to the top end, and systematically larger yields when restricting the analysis to houses bought-to-let.

1. Introduction

Rental yields[1] are an important measure for investors to base their investment decision on as well as policy makers, central bankers, and researchers to assess the state of housing markets.

Rental yields are widely used to detect disorders in housing markets (see for instance Weeken, 2004; Fox and Tulip, 2014; OECD, 2016) as they are linked to expected capital gains via the user cost formula for housing (Poterba, Jan 1992, Himmelberg et al., 2005).

In equilibrium, the cost of owning, i.e., the user cost of housing, should equal the return on owning: the rent. As Fox and Tulip (2014) put it: “Given the supply of housing is fixed in the short run, prices are determined by how much buyers are willing to pay. Hence a comparison of the costs of home ownership with the costs of the nearest alternative [i.e., rents] seems central to a measure of overvaluation.”

By rearranging the user cost formula and plugging in rental yields, it is possible to measure expected capital gains (McCarthy and Peach, 2004; Fox and Tulip, 2014; Hill and Syed, 2016), which may assist in the early detection of housing bubbles. Stiglitz (1990) defines asset bubbles as follows: “[I]f the reason that the price is high today is only because investors believe that the selling price is high tomorrow – when *fundamental* factors do not seem to justify such a price – then a bubble exists.” Rapidly increasing expected capital gains, that are noticeably higher than long-term averages, may hence indicate irrational exuberance.[2]

Investors evaluate a property's potential return (and the housing market in general) using the *net rental yield*, i.e., *gross rental yield* minus costs such as maintenance costs or interest payments. Forward-looking present value models predict that low current rental yields signal higher future capital gains as well as increasing rents and thus attractive investment opportunities (see Clark, 1995; Capozza and Seguin, 1996).

Whereas *average* rental yields are a good starting point to assess investment opportunities or the state of a housing market, they may however obscure substantial cross-sectional variation. There are theoretical arguments (see Section 2) why the ingredients in the user cost formula are expected to vary across the distribution and hence, in equilibrium, such variation directly translates into cross-sectional variation of rental yields.

This paper therefore proposes a framework to estimate quality-adjusted rental yields across the distribution and documents systematic cross-sectional variation in rental yields which is in-line with predictions from the user cost formula. Additionally, this paper measures the evolution of rental yields over time and space.

In general, there are two alternative ways to look at aggregate average rental yields which offer two different approaches how to generalize average rental yields to quantile-specific rental yields. First, one may aim to calculate rental yields separately for individual houses. As there are only few houses for which price and rent information is available, one may choose to restrict the analysis on this sub-sample of

E-mail address: sofie.waltl@liser.lu.

[1] Rental yields are defined as the ratio of annual rent over sales price. Sometimes this ratio is also referred to as *gross* rental yield as opposed to *net* rental yields which take into account housing related costs. The literature also studies reciprocal rental yields: price-to-rent ratios.

[2] For instance, the *UBS Global Real Estate Bubble Index*, which is “designed to track the risk of housing bubbles in global financial centers” (see Holzhey and Skoczek, 2016, page 4) includes reciprocal rental yields as one of five components.

https://doi.org/10.1016/j.regsciurbeco.2017.10.011
Received 8 February 2017; Received in revised form 18 October 2017; Accepted 26 October 2017
Available online 28 October 2017
0166-0462/

observations, or impute missing prices or rents. One may finally obtain an aggregate measure of the average rental yield by taking the average over all individual ratios. Quantile-specific rental yields would be obtained by evaluating the distribution of rental yields at different quantile levels.

Second, one may think of an average rental yield as the ratio of an average annual rent over an average sales price. This understanding seems to be very common among real estate agents and investment advisers. A generalization of this concept is to match quantiles of the rent distribution to the same quantiles of the price distribution.

In this paper, I will follow the second approach as it has some appealing advantages in terms of interpretation: A rental yield for a high quantile level is associated with high prices *and* high rents and vice versa. Such kind of interpretation is *in principal* not possible for the first approach where a high quantile level neither reflects a high price nor a high rent but only indicates a *high ratio*.[3] When explaining quantile-specific rental yields through quantile-specific ingredients in the user cost formula the relationship between rental yields, and the price and rent distribution is, however, essential.

Although the basic concept of a rental yield is straight-forward, there are crucial measurement challenges. Similar as to when constructing house price indices, it is important to control for differences in house characteristics to compare like with like. In the house price index literature this kind of "quality-adjustment" is usually performed by applying repeat-sales or hedonic methods (see de Haan and Diewert, 2013). When constructing rental yields an additional dimension of quality-adjustment should be considered: Next to quality differences *within* houses sold and *within* houses rented, there is possibly also a mismatch *across* houses sold and rented. For instance, houses sold may be on average larger than houses rented leading to biased results.

Several articles construct rental yields by comparing rental indices with house price indices (see Fu and Ng, 2001; Himmelberg et al., 2005; Gallin, 2008; Campbell et al., 2009; Duca et al., 2011). Whereas such a procedure accounts for quality differences within sales and within rents, it ignores quality differences across sales and rents as quality-adjustment is most probably performed in different ways for rental and sales price indices. Furthermore, using indices only allows changes of rental yields, but not levels, to be measured.

Alternatively, one may construct rental yields by taking the ratio of the average observed rent over the average observed sales price. Such kind of calculations ignore quality differences and are often found on real estate agents or financial advisers websites. Davis et al. (2008) calculate historic rental yields based on average sales prices and average imputed rents.

Hill and Syed (2016) use a hedonic imputation approach to account for quality differences in rental yields. If a house was sold but not rented they treat the rental price of this particular dwelling as missing and vice versa. They estimate separate hedonic models for the rental and the sales data set and use these models to impute missing rental and sales prices. Ultimately, they gain estimates for the rental and sales price for all dwellings in their data set and calculate dwelling-specific quality-adjusted price-to-rent ratios. They calculate the median over all these ratios to obtain an aggregate measure.

Fox and Tulip (2014) use a data set that consists of observed or imputed rent and sales prices for identical dwellings, and construct average rental yields based on them. Imputations rely on hedonic methods or on extrapolated prices using price and rent indices.

Bracke (2015) adapts the repeat-sales idea and restricts the sample to those dwellings having been sold and rented within a short period of time. Bracke hence uses an exact matching procedure to create the sample on which he performs his analysis.

Smith and Smith (2006) also use a matching approach but allow next to exact matches also pairs of observations that are, though not identical, similar in their characteristics. Both Bracke, and Smith and Smith use conservative matching approaches which come at the cost of strikingly small sample sizes that may be subject to sample selection bias (see Section 4.5 for a discussion). In the case of Smith and Smith, samples consist of 100 observations only.

There is a trade-off between aiming for a good match between houses sold and rented in terms of *observed and unobserved* characteristics, and avoiding a sample selection bias. When relying on exact matches, i.e., houses that were sold and rented within a short period of time, there is per construction no quality-mismatch. However, such a conservative approach may introduce a sample selection bias.[4] On the other hand, neglecting the fact, that houses sold and rented tend to be different in their characteristics, may lead to noisy estimates. This paper thus aims to find a compromise between these two competing goals by suggesting a two-step procedure labelled *Marginal Densities Matching* (MDM) approach.

In the first step, houses rented are matched to houses sold based on their characteristics using propensity score matching. The second step constructs samples from the *marginal* sales and rental price distribution net of house characteristics using penalized quantile regression in combination with a sampling algorithm proposed by Machado and Mata (2005). These samples are used to calculate quantile-specific rental yields that are fundamentally controlled for differences in house characteristics.

This paper thus contributes to the yet very sparse literature on measurement issues related to rental yields and is the first to develop a method to measure quality-adjusted rental yields cross-sectionally. Furthermore, it is connected to the infant literature measuring intra-urban, cross-sectional house price movements (see McMillen, 2014; Waltl, 2016b; Zhang and Yi, 2017).

The method is applied to micro-data on house transactions and asking rents from Sydney, Australia, between 2004 and 2014. Such comprehensive data on rents yet are very rare which may explain the gap in the literature on techniques to accurately measure rental yields.

As predicted by the user cost formula, rental yields are consistently downwards-sloping when moving from the low end of the market to the top end. Refraining from quality-adjustment leads to noisy results and on average lower rental yields. This paper documents systematically larger rental yields when relying on houses that are sold and rented within a short period of time. This deviation may either stem from systematic deviations in unobserved characteristics (affecting MDM results) or a sample selection bias (affecting exact matches). For the Sydney data, results based on exact matches are on average 6% to 8% higher than MDM results.

The remainder of the paper is structured as follows. Section 2 elaborates on expectations about cross-sectional variation in rental yields based on the user cost formula. Section 3 develops the methodology to construct quantile-specific and quality-adjusted rental yields. Section 4 describes the data set and presents empirical results. Finally, Section 5 concludes.

2. Expectations about cross-sectional variation

Poterba (Jan 1992) and Himmelberg et al. (2005) argue that in equilibrium the expected annual cost of owning should equal the annual cost of renting and hence compare

[3] However, it is possible to relate individual rental yields separately to the sales *or* the rental price distribution. These results are discussed in Section 4.6.

[4] For repeat-sales house price indices, which rely on price developments of identical dwellings, a sample selection bias has been well documented in the literature (see Clapp and Giaccotto, 1992; Wallace and Meese, 1997; Steele and Goy, 1997). In the repeat-sales case, there is a bias towards dwellings of lower quality: an Akerlof-type lemons bias.