Public Disclosure Authorized

Public Disclosure Authorized

Policy Research Working Paper 8983

Evaluating the Accuracy of Homeowner Self-Assessed Rents in Peru

Lidia Ceriani
Sergio Olivieri
Marco Ranzani

WORLD BANK GROUP
Poverty and Equity Global Practice
August 2019

Policy Research Working Paper 8983

Abstract

Attributing a rental value to the dwellings of homeowners is essential in various contexts, including distributional analysis and the compilation of national accounts, consumer price indexes (CPIs), and purchasing power parity indexes. One of the methods for making the attribution is to use homeowner estimates of the market rental value they would pay (receive) for their dwellings if these were rented. This is known as homeowner self-assessed rent. However, homeowner estimates may not be accurate because of the way questions aimed at soliciting such information are phrased, the sentimental attachment of the homeowners to the properties, lack of information about rental markets, and other reasons. Yet, researchers and practitioners often neglect to ascertain the accuracy of homeowner assessments. This study argues that comparing unconditional or conditional means may be misleading if one has not ascertained whether the observable characteristics of homeowner and tenant dwellings are similar. Using Peruvian data from 2003 to 2017, the study tests the accuracy of self-assessed rental values with matching estimators. In Metropolitan Lima, homeowners typically provide accurate estimates of the rental market values of their dwellings. In rural areas, market rental values are underestimated by homeowners in more instances. The direction and magnitude of the inaccuracies in Metropolitan Lima and in rural areas are comparable and range between −25 percent and −20 percent.

This paper is a product of the Poverty and Equity Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at mranzani@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

Evaluating the accuracy of homeowner self-assessed rents in Peru

Lidia Ceriani[a] Sergio Olivieri[b] Marco Ranzani[b]

Updated August, 2020

JEL: O18; R31.

Keywords: Housing; Imputed Rent; Peru

The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. The authors wish to thank Erich Battistin, Dean Jolliffe, Peter Lanjouw, Kristen Himelein, and Carlos Rodríguez-Castelán for their comments on previous versions of the paper.
a. Georgetown University.
b. Poverty and Equity Global Practice, World Bank.

1. Introduction

Attributing values to the flow of services households derive from their dwellings, such as rental values, is a knotty issue that recurs in various contexts, including the compilation of national accounts (Heston 1994), the estimation of consumer price indexes (CPIs) (Lebow and Rudd 2003), the compilation of purchasing power parity indexes (Deaton and Heston 2010), and the analysis of poverty and inequality (Balcazar et al. 2017; World Bank 2020). The key problem is calculating rental values for those households that do not rent their dwellings because they own the dwellings or, for instance, because the dwellings are provided free by relatives or employers. In such cases, the rental values must be estimated.

Despite the prominence of the issue in various contexts, the jury is still out on the best practice for imputing rental values across homeowners. Three main approaches to rent imputation are typically adopted in the literature: hedonic models, nonhedonic models (the rent-to-value approach, the user-cost approach, and the rental equivalence approach), and self-assessment. The hedonic theory of consumption implies that the implicit prices of goods are a function of the characteristics of the goods (Lancaster 1966). Hence, rents reported by tenants are estimated econometrically as a function of the observable characteristics of the dwellings, and the parameters are used to predict implicit rents across homeowners.

In nonhedonic models, the second approach, the implicit rent is understood as the rate of return that would have been obtained by an owner if the equity in the home had been invested in an interest-bearing account. (See Balcazar et al. 2017, for a comprehensive review of imputation methods, ranging from parametric to nonparametric models, including rent-to-value and user-cost approaches.)

The third approach relies on homeowners' subjective evaluations of the market rental values of their residences. This method has been widely used. For example, half the 21 economies in the Asia and Pacific region that participated in the International Comparison Program for the computation of 2011 purchasing power parities use self-assessed rental values to impute rents going to homeowners (ADB 2014). The African Development Bank also presents equivalent results for countries that include rents in national accounts: half the countries use self-assessment (AfDB 2013). The weights of the CPI in the United States are computed based on the values of dwellings found in the Consumer Expenditure Survey. The rental values of owner-occupied dwellings revealed in the survey depend on owner estimates (Lebow and Rudd 2003). According to a recent review of national methodologies used to produce official poverty estimates (World Bank 2020), poverty estimates in 16 percent of the countries surveyed include rental values that are self-assessed by homeowners.

Using homeowner assessments relies on the assumption that homeowners can give an accurate estimate of the rental market values of their dwellings, perhaps with the help of interviewers, as suggested by Garner and Kogan (2007). This information is frequently collected in household budget surveys, in which homeowners are often asked to estimate the rents they would pay to live in their current dwellings. However, owner-occupiers may not be able to provide accurate estimates of the rental prices for their dwellings, and the direction and magnitude of the inaccuracies become an empirical issue. In countries in which rental markets are thin because of high rates of homeownership, one may reasonably expect that homeowners might merely lack information about rental prices. In other circumstances, homeowners might overestimate the true rental values of their dwellings compared with rented dwellings having similar characteristics because of their attachment to specific features of their neighborhoods or of their residences, particularly if they designed or built the homes (Frick et al. 2010; Heston and Nakamura 2009). Extensive evidence on the United States suggests that self-assessed rental

equivalences may be overestimated. Using a national sample of the United States, Goodman and Ittner (1992) explore the accuracy of owner estimates of house values relative to the sales prices of the same properties. They find that the median homeowner in the mid-1980s overvalued the home by about 6 percent. Garner and Short (2001) find that housing costs based on self-reported rental equivalence resulted in higher estimates (almost 15 percent) than those based on a hedonic model. Garner, Short, and Kogan (2006) state that the median reported rental equivalence is higher in the U.S. Consumer Expenditure Survey than in estimates based on the same data using the hedonic model, the rent-to-value model, or the payment approach. The way the information on self-assessed rental values is elicited in surveys may also matter in explaining the inaccuracy of homeowners. Phrasing the question to homeowners as "How much would you receive as a rent if you were to lend your apartment?" as opposed to "How much would you pay if you were to rent the dwelling you are currently living in?" might be one way to mitigate self-assessment bias because of misreporting. In fact, the theoretical and experimental literature on the difference between willingness to accept and willingness to pay for the same good concludes that measures of the former are usually superior to measures of the latter (Fehr, Hakimov, and Kübler 2015; Hanemann 1991; Tunçel and Hammit 2014).

Even if homeowners were able to provide accurate estimates of the rental market values of their dwellings, a comparison of average values, either unconditional or conditional on a set of observable characteristics of dwellings, might lead to incorrect inferences. In countries in which rental markets are developed and thick, and information about market rental values is available, the characteristics of the dwellings inhabited by homeowners and by tenants might nonetheless differ considerably. Attention should therefore be paid to the selection among comparison groups to avoid comparing the incomparable. This matters both in the case of national accounts and country-level CPIs where obtaining correct estimates of average housing values serves the purpose and in the case of the analysis of household living standards wherein the construction of income or consumption aggregates requires the imputation of values for each household.

Despite the existing empirical evidence on the inaccuracy of homeowner self-assessed rental values described in section 2, there is no clear methodological recommendation on how to test accuracy. In developing countries in particular, analyses are hampered by the lack of up-to-date, complete administrative data and alternative surveys on the market for dwellings. Practitioners must therefore rely on the data collected in household budget surveys to test the accuracy of homeowner self-assessed rents.

The study illustrated in this paper proposes the use of matching estimators to assess the accuracy of homeowner self-assessed rents based on information typically available in household budget surveys. Matching estimators allow the identification of groups of tenants who live in dwellings that are as similar as possible in observable characteristics to the dwellings inhabited by homeowners. Thus, the comparison between homeowner self-assessed rents and market rents paid by tenants occurs among dwellings that share similar observable characteristics.

The rest of the paper is organized as follows. Section 2 provides a short review of studies that test the accuracy of homeowner self-assessed rents. Section 3 describes recent trends in the housing market in Peru. Section 4 presents the methodology used to ascertain the accuracy of homeowner self-assessed rents. Section 5 describes the microdata used in the analysis. Section 6 presents descriptive statistics and discusses the results. Section 7 offers final remarks.

2. Review of the Literature

Most of the literature on the accuracy of homeowner self-assessed housing values concentrates on the United States and compares self-assessed home values found in surveys with external information from professional appraisers, market transactions, tax assessments, other surveys, and so on. In their seminal paper, Kish and Lansing (1954) compare the homeowner estimates in the 1950 Survey of Consumer Finances and estimates of professional appraisers who were asked to evaluate the same properties. They found a large bias between the homeowner and professional evaluations, but, on average, the bias was negligible. Using a similar approach and similar data, Kain and Quigley (1972) find comparable results. By contrast, Ihlanfeldt and Martinez-Vazquez (1986) find that the owner estimates tend to be higher than the appraised values.

In a rare study on a developing country, Gonzalez-Navarro and Quintana-Domeque (2009) compare data from a household survey and appraisals of the same homes by a real estate agent in the city of Acayucan, Veracruz, central Mexico. They find large variability in the bias between homeowner and professional appraisal valuations. Only 11 percent of the estimates of survey respondents were within 10 percent of the appraiser values; 25 percent of the respondents believed their homes were worth at most 70 percent of the appraised values; and 35 percent of the respondents believed their homes were worth at least 150 percent of the appraised values. The length of tenure appears positively correlated with the bias: owners with long tenure overestimated the values of their homes, while families with tenure of less than two years seem to have reasonably accurate and unbiased estimates of the values of their homes.

Recent contributions exploiting data on housing transactions include Bigelow, Ifft, and Kuethe (2020), Chan, Dastrup, and Ellen (2016), and Corradin, Fillat, and Vergara-Alert (2017). Bigelow, Ifft, and Kuethe (2020) concentrate on farmland owners in New York State. They advocate comparisons across both self-estimates and market transactions because of key differences in the way farmers and market participants perceive the value of farmland. They point to the importance of taking into account the level of activity or thickness of the market, where thickness refers to a level of transactions that is above the median. For instance, landowners in thick markets seem generally to overestimate the value of their land. Comparing self-reported valuations and market selling prices by zip code over time (1984–2013 in Corradin, Fillat, and Vergara-Alert 2017, and 1997–2011 in Chan, Dastrup, and Ellen 2016), Corradin, Fillat, and Vergara-Alert (2017) and Chan, Dastrup, and Ellen (2016) reach a similar conclusion: the average value of housing price misperception is countercyclical if this value is compared with the housing market cycle. Tur-Sinai, Fleishman, and Romanov (2020) conduct a similar exercise on Israeli data at the census tract level. They investigate the bias size of subjective valuations against transaction prices across a distribution of dwelling prices and find that the effect of the bias is opposite at the opposite ends of the distribution. Homeowners living in the most inexpensive dwellings tend to give upward-biased estimates, while homeowners living in the most expensive dwellings tend to understate the values of their homes.

Other approaches compare information found in different surveys. Several authors compare the approach to homeowner dwellings adopted by the U.S. Bureau of Labor Statistics for the CPI with the approach adopted by the Bureau of Economic Analysis for personal consumption expenditures. While the CPI calculations of the former rely on self-reported homeowner estimates, the latter uses rent-to-value ratios. Lebow and Rudd (2003), for instance, suggest that homeowners may give quite inaccurate estimates of the rent values of their homes, and this may create large biases in the computation of the CPI given the large weight attached to the imputed rent category. Garner et al. (2006) find that, in 1992, the estimates of dwelling services of renters and owners

were about 9 percent higher in the Bureau of Economic Analysis data relative to the data of the Bureau of Labor Statistics. In addition, they find that the two data series consistently grew apart from 1992 to 2000. As underlined by Garner and Short (2001), the reason for this mismatch may be that respondents in both cases have mistaken and biased ideas about the market values of their dwellings or that the reported rental equivalence values are likely to be capturing variations in housing and neighborhood quality that hedonic approaches do not capture. Corradin, Fillat, and Vergara-Alert 2017 make an argument that is similar to the latter point. Moreover, owners may express above-market evaluations of their dwellings because of a special attachment for specific features of their homes, especially if they designed or built the homes. Heston and Nakamura (2009) call this the owner pride factor. Van der Cruijsen, Jansen, and van Rooij (2018) provide a review of possible behavioral reasons beyond the bias of homeowner valuations. They compare the self-assessed homeowner valuations with appraisals of the same housing administrated by the municipality to which the housing belongs for tax purposes. They find that the median respondent overestimates the true value by 11 percent.

Sometimes, the same survey supplies enough information to check the accuracy of homeowner self-assessments. Van der Cruijsen, Jansen, and van Rooij (2018) and Gao and Liang (2019) compare the two sets of values reported by respondents in, respectively, the DNB Household Survey in the Netherlands and the China Household Finance Survey: the self-assessed current value and the purchase price for the same dwelling, which is updated using an evaluation of the fluctuations in housing prices in the same neighborhood or region over time. Both contributions find that households systematically overestimate their home values. Benítez-Silva et al. (2015) exploit the longitudinal panel structures of the Health and Retirement Study and the American Housing Survey and are able to compare the transaction prices at a given time and self-reported housing values before the sales of the same housing. They conclude that homeowners, on average, overestimate the current values of their properties by around 8 percent. Arévalo and Ruiz-Castillo (2006) compare self-assessed values with the results of a hedonic model on the same dwellings in Spain, calibrating the model on the reported rents paid by tenants in the same survey. They find that, on average, self-assessed values are close to the results of the hedonic model.

This paper proposes the use of matching estimators to assess the accuracy of homeowner self-assessed rents based only on the information available in household budget surveys, thus departing from most of the existing literature. This represents an important advantage because researchers and practitioners working on rent imputation in developing countries are left with few sources of information, and there are often no alternative surveys, reliable administrative data on market transactions or tax assessments, or data collected through professional appraisals. The paper argues that ensuring homeowner and tenant dwellings exhibit similar characteristics is key for an accurate comparison, and attention should therefore be paid to the choice of the comparison groups to avoid comparing the incomparable.

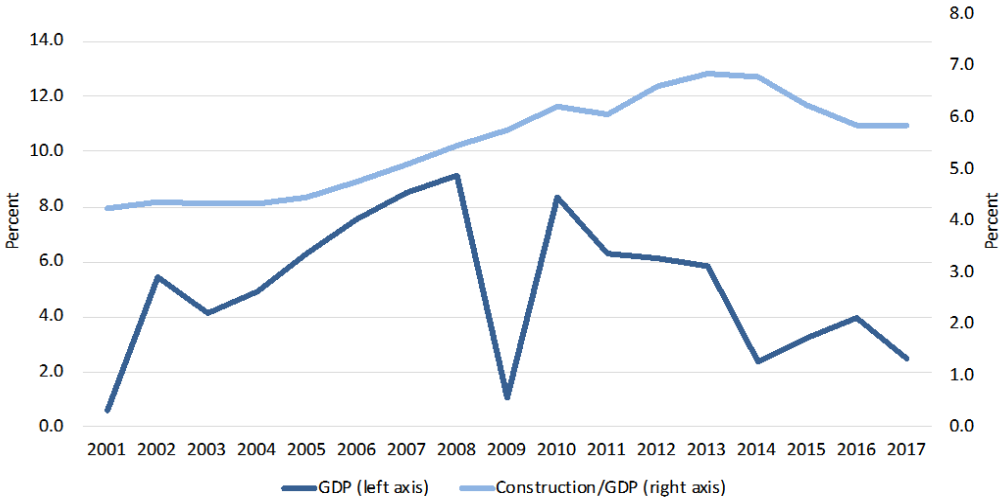
3. The Housing Market in Peru

This section briefly describes trends in the housing market in Peru in the recent past. The housing market has been dynamic. The dynamism has been driven by a rapid process of urbanization, changes in policies, and robust growth in the construction sector. The current process of urbanization started in the mid-1950s, when the rural population began moving from the Andes toward Lima Province, in which the capital is located. The urban population grew from about 47 percent of the country's population in 1960 to almost 78 percent in

2017.[1] At the beginning of the urbanization process, the government addressed the growing demand for housing in urban settlements by allowing rural migrants to occupy peripheral land and by recognizing the legal status of existing informal settlements. Efforts to regularize and upgrade these informal settlements—the barriadas—were undertaken during the 1970s, but were halted first by the economic crises of the 1980s and then by the new constitution approved in 1993, which put an end to the recognition of housing rights and dismantled all related institutions. In the early 2000s, the Ministry of Housing was reestablished, and the government instituted the Mivivienda Fund, which is the main public institution for the provision of social housing. The fund was reorganized in 2002 according to the National Housing Plan for 2003–07. In 2006, the 2006–11 National Housing Plan shifted the focus of social housing efforts toward the poorest segments of the population, that is, people with low or extremely low incomes.[2]

Thanks to these public housing policies, combined with sustained economic growth and an expanding urban population, the Peruvian housing market grew substantially in 2005–13. While the economy was growing at an impressive average annual rate of about 7 percent during those years, the construction sector posted an average growth rate at above 12 percent. These accomplishments are even more striking given that the global financial crisis of 2008–09 occurred during the period. The contribution of the construction sector to gross domestic product (GDP) rose from 4.1 percent in 2001 to 5.1 percent in 2007 and peaked at 6.8 percent in 2013 and 2014 (Figure 1).

Figure 1. GDP annual growth rate and contribution of the construction sector to GDP, 2001–17



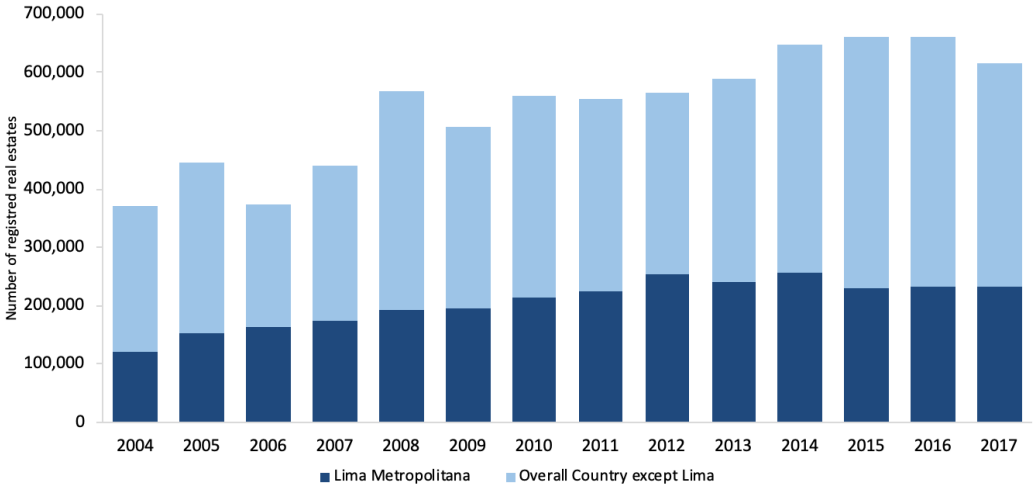
Source: Data of the National Institute of Statistics and Informatics.
Note: The data for 2016 and 2017 are provisional.

Among all regions in Peru, Metropolitan Lima—the conurbation of the cities of Callao and Lima—has been the most dynamic. The population of Metropolitan Lima increased from 900,000 (15 percent of the total population of the country) in 1940 to 10.5 million (36 percent of the total population) in 2017, the year of the latest available census.

[1] Data of WDI (World Development Indicators) (database), World Bank, Washington, DC, http://data.worldbank.org/products/wdi.
[2] The socioeconomic levels are defined by the Peruvian Association of Market Research Firms (Asociación Peruana de Empresas de Investigación de Mercados). For an extensive discussion on the history of progressive housing in Peru, see Apeim (2007) and Fernández-Maldonado (2010).

As a consequence of the progressive public housing policies, the units of real estate registered in the cadaster of Metropolitan Lima, which accounts for about 40 percent of all registered units in the country, rose by more than 160 percent, from 88,715 units in 2001 to over 233,000 units in 2017, reaching a peak of 256,271 units in 2014 (Figure 2).

Figure 2. Number of units of real estate registered in the National Buildings Cadaster, 2001–17



Source: Based on Urban Land Registry data, National Office of Public Registries, https://www.sunarp.gob.pe/estadisticas.asp.

The revamping of public housing policies during the second half of the 2000s led to an expansion in the number of individuals with access to credit to finance housing purchases or renovation. Public housing policies during this period were focused on expanding access to credit among the poorest socioeconomic segments of the population, which represented for more than 50 percent of the population of Metropolitan Lima in 2007. In Metropolitan Lima, the number of loans sponsored by the Mivivienda Fund alone rose by more than 100 percent in 2008–16. The number of disbursements of Mivivienda Housing Subsidies (Bono Familiar Habitacional) grew by about 200 percent in 2004–16 and reached 4,351 disbursements in 2016.

4. Empirical Strategy

Several approaches are typically used to estimate implicit rental values for homeowners and nonmarket tenants. Among these methods, the rental values reported by homeowners should they rent out the dwellings they live in are used as a good approximation of the values of their consumption of housing services. The approach based on the use of self-assessed imputed rents relies on the assumption that homeowners and nonmarket tenants are informed and able to estimate the values of their dwellings on the rental market.

Oftentimes and in some countries more than in others, homeowners and tenants live in dwellings with different characteristics and spatially clustered in different locations. For example, in many countries, tenants are concentrated in urban areas and live in small apartments. In addition, homeowners might overestimate the rental values of their dwellings because of their attachment to the dwellings (Heston and Nakamura 2009). The rental market might also be thin, and, therefore, sufficient information flows about rental market prices is unavailable to help homeowners make informed guesses.

For all these reasons and particularly because of the differences in the distribution of the observable characteristics of dwellings, this study argues that the self-assessed rental prices reported by homeowners should be carefully analyzed before being used as estimates of the rental market prices of homeowner dwellings. It proposes the use of matching estimators to ascertain the accuracy of homeowner self-assessed values by comparing homeowner estimates and tenant rental prices across dwellings with similar characteristics.

Matching estimators and, more precisely, propensity score matching estimators have been widely used to correct for differences in observable characteristics in nonexperimental settings to establish casual inference. Matching estimators construct "the correct sample counterpart for the missing information on the treated outcomes had they not been treated by pairing each participant with members of the nontreated group" (Blundell and Costa Dias 2009, 593). The fundamental assumption is that the set of observable characteristics contains all the relevant information about the potential outcome in the absence of treatment that was available to individuals at the time of deciding whether to be treated or untreated. In other words, the researcher has all the information that affects participation and outcomes among the untreated. This is known as the conditional independence assumption. It implies that unobservables in the outcome equation are orthogonal to treatment, conditional on observable characteristics, that is, there is no selection on unobservables.

Matching estimators seek one (or a set of) untreated observation(s) with the same realization of X for each treated observation. The outcome among the untreated so identified will be a good predictor of the unobserved counterfactual. This is possible only if observable characteristics in X do not predict participation perfectly. Space thus exists for unobserved factors to affect treatment status. This means one observes the treated and the untreated with similar characteristics. Because the curse of dimensionality—the potentially large number of observable characteristics in X—may be an issue in the implementation of matching estimators, a common alternative proposed in the literature is to match on a function of X rather than on the set of X. The analysis uses the linear prediction of the probability of participation given the set of characteristics in X or the propensity score $P(X)$ (Rosenbaum and Rubin 1983) that is obtained by estimating a logit model. Under the conditional independence assumption and overall and common support assumptions, the matching estimator is derived by averaging the outcome differences among the treated $R(1)$ and the untreated $R(0)$ with similar characteristics in X using the weights of the distribution of X among the treated. Formally, this gives the following:

$$\Delta^{PSM} = E_{P(X)|D=1}\{E[R(1)|D=1, P(X)] - E[R(0)|D=0, P(X)]\} \qquad (1)$$

The matching estimator is the average difference in outcomes over the common support, weighted by the distribution of X among the treated. The following step is to establish a metric of proximity between the propensity scores for the treated and control observations and a set of weights to associate the selected set of untreated observations to each treated observation. Among the possible algorithms the literature has proposed, this analysis applies three methods: (1) kernel, (2) nearest neighbor, and (3) radius matching.[3]

[3] Nearest neighbor matching is performed by pairing each of the treated with the closest (or a certain number of the closest) untreated observation(s) in terms of propensity score. Allowing the choice of more than one nearest neighbor is important because the nearest neighbor might be remote in terms of propensity score and might not be a good match. Kernel matching does not restrict the set of matching partners, but uses as controls a weighted average of all observations in the comparison group, where a larger weight is given to observations that are closer to the treated observation in terms of propensity score. For a good guide to the implementation of propensity score matching, see Caliendo and Kopeinig (2008).

The choice of the appropriate matching variables should be dictated by economic theory and eventually adjusted based on information available in the data. Rosenbaum and Rubin (1983) argue that only variables measured before the treatment should be included to avoid any endogeneity with respect to the exposure to the treatment. Lechner (2008) shows that this condition can be relaxed as long as the effect of the treatment on the covariates is nonsystematic; variables observed after the treatment thus only induce a measurement error in X. In the empirical application in the study, the choice of the appropriate variables is guided by the hedonic theory of consumption (Freeman 2003; Lancaster 1966), which shows how housing can be thought of as a commodity composite of different characteristics, and by consolidated empirical evidence. (See Hill 2013 for a survey of hedonic models for housing prices.) This includes location of the dwelling; structural attributes of the dwelling, such as whether it is a detached home or an apartment, type of construction, age of the dwelling, dimensions and number of rooms, and so on; and neighborhood characteristics, such as the quality of schools, accessibility of public transport, proximity of streets, crime rate, poverty rate, traffic congestion, and so on.

5. Data

The study relies on data from the 15 rounds of the National Household Survey on Living Conditions and Poverty (Encuesta Nacional de Hogares sobre Condiciones de Vida y Pobreza) over 2003–17. This long time span includes the period of the rapid expansion of the housing market and the implementation of the progressive housing policies of the second half of the 2000s. Over the period, the survey was collected using the same methodology, including the questionnaire and sampling design, though there was an update of the sampling frame following the 2007 population census.

The National Household Survey is a continuous survey that collects detailed information on demographics, education, health, labor market status, farm and nonfarm income, participation in social programs, and perceptions of governance, democracy, and transparency.[4] The survey is the official source of information on monthly indicators on trends in poverty, inequality, and living conditions among households. It is representative across the 25 regions, plus Lima Province, which is separate from the Lima Region. Within each region, the survey is representative across urban and rural areas. The survey sampled between 21,919 households, in 2003, and 36,996 households, in 2017.

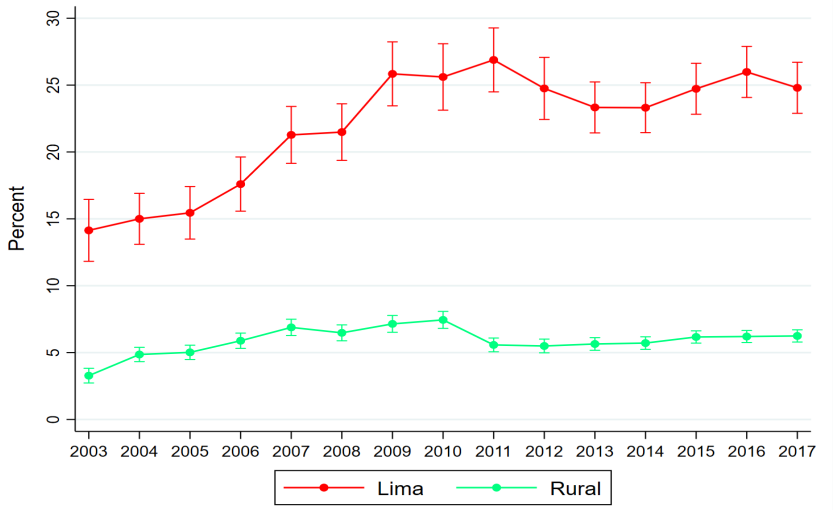
The first section of the questionnaire collects information on the dwelling characteristics of households, including type of dwelling, building materials, number of rooms, ownership status, renovations recently undertaken and the associated means of financing, and utilities and other facilities serving the dwelling as well as the cost of these. The ownership status distinguishes the following six categories: (1) tenants, (2) outright owners, (3) owners by squatting, (4) owners paying mortgage, (5) individuals living for free in an accommodation provided by an employer, and (6) accommodation provided by any other individual or entity.[5] Tenants report the monthly rent paid, while all other individuals are asked how much they could make if they were to rent the dwelling they live in.

[4] The National Institute of Statistics and Informatics, the national statistics office, conducted a first survey of living standards in 1985 with the support of the World Bank. Subsequently, it undertook a permanent annual national survey on living conditions and poverty in 1995. In 1997, the methodology was improved with the support of the Inter-American Development Bank, the United Nations Economic Commission for Latin America and the Caribbean, and the World Bank. Starting in 2004, the methodology has been updated, including a change in expansion factors following the 2007 population census.

[5] Since the beginning of the urbanization process in Peru, in the mid-1950s, the predominant method used by rural migrants to acquire housing has been squatting on peripheral land. For instance, see Fernández-Maldonado (2010); Gwinner (2007).

The study makes use of tenants and owners, both outright and paying mortgages, but leaves out other ownership status categories. The choice is motivated by the goal of assessing the accuracy of homeowner self-assessed rental values relative to the rental market prices paid by tenants. The study therefore excludes from the sample homeowners who do not pay for the housing in which they live and subsidized tenants who obtain their dwellings free or at subsidized prices. The study also restricts the sample to households with heads ages 25–64.[6] The analysis retains households in Metropolitan Lima and households residing in rural areas of the country to ascertain whether living in areas with more highly developed and thicker rental markets makes a difference in the ability of homeowners to assess the rental market values of their dwellings. The share of tenants in Metropolitan Lima rose from an average of 14 percent in 2003 to about 25 percent in 2017, an increase of about 11 percentage points, for the reasons described in section 2 (Figure 3). Over the same period, the rental market expanded in rural areas, too. However, the growth in the share of rental market tenants was limited, from 3.3 percent in 2003 to 6.2 percent in 2017.

Figure 3. Share of rental market tenants, Metropolitan Lima and rural areas, 2003–17



Source: Based on National Household Survey data.

6. Results

This section opens with a description of the differences between the characteristics of homeowner and tenant dwellings. It then shows the unconditional differences between the self-assessed rental values of homeowners and the market rents reported by tenants. Finally, it measures the accuracy of self-assessed rental values using matching estimators that correct for the lack of common support in dwelling characteristics.

Descriptive Statistics

In the spirit of the hedonic theory of consumption (Freeman 2003; Lancaster 1966) and in line with the literature on rent imputation, the following characteristics of dwellings are considered: the number of rooms;

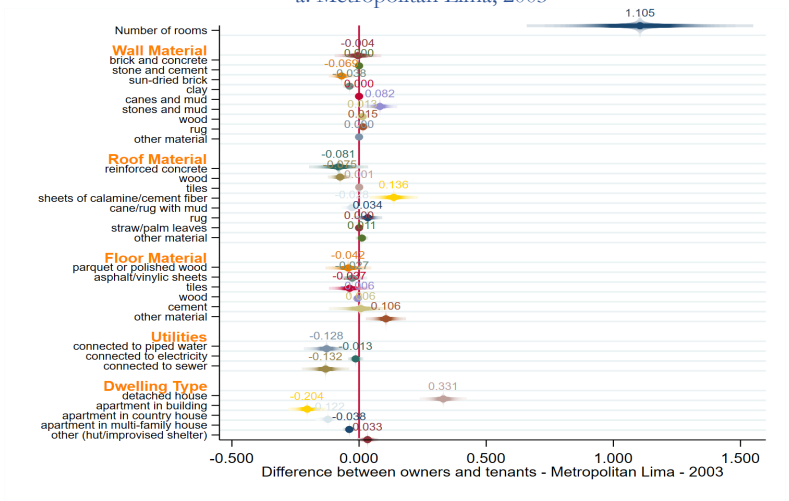
[6] Furthermore, it excludes all households on which the self-assessed rents or market rents are missing, accounting for less than 3 percent of the total 2003–17 sample. Less than 0.3 percent of the total sample is dropped because information on dwelling characteristics is missing.

the materials used in the construction of external walls, roofs, and floors; dummies to indicate the absence or presence of access to water, electricity, and sewerage by the dwelling; a set of categorical variables for the type of dwelling, including detached house, apartment in building, apartment in detached house with independent water access and sewerage, apartment in detached house with shared water services; and hut or other type of dwelling.

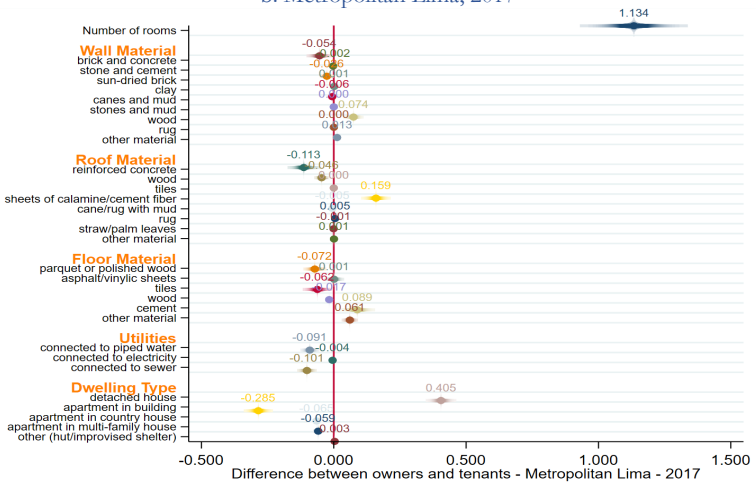
Figure 4, panels a and b, illustrate the differences between tenants and owners in the characteristics of their dwellings in the last year in the dataset. (Annex A, Figure A.1 shows the differences in the remaining years.) In Metropolitan Lima, owners live in dwellings that have an average of one extra room relative to tenants. The dwellings are, however, typically built with lower-quality construction materials. For example, the exterior walls of about 82.8 percent of homeowner dwellings are built with bricks and concrete, while the corresponding share of tenant dwellings is more than 88.0 percent. Among homeowner dwellings, 10.8 percent have wooden exterior walls; this is true of only 3.4 percent of tenant dwellings. Similarly, the material used in roofs in dwellings inhabited by owners is predominantly reinforced concrete (67.5 percent), followed by sheets of calamine and cement fiber (26.7 percent); in the case of tenant dwellings, more than 78.0 percent of the roofs are made of reinforced concrete, followed by 8.2 percent that are made of wood. For the construction materials of floors, 53.0 percent of the dwellings inhabited by homeowners use cement, compared with 44.3 percent of rental dwellings. The second most widely used flooring material is tiles: 20.2 percent in the case of homeowner dwellings and 26.4 percent in the case of tenant dwellings, followed by parquet or polished wood in 9.8 percent and 17.0 percent of the cases. Tenant dwellings are more likely to be connected to piped water and sewerage. Among tenant dwellings, 99.2 percent are connected to piped water, compared with only 90.0 percent of homeowners. Virtually all rental dwellings are connected to sewerage, compared with about 89.1 percent of dwellings inhabited by the homeowners. No significant difference is found in access to electricity. Homeowners live mainly in detached homes (97 percent), while tenants are more likely to live in apartments (45 percent).

Figure 4. Differences in dwelling characteristics, homeowners and tenants, Metropolitan Lima and rural areas, 2003 and 2017

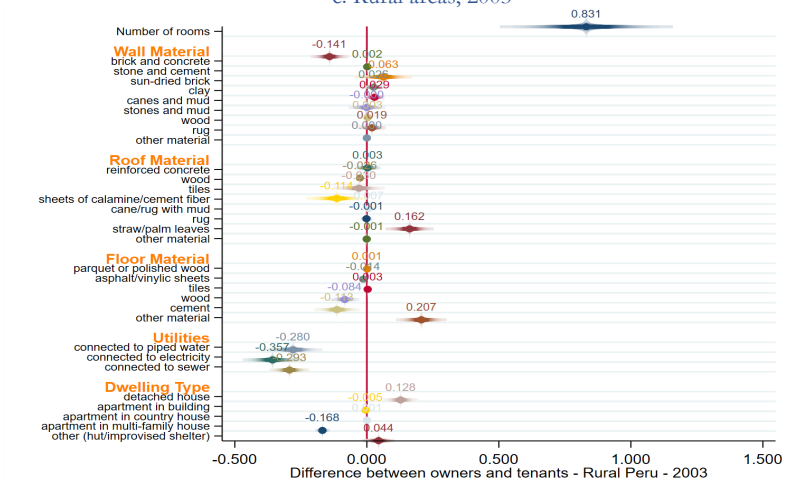
a. Metropolitan Lima, 2003



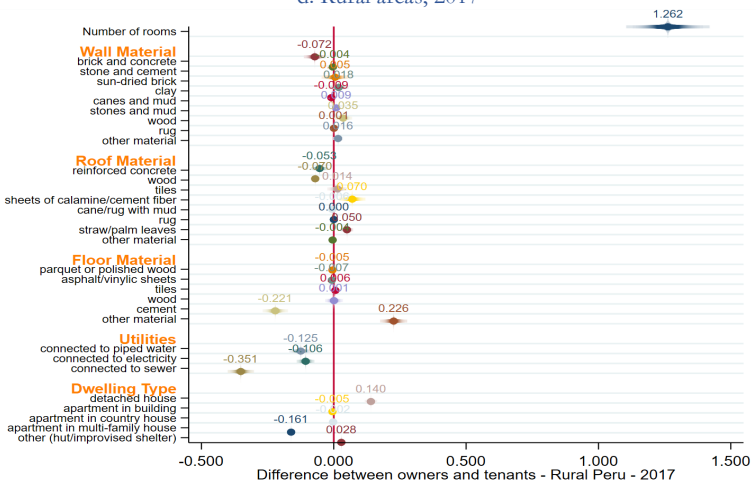
b. Metropolitan Lima, 2017



c. Rural areas, 2003



d. Rural areas, 2017



Source: Based on National Household Survey data.

Although the average dwellings in rural areas are lower in quality relative to the average dwellings in Metropolitan Lima, differences may still be detected in the dwellings inhabited by homeowners and tenants. Homeowner dwellings have on average of 3.2 rooms, whereas rental dwellings have around one room less (2.0 rooms). The exterior walls of homeowner dwellings are built of sun-dried bricks in 44.0 percent of the cases, followed by bricks and concrete at 19.6 percent and clay at 16.3 percent. Among tenant dwellings, the predominant material used for exterior walls is sun-dried bricks (43.4 percent), followed by bricks and concrete at 26.8 percent and clay at 14.5 percent. Roofs are predominantly built with sheets of calamine and cement fiber in the case of both homeowner (63.4 percent) and tenant (56.4 percent) dwellings. Reinforced concrete is used in 16.0 percent of tenant dwellings, but only in 10.6 percent of homeowner dwellings. The use of wood also varies. Virtually no homeowner dwellings use wood, compared with 7.5 percent of tenant dwellings. More than one in every two tenant dwellings have cement floors, while this is true of fewer than one in three homeowner dwellings. Tenants live in dwellings that are more likely to be connected to piped water (88.6 percent, compared with 76.1 percent among homeowner dwellings) and sewerage (71.9 percent, compared with 36.8 percent among homeowners, a difference of 35 percentage points). The difference between homeowner and tenant dwellings in the access to electricity is about 11 percentage points. Homeowners live predominantly in detached homes (95 percent), while tenants live in detached homes (81 percent) and apartments in multifamily homes (28 percent).

Between 2003 and 2017, there was general improvement in the quality of dwellings in Metropolitan Lima and in rural areas. For example, among dwellings inhabited by homeowners or tenants, a growing share in rural areas had exterior walls made of brick and cement; a growing share in urban areas had floors covered with tiles; and a growing share in both areas had roofs made of reinforced concrete and connections to piped water, electricity, and sewerage. However, the improvements were not uniform among tenants and homeowners, and the differences in dwelling characteristics widened in certain dimensions. For instance, the difference in the number of rooms between homeowner- and tenant-inhabited dwellings increased by about a half room in rural areas. The share of dwellings connected to utilities rose by about 3.0 percentage points in access to sewerage and 3.7 percentage points in access to piped water between homeowners and tenants, respectively, in Metropolitan Lima. In rural areas, the gap in access to piped water widened by 15.5 percentage points and in access to sewerage by 25.0 percentage points.

The rental market expanded considerably, particularly in Metropolitan Lima. This may have generated more and better information about the rental values of dwellings. Yet, the growth of the market does not guarantee that the average difference between the self-assessed rental values reported by homeowners and market rents is statistically equal to zero. The characteristics of homeowner- and tenant-inhabited dwellings that contribute in determining rental price assessments may continue to differ over time no matter how quickly a country develops or urbanizes and no matter how much rental markets grow. This points to the importance of comparing dwellings with similar characteristics to ascertain the accuracy of homeowner self-assessed rental values.

The rental value reported by homeowners based on their own estimates of the amounts they could make if they rented out the dwellings they live in is used as an approximation of the values of their consumption of housing services. Table 1 shows the unconditional average self-assessed rental values reported by homeowners and the difference with respect to the average market rents paid by tenants for each year covered in the analysis and, separately, for Metropolitan Lima and rural areas. Self-assessed and market rents declined because of the 2008–

09 global financial crisis in Metropolitan Lima, but not in rural areas. Both self-assessed rental values and reported market rents increased by around 3 percent a year in Metropolitan Lima and by about 7 percent a year in rural areas. The unconditional difference varies year-by-year and ranges between S/. −696 in 2017 and S/. 2,287 in 2007 in Metropolitan Lima, which means that, in some years, homeowners seem to have underestimated the rental market values of their dwellings, while, in other years, they seem to have overestimated the values. Meanwhile, in rural areas, homeowners seem to have always underestimated the rental values of their dwellings relative to the rents paid by market tenants. In rural areas, the average unconditional inaccuracy seem to have expanded during the most recent years.

Table 1. Average self-assessed monthly rents and unconditional difference with market rents paid by tenants, 2003–17

	Metropolitan Lima			*Rural areas*		
	Self-assessed rent	*Delta*	*P-value*	*Self-assessed rent*	*Delta*	*P-value*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
2003	5,081	−489	0.378	480	−218	0.000
2004	7,125	211	0.717	719	−316	0.000
2005	7,456	2,229	0.001	701	−254	0.000
2006	7,589	1,717	0.005	725	−285	0.000
2007	7,634	2,287	0.000	915	−262	0.000
2008	6,469	1,546	0.000	985	−209	0.021
2009	6,720	1,699	0.000	1,037	−177	0.012
2010	6,496	673	0.111	1,042	−309	0.000
2011	7,012	1,191	0.003	1,180	−125	0.184
2012	7,190	1,201	0.003	1,203	−298	0.001
2013	7,013	−17	0.962	1,223	−369	0.000
2014	7,512	319	0.370	1,286	−347	0.000
2015	7,969	924	0.014	1,256	−432	0.000
2016	8,170	188	0.606	1,293	−476	0.000
2017	7,956	−696	0.072	1,302	−545	0.000

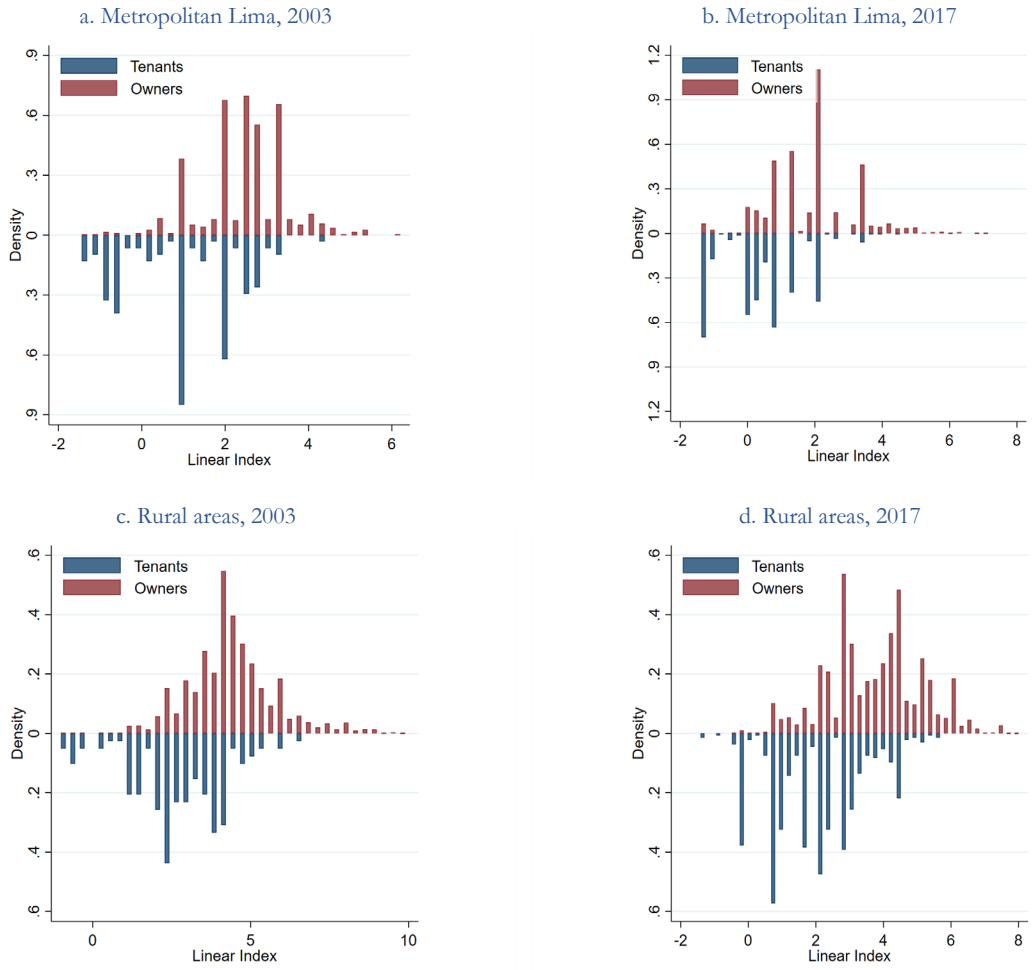
Source: Based on National Household Survey data.
Note: Monetary values are expressed in Peruvian sol in 2017 prices using CPIs of the WDI database.

Propensity Score Matching Results

The conditional independence assumption requires that the selection be solely based on observables. Any differences in outcomes between treated and control individuals with the same values of observable characteristics are thus ascribable to treatment. The analysis is guided by the hedonic theory of consumption, and the set of covariates includes the following: quartiles of the total number of rooms; a dummy for dwellings with external walls made of low-quality materials, including clay, canes and mud, stones and mud, wood, or rug; a dummy for dwellings with floors made of low-quality materials, including dirt; a dummy for dwellings with roofs made of low-quality materials, including cane and rug with mud, rug, or straw and palm leaves; a dummy for public water network access inside the dwelling or outside the dwelling, but inside the building; a dummy for access to electricity; a dummy for public sewerage access inside the dwelling or outside the dwelling, but inside the building; a set of five dummies for dwelling type, that is, detached house, apartment in a building, apartment in a detached house with independent water access and sewerage, apartment in a detached house with shared water services, and a hut or other type of dwelling; and a set of three dummies capturing geographical location outside Metropolitan Lima (the Costa, Selva, or Sierra). A logit regression of the probability of being a homeowner is estimated over the covariates listed above separately for each year and for Metropolitan Lima and rural areas.

In addition to the conditional independence assumption, propensity score matching estimators require the satisfaction of the overlap and common support condition. Figure 5 illustrates the density distributions of the linear index of the propensity scores of homeowners and tenants in Metropolitan Lima (panels a and b) and in rural areas (panels c and d) for the first and last year of the period of analysis. As argued by Lechner (2000), if the observations are concentrated in the tails of the propensity score distribution, the linear index may be the preferred choice for a visual inspection of the densities aimed at identifying a lack of common support as well as to perform matching. In the case under analysis, the common support condition is unlikely to be met because most of the density bins of the linear index among homeowners lack support among tenants in the upper tail of the index distribution in Metropolitan Lima and in rural areas. The lack of common support is not attenuated over time, as highlighted by the lack of support in 2017 (panels b and d).

Figure 5. Distribution of the linear index of the propensity score, by ownership status, Metropolitan Lima and rural areas, 2003 and 2017



Source: Based on National Household Survey data.

To impose the common support condition, the analysis relies on two methods, as follows: (1) a comparison of minima and maximums and (2) trimming. The first method is based on comparing the minima and maximums of the linear index in both groups and on dropping homeowners whose index is higher than the maximum or less than the minimum index of tenants. The second method consists of estimating the density distribution of

the linear index in both groups and dropping the share—2 percent in the case here—of homeowner observations in which the linear index density of the tenant observations is the lowest.

To assess the quality of the matching exercise, the analysis considers two balancing tests. First, as proposed by Sianesi (2004), the propensity score is reestimated only on treated and matched untreated, and the pseudo-R2s are compared before and after matching. After matching, no systematic differences should be detected in the distribution of covariates between the treated and control groups and the pseudo-R2 should be fairly low. Second, the analysis uses Rosenbaum and Rubin's (1983) absolute standardized difference (B) of the means of the linear index of the propensity score in the treated and (matched) nontreated group. Rubin (2001) recommends that B be less than 25 if the matched sample is to be considered sufficiently balanced. In Metropolitan Lima, kernel and radius matching is able to balance the distribution of the covariates in the homeowner and tenant groups as indicated by the balancing tests (Table 2). The pseudo-R2s of a regression estimated on the matched sample is low in all years but two, namely, 2007 and 2010, in the case of kernel matching. A similar conclusion can be derived by looking at the absolute standardized difference, which is below the 25 threshold with the same exceptions. Nearest neighbor matching does not achieve a balance in 40 percent of the years considered, namely, in 2003–04, 2007, 2009–10, and 2011. By contrast, in the case of rural areas, the matching procedure does not achieve a balance in the distribution of covariates in the majority of the years considered in the analysis if kernel and nearest neighbor marching is used. In the case of radius matching, which imposes a caliper, the balance is obtained; both the pseudo-R2s are virtually zero in all years; and the absolute standardized difference is below 25.

Table 2. Balancing tests before and after matching, Metropolitan Lima and rural areas, by year, 2003–17

	Metropolitan Lima											
	Kernel				*Nearest neighbor*				*Radius*			
	R2 before	*R2 after*	*Rubin's B before*	*Rubin's B after*	*R2 before*	*R2 after*	*Rubin's B before*	*Rubin's B after*	*R2 before*	*R2 after*	*Rubin's B before*	*Rubin's B after*
2003	0.213	0.008	121.7	21.6	0.213	0.039	121.7	45.9	0.213	0.000	121.7	0.0
2004	0.268	0.004	138.8	14.8	0.268	0.046	138.8	49.0	0.268	0.000	138.8	0.0
2005	0.183	0.002	111.7	9.4	0.183	0.009	111.7	21.9	0.183	0.000	111.7	1.2
2006	0.227	0.008	128.4	20.9	0.227	0.008	128.4	20.9	0.227	0.000	128.4	0.0
2007	0.243	0.016	134.3	29.2	0.243	0.021	134.3	33.5	0.243	0.000	134.3	5.2
2008	0.272	0.008	142.1	20.6	0.272	0.011	142.1	24.6	0.272	0.000	142.1	2.9
2009	0.257	0.008	137.4	21.1	0.257	0.019	137.4	31.8	0.257	0.000	137.4	0.0
2010	0.240	0.016	129.9	29.2	0.240	0.039	129.9	46.4	0.240	0.000	129.9	0.0
2011	0.215	0.002	122.3	10.8	0.215	0.006	122.3	18.9	0.215	0.001	122.3	7.4
2012	0.244	0.001	135.6	7.6	0.244	0.021	135.6	33.4	0.244	0.000	135.6	0.0
2013	0.219	0.003	125.5	12.2	0.219	0.005	125.5	17.0	0.219	0.000	125.5	0.0
2014	0.195	0.003	117.1	12.3	0.195	0.005	117.1	16.5	0.195	0.000	117.1	1.7
2015	0.214	0.004	121.7	15.2	0.214	0.004	121.7	14.6	0.214	0.000	121.7	0.6
2016	0.167	0.008	104.4	21.0	0.167	0.006	104.4	17.7	0.167	0.000	104.4	0.0
2017	0.206	0.001	117.8	8.0	0.206	0.007	117.8	20.1	0.206	0.000	117.8	0.0
	Rural areas											
	Kernel				*Nearest neighbor*				*Radius*			
	R2 before	*R2 after*	*Rubin's B before*	*Rubin's B after*	*R2 before*	*R2 after*	*Rubin's B before*	*Rubin's B after*	*R2 before*	*R2 after*	*Rubin's B before*	*Rubin's B after*
2003	0.174	0.025	117.2	37.5	0.174	0.055	117.2	55.3	0.174	0.002	117.2	9.1
2004	0.211	0.015	127.3	29.4	0.211	0.030	127.3	41.5	0.211	0.000	127.3	2.0
2005	0.222	0.011	135.9	24.4	0.222	0.023	135.9	35.6	0.222	0.001	135.9	9.0
2006	0.187	0.011	126.8	24.7	0.187	0.017	126.8	30.7	0.187	0.007	126.8	19.5
2007	0.293	0.021	160.2	34.4	0.293	0.089	160.2	72.7	0.293	0.002	160.2	11.2
2008	0.244	0.012	142.3	25.5	0.244	0.019	142.3	32.9	0.244	0.000	142.3	0.9
2009	0.251	0.018	145.2	31.9	0.251	0.036	145.2	45.5	0.251	0.001	145.2	5.3
2010	0.258	0.011	148.1	25.1	0.258	0.027	148.1	39.3	0.258	0.000	148.1	3.2
2011	0.272	0.007	157.3	19.3	0.272	0.051	157.3	52.2	0.272	0.000	157.3	1.3
2012	0.229	0.007	136.9	20.2	0.229	0.009	136.9	22.4	0.229	0.005	136.9	17.1
2013	0.216	0.014	136.7	28.3	0.216	0.019	136.7	32.5	0.216	0.001	136.7	5.9
2014	0.235	0.014	144.1	28.0	0.235	0.030	144.1	41.0	0.235	0.001	144.1	8.4
2015	0.236	0.010	139.7	23.9	0.236	0.015	139.7	28.9	0.236	0.000	139.7	4.0
2016	0.187	0.004	123.4	15.7	0.187	0.011	123.4	24.2	0.187	0.000	123.4	4.7
2017	0.195	0.006	128.2	18.4	0.195	0.022	128.2	34.9	0.195	0.001	128.2	5.9

Source: Based on National Household Survey data.

The fact that nearest neighbor matching based on the closest neighbor in terms of the linear index of the propensity score fails to achieve the balance in the distribution of covariates is not surprising because the single tenant observation chosen as a matching partner for a homeowner might be different in terms of observable characteristics. By contrast, radius matching overcomes this risk by imposing a tolerance level on the maximum distance in terms of the propensity score for the comparison group observations that are used. This is reflected in the high degree of balance of the covariates in the matched samples throughout the period of analysis.

Kernel matching might not readily achieve a good balance because it makes use of all the observations in the tenant group that lie in the region of the common support and weights them according to the distance between each tenant and the homeowner for whom the counterfactual is estimated.

Next, the analysis turns to the estimates of the accuracy of homeowner self-assessments based on the kernel, nearest neighbor, and radius matching that are reported in Table 3. In Metropolitan Lima, homeowners are able to provide accurate estimates of the rental market values of their dwellings in every year of the survey, with the exception of the two most recent years (2016 and 2017) when kernel and radius matching algorithms are used. In 2016 and 2017, homeowners underestimated the market values by about 20 percent. In rural areas, the self-assessed rental values estimated by homeowners seem to have been less accurate in more instances, precisely in the most recent years between 2014 and 2017 in the case of kernel and radius matching and in 2016 only in the case of nearest neighbor matching. Considering that the balance between homeowners and tenants is achieved over the entire period only with radius matching, one may conclude that homeowners provided inaccurate estimates of the market rents of their dwellings in 2012 and between 2014 and 2017. The absolute size of the inaccuracy ranges, on average, from −24 percent between 2014 and 2017 to +14 percent in 2012. Thus, with one exception, the direction of the inaccuracy is the same in Metropolitan Lima and in rural areas, that is, homeowners underestimate the market rental values of their dwellings by about −25/−20 percent.

Table 3. Estimates of the average accuracy of homeowner self-assessed rents based on matching estimators, Metropolitan Lima and rural areas, by year, 2003–17

	Metropolitan Lima						*Rural areas*					
	Kernel		*Nearest-neighbor*		*Radius*		*Kernel*		*Nearest-neighbor*		*Radius*	
	Difference	*P-value*	*Difference*	*P-value*	*Difference*	*P-value*	*Difference*	*P-value*	*Difference*	*P-value*	*Difference*	*P-value*
2003	−1326	0.328	−4793	0.072	−1363	0.370	−27	0.649	2	0.989	−31	0.462
2004	−628	0.479	76	0.976	−643	0.522	−73	0.301	−29	0.758	−76	0.209
2005	677	0.277	1840	0.141	696	0.158	59	0.226	84	0.279	11	0.737
2006	245	0.769	927	0.610	461	0.627	−22	0.828	88	0.558	−183	0.237
2007	461	0.652	2809	0.240	568	0.617	27	0.654	97	0.271	82	0.039
2008	−73	0.905	1073	0.499	39	0.952	70	0.379	156	0.109	75	0.134
2009	437	0.597	1196	0.549	714	0.376	−123	0.281	−26	0.859	−135	0.085
2010	−818	0.269	−2006	0.239	−799	0.200	−11	0.911	−70	0.527	−42	0.561
2011	290	0.593	2399	0.203	222	0.687	55	0.525	−17	0.869	−1	0.992
2012	−207	0.712	−1416	0.426	−118	0.858	−16	0.936	−94	0.833	185	0.044
2013	−594	0.510	716	0.830	−796	0.437	86	0.333	−135	0.216	−22	0.800
2014	−342	0.617	−2608	0.227	−335	0.628	−542	0.011	−723	0.038	−374	0.054
2015	−40	0.966	−1554	0.535	−172	0.824	−307	0.033	−237	0.164	−276	0.041
2016	−1890	0.034	−1806	0.588	−1954	0.052	−325	0.003	−220	0.143	−251	0.007
2017	−2191	0.001	−1657	0.532	−1891	0.008	−419	0.001	−297	0.116	−368	0.000

Source: Based on National Household Survey data.

Note: Kernel matching is performed using the Epanechnikov Kernel and a bandwidth of 0.06. Nearest neighbor matching is performed on the closest neighbor with replacement. Radius matching imposes a caliper of 0.001. Standard errors are bootstrapped with 100 replications.

In addition to the unconditional mean values illustrated above in this section, another simple way to compare homeowner self-assessments with the market rents paid by tenants is to estimate an ordinary least squares regression that provides the average size of the inaccuracy, conditional on a set of covariates. The analysis estimates this regression separately for each year and in Metropolitan Lima and in rural areas using the same set of covariates used in the matching exercise without the imposition of the overlap and common support conditions (Annex A, Table A.1). In Metropolitan Lima, a statistically significant difference is found between homeowner self-assessed values and the market rents paid by tenants in 6 of 15 years, and these years do not overlap with the survey years in which a statistically significant inaccuracy was detected through matching. Similarly, in the case of rural areas, ordinary least squares estimates find a statically significant difference between what estimated by homeowners and what tenants pay in 7 out of 15 years and only in 2012 and 2017 both ordinary least squares and matching estimators identified a statistically significant difference.

7. Conclusions

This paper proposes the use of matching estimators to test the accuracy of the value reports of homeowners by making use of information from a typical household budget survey. The lack of supplemental sources, such as administrative data from national cadastres, is a common issue faced by researchers and practitioners working in developing countries.

The findings indicate that, in 2003–17, homeowners in Metropolitan Lima, on average, provided, with the exception of the two most recent years, accurate estimates of the housing services values of the dwellings they were living in, compared with tenants living in dwellings with observably similar characteristics. In rural areas, two matching algorithms, namely, kernel and nearest neighbor, do not achieve a good balance of the matched sample in the majority of the years considered. Homeowners did not provide accurate market rental values of their dwellings between 2014 and 2017 and underestimated the rental value on average by about −24 percent, a magnitude similar to that detected in Metropolitan Lima in 2016 and 107 (−20 percent).

As a note of caution, matching algorithms differ in the number of tenant observations used and in the way they are used to estimate the counterfactual. For these reasons, they also differ in terms of the capacity to achieve a balance in terms of dwelling characteristics. This is reflected in the comparability of the matched sample and ultimately in the reliability of the accuracy test. Hence, should practitioners resort to matching algorithms to assess the accuracy of homeowner self-assessed rental values, as the paper argues, attention should be paid to the choice of the algorithm as well as to the satisfaction of the overlap and common support conditions. This is key to obtaining a balanced matched sample and to proceed to a comparison using comparable observations.

This paper offers a cautionary tale about the suitability of self-assessed rental values in obtaining a good approximation of market rental values and argues that adequate methods, such as matching estimators, should be used to avoid comparing the incomparable given that considerable differences exist in terms of dwelling characteristics between homeowners and tenants in many developing countries. Homeowners may or may not provide accurate estimates of the rental market prices of the dwellings they live in for several reasons, including the way questions during the collection of such information are phrased, sentimental attachment to the properties, and a lack of information because of spatially clustered or thin rental markets.

Self-assessment rental values are often used for the construction of important economic indicators, such as CPIs (as measures of inflation), national accounts, and income and consumption aggregates for the analysis of household living standards. Further research should concentrate on how household budget surveys, which are the typical source of information for these matters, might be improved to capture rental market values more accurately. Possible ways forward might include the oversampling of atypical housing types (including the oversampling of tenants), expert assessments of the dwellings of the sampled households, and linking sampled populations to administrative data containing dwelling prices.

References

ADB (Asian Development Bank). 2014. *ICP 2011: 2011 International Comparison Program in Asia and the Pacific; Purchasing Power Parities and Real Expenditures*. Mandaluyong City, the Philippines: ADB.

AfDB (African Development Bank). 2013. "The Reliability of Economic Statistics in Africa: Special Focus on GDP Measurement." June, Statistical Capacity Building Division, Statistics Division, AfDB, Abidjan, Côte d'Ivoire.

Apeim (Asociación Peruana de Empresas de Investigación de Mercados, Peruvian Association of Market Research Firms). 2007. "Niveles Socioeconómicos en el Perú 2007–2008." November, Apeim, Lima.

Arévalo, Raquel, and Javier Ruiz-Castillo. 2006. "On the Imputation of Rental Prices to Owner-Occupied Housing." *Journal of the European Economic Association* 4 (4): 830–61.

Balcázar, Carlos Felipe, Lidia Ceriani, Sergio Olivieri, and Marco Ranzani. 2017. "Rent-Imputation for Welfare Measurement: A Review of Methodologies and Empirical Findings." *Review of Income and Wealth* 63 (4): 881–98.

Benítez-Silva, Hugo, Selçuk Eren, Frank Heiland, and Sergi Jiménez-Martín. 2015. "How Well Do Individuals Predict the Selling Prices of Their Homes?" *Journal of Housing Economics* 29 (September): 12–25.

Bigelow, Daniel P., Jennifer Ifft, and Todd Kuethe. 2020. "Following the Market? Hedonic Farmland Valuation Using Sales Prices versus Self-Reported Values." *Land Economics* 96 (3): 418–40.

Blundell, Richard, and Mónica Costa Dias. 2009. "Alternative Approaches to Evaluation in Empirical Microeconomics." *Journal of Human Resources* 44 (3): 565–640.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22 (1): 31–72.

Chan, Sewin, Samuel Dastrup, and Ingrid Gould Ellen. 2016. "Do Homeowners Mark to Market? A Comparison of Self-Reported and Estimated Market Home Values during the Housing Boom and Bust." *Real Estate Economics* 44 (3): 627–57.

Corradin, Stefano, José L. Fillat, and Carles Vergara-Alert. 2017. "Portfolio Choice with House Value Misperception." Working Paper 17–16 (October), Federal Reserve Bank of Boston, Boston.

Deaton, Angus S., and Alan Heston. 2010. "Understanding PPPs and PPP-Based National Accounts." *American Economic Journal: Macroeconomics* 2 (4): 1–35.

Fehr, Dietmar, Rustamdjan Hakimov, and Dorothea Kübler. 2015. "The Willingness to Pay–Willingness to Accept Gap: A Failed Replication of Plott and Zeiler." *European Economic Review* 78 (August): 120–28.

Fernández-Maldonado, Ana M. 2010. "Recent Housing Policies in Lima and Their Effects on Sustainability." Paper presented at the International Society of City and Regional Planners' 46th ISOCARP Congress, Nairobi, September 19–23.

Freeman, A. Myrick, III. 2003. *The Measurement of Environmental and Resource Values: Theory and Methods*, 2nd ed. Washington, DC: Resources for the Future.

Frick, Joachim R., Markus M. Grabka, Timothy M. Smeeding, and Panos Tsakloglou. 2010. "Distributional Effects of Imputed Rents in Five European Countries." *Journal of Housing Economics* 19 (3): 167–79.

Gao, Nan, and Pinghan Liang. 2019. "Home Value Misestimation and Household Behavior: Evidence from China." *China Economic Review* 55 (June): 168–80.

Garner, Thesia I., George Janini, William Passero, Laura Paszkiewicz, and Mark Vendemia. 2006. "The CE and the PCE: A Comparison." *Monthly Labor Review* 129 (9): 20–40.

Garner, Thesia I., and Uri Kogan. 2007. "Comparing Approaches to Value Owner-Occupied Housing Using U.S. Consumer Expenditure Survey Data." Paper presented at the Allied Social Science Association–Society of Government Economists' Annual Meeting, Chicago, January 7.

Garner, Thesia I., and Kathleen S. Short. 2001. "Owner-Occupied Shelter in Experimental Poverty Measures." Paper presented at the Southern Economic Association's Annual Meeting, Tampa, November 17–19.

Garner, Thesia I., Kathleen S. Short, and Uri Kogan. 2006. "What Do We Know about the Value of Owner Occupied Housing Services? Rental Equivalence and Other Approaches." Paper presented at the Southern Economic Association's Annual Meeting, Charleston, SC, November 18–21.

Gonzalez-Navarro, Marco, and Climent Quintana-Domeque. 2009. "The Reliability of Self-Reported Home Values in a Developing Country Context." *Journal of Housing Economics* 18 (4): 311–24.

Goodman, John L., Jr., and John B. Ittner. 1992. "The Accuracy of Home Owners' Estimates of House Value." *Journal of Housing Economics* 2 (4): 339–57.

Gwinner, William Britt. 2007. "Housing." In *An Opportunity for a Different Peru: Prosperous, Equitable, and Governable*, edited by Marcelo M. Giugale, Vicente Fretes-Cibils, and John L. Newman, 349–62. Washington, DC: World Bank.

Hanemann, W. Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 81 (3): 635–47.

Heston, Alan W. 1994. "A Brief Review of Some Problems in Using National Accounts Data in Level of Output Comparisons and Growth Studies." *Journal of Development Economics* 44 (1): 29–52.

Heston, Alan W., and Alice O. Nakamura. 2009. "Questions about the Equivalence of Market Rents and User Costs for Owner Occupied Housing." *Journal of Housing Economics* 18 (3): 273–79.

Hill, Robert J. 2013. "Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy." *Journal of Economic Surveys* 27 (4): 879–914.

Ihlanfeldt, Keith R., and Jorge Martinez-Vazquez. 1986. "Alternative Value Estimates of Owner-Occupied Housing: Evidence of Sample Selection Bias and Systematic Errors." *Journal of Urban Economics* 20 (3): 356–69.

Kain, John F., and John M. Quigley. 1972. "Note on Owner's Estimate of Housing Value." *Journal of the American Statistical Association* 67 (340): 803–06.

Kish, Leslie, and John B. Lansing. 1954. "Response Errors in Estimating the Value of Homes." *Journal of the American Statistical Association*, 49 (267): 520–38.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74 (2): 132–57.

Lebow, David E., and Jeremy B. Rudd. 2003. "Measurement Error in the Consumer Price Index: Where Do We Stand?" *Journal of Economic Literature* 41 (1): 159–201.

Lechner, Michael. 2000. "An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany." *Journal of Human Resources* 35 (2): 347–75.

Lechner, Michael. 2008. "A Note on Endogenous Control Variables in Causal Studies." *Statistics and Probability Letters* 78 (2): 190–95.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.

Rosenbaum, Paul R., and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *American Statistician* 39 (1): 33–38.

Rubin, Donald B. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology* 2 (3): 169–88.

Sianesi, Barbara. 2004. "An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990s." *Review of Economics and Statistics* 86 (1): 133–55.

Tunçel, Tuba, and James K. Hammitt. 2014. "A New Meta-Analysis on the WTP/WTA Disparity." *Journal of Environmental Economics and Management* 68 (1): 175–87.

Tur-Sinai, Aviad, Larisa Fleishman, and Dmitri A. Romanov. 2020. "The Accuracy of Self-Reported Dwelling Valuation." *Journal of Housing Economics* 48 (June), 101660.

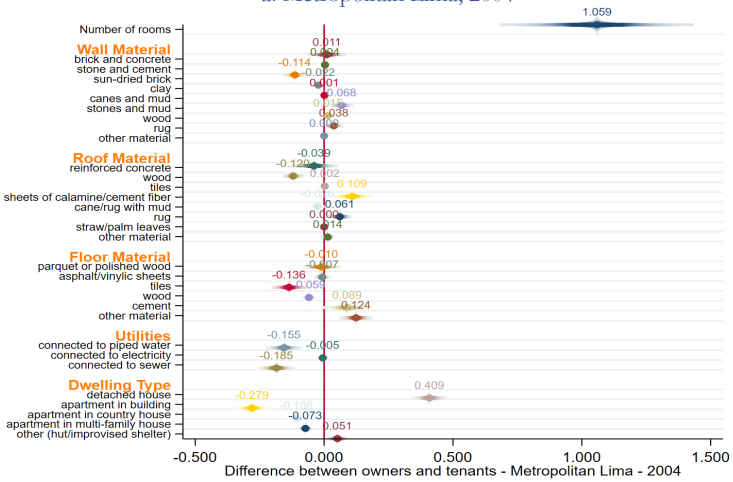
van der Cruijsen, Carin, David-Jan Jansen, and Maarten van Rooij. 2018. "The Rose-Tinted Spectacles of Homeowners." *Journal of Consumer Affairs* 52 (1): 61–87.

World Bank. 2020. On the Construction of a Welfare Indicator for Inequality and Poverty Analysis. Internal Draft, March.

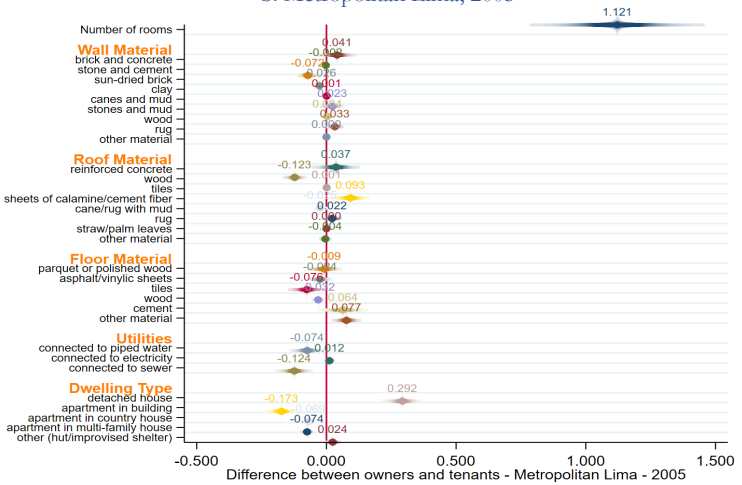
Annex A

Figure A 1. Differences in dwelling characteristics, homeowners and tenants, Metropolitan Lima and rural areas, 2004 and 2005

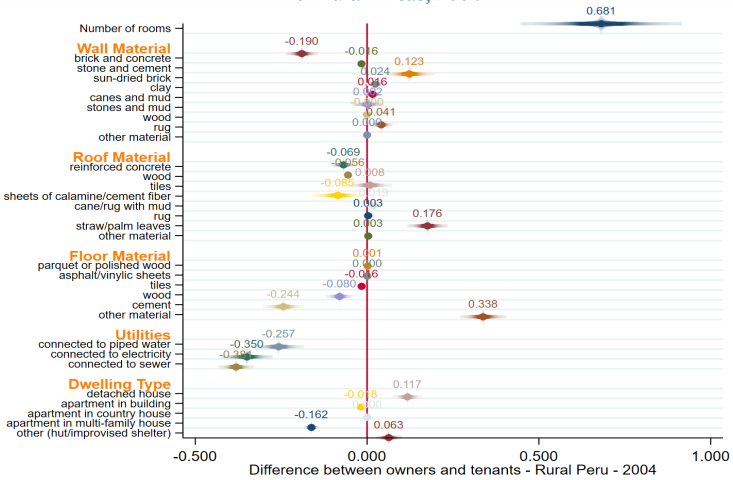
a. Metropolitan Lima, 2004



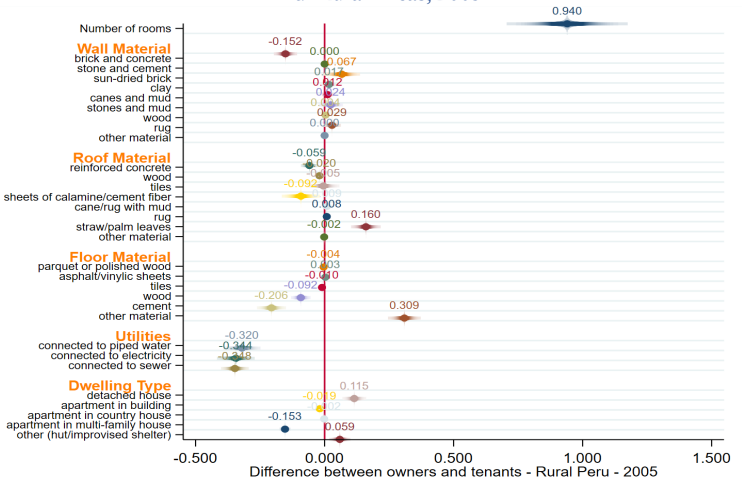
b. Metropolitan Lima, 2005



c. Rural Areas, 2004



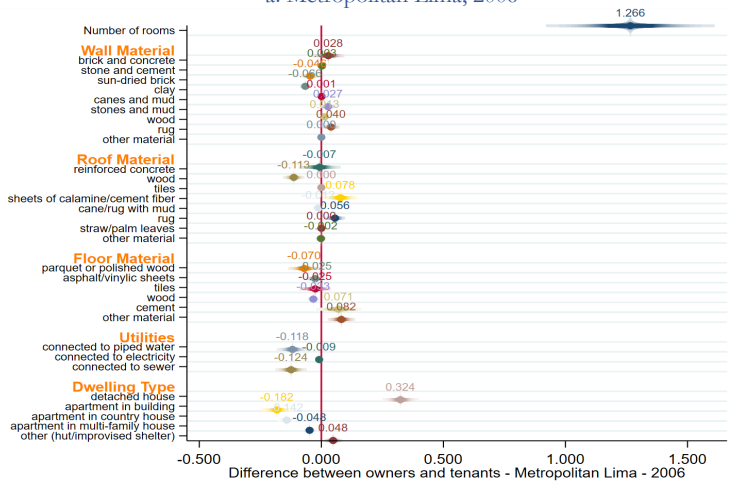
d. Rural Areas, 2005



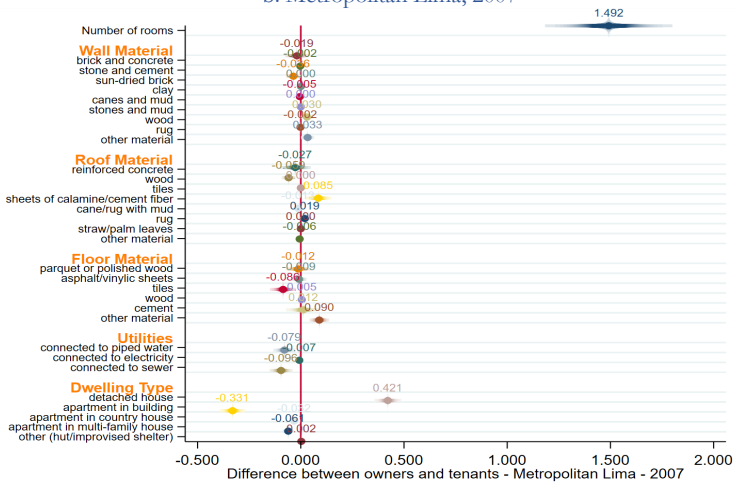
Source: Based on National Household Survey data.

Cont. Differences in dwelling characteristics, homeowners and tenants, Metropolitan Lima and rural areas, 2006 and 2007

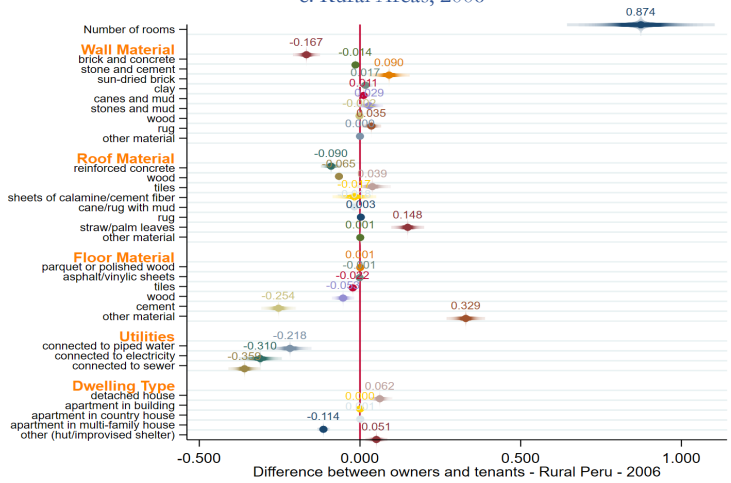
a. Metropolitan Lima, 2006



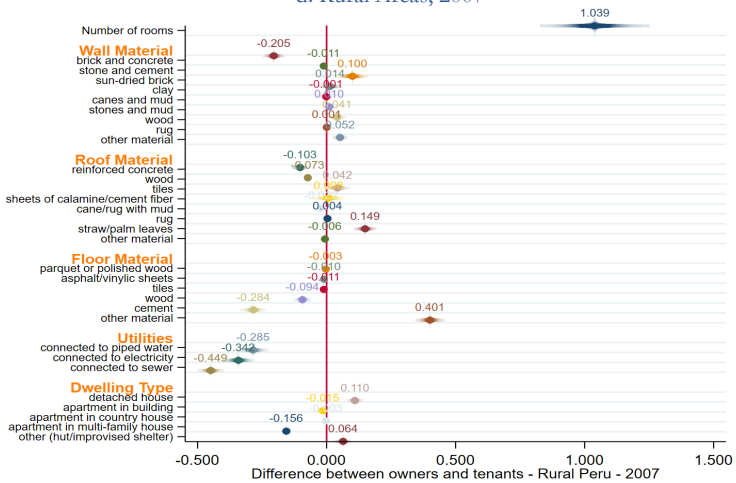
b. Metropolitan Lima, 2007



c. Rural Areas, 2006



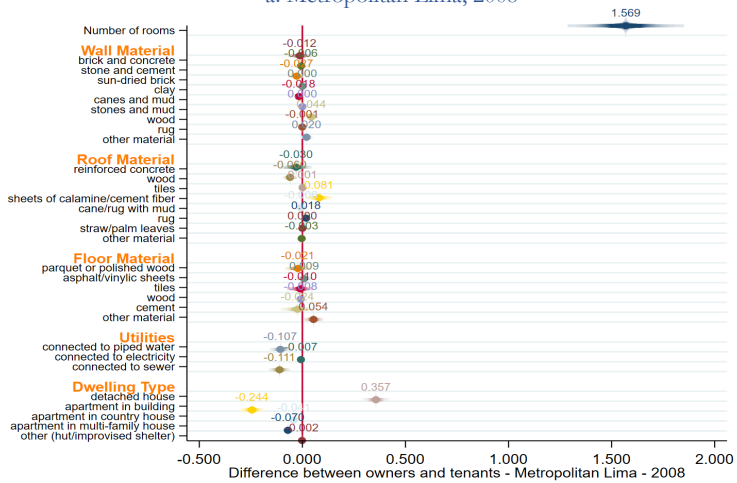
d. Rural Areas, 2007



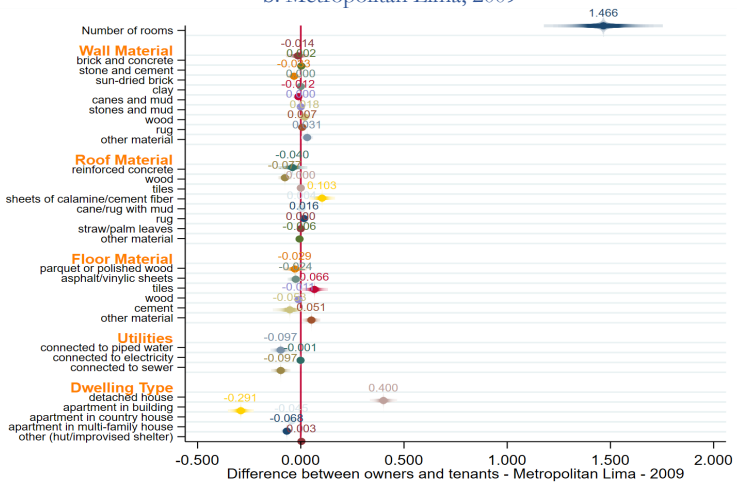
Source: Based on National Household Survey data.

Cont. Differences in dwelling characteristics, homeowners and tenants, Metropolitan Lima and rural areas, 2008 and 2009

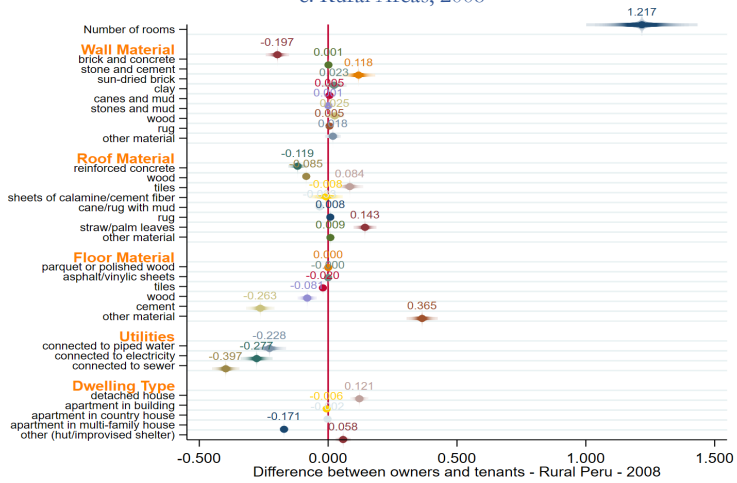
a. Metropolitan Lima, 2008



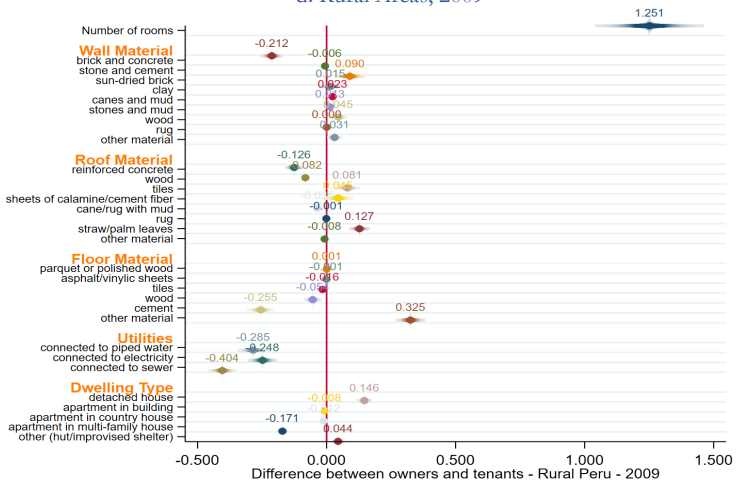
b. Metropolitan Lima, 2009



c. Rural Areas, 2008



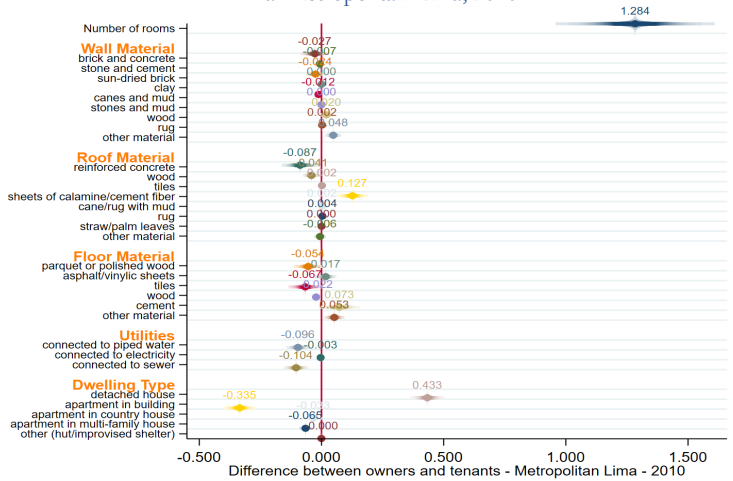
d. Rural Areas, 2009



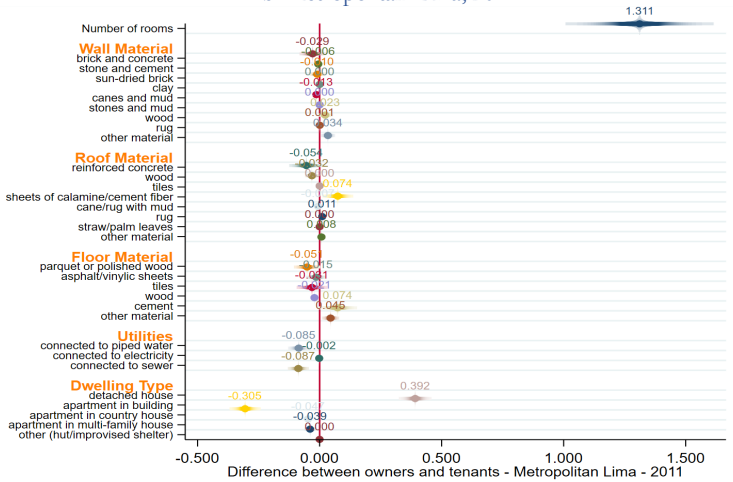
Source: Based on National Household Survey data.

Cont. Differences in dwelling characteristics, homeowners and tenants, Metropolitan Lima and rural areas, 2010 and 2011

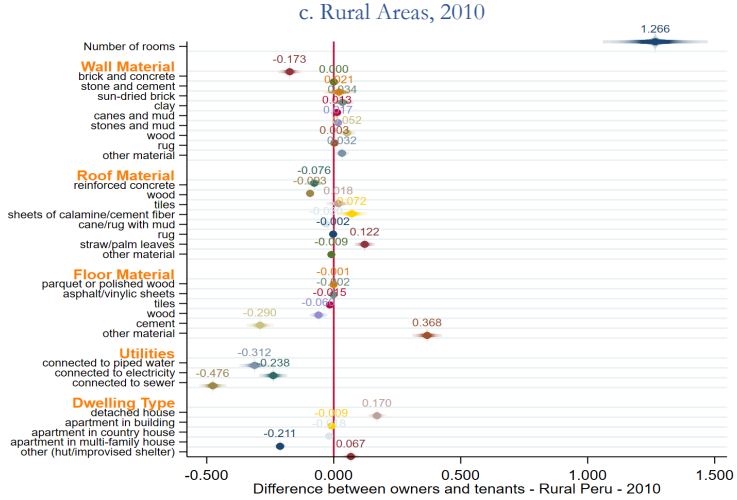
a. Metropolitan Lima, 2010



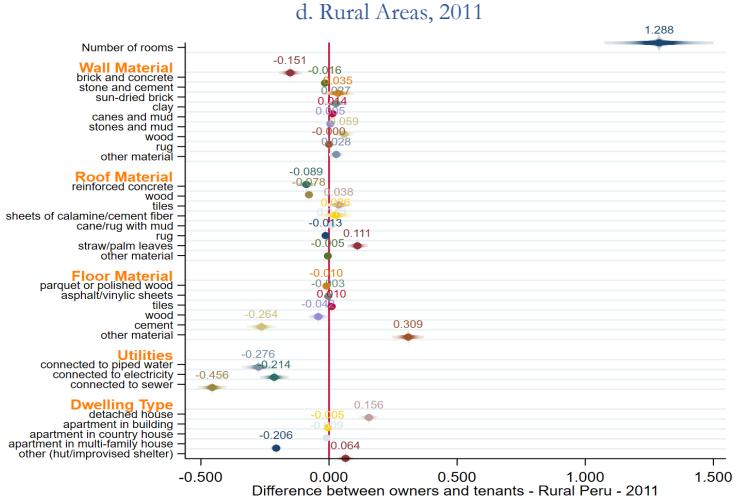
b. Metropolitan Lima, 2011



c. Rural Areas, 2010



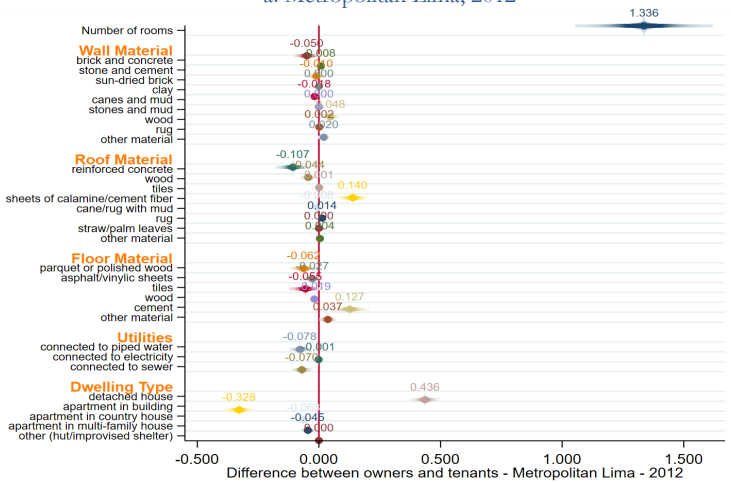
d. Rural Areas, 2011



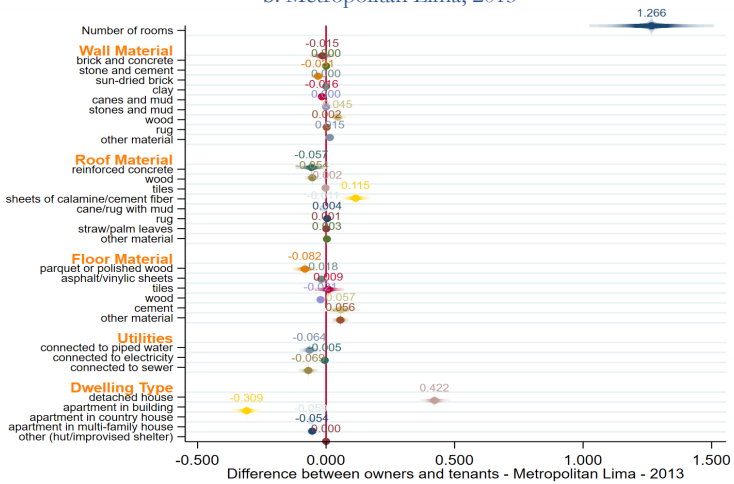
Source: Based on National Household Survey data.

Cont. Differences in dwelling characteristics, homeowners and tenants, Metropolitan Lima and rural areas, 2012 and 2013

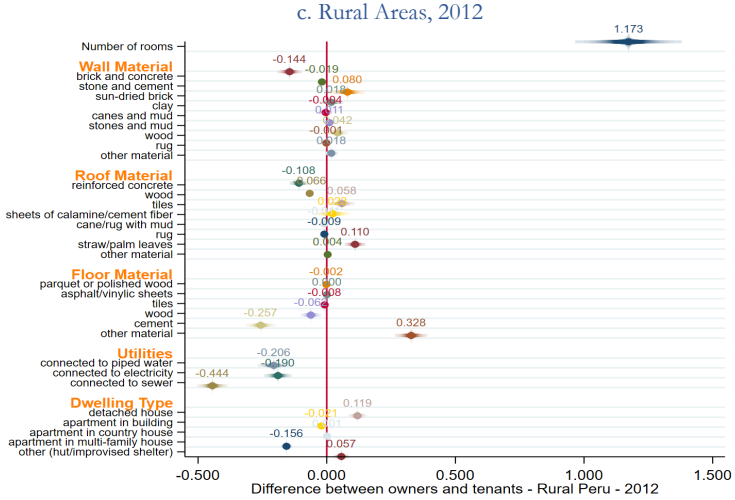
a. Metropolitan Lima, 2012



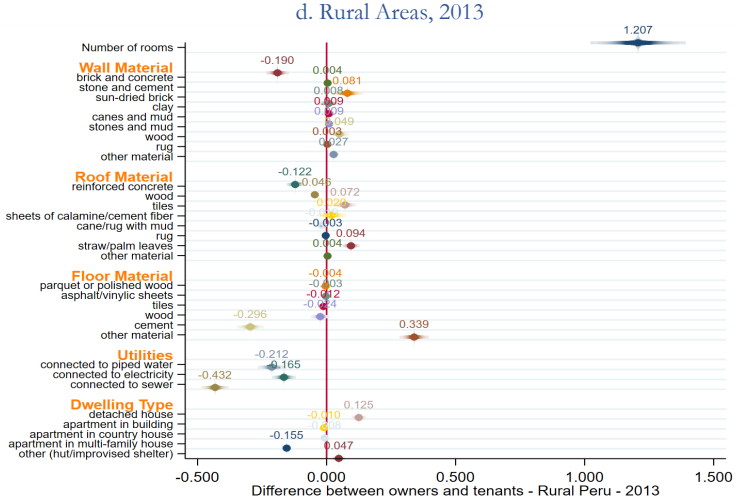
b. Metropolitan Lima, 2013



c. Rural Areas, 2012



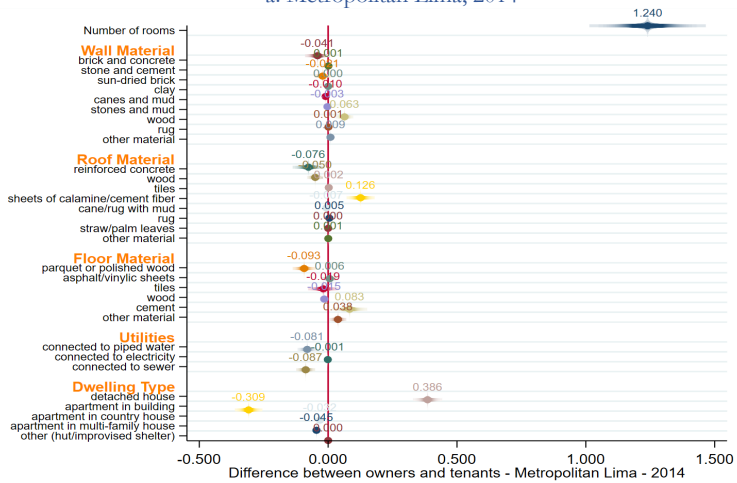
d. Rural Areas, 2013



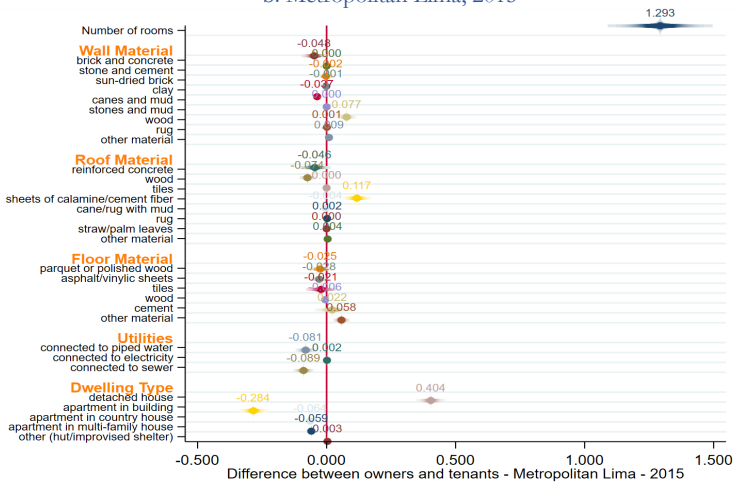
Source: Based on National Household Survey data.

Cont. Differences in dwelling characteristics, homeowners and tenants, Metropolitan Lima and rural areas, 2014 and 2015

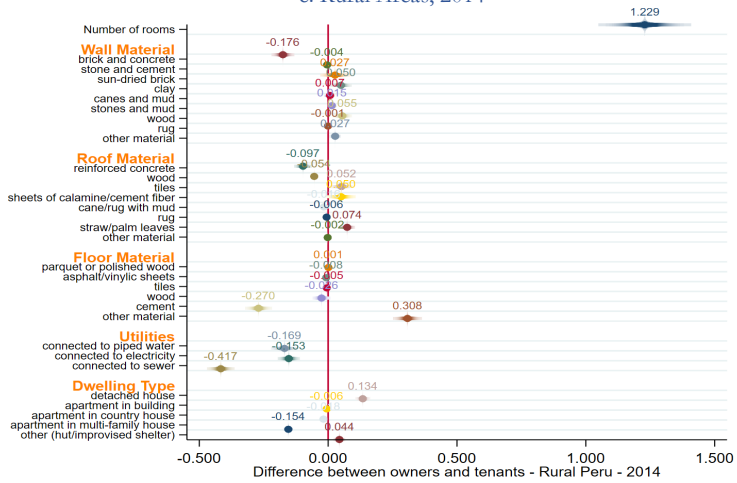
a. Metropolitan Lima, 2014



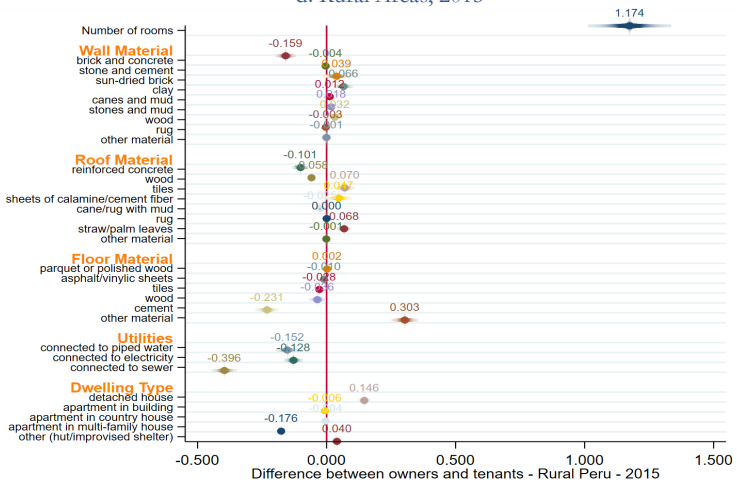
b. Metropolitan Lima, 2015



c. Rural Areas, 2014



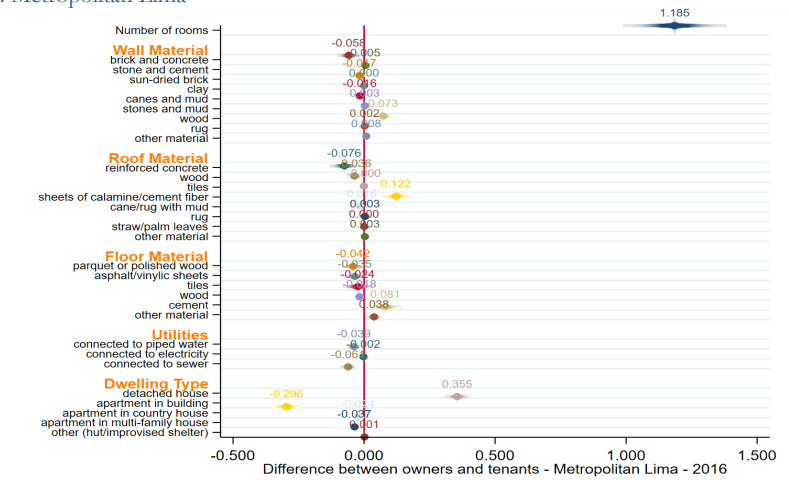
d. Rural Areas, 2015



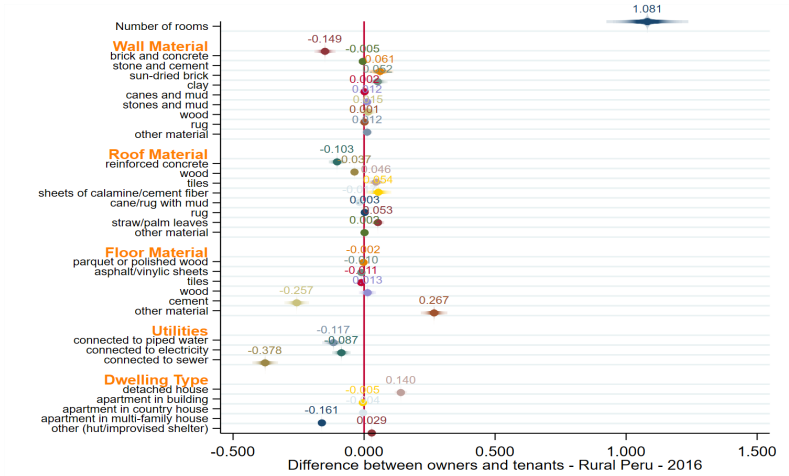
Source: Based on National Household Survey data.

Cont. Differences in dwelling characteristics, homeowners and tenants, Metropolitan Lima and rural areas, 2016

a. Metropolitan Lima



b. Rural Areas

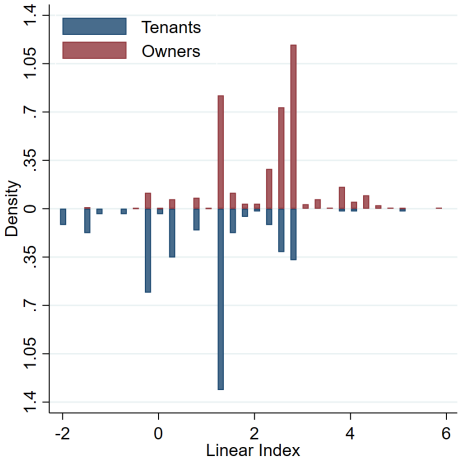


Source: Based on National Household Survey data.

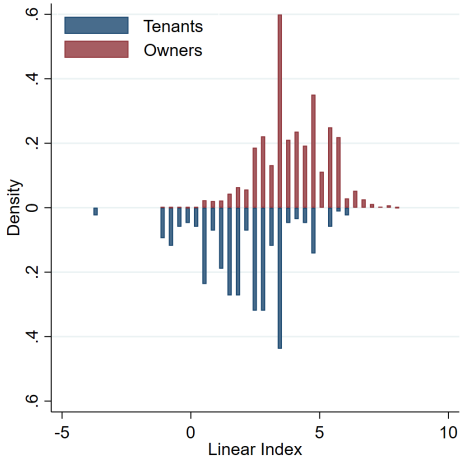
Figure A 2. Distribution of the linear index of the propensity score, by ownership status, Metropolitan Lima and rural areas, 2004 and 2005

a. Metropolitan Lima, 2004

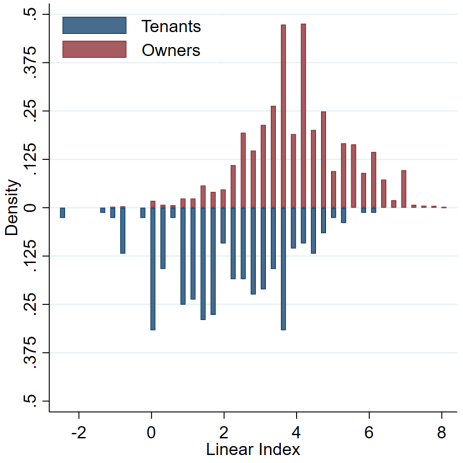
b. Metropolitan Lima, 2005



c. Rural Areas, 2004



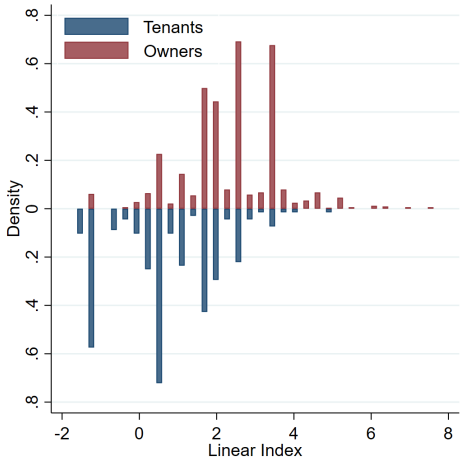
d. Rural Areas, 2005



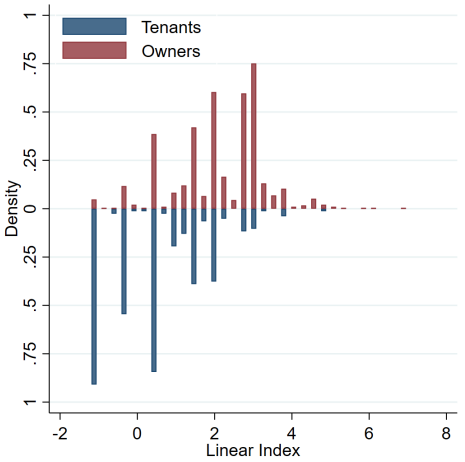
Source: Based on National Household Survey data.

Cont. Distribution of the linear index of the propensity score, by ownership status, Metropolitan Lima and rural areas, 2006 and 2007

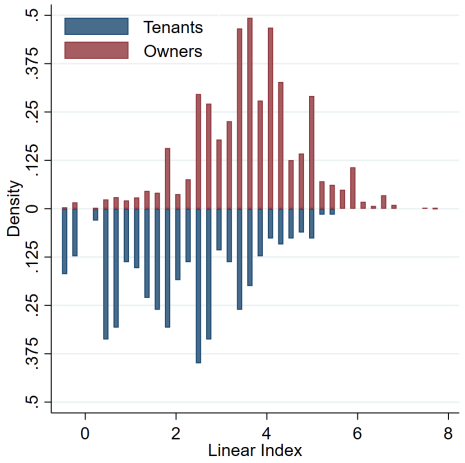
a. Metropolitan Lima, 2006



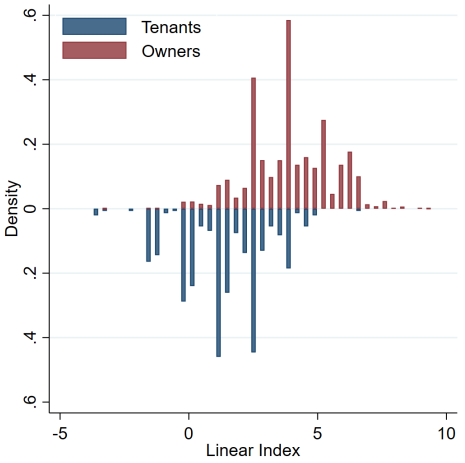
b. Metropolitan Lima, 2007



c. Rural Areas, 2006



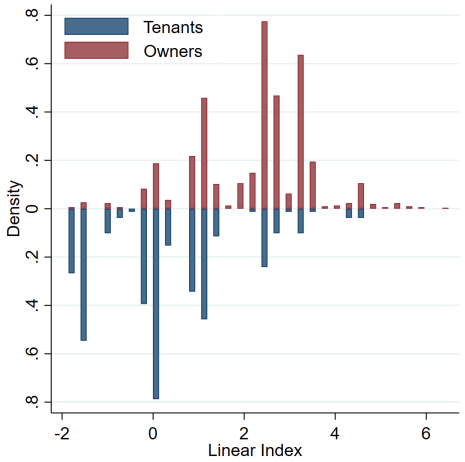
d. Rural Areas, 2007



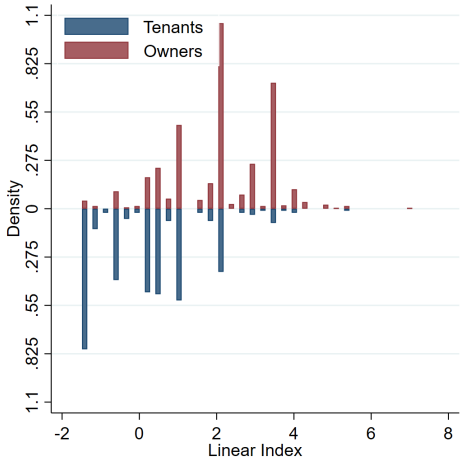
Source: Based on National Household Survey data.

Cont. Distribution of the linear index of the propensity score, by ownership status, Metropolitan Lima and rural areas, 2008 and 2009

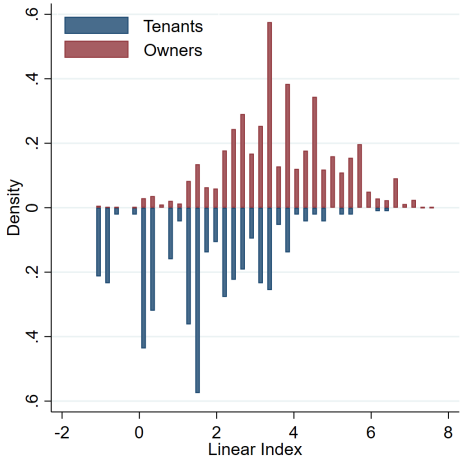
a. Metropolitan Lima, 2008



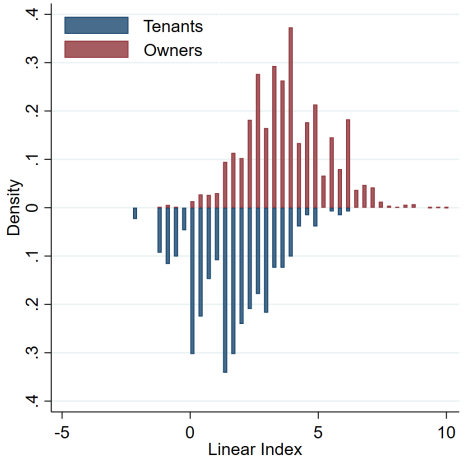
b. Metropolitan Lima, 2008



c. Rural Areas, 2008



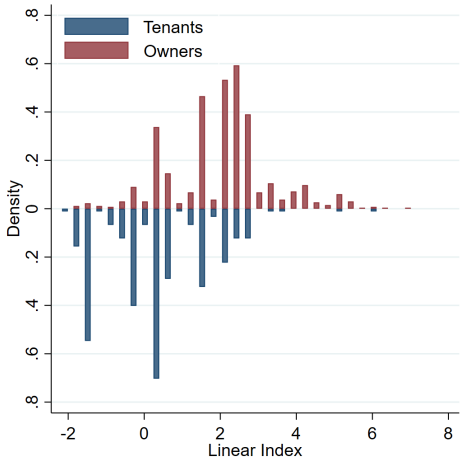
d. Rural Areas, 2009



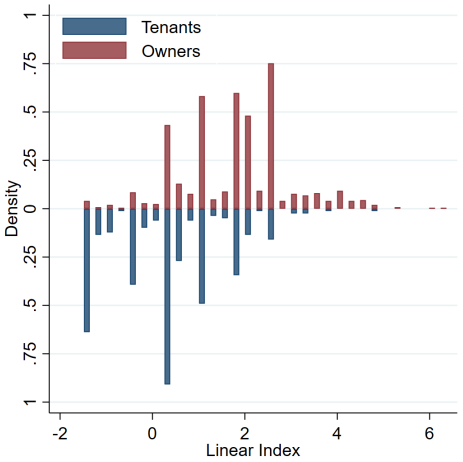
Source: Based on National Household Survey data.

Cont. Distribution of the linear index of the propensity score, by ownership status, Metropolitan Lima and rural areas, 2010 and 2011

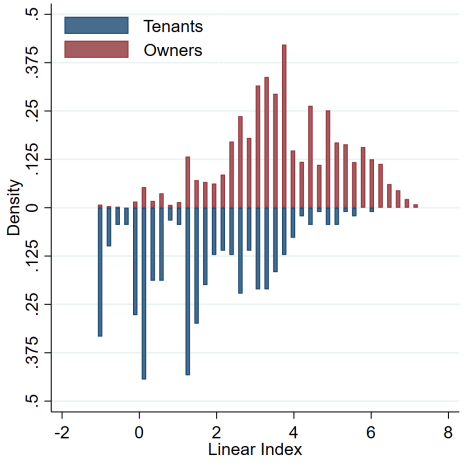
a. Metropolitan Lima, 2010



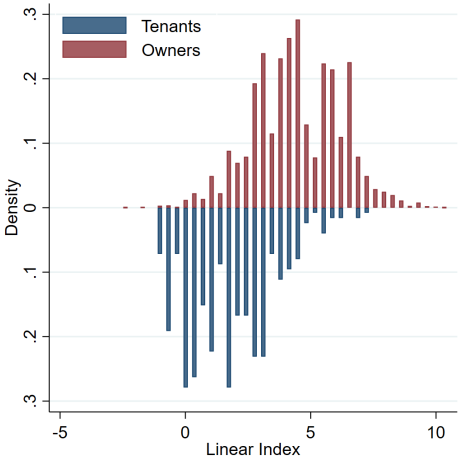
b. Metropolitan Lima, 2011



c. Rural Areas, 2010



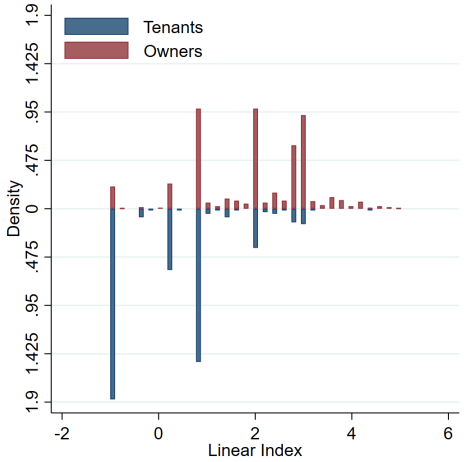
d. Rural Areas, 2011



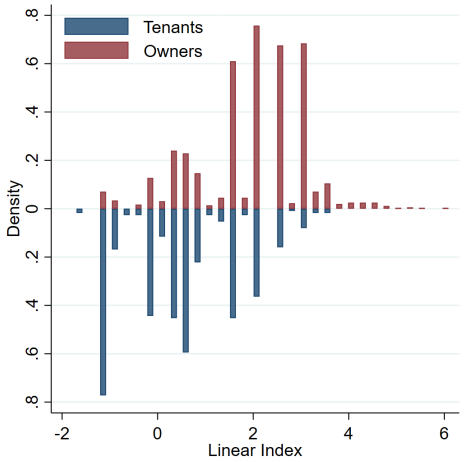
Source: Based on National Household Survey data.

Cont. Distribution of the linear index of the propensity score, by ownership status, Metropolitan Lima and rural areas, 2012 and 2013

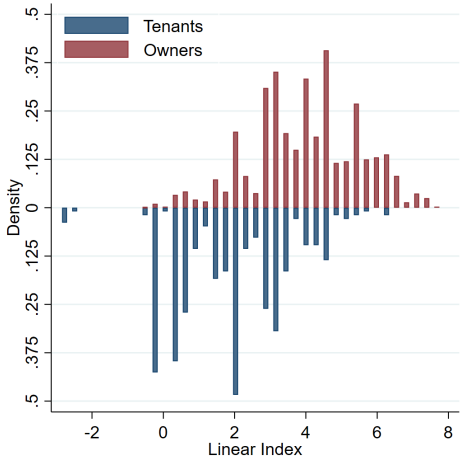
a. Metropolitan Lima, 2012



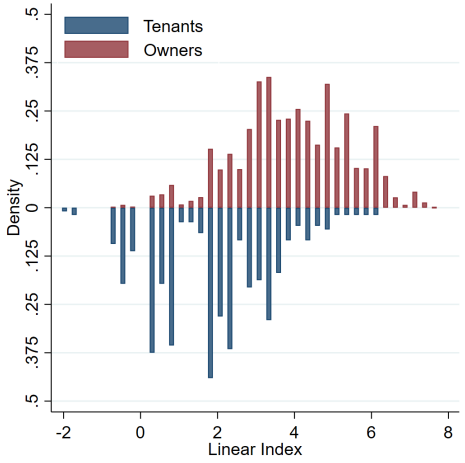
b. Metropolitan Lima, 2013



c. Rural Areas, 2012



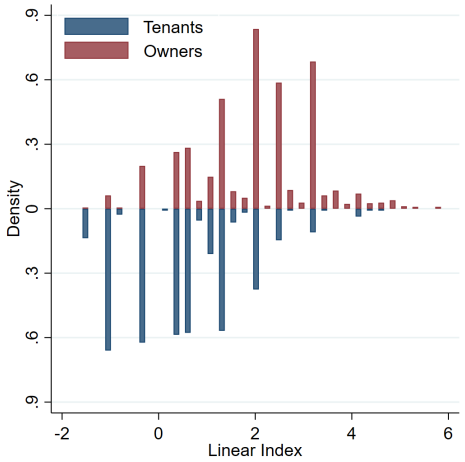
d. Rural Areas, 2013



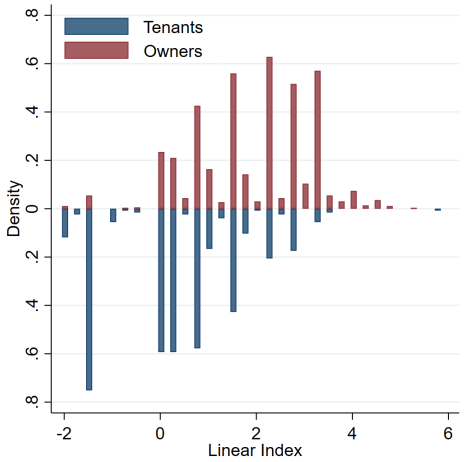
Source: Based on National Household Survey data.

Cont. Distribution of the linear index of the propensity score, by ownership status, Metropolitan Lima and rural areas, 2014 and 2015

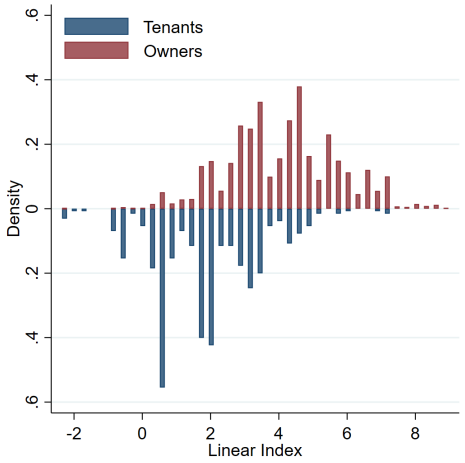
a. Metropolitan Lima, 2014



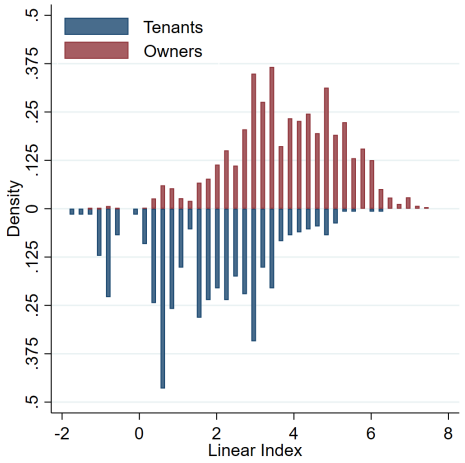
b. Metropolitan Lima, 2015



c. Rural Areas, 2014



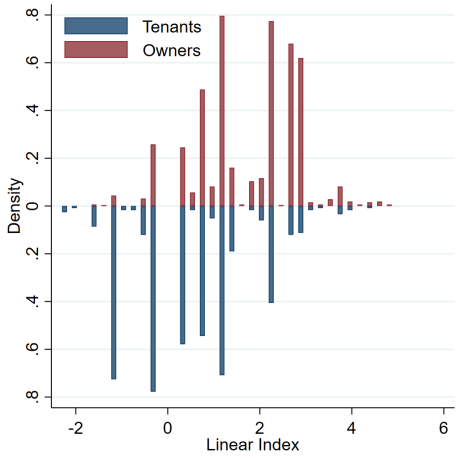
d. Rural Areas, 2015



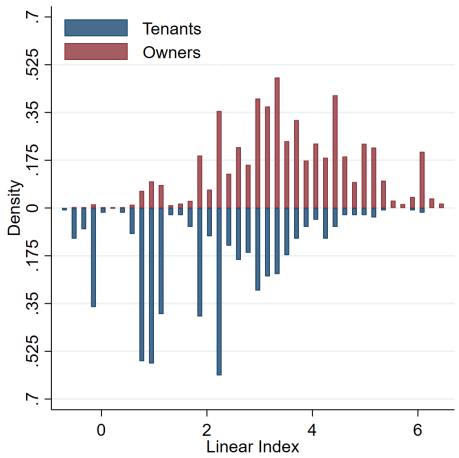
Source: Based on National Household Survey data.

Cont. Distribution of the linear index of the propensity score, by ownership status, Metropolitan Lima and rural areas, 2016

a. Metropolitan Lima, 2016



b. Rural Areas, 2016



Source: Based on National Household Survey data.

Table A 1. Estimates of the average accuracy of homeowner self-assessed rents based on OLS, Metropolitan Lima and rural areas, by year, 2003-17

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
							Metropolitan Lima								
owner	-1,078	3.678	1,704***	881.7	884.3**	345.3	695.7**	462.6	680.1*	1,078**	-281.8	140.9	969.5**	-95.01	-410.1
	(855.5)	(568.6)	(622.8)	(555.2)	(404.4)	(362.6)	(294.4)	(392.3)	(398.4)	(424.5)	(402.3)	(318.5)	(390.2)	(368.1)	(411.4)
Observations	870	1,349	1,302	1,359	1,424	1,450	1,288	1,188	1,322	1,329	1,894	1,975	1,971	2,032	1,967
Adjusted R-squared	0.383	0.357	0.271	0.322	0.185	0.257	0.263	0.257	0.239	0.265	0.261	0.274	0.324	0.279	0.329
							Rural Areas								
owner	-59.13	30.14	60.87	76.13	319.4***	249.9***	247.9***	209.3**	531.4***	286.2**	170.8	101.8	-81.81	-162.4	-384.9**
	(75.06)	(65.68)	(53.02)	(76.17)	(66.55)	(79.16)	(69.65)	(84.53)	(142.3)	(116.7)	(106.1)	(155.9)	(96.42)	(137.8)	(158.7)
Observations	4,035	6,206	6,392	6,511	6,635	6,582	6,444	6,569	7,645	7,596	9,173	9,395	10,516	10,964	10,728
Adjusted R-squared	0.363	0.357	0.417	0.379	0.347	0.202	0.303	0.304	0.236	0.294	0.287	0.234	0.278	0.249	0.267

Source: Based on National Household Survey data.
Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1