cemmap

centre for microdata methods and practice

The Housing Stock, Housing Prices, and User Costs: The Roles of Location, Structure and Unobserved Quality

Jonathan Halket
Lars Nesheim
Florian Oswald

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP5/20

The Housing Stock, Housing Prices, and User Costs: The Roles of Location, Structure and Unobserved Quality

Jonathan Halket*
Lars Nesheim†
Florian Oswald‡

28th January 2020

Abstract

Which housing characteristics are important for understanding homeownership rates? How are housing characteristics priced in the rental and owner-occupied markets? And what can the answers to the previous questions tell us about economic theories of homeownership? Using the English Housing Survey, we estimate a selection model of the allocation of properties to the owner-occupied and rental sectors. Structural characteristics and unobserved quality are important for selection. Location is not. Accounting for selection is important for estimates of rent-to-price ratios and can explain some puzzling correlations between rent-to-price ratios and homeownership rates. These patterns are consistent with, among others, hypotheses of contracting frictions in the rental market likely related to housing maintenance.[1]

Keywords: Homeownership, Rents, Housing Supply, Selection
JEL Classification: C31, R31.

*University of Essex, Texas A&M University and CeMMAP, Institute for Fiscal Studies
†University College London and CeMMAP, Institute for Fiscal Studies
‡Sciences Po Paris

[1]We thank the editor and four anonymous referees for helpful comments on an earlier draft. This research was supported by the UK Economic and Social Research Council through the Centre for Microdata Methods and Practice (CeMMAP grant number ES/I034021/1), through an ESRC transformative research grant, grant number ES/M000486/1, and ESRC Large Research Grant ES/P008909/1; The work

1 Introduction

In this paper, we ask a very simple set of questions: Which housing characteristics are important for the allocation of properties to the rental and owner-occupied markets? Conditional on that allocation, how are housing characteristics priced in the rental and owner-occupied markets? And what can the answers to the previous questions tell us about economic theories of homeownership?

The housing stock is one of the most important real assets in any economy and so the allocation and pricing of that stock is an important determinant of both economic performance and consumer welfare. Yet, surprisingly little is known empirically about the factors that determine whether a housing unit is owner-occupied or is owned by a landlord. Nor do we know which housing characteristics are valued more highly by owner-occupiers and which are valued more highly by landlords nor much about how houses' relative asset values and yields vary with their characteristics.

Our approach to answer these questions is straightforward. We model an economy as endowed with a set of properties. Landlords and owner-occupiers take prices and rents as given. A property ends up in either the rental or owner-occupied part of the housing market depending on which part of the market values the property more. We then use a rich micro data set to estimate both a simple selection model of *properties* (not households) into the rental and owner-occupied housing markets and the properties' rents and prices, respectively, exploiting the tremendous variation in dwelling characteristics, prices, rents and ownership rates within the greater London metropolitan area between 2008-2012.

We find that landlords and owner-occupiers value observable locational characteristics similarly but observable physical characteristics, like the size of the house, differently, even though rent-to-price ratios vary systematically along all dimensions. Also some unobservable characteristics are valued differently and others are valued similarly.

We then analyze the joint implications of these results in the context of the simplest possible dynamic setting.

More specifically we find that:

1. Observable physical characteristics of a property like dwelling type and size are by far the most important determinants of the probability of being in the owner-occupied sector. Housing units with high value physical attributes (large or more detached dwellings) are more likely to be owner-occupied. At the same time, these "large" properties have higher rent-to-price ratios than smaller properties. From a consumer's perspective, this is intuitive. Owner-occupancy is more likely to be "purchased" when its relative price is lower. However, from a supply-side perspective, these empirical facts have more interesting implications. Despite relatively high gross rental yields, landlords are less likely to buy and let out properties with high value physical characteristics. Revealed preference together with these facts, which are fairly stable over time despite large changes in property prices, imply that rental sector user costs of housing capital increase with property size faster than owner-occupied user costs.

2. Location is relatively unimportant for the likelihood of being owner-occupied after controlling for physical characteristics. However, rent-to-price ratios vary significantly with location. Some features of the rent-to-price patterns are unstable over time which may reflect time and location dependent expectations of capital gains.

3. Modeling and measuring differences in unobserved quality is essential for understanding which properties become rentals and at what price. Selection into sectors depends on both observed and unobserved characteristics. In particular, we find that the data reject uni-dimensional models of unobserved quality and that the different elements of unobserved quality are differentially correlated with selection. On average, rentals have lower unobserved "rental" quality and lower unobserved "owner" quality. These findings imply that imputed rents and rent-to-price measures that do not control for selection are biased.

These three facts quantify an intuitive relationship between a house's attributes, its prices in the rental and owner-occupied sectors and its propensity to be rented. For instance, gross yields that decline with the value of the underlying land are consistent with a user-cost decomposition where maintenance costs[2] increase primarily with the value of the structure. Meanwhile, selection based on the value of the structure even as gross yields increase with physical value is consistent with a theory where agency problems increase maintenance costs in the rental sector relative to the owner-occupied sector; particularly for larger properties and for properties that are more detached. Similarly, maintenance of some unobservable attributes, like a nice kitchen or a jacuzzi may be particularly hard to contract upon for landlords and thus increase the likelihood that properties with those characteristics appear in the owner-occupied sector, while other unobservable attributes, like features of the property's vicinity or qualities of its view may be value enhancing but don't pose contracting problems and therefore don't affect selection. As we discuss below, the enumerated facts above may also be consistent with other, non-maintenance cost explanations.

1.1 Relation to the literature

An extensive household tenure-choice literature studies demand-side selection into home-ownership. This literature studies who chooses to own houses and how household tenure choice interacts with decisions such as savings decisions. For example, see Rosen (1979), Goodman (1988), Kan (2000), or Campbell and Cocco (2007). This literature also studies how observable and unobservable household factors affect selection into and welfare from homeownership. For instance, Diaz and Luengo-Prado (2008) and Blow and Nesheim (2009) examine how the shadow price of housing for owner-occupiers can differ systematically from rental costs due to differences in the Lagrange multipliers from households' constrained maximization problems. In general, the multipliers are functions of households' current assets, income and other state variables. In an accessible overview of the user-cost literature Dıaz and Luengo-Prado (2011) report recent user-cost estimates and ownership implications for the US, while Diewert (2003) explores the implications of user-cost estimates for official price indices, and Gallin (2008) estimates how well rent-to-price ratios predict both rents and prices over a four-year horizon.

[2]Throughout we refer to the costs of upkeep and care of a structure costs as maintenance costs. They are the costs required to keep a property at constant quality.

In contrast, in this paper we study supply-side selection into the owner-occupied sector. That is, we study what aspects of housing units explain why some units are more likely to end up in the rental sector while others are more likely to end up in the owner-occupied sector. Glaeser and Shapiro (2003) and Amior and Halket (2014) observe that there is a striking difference in the homeownership rates of single versus multi-family housing units throughout the US. This leads the former to hypothesize that “homeownership is particularly correlated with housing structure.” Here we test this hypothesis on richer, more detailed data from England. Our data allow us to control for variation in location within a metropolitan area as well as many physical characteristics. Moreover, our data contain a rich enough sample of rental properties to gain good estimates of how prices and rents vary within a metropolitan area. We show that not only is ownership correlated with structure as found in Glaeser and Shapiro (2003) and Amior and Halket (2014) but also it is correlated with many physical attributes. Moreover, relative variation in prices and rents can point to an explanation as to why.

What explains these correlations? The technology that provides housing services is the same regardless of housing sector: in both sectors, a physical housing unit provides the services. However, there may be sectoral differences in the costs of capital or in operating costs. For instance, it has long been suggested that the contracts that govern the provision of housing services in the rental and owner-occupied markets may be constrained by tenure-dependent information frictions. A literature going back to Sweeney (1974) and Henderson and Ioannides (1983) has argued that contracting frictions in the rental sector result in higher maintenance costs and less upkeep and investment. Because of monitoring costs, tenants under-invest in maintenance resulting in a higher depreciation rate in the rental sector. In this literature, it is assumed but often unstated that these differential costs are likely to be correlated with the physical characteristics of the property. As a result, a housing unit with higher rental-specific maintenance costs, or for which rental contracting frictions are greater, is more likely to be found in the owner-occupied sector.

Another possible explanation is that landlords with lower costs of capital (perhaps due to tax wedges) sort into physically smaller houses but not into high value locations based on property size related cost advantages. We are unaware of any existing theories that explicitly provide this exact prediction. Coulson and Fisher (2014) argue that there may be some endogenous differences in ownership structure (and thus perhaps financing costs) based on variation in maintenance technologies across ownership structures and the physical size of the building.

Empirical support for these theories has been scant. Linneman (1985) notes that the "efficiency" of landlord provided housing services is an important factor determining ownership rates. Casas-Arce and Saiz (2010) examine how different jurisdictions' legal systems and propensities to enforce contracts affect ownership propensities. Hanson (2012) looks at how the mortgage interest tax deduction interacts with home sizes to affect ownership rates in the US. Hilber (2005) examines neighborhood externality risks in the AHS and finds that they are negatively correlated with homeownership. Coulson and Fisher (2014) study how property size affects a building's management structure. Harding et al. (2000) find that homeowners that are more likely to default on their mortgage may under-maintain their house.

In the corporate finance literature, studies of selection due to various contracting or tax frictions and the effects of selection on the distribution of observed returns are common. Among many others, see Prabhala (2008). In the housing literature, despite the long-standing theories discussed above and the cautionary warning in Glaeser and Gyourko (2007) about how the "unobserved costs of home owning such as maintenance" can bedevil estimates of rents and prices based on the indifferences of households to various houses, there are no studies that control for supply-side selection when estimating hedonic rent and price functions. The closest perhaps is Heston and Nakamura (2009) which uses a small sample of federal employee data in several small markets to show that owner-occupied housing units are 15 percent more valuable than observably equivalent rentals. The result is based on self-reported estimates of rental flows obtained from owner-occupiers. We are the first to (a) estimate a selection model of hedonic housing prices and rents for a large housing market, (b) to estimate the importance of unobserved quality in this market, and (c) to analyze the implications of these models for sector specific user costs. Furthermore, we use repeated cross-sections sampled from periods of both housing price decline and boom in England and find that the estimated relationship between physical attributes, selection and rent-to-price remains very stable.

In Section 6 we show how to use the results from the selection model, coupled with additional identifying assumptions, to obtain estimates of how much the user-costs of rent and owner-occupied housing vary with the characteristics of a property. These estimates are interesting in their own right. In addition, if we assume that a single riskadjusted discount rate prices all housing in the market as in Epple et al. (2013), then we can say more. In this case, if different housing units have different rent-to-price ratios, then these differences must be due either to differing expectations about the future or to differences in the costs of renting out the property. We argue that expectations about the future, given rents and prices, should not systematically affect selection into renting. This enables us to use our selection model results to estimate the potential size of the moral hazard problem described above. These additional findings rely on fairly strong assumptions. We cannot rule out other potential mechanisms such as variation in the effective discount rates that price various types of housing which in principle could generate these same findings. Our data do not include measures of maintenance or other variables that would allow us to test one mechanism against another.

Our findings have important implications for macro models of the housing market. In the UK and the US more than 60 percent of households tie up a large part of their financial portfolio in a single, risky, illiquid asset; housing. Why is this the case? Why don't financial and rental housing markets provide contracts that enable households to enjoy the consumption flow from a rented, three bedroom, detached, 120 square-meter house in the suburbs while enjoying the dividend flows and potential capital gains of the wider financial market?

Numerous theoretical explanations have been proposed to explain the high market share of homeownership, including insurance motives, tax considerations, a "warm glow" from housing and a variety of contracting problems between renters and landlords. For instance Chambers et al. (2009b,a); Chen (2010); Oswald (2019) incorporate higher maintenance costs for rentals and a limited supply of "high value" rentals into their models. Often these two elements are the primary factors that explain why so many households choose to own in these models. We show that selection due to unobservables can simultaneously explain both higher maintenance costs and the limited supply of *physically* valuable rentals. Our estimates imply that, potentially, differential maintenance costs are large and consistent with the less direct estimates/calibrations in the macro models.

Finally, accounting for unobserved quality can potentially explain two curious features in the raw data. Firstly, several studies have attempted to measure maintenance costs to determine whether rental properties do indeed have higher observed maintenance costs. These studies have been inconclusive. Differences between rental and owner-occupied maintenance expenditures in national accounts data are small. There are several studies that attempt to measure at the disaggregate level whether rentals have higher maintenance costs. Their findings are mixed. Galster (1983) estimates that owner-occupiers occupy better properties and better maintain them. Shilling et al. (1991) estimates a hedonic model of sales prices for rental and owner-occupied single-family property in a single parish in Louisiana and finds that rentals depreciate faster. Malpezzi et al. (1987) estimates hedonic models of rents and prices from the AHS and finds that rents decline evenly with age whereas prices decline at a declining rate. However, Gatzlaff et al. (1998) finds limited evidence of differential maintenance by comparing appreciation rates of rentals and owner-occupied housing units.

Using the assumptions in Section 6.3.3, our findings on selection can rationalize these mixed results. Because of selection on unobservables, cross-tenure comparisons of maintenance costs for observationally similar housing units will not necessarily reveal evidence of differential maintenance costs. Properties with higher maintenance costs (more contracting issues) select into owner-occupancy. We show that not properly accounting for selection on unobservables leads to biased measures of the difference in maintenance costs. Indeed, under some assumptions about interest rates and capital gains, our estimates suggest (as an upper bound) that rental property maintenance costs for large rental properties can be as much as 40 percent higher. This evidence of large differential maintenance costs is consistent with the theory of contracting frictions outlined above.

Secondly, several studies (see Verbrugge, 2008; Landvoigt et al., 2015; Epple et al., 2013; Verbrugge and Poole, 2010; Heston and Nakamura, 2009; Garner and Verbrugge, 2009) using US data have found that rent-to-price ratios decline with property prices. If we estimate rent and price functions using UK data without accounting for selection, we also find that rent-to-price ratios decline with property prices. The more expensive a property is, the lower is its predicted rent-to-price ratio. As ownership rates are (unconditionally) increasing in dwelling prices, this could lead one to the curious conclusion that households tend to own[3] homes that have low rent-to-price ratios. This is a bit puzzling from the household's perspective. We show in Section 5.3.1 that this uncon-

[3]Throughout we use "households owning a home" (or equivalent) interchangeably with owner-occupancy. We never mean households' propensity to own a home it does not occupy as its primary residence (e.g. investment or vacation homes).

ditional correlation between ownership and rent-to-price ratios breaks down once one controls for selection. Households actually tend to own housing units whose physical characteristics imply *high* rent-to-price ratios.

Sections 2 and 3 introduce the model and data, respectively. Section 4 explains our estimation procedure. Section 5 discusses the results from our model, including measures of bias. Section 6 discusses the implications of our model results for user-costs and maintenance.

2 Model

A property has observed characteristics $z \in \mathbf{R}^n$ and unobserved characteristics $\varepsilon \in \mathbf{R}^3$. Observed characteristics include location, size, dwelling type, and a number of additional property characteristics (number of rooms, quality, etc.) detailed in Section 4.1. We assume that the value of unobserved characteristics is completely captured by a vector of three factors that affect rental-sector rents, owner-sector prices, and selection into sector. The model allows for differential valuation of characteristics across sectors, imperfect correlation between sector allocation and prices and rents, and unobserved factors that affect sector allocation but not prices or rents.

2.1 Rent and price equations

If a dwelling unit is in the rental sector, its rent is observed. Let log annual rent be given by

$$\ln R(z, \varepsilon_r) = \alpha z + \varepsilon_r. \tag{1}$$

This is a log linear approximation to the true hedonic rent function. The parameter vector α measures the percentage impact of observed characteristics on rents and ε_r captures the impact of unobserved characteristics on rents.

Similarly, if a dwelling unit is in the owner-occupied sector, its value is observed. Let the log value in the owner-occupied sector be

$$\ln \pi_o(z, \varepsilon_o) \quad = \quad \beta z + \varepsilon_o. \tag{2}$$

This is a log linear approximation to the hedonic price function. The parameter vector

β captures the percentage impact of observed characteristics on prices in the owner-occupied sector. The variable ε_o captures the impact of unobservable characteristics on owner-sector prices. We allow ε_o and ε_r to be correlated.

If a property is in the rental sector, its value is not observed. Nevertheless, let the log value of a dwelling in the rental sector be

$$\ln \pi_r(z, \varepsilon_o - \varepsilon_s) \quad = \quad (\beta - \gamma)z + \varepsilon_o - \varepsilon_s. \tag{3}$$

The parameter vector γ captures the net loss or gain in willingness-to-pay by a landlord relative to an owner-occupier. Positive elements of γ correspond to characteristics that generate a net loss in the rental sector and negative elements correspond to characteristics that reflect a net gain. Similarly, $\varepsilon_o - \varepsilon_s$ reflects how unobservable factors affect the net gain or loss in relative values. We discuss the relationship between rental sector prices and rents in Section 6.1.

2.2 Allocation of properties to sectors

In equilibrium, a property is owned by the agent (landlord or owner-occupier) who has the highest value. Thus, for each property, choices of landlords and owner-occupiers determine which sector the property is in. If a property is in the rental sector, we observe rent R but not value π_r. If it is in the owner-sector, we observe value π_o but not the implicit utility of occupancy. This is the censoring problem in our application.

Mathematically, a housing unit is observed in the owner-occupied sector if

$$\ln \pi_o(z, \varepsilon_o) \geq \ln \pi_r(z, \varepsilon_o - \varepsilon_s). \tag{4}$$

That is, if

$$\gamma z + \varepsilon_s \geq 0 \tag{5}$$

The values π_o and π_r are conditioned on sector. They measure willingness to pay of buyers in each sector. The unconditional price of a property is the price of the property in the market where it is most valuable. That is,

$$P(z, \varepsilon) = \max_{\{own,rent\}} \{\pi_o(z, \varepsilon_o), \pi_r(z, \varepsilon_o - \varepsilon_s)\} \tag{6}$$

The parameters of the value and rent functions may vary over time. We leave their dependence on t implicit.

Let $\varepsilon = (\varepsilon_r, \varepsilon_o, \varepsilon_s)$. Under the assumption that $\varepsilon \sim N(0, \Sigma)$ independent of z, this is a standard Type 5 Tobit model (Amemiya, 1985). In the Type 5 Tobit model, as long as the variables in z are linearly independent, the normalized parameters $(\alpha, \beta, \widetilde{\gamma})$ and $(\sigma_{rr}, \sigma_{oo}, \widetilde{\sigma}_{rs}, \widetilde{\sigma}_{os})$ are identified where $\widetilde{\gamma} = \frac{\gamma}{\sqrt{\sigma_{ss}}}$, $\widetilde{\sigma}_{rs} = \frac{\sigma_{rs}}{\sqrt{\sigma_{ss}}}$ and $\widetilde{\sigma}_{os} = \frac{\sigma_{os}}{\sqrt{\sigma_{ss}}}$. We use the notation σ_{ij} to denote the (i, j) element of the covariance matrix Σ. As in all discrete choice models, the parameter σ_{ss} is not identified because sector choice is a binary choice. Also, σ_{ro} is not identified because π_o and R are never simultaneously observed.

For most of our analysis, these normalizations play no role. Our results on selection and relative rents and prices are fully identified regardless of the normalization and the quantitative results are invariant to the normalization. In Section 6, these normalizations do play a role. Values of the parameters $(\sigma_{ro}, \sigma_{ss})$ affect our estimates of user costs and contracting costs. We discuss identification of these additional parameters in Section 6.2.

The model has several important features. First, the values of unobserved characteristics in the owner-occupied sector is not restricted to be perfectly correlated with their values in the rental sector. Second, the impact of unobserved characteristics on selection is not restricted to be perfectly correlated with owner-occupied prices nor with rents. Third, the correlation of ε_r and ε_s may differ from the correlation between ε_o and ε_s. Economically, selection into sectors may be differentially related to unobservables that have high value in the rental sector, ε_r, versus unobservables that have high value in the owner-sector, ε_o.

In the next two sections we discuss the data and estimation results. Then in Section 6, after discussing estimates of (π_o, R, π_r) and the selection of properties into the two sectors, we discuss the extent to which the estimates are consistent with how various components of user costs may vary across properties and across the two sectors.

3 Data

We use data from the "secure access" version of waves 2011-2014 of the English Housing Survey (EHS). The secure access version of the EHS contains detailed information on a large sample of properties including information on postcode, value, rent, and a large set of property characteristics. The sample is not a simple random sample but is constructed as follows.

The EHS uses a complex multistage methodology. Each wave comprises two surveys which are then combined to produce two samples. Each sample is constructed using data from surveys from multiple waves. In each wave, the EHS team conducted a "household survey" and a "physical survey". For example, to construct the 2011 wave, the EHS team sampled approximately 17,500 households in the financial year 2008/2009 (April 2008 - March 2009). These households were drawn from the list of addresses held by Royal Mail. At each sampled address, one dwelling was sampled. At each dwelling, one household was sampled.

Respondents from this selection (approximately 17,000) comprised the household interview sample. The EHS team then chose a subsample of these dwellings (approximately 8,000 in 2008/2009), including vacant ones, and performed a physical inspection. This is called the "physical survey." The subsample was constructed from the 17,500 by including all social housing [4] addresses and taking a subsample of private addresses. Private rental properties were over-sampled. Finally, to construct the final "housing stock" sample, the EHS team combined data from two physical surveys. For instance, the housing stock sample in the 2011 wave is comprised of the physical surveys from 2008/2009 and 2009/2010. Weighting for the final sample is based on this two year sampling window.

We focus discussion on the 2011 wave of the EHS. While we also analyze the 2012, 2013, and 2014 waves, these later waves have some limitations. In the later waves, property prices were top-coded at £1,000,000. Also, due to budget cuts, the later waves used smaller samples and collected information on a smaller range of topics. Despite these limitations, our results are robust across waves. Because the samples in each wave use data from a two year span, the samples overlap. For instance, the samples in the 2011 and 2012 waves both contain the same data collected in 2009/2010 as part of their samples.

Owner-occupied property values recorded in the survey were obtained in one of two ways. For 82.7% of properties, owners self-reported an estimate of the current market

[4] Social housing units are public housing units that are owned or regulated by the government. They consist of Local Authority (LA) provided housing and housing provided by registered social landlords (RSL). RSL's are non-profit organisations that provide low-cost housing. Rents for all social housing units are regulated by the government and highly subsidised.

value. For the remainder of owner-occupied properties, a professional surveyor valued the property on-site. Rental sector rents were self-reported by tenants.

For each dwelling, the secure access version of the EHS reports the full postcode[5] We match each postcode with its geographic coordinates using the Office for National Statistics's Postcode Directory for 2013. Because postcodes can change over time, there are a few unmatched postcodes. In 2014, there is 1 unmatched owner-occupied property out of 5,184 and 2 unmatched private rentals out of 2,683 for all of England. In 2011, there are none. The numbers of unmatched properties for other waves are similar. For these unmatched properties, we impute coordinates using the mean geographic coordinates of all postcodes sharing the same postcode district (postcodes are grouped geographically and the first three to four characters of a 7-8 character postcode are its postcode district). For each property, we then convert its geographic coordinates to polar coordinates (d, θ) centered around Trafalgar square. That is, for each property we compute d, the Euclidean distance from Trafalgar Square and θ, the angular distance from due east measured in radians. That is, $\theta = 0$, is east, $\theta = 0.5\pi$ is south, etc.

Our analysis focuses on the subsample of dwellings within 140 km of Trafalgar Square in London. We label this region "Greater London." It is worth noting that this 140 km circle extends beyond the London "travel to work area" (TTWA) defined by UK statistical authorities. The TTWA, by definition, includes only 75% of the economically active population. We extend it to ensure that our sample of properties in less dense areas is large enough while still retaining a reasonable commute to London and thus arguably a single economic market. For example, a train ride from Ipswich, at the outer edge of our Greater London region, to London Liverpool St Station takes a little over an hour.[6] Our results are robust to alternative definitions of Greater London and are robust to including the full sample for all of England. We use the Greater London subsample because, in our view, it constitutes a single economic market, and because distance from London has such a large effect, that focusing on London facilitates visual interpretation of our results using graphs.

We present summary statistics for the owner-occupied, private rental and social housing sectors. However, when we estimate the model, we restrict the analysis to

[5]The average size of a postcode in the greater London area (our focus, as explained below) is 0.09 square km.

[6]http://ojp.nationalrail.co.uk/service/timesandfares/IPS/LST/tomorrow/0715/dep

private sector housing. In the private sector, investors may buy and sell properties freely and prices are determined by the market. In the social sector, supply is largely determined by political forces, not by the choices of investors. In addition, prices in this sector are subsidised and highly regulated.

Table 1 displays the overall market shares of owner-occupied housing, private rentals, and publicly assisted housing in England and Greater London. In England in Wave 2011 (2008-2010), 67.9 percent of housing units were owner-occupied units while 14.3 percent were private rentals and 17.8 percent were social housing. The market shares in Greater London are similar, 66.6 , 15.7 and 17.7 percent respectively. Greater London has slightly more private rentals and fewer owner-occupiers. By Wave 2014 (2011-2013), the share of private rentals has increased by three percentage points at the expense of owner-occupancy.

Table 2 shows how market shares vary with distance from the center of London. Within 10 km of the center, the owner-occupied share is 37.9 percent while the private rented and social housing shares are 23.7 and 38.4 percent respectively. The owner-occupied share increases with distance. More than 50 km from the center, the owner-occupied share is 72.9 percent while the rental and social housing shares decline to 13.4 and 13.7 percent respectively. We will see that these patterns with respect to distance do not persist after controlling for structural characteristics.

Tables 3 and 4 show how market shares vary with size and dwelling type. Large properties are much more likely to be owner-occupied; 90.1 percent of properties larger than 100 square meters are owner-occupied versus only 33.1 percent for properties less than 50 square meters. Similarly, semi-detached and detached houses and bungalows are much more likely to be in the owner-occupied sector while converted flats and dwellings in multi-unit structures are more likely to be in the rental sector. 73.9, 94.4 and 76.8 percent of the former dwelling types are owner-occupied while only 39.3, 32.3 and 20.7 percent of the latter types are owner-occupied. As with the previous literature, we find that homeownership is particularly correlated with structure. We will see that these patterns hold up even after controlling for location and other property characteristics.

Table 5 displays summary statistics by sector for the other main variables that are used in our empirical analysis below. One can see that there are substantial differences across sectors. Owner-occupied properties are bigger and have more rooms. They have rear plots that are nearly twice as large as rental properties (8.2 vs 15.4 sq. m.). They are more likely to have 3 or more bedrooms (74 vs 39 percent), 2 or more bathrooms (30 vs 14 percent), and 2 or more living rooms (47 vs 18 percent). Owner-occupied properties also have slightly higher investments in energy efficiency. 31 percent of owner-occupied properties have cavity loft insulation versus 15 percent of rental properties. 80 percent have high degrees of double glazing (more than 80 percent of windows double glazed) versus 67 percent of rentals. Rentals do have slightly higher energy efficiency scores but that could be because they are smaller; it is cheaper to heat and light a small property. Finally, owner-occupied properties are more likely to have access to off-street parking (77 vs 47 percent) and more likely to have no litter in the neighborhood (82 vs 69 percent).

4 Estimation

As noted previously, the model from Section 2 is a Type 5 Tobit model. The estimating equations are equations (1), (2) and

$$\ln \pi_o - \ln \pi_r = \gamma \mathbf{z} + \varepsilon_{\mathbf{s}}. \tag{7}$$

As discussed in (Amemiya, 1985; Lee and Trost, 1978; Willis and Rosen, 1979) , the two pairs of equations, (1) and (7) and (2) and (7), can be estimated separately at the cost of efficiency. Each pair of equations is a Type 2 Tobit model, and so, the parameters in each pair can be estimated either using maximum likelihood estimation or Heckman's two-step procedure. Due to computational constraints in the secure data laboratory, we chose to estimate the two pairs of equations separately using maximum likelihood estimation. See Amemiya (1985) or Wooldridge (2010) for details of the Type 2 Tobit likelihood function. In our case, separate estimation of the two pairs of equations produces statistically indistinguishable estimates of the common parameters γ.

Using our data, the two-step Heckman procedure fails to produce reliable estimates because the inverse Mills ratio is nearly collinear with the other regressors. In contrast, maximum likelihood estimates remain robust because the score of the log likelihood function exploits cross-equation restrictions.

In Section (4.1) we discuss our baseline specification. This is our preferred specification based on model fit and parsimony. To evaluate the robustness of results based on our baseline specification, we also analyse several alternative specifications. These are discussed in Section (5.4).

4.1 Baseline specification: Model 1

We label our baseline specification Model 1. In the baseline specification, z includes nonparametric functions of location and size and indicator variables for dwelling type. We describe the nonparametric functions in more detail below. In addition, the specification includes dwelling age, parking availability, litter in the local neighbourhood, rear plot depth, and several measures of energy efficiency.[7] For our baseline specification, we do not include the number of bedrooms separately from dwelling size because number of bedrooms is highly collinear with property size. We discuss the impact of bedrooms further in Section (5.2.2). For owner-occupied properties, we also include an indicator for whether the property price is self-reported or not. For rentals, we include measures of whether the rental was furnished by the landlord and an indicator for whether the rental was self-reported to be at market rent.[8]

Dwelling size is measured in square meters. As discussed in Section 3, location coordinates are measured in polar coordinates, distance d from Trafalgar square and angular distance θ from due East. The baseline model includes 3rd order Chebyshev polynomials in size, 7th order Chebyshev polynomials in distance, and 3rd order Fourier series in the angular distance. The model also includes interactions between the distance terms and the angular distance terms. We considered models with higher order polynomials in size and distance, higher order Fourier terms for angular direction, and additional interactions between distance and direction. Our baseline specification was chosen as the specification that minimises the Bayes Information Criterion (BIC). Adding additional polynomials terms added substantial variability to estimates of the effects of size, distance, and direction without substantially changing other results.

[7] Each property has an energy efficiency rating calculated by the surveyor using the Standard Assessment Procedure (SAP05 or SAP09). This rating is based on an estimate of each dwelling's energy cost per square meter. It takes account of the cost of space and water heating, ventilation, and lighting. We also include indicators for the age of the heating system and extent of insulation and double glazing.

[8] In the sample, 55 percent of rentals are unfurnished and 7 percent are reported to be not let by private, arms-length landlords. For the former properties, rents are slightly lower because furniture is not provided. For the latter, rents are lower because the landlord is a friend, relative, or employer. We capture both these additional observable elements of rental contracts through indicators that affect rent but not selection.

The nonparametric functions of location are important to capture complex spatial patterns in prices, rents and selection. While they don't capture all such patterns, because the sample is finite, they do allow the model to be as flexible as possible given the sample size. The distance variables capture the impact of distance from London on property prices and rents and on selection into the owner-occupied sector. The angular distance variable θ captures variation in outcomes that depends on direction of travel when moving away from the center. For example, as seen in Figure 1, the rate of decline of prices with distance is higher heading north or west than east or south. Interactions between distance and direction capture spatial patterns that might be caused by the location of motorways or rail lines, employment or shopping subcenters, or more generally local amenities or disamenities. For example, prices are higher near motorways and rail lines and near employment subcenters.

Results for Model 1 are reported in Table 6 and displayed in Figures 1-7.

4.2 Alternate specifications: Models 2-5

We considered several alternate specifications to gauge the robustness of our results to model specification. We label these alternate specifications Model 2-5. Each of these specifications adds a set of additional property and locational characteristics to our baseline model to see whether additional variables change our results. We report a subset of results from these in Table 7. Further results for all models are reported in the Supplementary Appendix.

Model 2 adds bedrooms, bathrooms, living rooms, number of big kitchens and fireplaces and indicators for attics, balconies and basements to Model 1. Model 3 adds to Model 2 measures of housing quality defined as the first 5 principal components of a set of more than 60 measures of property quality. Model 4 adds to Model 3 the first 3 principal components from a set of more than 40 measures of neighborhood quality. Finally, Model 5 adds the first four principal components from a set of six additional neighbourhood variables.

The additional variables included in Models 3-5 are defined as follows. The EHS contains more than 60 measures of property quality including: measures of whether the property meets “decent homes criteria” (these are criteria set by the government to measure thermal comfort, "reasonable" state of repair, etc), the estimated cost to upgrade to meet those criteria, several measures of accessibility (for wheelchair users or disabled), indicators of various problems (e.g. rising damp, inadequate lighting or ventilation, etc), several measures of interior stair features, several measures of dwelling security, and several measures of health and safety problems. All variables are assessed by a professional surveyor. We conducted a principal component analysis of these quality measures and based on a scree test selected the first 5 principal components to include in the model. These 5 principal components account for 49.8 percent of the variation in these quality variables. These variables are included in Models 3-5.

Similarly, the EHS contains more than 40 measures of neighbourhood quality, including measures of neighbourhood noise (e.g. neighbours, traffic noise, trains, planes, etc....), neighbourhood problems (e.g. litter, graffiti, vandalism, air quality, etc.) and neighbourhood quality (e.g. visual quality as rated by surveyor, whether resident feels safe, etc....). As above, all variables are assessed by a professional surveyor. We conducted a principal components analysis of these neighbourhood variables and based on a scree test selected the first 3 principal components to include in the model. The first 3 principal components accounted for 61.5 percent of the variance of the neighbourhood variables. These variables are included in Models 4-5.

Finally, Model 5 included an additional set of neighbourhood variables derived from six additional measures of the local neighbourhood including urban nature, and the density, type and age of neighboring structures. We conducted a principal components analysis of these variables. Based on a scree test, we selected the first 4 principal components that account for 57 percent of the variation of these variables.

Most of the estimates of common parameters are statistically indistinguishable across the specifications in Models 2-5. In particular, as seen in Table 7, our estimates of ρ_{rs} and ρ_{os} are statistically indistinguishable across specifications. For that reason, we don't discuss Models 3-5 further here.

However, because the numbers of various rooms are highly correlated with size and dwelling type, the Model 2 results do have some notable differences from Model 1 in these dimensions. For this reason, we discuss these results more in Section 5.2.2 and show that the important qualitative results are similar across the two models. Despite the differences, we prefer the more parsimonious Model 1 because the Model 2 results are less precise due to the high correlation between numbers of rooms and size.

4.3 Exclusion restrictions

Models 1-5 do not employ any exclusion restrictions. The EHS contains information on a large range of property characteristics and neighbourhood characteristics. After searching through the variables in the survey, we concluded that all of the variables detailed above provide measures of housing characteristics that potentially affect prices, rents and allocation to sectors. Therefore, in our view, none could be excluded from the pricing equations *a priori*. So, we sought for additional variables that could be plausibly excluded from either the rent or the value equation. However, any variable that affects the present value of a property will likely also affect both the owner-occupied value π_o and the rent R. If one observed some measure of switching costs that is uncorrelated with ε_r and ε_o, one could use such a measure as an instrument for selection. Lagged sector choices do not qualify because they are correlated with ε_r and ε_o.

One possibility would be historical events or political or legal restrictions that affected a property's sectoral allocation but not its value. Variables proxying for such events or restrictions could be included in the selection equation and legitimately could be excluded from the pricing equations. For example, the UK government introduced a "Build to Rent" scheme in September 2012 that subsidized construction of affordable rental properties under some restrictions. We investigated using local measures of the effects of this scheme as variables to shift sectoral allocations. However, there were two problems. First, none of the subsidised properties were completed until 2015, outside the period covered by our sample. Second, while the details of the scheme varied by locality, many, if not all, schemes included restrictions that would likely affect both rents and values hence violating the exclusion conditions.

Despite these considerations, geographic indicators might serve as useful proxies for local government policies that affect sector allocations. To test this, we considered four additional specifications that use geographic indicators as proxies for unobserved local policies. We label these models Model 6 - Model 9. Model 6 includes county fixed effects in the selection equation. It would have been better to use local authority or postcode district fixed effects, but this was not possible given our sample size. If local policies on sector choice vary across counties but market conditions do not (conditional on distance, direction and location quality), then it would be valid to exclude county fixed effects from the pricing equations. Model 7 includes a variable that measures the fraction of dwellings in the local authority area that are "right-to-buy" dwellings. Right-to-buy dwellings are dwellings that were formerly social housing but that have been sold. The fraction that are right-to-buy may reflect historical local political decisions that are independent from a dwelling's characteristics. Model 8 includes the predicted share of social housing in the local authority. This may be a valid instrument for reasons similar to those above. Finally Model 9 includes a measure of the predominant "tenure" of properties in the local authority area (i.e. privately built, local authority built, etc.). As above, this variable may be correlated with historical local policies that are independent of dwelling characteristics.

Results from Models 6-9 are detailed in the Supplementary Appendix. For Model 6, the county fixed effects are jointly significant but do not change any of our results. For Models 7-9, the additional variables have no first stage predictive power and so change nothing. We also tried combinations of these variables. Results were unchanged in all specifications.

5 Results

Parameter estimates for location, size and dwelling type are not reported in Table 6 because the parameter values are difficult to interpret. To discuss results in these dimensions, we plot predicted values, rents and ownership rates in Figures 1-6. We discuss these results in the next subsections.

The remaining parameter estimates are detailed in Table 6. For the most part, the parameter estimates are economically plausible and consistent with our main results. Rents increase faster than prices with respect to size of the property, with similar rates of increase with respect to other structural characteristics. At the same time, ownership rates increase despite increasing rent-to-price ratios. Better insulation or a newer heating system raises rents more than values and increases the probability of being in the owner-occupied sector. Off-street parking raises rents and values more or less equally and also increases the probability of being in the owner-occupied sector. Minor litter lowers rents more than values and lowers the probability of being in the owner-occupied sector. Rents increase faster than values with rear plot depth while ownership rates increase.

A few parameter estimates are either insignificant or counterintuitive. More energy efficient properties have lower prices, lower rents and are more likely to be rental properties. The size of these effects are small. Properties with more double glazing are more likely to be owner-occupied but have 3.9% lower values.

5.1 Location

Figure 1 shows how prices and rents decline with distance conditional on direction. Pointwise confidence bands are illustrated with shaded areas. There is one panel for each angle of the compass. In all cases, property values and rents decline dramatically as one moves away from the center and values decline faster than rents. Moving an otherwise identical owner-occupied property North from the center for a distance of 10 km implies that its value will fall by about 63%. For the same change in northerly distance, rents decline by 45% relative to the central rental property.[9] In both sectors, the hedonic functions flatten out appreciably at distances greater than 15 km. Regardless of direction, the qualitative pattern is the same, however, values in owner-occupied sector seem to decline fastest when going North.

These results are obtained without controlling for overall lot size. Our data do not include information on overall lot size, though we have included a measure of the rear plot depth. If anything, this likely biases the estimated slopes toward zero. Lot sizes are probably larger further away from the city center where land is cheaper.

Figure 2 shows the estimated relationship between distance and the owner-occupancy rate when moving away from central London. Again, there is one panel for each direction of the compass. The "unconditional" lines plot the estimated relationship between distance and ownership when no other correlates are included. The "conditional" lines plot the relationship with distance holding other characteristics fixed.[10] As in Table 2, the unconditional line shows that owner-occupancy is far more prevalent 20 km outside of London than inside the city center. However, the conditional line shows that, once other housing unit characteristics are controlled for, distance essentially plays no role in selection into owner-occupancy. Owner-occupancy as a function of distance is essentially flat at around 80-81 percent.

[9] This is calculated as percentage change between central price and price at 10km, where log price is equal to 0 in the first panel of figure 1, i.e. $\%\Delta = 100\frac{\exp(0)-\exp(1)}{\exp(1)} = -63\%$.

[10] The figure shows deviation of the predicted ownership rate from the sample average as distance varies. To achieve this, other characteristics z_2 are held fixed so that $\widehat{\gamma}_2 z_2 = \Phi^{-1}(0.809) - \widehat{\gamma}_1 \bar{z}_1$ where $\bar{z}_1$ is the vector of mean values of the distance polynomials and 0.809 is the sample average ownership rate.

In summary, both rents and prices fall with distance. Rents relative to prices rise with distance but distant housing units are not more likely to be found in the rental sector. Why then don't investors in properties far from the center convert more properties into rental units? It must be that it is not profitable to do so. Why might this be the case? Here we propose some theoretical explanations based on our discussion in Section 6.1.

One possible explanation is that even though rents rise relative to values as distance grows, maintenance costs relative to rents also rise as the value of the location falls. This point is best illustrated by thinking of the value of a property as being composed of the value of land and the value of built structure. Structure requires more maintenance than land. So, maintenance costs are increasing in the proportion of the value that is structural. Given that the value of land decreases quickly as distance increases and the other characteristics of the house are being held constant in these plots, the proportion of value that is structural increases with distance. A second possible explanation is that vacancy costs in the rental sector are higher further out because the rental markets there are thinner, resulting in longer expected vacancy durations and higher equilibrium rents. A third possible explanation is that properties close to the city center had higher expected capital gains during the period of our study. In any case, no matter the cause, our results showing no selection with respect to distance indicate that the *relative* value of a house on the two sectors (π_r versus π_o) is not affected by distance; the higher costs or lower expected capital gains of suburban properties is capitalized into rents.

Some limited evidence on these points can be obtained by studying changes in the hedonic functions over time. Figure 3 shows estimated hedonic prices and rents versus distance for all four waves of the EHS, 2011-2014. The functions become slightly steeper over time and the rent-to-price ratio is unstable. In the most recent wave, 2014, the rental function is in some areas slightly steeper with respect to distance than the price function, the opposite of the earlier waves. These facts suggest that perhaps (a) expected capital gains are location-dependent and (b) the relative difference in expected capital gains across locations may change at a fairly high frequency. Though we do not have good data on household expectations by location, it is at least true that ex-post capital gains have varied widely over this same time period. For example, our calculations using UK Land Registry data show that, from 2009 to 2014, property prices within 10km of London increased 42 percent while prices further out increased only by 22 percent.[11]

[11] Further details on these calculations are available from the authors upon request.

5.2 Structure

5.2.1 Size

Figure 4 (left panel) shows how rents and prices change with respect to the total floor space of the property. Compared to a baseline size of 40 square meters, an otherwise identical 100 square meter owner-occupied property will have a 73 percent higher value, whereas the same increase in size implies a 101 percent increase in rent in the rental sector. As a result, rent-to-price ratios increase with size. The right panel of Figure 4 shows how size affects the probability of being owner-occupied. Again, we compare the results from the selection model to an "unconditional" probit of ownership on size. The effects are dramatic. Unlike location and like dwelling type, size is hugely important for explaining variation in selection, even after controlling for other covariates. The market share of the owner-occupied sector increases from about 70 percent for properties 50 square meters in size to almost 80-90 percent for properties larger than 100 square meters. In summary, the rent-to-price ratio increases with size while the allocation of properties to the rental sector decreases with size.

5.2.2 Rooms vs size

As discussed in Section 4.2, we explored alternative, encompassing specifications where we added variables that were omitted from the baseline specification in order to check the robustness of our results on unobserved heterogeneity (discussed below). Figure 5 plots results from Model 2 showing how our results on size are affected by adding bedrooms, living rooms, bathrooms, kitchens, fireplaces, attic, basement and balcony to the baseline model. We refers to these variables jointly as "rooms." The top left panel of Figure 5 plots the effects of size on log rents and log prices holding number of rooms fixed. The top right panel plots the effects of number of bedrooms on log rents and log prices holding size fixed. The bottom panel shows the effects of size plus $E(rooms|size)$. That is, it shows the "total" effect of size on predicted log rents and log prices, including both the direct effect (top left panel) and the indirect effect (effect due to changes in expected number of rooms as size varies).

Looking at the top left panel, in contrast to Figure 4, size hardly affects rents at all for properties less than 80 square meters when number of rooms is held fixed. As a result, for small properties, prices increase faster than rents when number of rooms is held fixed. However, for larger properties, rents increase faster than values with size. Looking at the top right panel, rents increase much faster with number of bedrooms than values. Overall then, the impact of size and bedrooms on prices and rents is more nuanced than that illustrated in Figure 4.

However, number of rooms and size are highly correlated in the data; especially for smaller properties. This is common sense. It is difficult to have three bedrooms, two living rooms or a big kitchen with only 60 square meters of floor space. Taking this into account, the bottom panel of Figure 5 shows the total effects of size including the direct effects and the indirect effects due to changes in the various room indicators. These results are statistically and economically similar to those from the base specification plotted in Figure 4. One can see that including rooms in the model does not change the qualitative predictions of the model.

5.2.3 Dwelling type

Figure 6 shows how prices, rents and ownership vary with structure type in our baseline model. Both rents and values are higher for more "detached" properties (detached properties, semi-detached properties, and bungalows). Based on the point estimates, the rent-to-price ratio is highest for semi-detached and detached properties, declines by 2.8 percent for bungalows, 7.4 percent for converted flats, 10.8 percent for low rise flats, and 18.3 percent for high rise flats. Of these, the decline for low rise flats is statistically significant at the 10 percent level. The evidence is somewhat weak due to large standard errors, but all is consistent with more detached properties having higher rent to price ratios.

At the same time, more detached properties are more likely to be owner-occupied. These results are stable over waves.[12] As with size above and in contrast to location, higher values and rents are positively correlated with owner-occupancy. In contrast to location, the conditional relationship between dwelling type and predicted ownership is qualitatively similar to the unconditional relationship. The unconditional relationship is detailed in Table 4. Excluding the social housing sector, 95 percent of detached, 85

[12]Results from other waves on dwelling type and size as well as other unreported results are available upon request.

percent of semi-detached properties and 93.9 percent of bungalows are in the owner-occupied sector while dwellings in multi-unit structures (converted flats, low rise and high rise) have ownership rates that vary between 44.8 percent and 54.7 percent. Physical features are important determinants of selection into the owner-occupied sector. Figure 6 shows that conditional on location and other characteristics, the average predicted ownership rate is between 80 percent and 90 percent for semi-detached, detached and bungalows and falls to around 60 percent for dwellings in multi-unit structures.

The pattern is similar to the stylized fact documented in Glaeser and Shapiro (2003) that, in the US, housing units in multi-unit structures are extremely likely to be rented (85.9 percent in their study) whereas single-unit housing is very likely to be owned (85.5 percent in their study). Unconditional ownership rates do not vary quite as much in England across structure types. Conditioning narrows the difference still further.

In summary, as with property size, we find that the rent-to-price ratio is higher for the higher valued version of the characteristic (detached) but that the allocation of detached properties to the rental sector is lower. One caveat to these results is the following. In England property ownership predominantly takes one of two forms, freehold or leasehold. Freehold ownership is ownership in perpetuity. Leasehold ownership is ownership of a long lease (for example 75 years or 99 years).[13] It is clear that property prices should depend on the freehold or leasehold status of the property. Unfortunately the EHS only records information on the type of holding for owner-occupied properties. For the Greater London subsample, leaseholds comprise only 11 percent of owner-occupied properties. In addition, holding type is highly correlated with dwelling type. In the EHS sample, fewer than 23 percent of flats are freeholds while nearly 94 percent of detached houses are freeholds. Giglio et al. (2015) finds that leasehold flats sell for a noticeable duration-dependent discount compared to otherwise identical freeholds. So, some of the decline in prices that is captured by dwelling type may in fact be due to the higher prevalence of leaseholds for flats. However they also find that the type of holding does not affect rents. So the decline in rent-to-price ratios for flats is likely *understated* by omitting holding type. When an indicator for ownership type is included in the property price equation, the parameter estimate is 0.107 (0.042). Freehold status increases property prices nearly 10.7 percent relative to leasehold status. The estimates in Figure 6 are robust to including an indicator for leasehold in our estimation of equation (2).

[13] A third form of ownership, "commonhold", exists but is almost never used due to legal uncertainties.

5.3 Unobserved quality

Let $\rho_{rs} = corr(\varepsilon_r, \varepsilon_s)$ and $\rho_{os} = corr(\varepsilon_o, \varepsilon_s)$. Table 6 reports estimates of these error correlations from the selection model. For the rental sector, ρ_{rs} is 0.95. Properties that are likely to be in the rental sector (low ε_s) have lower unobserved rental quality (ε_r). For the owner-occupied sector, ρ_{os} is 0.67. Selection into the owner-occupied sector is correlated with unobserved owner-occupied quality (ε_o). Both correlations are high but are robust and precisely estimated. Table 7 shows estimates of ρ_{rs} and ρ_{os} obtained from Models 1-5. Adding a large range of additional variables has no impact on our estimates of these correlations. In addition, Models 6-9, reported in the Supplementary Appendix, produce virtually identical estimates. We conclude that a strong and robust pattern of selection on unobservables affects properties in the London housing market.

Significant error correlations have important implications for the hedonic estimates. Figure 7 shows, for an average housing unit, how the predicted average unobserved owner-occupied quality (ε_o) and rental quality (ε_r) vary with distance from London, size, and dwelling type. The top left panel in the figure shows that average owner-quality (ε_o) conditional on being in the owner-occupied sector is positive but does not vary with location. It also shows that average rental-quality (ε_r) conditional on being in the rental sector is negative and also does not vary with location. The results imply that (on average) owner-occupied properties have values that are 5 percent higher than a randomly selected property with identical observable characteristics and that rental properties have rents that are 20 percent lower than a randomly selected identical property. The top right panel shows similar estimates of unobserved quality as a function of size. In both sectors, unobserved quality decreases with size since owner-occupancy rates increase with size. Because the error correlation is larger in the rental sector, unobserved quality in the rental sector declines faster. Large housing units in the rental sector are likely to be of much lower unobserved rental quality. The average rental quality difference between a 50 square meter rental property and a 100 square meter rental property is almost 5 percent compared with an approximate 2.5 percent difference in similar sized owner-occupied properties. Finally, the lower panel shows similar results for unobserved quality versus dwelling type. More detached properties have lower unobserved quality and the decline is stronger for rental sector properties.

One way to explain these results is as follows. Some amenities affect users' enjoyment of a property but also have high maintenance costs. Such amenities raise rents, are negatively correlated with selection into the rental sector, and have a smaller net impact on property prices (the high maintenance cost partially offsets the use value when capitalized into prices). Examples of such amenities could be a jacuzzi, nice countertops and cabinets in the kitchen or a built-in stereo system. Other amenities could affect both prices and rents but have little effect on selection, perhaps because they raise no maintenance concerns. Examples could include aspects of the property's layout, architecture or view.

Other possible explanations include differential expected capital gains, costs of capital or differential vacancy costs. It is logically possible that properties with high unobserved rental quality also had expected rental sector capital gains that were low relative to owner-occupied capital gains. This would be counterintuitive and is unlikely since the option value of switching sectors is part of the expected rental sector gain. Furthermore, given that our findings are robust over several waves, it would imply that these expectations persisted for a number of years. Under rational expectations, this would imply a growing gap in values.

It is also possible that properties with high unobserved rental quality also incur higher rental sector opportunity costs of capital; perhaps either banks have some information about unobserved rental quality and judge such properties to be higher risk or riskier landlords sort into properties with higher unobserved rental quality. However, we are not aware of any literature or evidence providing support for this possibility. Finally, it is possible that markets for properties with high unobserved rental quality are thin and therefore have high expected vacancy costs.

Our results on unobserved quality offer a partial explanation for a puzzle in the housing literature; how to explain high homeownership rates. Many models of homeownership and housing demand use a preference specification which includes a preference for owning (i.e. a "warm glow" from owning, for example see Iacoviello and Pavan (2013) or Kiyotaki et al. (2011)). In calibration exercises in these models, such a preference for owning is often required to generate high homeownership rates. Our results show that an econometrician measuring demand for homeownership using only observable housing characteristics would indeed find a preference for owning. This is because rentals, on average, have lower unobserved quality, and, more importantly, the difference in unobserved quality between observably similar properties is increasing in the probability it is owner-occupied.

5.3.1 Bias in imputed rents and/or prices

The estimation results imply that hedonic estimates of rents and prices that do not control for selection are biased. We find that these biases are statistically and economically significant. Moreover, removing the bias helps explain a common and puzzling finding in the literature.

To illustrate the bias, we re-estimate α and β without controlling for selection. We then predict log rents and log values for all housing units in the sample, first using our original estimates and then also the biased estimates and plot the differences (the bias) between the biased and unbiased imputations. Figure 8 shows the bias in imputed log rent for owner-occupied properties and the bias in imputed value for rental properties. The figure shows how these biases vary with location, size and dwelling type. The first panel shows that the bias in imputed rents is 25 percent on average (here in westerly direction, but robust to other directions). The bias doesn't vary with location. However, the bias does vary significantly with property characteristics. The top right panel shows that it is less than 20 percent for small rental properties and increases to close to 30 percent for very large properties. The bottom panel shows that the bias in imputed rents is more than 20 percent for semi-detached and more than 25 percent for detached properties. The bias in imputed values moves in the opposite direction. The average bias is less than 10 percent in this case. The bias is high for small properties and low for large properties. It is low for detached properties and high for flats.

These results have important implications for statistical authorities and for real estate professionals. Imputed rents are used by statistical authorities to account for housing in price indices and are also considered by tax authorities when considering taxation of implicit income from owner-occupied properties. Imputed values of rental properties could be used to value rental properties that haven't recently transacted.

As a second illustration of the importance of the bias, the left panel in Figure 9 plots the biased predictions of rent-to-price ratios against the biased predictions of log price for both rentals and owner-occupied properties. It also plots the homeownership rate versus the predicted log price. One can see in the figure that homeownership rates increase with predicted log price. At the same time, the biased predictions of the rent-to-price ratio decline with the price of the home. This is a common finding: Verbrugge (2008); Heston and Nakamura (2009); Verbrugge and Poole (2010); Bracke (2015a); Epple et al. (2013) all find that rent-to-price ratios decline with prices while, similarly, Landvoigt et al. (2015) estimates that housing service flows rise less than one-for-one with property prices in the cross-section. Halket and Pignatti Morano di Custoza (2015) shows that some of this co-variation can be explained by differences in rental vacancy rates. The relationships shown in the left panel of Figure 9 and in these studies are puzzling from a certain angle and are a challenge for models which attempt to explain the distribution of household homeownership choices: why do so many households choose to buy expensive properties when seemingly equivalent rental properties are relatively cheap? Estimates from our selection corrected model provide the answer: The rentals are not equivalent. More expensive properties in the rental sector on average have lower unobserved quality. In the right panel of 9, one can see that unbiased predictions of the rent-to-price ratio increase with predicted log price except for the most expensive owner-occupied properties.

5.4 Model fit

The model fits prices in the owner-occupied sector better than it fits rents in the rental sector. In the owner-occupied sector, the R^2 is 0.686; the model explains 68.6 percent of the variance of log prices. In the rental sector, the model explains only 43.3 percent. A larger share of the variance in rents is due to unobserved characteristics. This could be due to more unobservable physical characteristics of the property being priced into rents than prices, perhaps due to higher expected depreciation of many unobserved components or to the shorter expected duration of renters (i.e. renters care about the color of the paint on the walls because they are not willing or perhaps contractually able to repaint the walls; owners are more willing to repaint). It may also be because rents further vary with various tenant and landlord characteristics, such as the age of the tenant or whether the landlord is a professional or corporate landlord. Adding characteristics of the tenant and landlord into our model would add further selection issues that are beyond the scope of this paper (see, e.g. Ackerberg and Botticini (2002)).

We can also look at how well the selection model performs in terms of classifying properties as rental or owner-occupied properties. Table (8) shows percentiles of the predicted ownership probabilities by sector. For owner-occupied properties, all models perform equally well. 90 percent of owner-occupied properties have predicted ownership probabilities higher than 0.63. For rental properties, the classification probabilities are less strongly separated. 50 percent of rental properties have predicted ownership probabilities higher than 62 percent to 66 percent. This is not surprising since 81 percent of properties are owner-occupied. For rental sector properties, the five models agree for all but the 1st percentile of rental properties.

6 Implications for user cost and rent-to-price ratio

6.1 User costs and the rent-to-price ratio

A hypothetical marginal investor's willingness to pay for a property in a sector, either $\pi_o(z,\varepsilon)$ or $\pi_r(z,\varepsilon)$, equals the property's stream of utility or rent flows, net of sector-specific maintenance, and discounted by the sector-specific opportunity cost or user cost of capital. A sector's user cost of capital is determined by the effective rate of interest, the cost of maintenance and expectations about future capital gains, taking into account taxes, transactions costs, inflation, risk, and any option value from switching sectors. In Appendix A we show that these relationships can be characterized by two Poterba-like user cost equations (Poterba, 1992):

$$\pi_o(z,\varepsilon) = \frac{U(z,\varepsilon)}{r_o(z,\varepsilon)+c_o(z,\varepsilon)-g_o(z,\varepsilon)} \tag{8}$$

$$\pi_r(z,\varepsilon) = \frac{R(z,\varepsilon)}{r_r(z,\varepsilon)+c_r(z,\varepsilon)-g_r(z,\varepsilon)} \tag{9}$$

where for each sector i, $r_i(z,\varepsilon)$ is the after-tax discount rate for the marginal investor in that property, $c_i(z,\varepsilon)$ is the cost of management and maintenance (including property taxes and amortized vacancy costs), and $g_i(z,\varepsilon)$ is the expected capital gain (including switching costs and any option value). $U(z,\varepsilon)$ is the utility flow for an owner-occupier.

Each element in these equations may vary both across property types and sector. Since tax policies and borrowing constraints frequently depend on tenure, there need not be a single discount rate that prices all property in equilibrium. For example, in our sample, mortgage interest payments are not deductible from taxable income in England for owner-occupiers but are for landlords. This may be reflected in differences between r_o and r_r. Capital gains are not taxed for owner-occupiers but are taxed for landlords. We assume this is subsumed in differences between g_o and g_r. Lettings are exempt from Value Added Taxes in the UK but *net* rental income may be subject to income taxes. Assuming a common income tax rate, this can be subsumed into $c_r(z,\varepsilon)$. Costs of vacancies in either sector can also be subsumed into c_i. Each element in the equations may vary across time as well. To keep notation simple, we do not denote any such dependence in the equations.

In general, for a property of type (z,ε), user costs in the two sectors will differ. Owners of properties will not in general be indifferent between the two sectors: properties with relatively high rental sector user costs will be selected into the owner-occupied sector. Only for owners of properties at the margin, where $P(z,\varepsilon) = \pi_o(z,\varepsilon) = \pi_r(z,\varepsilon)$, do the two user cost equations (8)-(9) collapse to the more familiar, single equation user cost formula (such as the one in Poterba (1992)). Elsewhere, for owners in the owner-occupied sector, $P(z,\varepsilon) = \pi_o(z,\varepsilon) > \pi_r(z,\varepsilon)$ while for rental sector owners $P(z,\varepsilon) = \pi_r(z,\varepsilon) > \pi_o(z,\varepsilon)$.

6.2 Identification of user cost distribution

We can use our results to obtain estimates of the user cost distribution. Let user costs be $u_o(z,\varepsilon) = \frac{U(z,\varepsilon)}{\pi_o(z,\varepsilon)}$ and $u_r(z,\varepsilon) = \frac{R(z,\varepsilon)}{\pi_r(z,\varepsilon)}$. The variable u_r also is the yield on a rental property (gross of maintenance expenses). If we assume that the service flows from dwelling (z,ε) are the same regardless of sector, then $R(z,\varepsilon) = U(z,\varepsilon)$. Combining this with equations (1) - (3) implies that

$$\ln u_o = (\alpha - \beta) z + \varepsilon_r - \varepsilon_o \tag{10}$$

$$\ln u_r = (\alpha - \beta + \sqrt{\sigma_{ss}}\widetilde{\gamma}) z + \varepsilon_r - \varepsilon_o + \sqrt{\sigma_{ss}}\widetilde{\varepsilon}_s \tag{11}$$

where $\widetilde{\varepsilon}_s = \frac{\varepsilon_s}{\sqrt{\sigma_{ss}}}$. Equations (10) and (11) imply that the ratio of user costs satisfies:

$$\ln u_r - \ln u_o = \sqrt{\sigma_{ss}}\widetilde{\gamma} z + \sqrt{\sigma_{ss}}\widetilde{\varepsilon}_s. \tag{12}$$

User costs and yields depend on the estimated parameters $(\alpha, \beta, \widetilde{\gamma})$ and on the unknown parameter σ_{ss}. In addition, the covariance matrix of the user costs depends on the unknown parameter σ_{ro}.

As noted in Section 2.2, $(\sigma_{ss}, \sigma_{ro})$ are not point identified. While σ_{ro} is not point identified, we show in Appendix B that it is partially identified. Let $\rho_{ro} = \frac{\sigma_{ro}}{\sqrt{\sigma_{rr}}\sqrt{\sigma_{oo}}}$ be the correlation between unobserved rental and owner qualities. We estimated $\rho_{rs} = 0.951$ and $\rho_{os} = 0.667$. Because the covariance matrix Σ must be positive semi-definite, we show in Appendix B that these estimates imply $\rho_{ro} \in [0.407, 0.867]$. Evidently and economically sensibly, many of the unobserved characteristics in our sample are valued both by renters and owner-occupiers.

The identified set for σ_{ss} cannot be pinned down in the same way. Additional information or additional assumptions are required. In Appendix B we show how information on rental sector average yields from Bracke (2015b) and assumptions about the relative variances of the unobservable components of rental and owner-occupied values can be used to calibrate a reasonable range of values for σ_{ss} and to further narrow the range of feasible values for ρ_{ro}. For example, given that ρ_{ro} is in the identified set above, to match Bracke's estimated average rental sector yield of 6.2 percent requires that $\sigma_{ss} \in [0.0001, 0.6]$. Varying σ_{ss} within this interval then pins down the the relative variances of the unobservable components and ρ_{ro}. For example, $\sigma_{ss} = 0.16$ implies that the relative variances of the unobservable components of rental and owner-occupied properties' values are equal and that rental sector yields are relatively high compared to yields of observably equivalent owner-occupied properties, while $\sigma_{ss} = 0.60$ implies that the former is 4.2 times greater and that rental sector yields are relatively low. While we cannot rule out that σ_{ss} takes on values smaller than 0.16, extremely small values are implausible since they imply unobservables play no role in selection (this is not mathematically impossible but seems unlikely given our probit model does not perfectly predict sector), high rental yields, and that ρ_{ro} must be near the top of the identified set.

In the remainder of this section, we discuss qualitative and quantitative implications of our results for user costs and maintenance costs using two alternative calibrations, $\sigma_{ss} = \{0.16, 0.6\}$. In each case, we fix ρ_{ro} so that average rental sector yields are 6.2 percent. The qualitative implications are robust to other reasonable choices for (σ_{ss}, ρ_{ro}). We consider a range of values in Appendix B. For example, smaller values of σ_{ss} shift the dashed curves in Figures 10-12 downward and also imply very low variances of the unobserved component of log rental sector values and high rental sector yields. Larger values of σ_{ss} shift the dashed curves up and imply very large variances of the unobserved component of log rental sector values as well as very low rental sector yields.

6.3 Estimates

6.3.1 User cost variation with respect to location

As discussed above, the rent-to-price ratio increases with distance in several waves but decreases with distance in 2014. At the same time, location is unimportant for selection except possibly for distances very close to the center. Using equation (12), these facts imply that differential users costs between the two sectors do not vary much with distance except for locations very close to the center. This can be seen in Figure 10 which shows average log user costs in each sector as a function of distance. The figure shows estimates for two distinct values of σ_{ss}, $\sigma_{ss} = 0.16$ and $\sigma_{ss} = 0.6$. At distances less than 10 kilometers, owner user costs increase with distance. The pattern is the same for both values of σ_{ss}. Rental sector user costs are higher than owner-sector user costs with a wedge that varies slightly with σ_{ss}; rental user costs are approximately 0.1 log points higher when σ_{ss} is large.

In other words, while the components of user costs (effective discount rates, maintenance costs, and expected capital gains) may vary with distance from London, they do not display significant differential variation across the two housing sectors. Considering the individual components of user costs, it is unlikely that the contributions of, e.g., discount rates and maintenance costs vary differentially in such a way to cancel one another out. On the other hand, as discussed in Section 5.1, it is likely that both rental and owner-occupied maintenance costs as a proportion of value rise with distance. Apparently though, in some time periods like those covered in the 2014 wave, this maintenance effect on the rent-to-price ratio is dominated by other effects. That is, during this period either the discount rate in London went up relative to outside London or relative expected capital gains fell in London.

6.3.2 Variation with respect to size and dwelling type

In Section 5, we found that the more detached and/or the larger a property is the higher is its rent-to-price ratio but the lower is its likelihood of being a rental. Detachedness and size are each positively valued and are each negatively correlated with being in the rental sector. Rents rise faster than π_o which in turn rises faster than π_r. So gross yields for landlords, R/π_r, increase when the value of a house's physical structure increases. Considering equations (8) and (9), this implies that either r_o/r_r or c_o/c_r decreases or g_o/g_r increases with detachedness and with size. Additionally since R/π_r increases, the user costs for landlords must be increasing.

Figures 11 and 12 show the resulting predictions for average log user costs as functions of size and dwelling type in the two sectors. Both figures show the same picture. User costs are higher for bigger properties and for more detached properties. Rental sector user costs increase more with both size and for detached properties. The differential change is larger when σ_{ss} is larger.

What explains these increases? Is it likely that g_o/g_r is dramatically different for detached houses than for dwellings in multi-unit structures or for 110 square meter flats versus 50 square meter flats? This ratio may vary slightly with these physical features due to sectoral differences in taxation of capital gains. For instance, capital gains below a certain threshold are tax exempt for owner-occupiers. However, any variation due to differential tax treatments should also be reflected in our findings with respect to location. This is not the case. Location is uncorrelated with ownership after controlling for observable and unobservable structural characteristics. As discussed above, this suggests that even if expected capital gains varies with location, g_o/g_r is roughly constant. Moreover, given that we find similar patterns with respect to physical features across all of our waves, any wedge in g_o/g_r would have to persist in a way that could imply divergent prices (or a violation of rational expectations).

In contrast, in the case of costs, it is theoretically plausible that rental costs c_r increase faster than c_o when size increases or when one compares detached houses to dwellings in multi-unit structures. This is the direct or indirect implication of the theories of Galster (1983), Henderson and Ioannides (1983) and Coulson and Fisher (2014). Our findings suggest that, if true, this differential increase in costs is large. We compute an estimate of an upper bound for this magnitude in Section 6.3.3.

In the case of interest costs, it is also theoretically plausible that interest costs for the marginal owner-occupier, r_o, decrease with size and detachedness. There is weak evidence in the EHS data that the loan-to-value ratio at time of purchase falls with size for owner-occupiers in our sample. Since smaller loan-to-value ratios imply lower interest rates in the UK market, this suggests that due to residential sorting the marginal purchasers of big owner-occupied properties are different from the marginal purchasers of small properties. These results are available upon request. Comparable evidence for landlords are not available in the EHS. It is possible that less constrained owner-occupiers tend to live in physically more valuable housing units (but not locationally more valuable housing units). In this case, the marginal owner-occupier's discount rate would decline with structure value. However absent increases in c_r or decreases in g_r, our results on gross rental yields R/π_r imply that the discount rate of the marginal landlord must increase with size. We are less aware of theories that provide this implication in equilibrium. Coulson and Fisher (2014) argue that differences in maintenance technologies (including informational aspects) can lead to different ownership structures and thus perhaps financing costs. So it is possible that contracting problems related to the physical structure leads to variation in and selection on both c_o/c_r and r_o/r_r. Further investigation is required to determine whether and to what degree mortgage costs vary across owner-occupiers, across landlords, and across different types of housing units.

6.3.3 Implication for maintenance costs

Based on our above estimates, we can back out an estimate of how sector affects maintenance costs for a property if we make additional assumptions about the discount rates and expected capital gains.[14] For instance, if we make the assumption that $r_r - g_r = r_o - g_o$, equation (12) becomes:

$$\frac{1+\frac{c_r}{r-g}}{1+\frac{c_o}{r-g}} = e^{\sqrt{\sigma_{ss}}\tilde{\gamma}_z - \sqrt{\sigma_{ss}}\tilde{\eta}_s}.$$

In Figures 13- 15, we calibrate $c_o = .017$ (consistent with measures of owner-occupied depreciation in Gatzlaff et al. (1998); Malpezzi et al. (1987); Amior and Halket (2014), among others) and set $r - g = 0.01$ for both sectors. For each value of $\sigma_{ss} \in \{0.16, 0.6\}$, we then plot the conditional and unconditional mean values of c_r as functions of location, size, and dwelling type. The conditional mean functions show the means conditional on being either in the rental or the owner-occupied sector. The unconditional average, $E(c_r)$, which measures the average cost of maintenance if all properties were

[14] One could also do a similar exercise for, say, discount rates. We choose to focus here on maintenance costs because such costs are part of national statistics and because our upper bound estimate may be helpful to researchers looking to calibrate a wedge in maintenance costs between the two sectors.

rented, is higher than c_o. In Figure 13 one can see that, regardless of σ_{ss}, it increases close to the center of London. It is about 1 percentage point higher at the center than distances greater than 5 kilometres. When $\sigma_{ss} = 0.16$, it is nearly 1 percentage point higher than c_o. When $\sigma_{ss} = 0.6$, it is 2 percentage points higher. The conditional mean, $E(c_r | owner)$, which measures the counterfactual cost of maintaining owner-occupied properties were they instead rented, is still higher, about 1 percentage point higher than the unconditional mean. In contrast, the conditional mean, $E(c_r | rent)$ is lower. As one would expect, properties in the rental sector have lower average renter-specific maintenance costs. This stems from the fact that these properties have lower unobserved rental quality.

Figure 14 shows analogous results for maintenance costs versus property size. In this case, the gap between c_r and c_o increases strongly with property size. c_r increases by about 1 to 2 percentage points when property size increases from 50 to 150 square meters, depending on σ_{ss}. The gap between the unconditional mean of c_r and $E(c_r | own)$ is a little less than 1 percentage point when σ_{ss} is low and nearly 2 percentage points when σ_{ss} is high.

Finally, Figure 15 shows how maintenance costs vary with property type. The gap between c_r and c_o is larger for detached properties, 1-1.5 percentage points larger when $\sigma_{ss} = 0.16$ and 3.5-6 percentage points when $\sigma_{ss} = 0.6$. And, like the previous findings, the gap between the conditional and unconditional mean is nearly twice as large when σ_{ss} is larger.

These findings illustrate how maintenance costs vary with sector and with property characteristics. In all cases, the magnitudes are economically substantial. The findings also illustrate how the magnitudes vary with assumptions about σ_{ss} and suggest why estimates of differential maintenance costs that do not control for selection fail to find higher costs in the rental sector. By assuming all differences between the sectors are due to contracting frictions, they provide an upper bound on the moral hazard problem for landlords and tenants.

7 Conclusion

Housing units are not randomly selected into a housing sector. Physical attributes including some that are unobservable in our data are important for selection. Location is not. These findings are consistent with theories of contracting frictions over maintenance and upkeep of the property but may also perhaps be explained by other theories. Most existing models of households' homeownership decisions, such as Landvoigt et al. (2015); Cocco (2005); Diaz and Luengo-Prado (2008); Henderson and Ioannides (1983), largely have abstracted away from explicit considerations of the multi-characteristic nature of housing units. To understand the puzzles of homeownership, our findings point to a need to examine both sector-specific housing costs as well as, on the demand side, to model household choices of both ownership and housing characteristics. Perhaps, households that have a higher demand for larger housing units or detached houses or housing units with high maintenance amenities are more likely to save for a down payment everything else equal.

The results also imply that properly accounting for the bias that selection imparts may encourage refinements in the construction of price indices both for housing and for consumer prices as well as national accounts. It may also help to better understand some of the relative movements of rents and prices over time such as those documented in Campbell et al. (2009).

References

ACKERBERG, D. A. AND M. BOTTICINI (2002): "Endogenous Matching and the Empirical Determinants of Contract Form," *Journal of Political Economy*, 110, 564–591.

AMEMIYA, T. (1985): *Advanced econometrics*, Harvard university press.

AMIOR, M. AND J. HALKET (2014): "Do Households Use Homeownership To Insure Themselves? Evidence Across U.S. Cities," *Quantitative Economics*, 5.

BLOW, L. AND L. NESHEIM (2009): "Dynamic housing expenditures and household welfare," CeMMAP working papers CWP04/09, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

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

——— (2015b): "How much do investors pay for houses?" Bank of England working papers 549, Bank of England.

CAMPBELL, J. Y. AND J. F. COCCO (2007): "How do house prices affect consumption? Evidence from micro data," *Journal of Monetary Economics*, 54, 591–621.

CAMPBELL, S. D., M. A. DAVIS, J. GALLIN, AND R. F. MARTIN (2009): "What moves housing markets: A variance decomposition of the rent-price ratio," *Journal of Urban Economics*, 66, 90–102.

CASAS-ARCE, P. AND A. SAIZ (2010): "Owning versus Renting: Do Courts Matter?" *Journal of Law and Economics*, 53, 137–165.

CHAMBERS, M., C. GARRIGA, AND D. SCHLAGENHAUF (2009a): "The Loan Structure and Housing Tenure Decisions in an Equilibrium Model of Mortgage Choice," *Review of Economic Dynamics*, 12, 444–468.

CHAMBERS, M., C. GARRIGA, AND D. E. SCHLAGENHAUF (2009b): "Accounting For Changes In The Homeownership Rate," *International Economic Review*, 50, 677–726.

CHEN, K. (2010): "A Life-Cycle Analysis of Social Security with Housing," *Review of Economic Dynamics*, 13, 597–615.

COCCO, J. (2005): "Portfolio choice in the presence of housing," *Review of Financial Studies*, 18, 535–567.

COULSON, N. E. AND L. M. FISHER (2014): "Houses, apartments and condos: The governance of multifamily housing," Tech. rep., Penn State University.

DEPARTMENT FOR COMMUNITIES AND LOCAL GOVERNMENT (2017): "English Housing Survey, 2008-2014: Secure Access. 6th Edition," Tech. Rep. SN: 6923, UK Data Service, http://doi.org/10.5255/UKDA-SN-6923-6.

DIAZ, A. AND M. J. LUENGO-PRADO (2008): "On the user cost and homeownership," *Review of Economic Dynamics*, 11, 584–613.

DIAZ, A. AND M. J. LUENGO-PRADO (2011): "The user cost, home ownership and housing prices: Theory and evidence from the US," .

DIEWERT, W. E. (2003): *The treatment of Owner Occupied Housing and other durables in a Consumer Price Index*, Department of Economics, University of British Columbia.

EPPLE, D., L. QUINTERO, AND H. SIEG (2013): "Estimating Hedonic Functions for Rents and Values in the Presence of Unobserved Heterogeneity in the Quality for Housing," Tech. rep.

GALLIN, J. (2008): "The long-run relationship between house prices and rents," *Real Estate Economics*, 36, 635–658.

GALSTER, G. C. (1983): "Empirical Evidence on Cross-Tenure Differences in Home Maintenance and Conditions," *Land Economics*, 59, pp. 107–113.

GARNER, T. I. AND R. VERBRUGGE (2009): "The puzzling divergence of us rents and user costs, 1980-2004: Summary and extensions," *Price and productivity measurement*, 1, 125–146.

GATZLAFF, D. H., R. K. GREEN, AND D. C. LING (1998): "Cross-Tenure Differences in Home Maintenance and Appreciation," *Land Economics*, 74, pp. 328–342.

GIGLIO, S., M. MAGGIORI, AND J. STROEBEL (2015): "Very Long-Run Discount Rates," *The Quarterly Journal of Economics*, 130, 1–53.

GLAESER, E. L. AND J. GYOURKO (2007): "Arbitrage in Housing Markets," NBER Working Papers 13704, National Bureau of Economic Research, Inc.

GLAESER, E. L. AND J. M. SHAPIRO (2003): "The Benefits of the Home Mortgage Interest Deduction," in *Tax Policy and the Economy, Volume 17*, National Bureau of Economic Research, Inc, NBER Chapters, 37–82.

GOODMAN, A. C. (1988): "An econometric model of housing price, permanent income, tenure choice, and housing demand," *Journal of Urban Economics*, 23, 327 – 353.

HALKET, J. AND M. PIGNATTI MORANO DI CUSTOZA (2015): "Homeownership and the scarcity of rentals," *Journal of Monetary Economics*, 76, 107–123.

HANSON, A. (2012): "Size of home, homeownership, and the mortgage interest deduction," *Journal of Housing Economics*, 21, 195–210.

HARDING, J., T. J. MICELI, AND C. SIRMANS (2000): "Do Owners Take Better Care of Their Housing Than Renters?" *Real Estate Economics*, 28, 663–681.

HENDERSON, J. V. AND Y. M. IOANNIDES (1983): "A model of housing tenure choice," *American Economic Review*, 73, 98–113.

HESTON, A. AND A. O. NAKAMURA (2009): "Questions about the equivalence of market rents and user costs for owner occupied housing," *Journal of Housing Economics*, 18, 273 – 279.

HILBER, C. A. (2005): "Neighborhood externality risk and the homeownership status of properties," *Journal of Urban Economics*, 57, 213–241.

IACOVIELLO, M. AND M. PAVAN (2013): "Housing and debt over the life cycle and over the business cycle," *Journal of Monetary Economics*, 60, 221–238.

KAN, K. (2000): "Dynamic modeling of housing tenure choice," *Journal of Urban Economics*.

KIYOTAKI, N., A. MICHAELIDES, AND K. NIKOLOV (2011): "Winners and Losers in Housing Markets," *Journal of Money, Credit and Banking*, 43, 255–296.

LANDVOIGT, T., M. PIAZZESI, AND M. SCHNEIDER (2015): "The Housing Market(s) of San Diego," *American Economic Review*, 1371–1407.

LEE, L.-F. AND R. P. TROST (1978): "Estimation of some limited dependent variable models with application to housing demand," *Journal of Econometrics*, 8, 357–382.

LINNEMAN, P. (1985): "An economic analysis of the homeownership decision," *Journal of Urban Economics*, 17, 230–246.

MALPEZZI, S., L. OZANNE, AND T. G. THIBODEAU (1987): "Microeconomic Estimates of Housing Depreciation," *Land Economics*, 63, 372–385.

OSWALD, F. (2019): "The effect of homeownership on the option value of regional migration," *Quantitative Economics*, 10, 1453–1493.

POTERBA, J. M. (1992): "Taxation and Housing: Old Questions, New Answers," *American Economic Review*, 82, 237–42.

PRABHALA, N. R. (2008): "Self-selection models in corporate finance," *Handbook of Empirical Corporate Finance SET*, 1, 37.

ROSEN, H. (1979): "Housing Decisions and the US Income Tax: An Econometric Analysis," *Journal of Public Economics*, 11, 1–24.

SHILLING, J. D., C. SIRMANS, AND J. F. DOMBROW (1991): "Measuring depreciation in single-family rental and owner-occupied housing," *Journal of Housing Economics*, 1, 368 – 383.

SWEENEY, J. L. (1974): "Housing unit maintenance and the mode of tenure," *Journal of Economic Theory*, 8, 111 – 138.

VERBRUGGE, R. (2008): "The Puzzling Divergence Of Rents And User Costs, 1980-2004," *Review of Income and Wealth*, 54, 671–699.

VERBRUGGE, R. AND R. POOLE (2010): "Explaining the Rent-OER Inflation Divergence, 1999-2007," *Real Estate Economics*, 38, 633–657.

WILLIS, R. J. AND S. ROSEN (1979): "Education and self-selection," *Journal of political Economy*, 87, S7–S36.

WOOLDRIDGE, J. M. (2010): *Econometric analysis of cross section and panel data*, MIT press.

A User costs in a two sector model with switching

Time is discrete. Define the flow value (in non-durable consumption units) from occupying a property of type (z,ε) in sector i as $U_i(z,\varepsilon)$. Assume sector-specific maintenance costs (including property taxes) are $c_i(z,\varepsilon)\pi_i(z,\varepsilon)$ and opportunity cost of capital is $r_i(z,\varepsilon)$. To simplify notation below, assume maintenance costs are paid at the end of each time period. Let $g_i^*(z,\varepsilon)$ be the stochastic sector-specific after-tax capital gains and costs of switching to sector j are $s_j(z,\varepsilon)\pi_j(z,\varepsilon)$. Then, the present value of a property in a sector i is

$$\pi_i(z,\varepsilon) = U_i(z,\varepsilon) - \frac{c_i(z,\varepsilon)\pi_i(z,\varepsilon)}{1+r_i(z,\varepsilon)} + \frac{E\left(\max_{i,j}\left\{[1+g_i^*(z,\varepsilon)]\pi_i(z,\varepsilon), \left[1+g_j^*(z,\varepsilon)-s(z,\varepsilon)\right]\pi_j(z,\varepsilon)\right\}\right)}{1+r_i(z,\varepsilon)}. \tag{13}$$

Thus, property value equals the current net utility flow plus the discounted expected future value. The future value includes an option value from the option to switch sectors.

Define the sector-specific expected capital gains function $g_i(z,\varepsilon)$ as

$$1+g_i(z,\varepsilon) = E\left(\max_{i,j}\left\{1+g_i^*(z,\varepsilon), [1+g_j^*(z',\varepsilon')-s_j(z,\varepsilon)]\frac{\pi_i(z,\varepsilon)}{\pi_j(z,\varepsilon)}\right\}\right).$$

Note that $g_i(z,\varepsilon)$ includes both expected capital gains and the option value of switching sectors net of switching costs. Using this definition, equation (13) can be rewritten as

$$\begin{aligned} U_i(z,\varepsilon) &= \left[\frac{c_i(z,\varepsilon)+r_i(z,\varepsilon)-g_i(z,\varepsilon)}{1+r_i(z,\varepsilon)}\right]\pi_i(z,\varepsilon) \\ &\approx [c_i(z,\varepsilon)+r_i(z,\varepsilon)-g_i(z,\varepsilon)]\,\pi_i(z,\varepsilon) \end{aligned}$$

The approximation becomes exact as the duration of the time period shrinks.

Finally, for a competitive landlord, rents equal the flow value of occupancy so $R(z,\varepsilon) = U_r(z,\varepsilon)$. Equations (8) and (9) follow.

B Set identification of ρ_{ro} and σ_{ss}

Let

$$C = \begin{bmatrix} 1 & \rho_{ro} & \rho_{rs} \\ \rho_{ro} & 1 & \rho_{os} \\ \rho_{rs} & \rho_{os} & 1 \end{bmatrix}$$

be the correlation matrix corresponding to the covariance matrix Σ. From the discussion in Section 2.2, two of the three correlations in the matrix are identified. We have $\rho_{rs} = \frac{\widetilde{\sigma}_{rs}}{\sqrt{\sigma_{rr}}}$ and $\rho_{os} = \frac{\widetilde{\sigma}_{os}}{\sqrt{\sigma_{oo}}}$. The remaining correlation parameter, ρ_{ro}, is not point identified. However, it is partially identified because C must be a valid correlation matrix: the absolute values of the individual correlations must be less than or equal to one and the determinant of C must be non-negative. This latter condition is

$$\begin{aligned} 1 + 2\rho_{ro}\rho_{rs}\rho_{os} - \rho_{ro}^2 - \rho_{rs}^2 - \rho_{os}^2 &\geq 0 \\ 1 - \rho_{rs}^2 - \rho_{os}^2 + \rho_{rs}^2\rho_{os}^2 &\geq (\rho_{ro} - \rho_{rs}\rho_{os})^2 \\ \left(1 - \rho_{rs}^2\right)\left(1 - \rho_{os}^2\right) &\geq (\rho_{ro} - \rho_{rs}\rho_{os})^2 \end{aligned}$$

which implies that

$$\rho_{rs}\rho_{os} - \sqrt{1 - \rho_{rs}^2}\sqrt{1 - \rho_{os}^2} \leq \rho_{ro} \leq \rho_{rs}\rho_{os} + \sqrt{1 - \rho_{rs}^2}\sqrt{1 - \rho_{os}^2}.$$

Combining these inequalities with the point estimates in Table 6, $\rho_{rs} = 0.951$ and $\rho_{os} = 0.667$, we obtain $\rho_{ro} \in [0.407, 0.867]$. Thus, we find that unobserved rental and owner quality are positively correlated and that the correlation is relatively strong, being at least 0.407.

The identified set for σ_{ss} cannot be pinned down in the same way. Additional information or additional assumptions are required. To point identify (or calibrate) both σ_{ss} and ρ_{ro} one would need to use two moments outside of our model. Below, we show how to use the relative variances of the unobservable components of rental and owner-occupied properties values and an estimate from Bracke (2015b) on the average rental sector yield in London.

The variance of the unobservable component of log rental sector value is

$$\mathrm{Var}\left(\varepsilon_o - \varepsilon_s\right) = \sigma_{oo} + \sigma_{ss} - 2\sigma_{os}.$$

A value for σ_{ss} can be obtained from knowledge of the ratio $\mathrm{Var}\left(\varepsilon_o - \varepsilon_s\right)/\mathrm{Var}\left(\varepsilon_o\right)$. The ratio measures the importance of unobservables in explaining the variance in landlords' valuations relative to owner-occupiers'. For example, $\sigma_{ss} = 0.160$ implies the ratio equals 1. The value $\sigma_{ss} = 0.315$ implies a ratio of 2, $\sigma_{ss} = 0.447$ implies a ratio of 3, and $\sigma_{ss} = 0.6$ implies a ratio of 4.2.

Since ε is normally distributed, our model implies that the conditional mean rental sector yield is

$$\begin{aligned} E(u_r|rental) &= \int_z E(u_r|z, \widetilde{\varepsilon}_s \leq -\widetilde{\gamma}z) f(z|rental)\,dz \\ &= \int_z \frac{\Phi(-\widetilde{\gamma}z - \Psi_{12})}{\Phi(-\widetilde{\gamma}z)} e^{(\alpha-\beta+\sqrt{\sigma_{ss}}\widetilde{\gamma})z+0.5\Psi_{11}} f(z|rental)\,dz \end{aligned} \tag{14}$$

where $\Psi_{12} = \text{Cov}(\widetilde{\varepsilon}_s, \varepsilon_r - \varepsilon_o + \sqrt{\sigma_{ss}}\widetilde{\varepsilon}_s)$ and $\Psi_{11} = \text{Var}(\varepsilon_r - \varepsilon_o + \sqrt{\sigma_{ss}}\widetilde{\varepsilon}_s)$. Expected yield depends on our parameter estimates $\widetilde{\gamma}$, our data $f(z|rental)$, and the two parameters (ρ_{ro}, σ_{ss}). Expected yields are decreasing in both ρ_{ro} and σ_{ss}.

Figure 3 in Bracke (2015b), based on data on rents and prices for the UK buy-to-let market, shows that the average rental sector yield in London in 2009-2014 was 0.062. Combining this estimate with equation (14), using the range of values $\rho_{ro} \in [0.407, 0.867]$, we find $\sigma_{ss} \in [0.0001, 0.6]$. If Bracke's number is an underestimate of rental sector yields, then the upper bound on σ_{ss} is lower. If Bracke's number is an overestimate, then σ_{ss} could be higher. However, as discussed above, larger values imply large values for the variance of the unobserved component of log rental values. While we cannot rule out that σ_{ss} takes on values smaller than 0.16, extremely small values are implausible since they imply unobservables play no role in selection (this is not mathematically impossible but seems unlikely given our probit model does not perfectly predict sector), high rental yields, and that ρ_{ro} must be near the top of the identified set. For these reasons, we consider non-negligible values of σ_{ss} to be the most plausible.

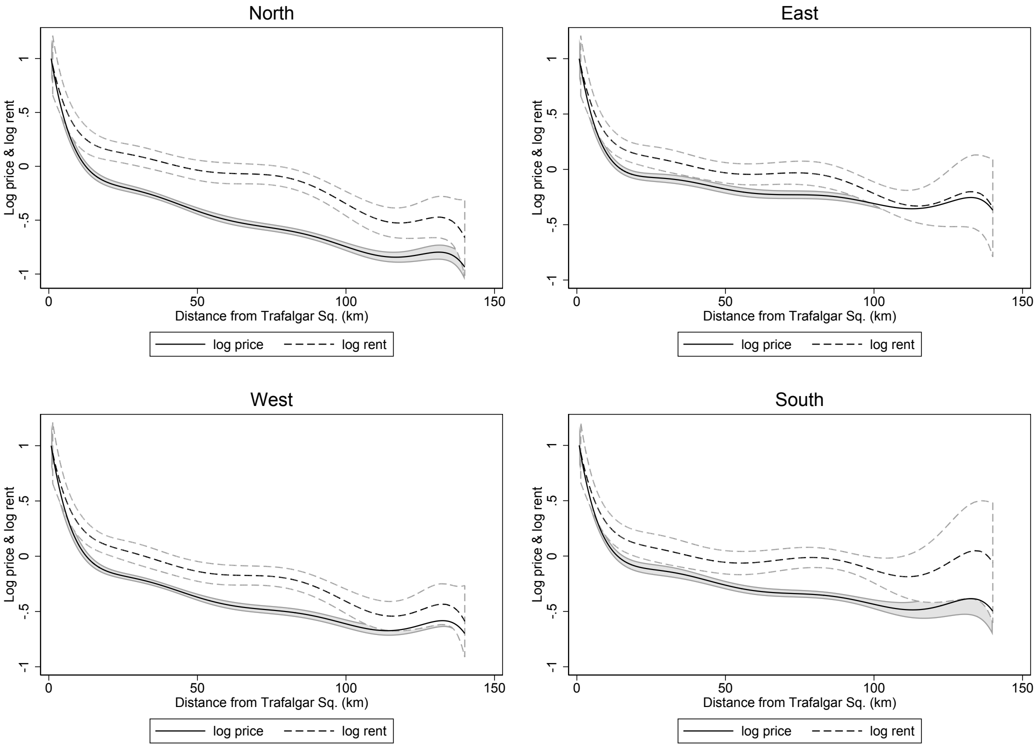
One can see how various values of (σ_{ss}, ρ_{ro}) affect predicted yields in the following table

	ρ_{ro}			
σ_{ss}	0.407	0.614	0.729	0.867
0.16	0.074	0.062	0.056	0.05
0.315	0.069	0.058	0.052	0.046
0.447	0.065	0.055	0.050	0.043
0.6	0.062	0.054	0.048	0.042

Yields are decreasing in both ρ_{ro} and σ_{ss}.

C Figures and Tables

Figure 1: Relative rents and prices vs. location



Log rents and prices have been normalized to equal 1 at the center of London.

Figure 2: Ownership vs. location

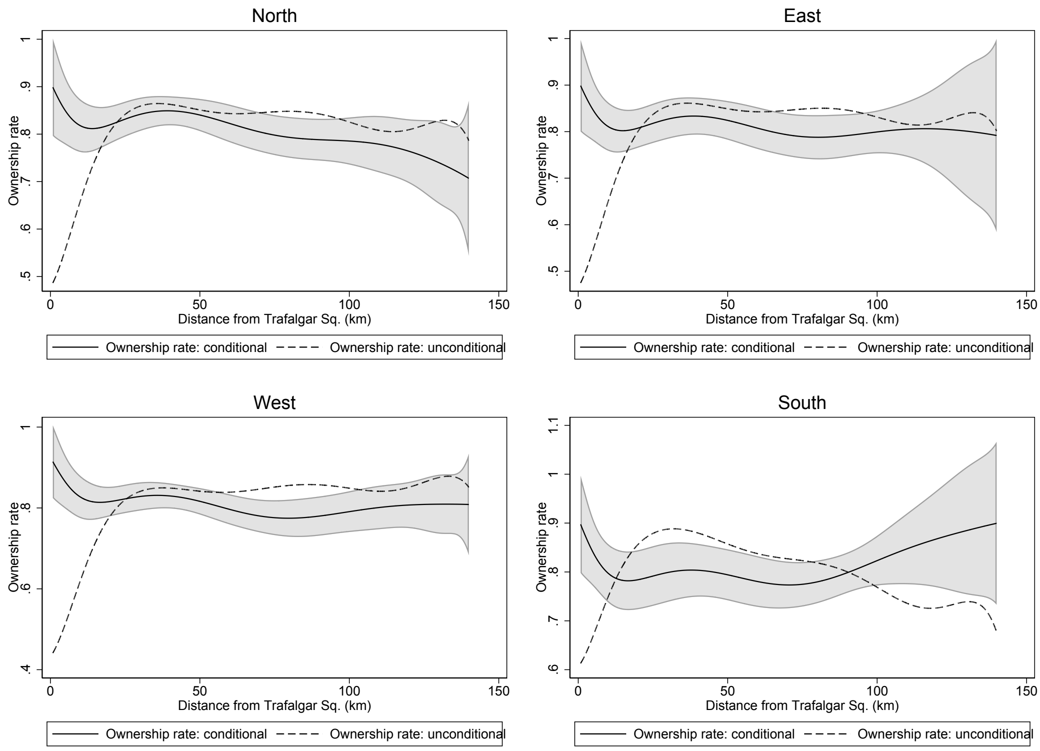
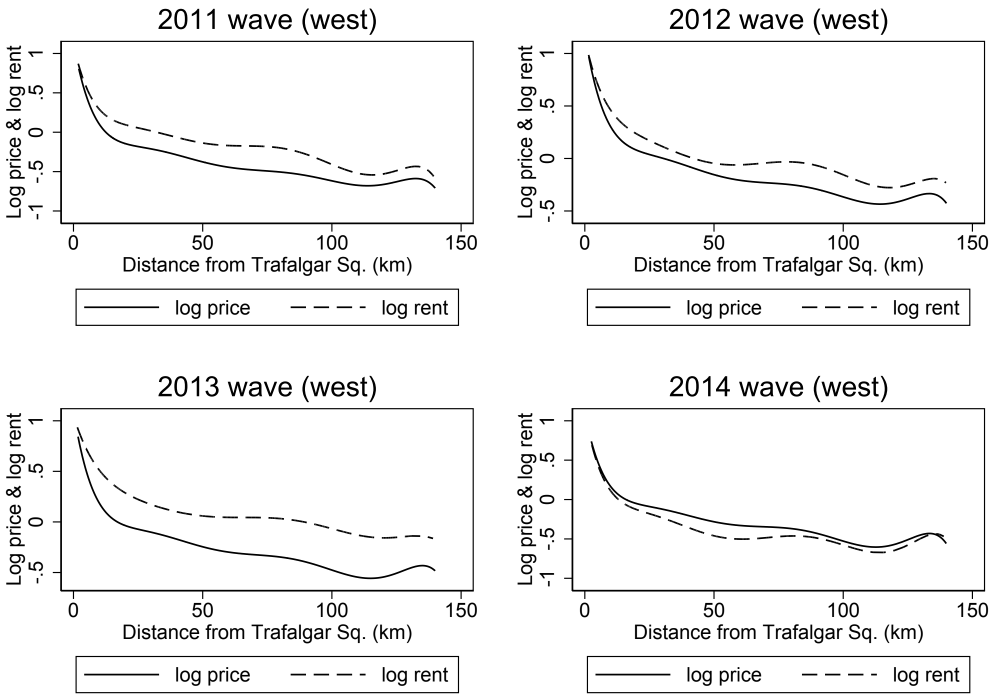
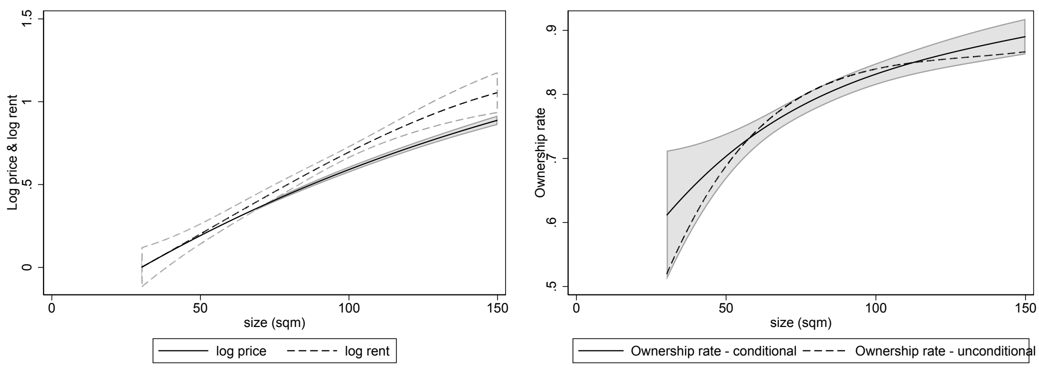


Figure 3: Relative rents and prices vs. location: time variation



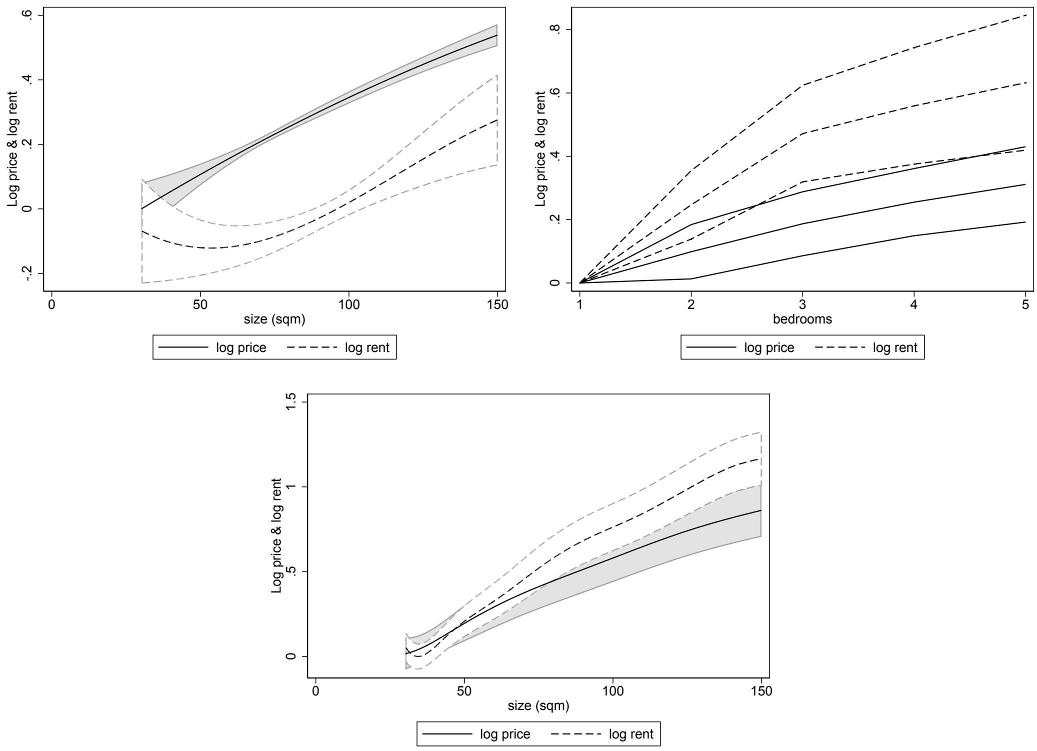
Log rents and prices have been normalized to equal 1 at the center of London. Confidence bands have been omitted in this figure to improve readibility. They are qualitatively similar to the ones displayed in figure 1.

Figure 4: Rents, prices and ownership vs. size



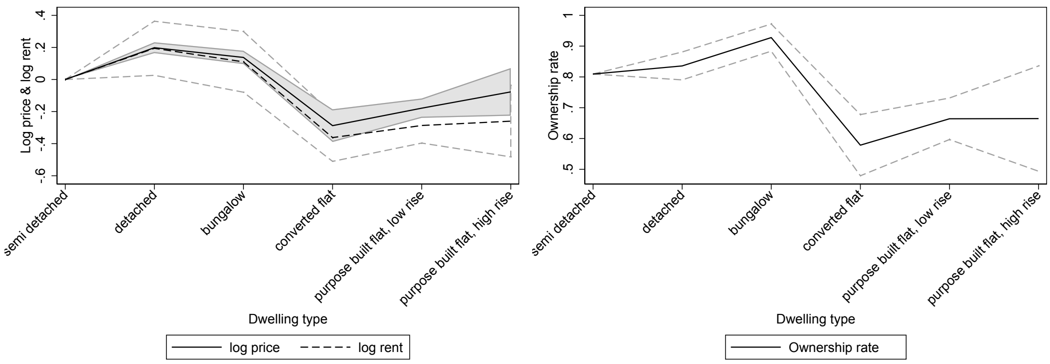
Log rents and prices have been normalized to equal 0 at 40 sqm.

Figure 5: Relative rents and prices vs. size, bedrooms and bedrooms + size



The top left panel shows how predicted prices and rent vary with size holding number of rooms fixed. The top right panel shows predicted prices and rents versus bedrooms holding size and number of other rooms fixed. The bottom panel shows predicted prices and rents versus size and $E(rooms|size)$. Log rents and prices have been normalized to equal 0 at 40 sqm and 1 bedroom, respectively.

Figure 6: Rent, price and ownership vs. dwelling type



Log rents and prices have been normalized to equal 0 for semi-detached housing.

Figure 7: Unobserved quality

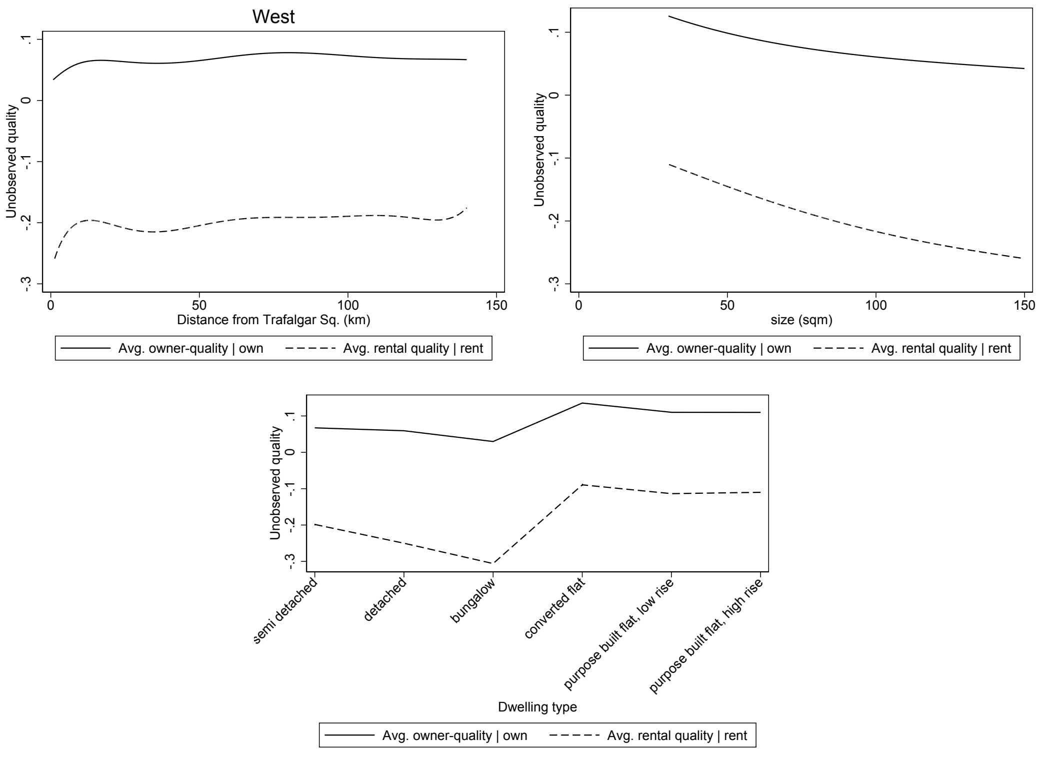


Figure 8: Bias in imputed rent

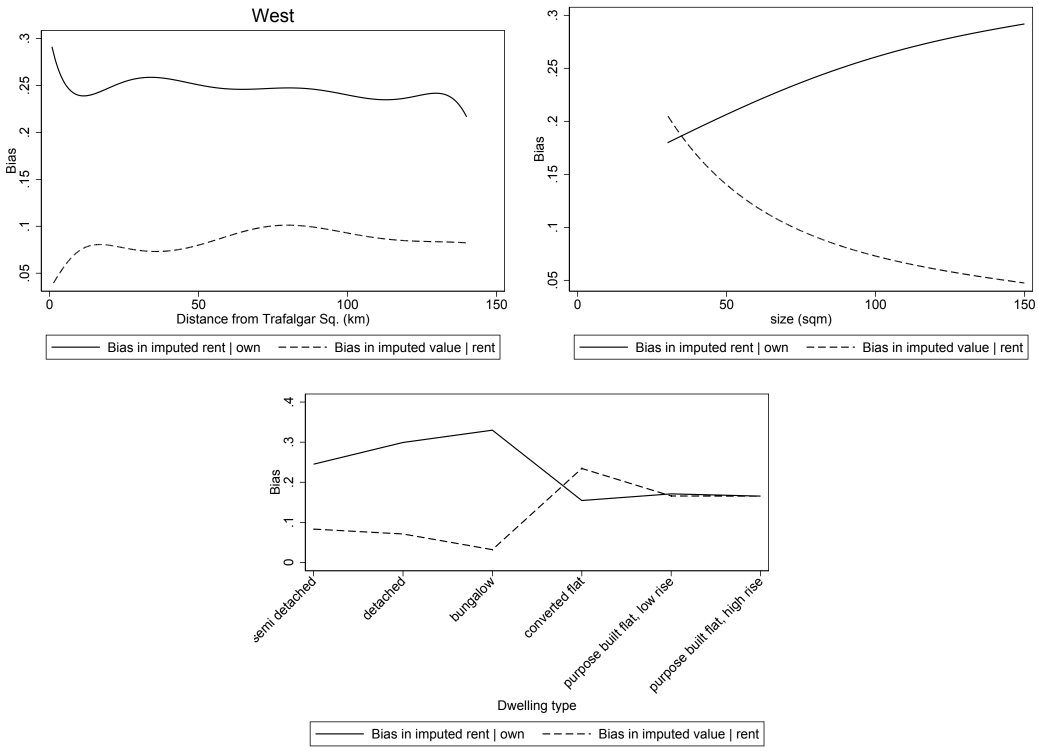


Figure 9: Rent-to-price ratios: biased (left panel) and unbiased (right panel)

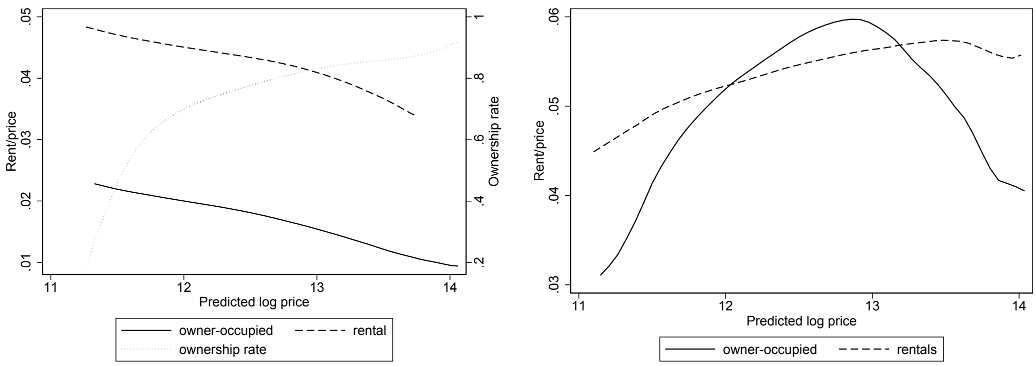


Figure 10: Log user cost vs distance

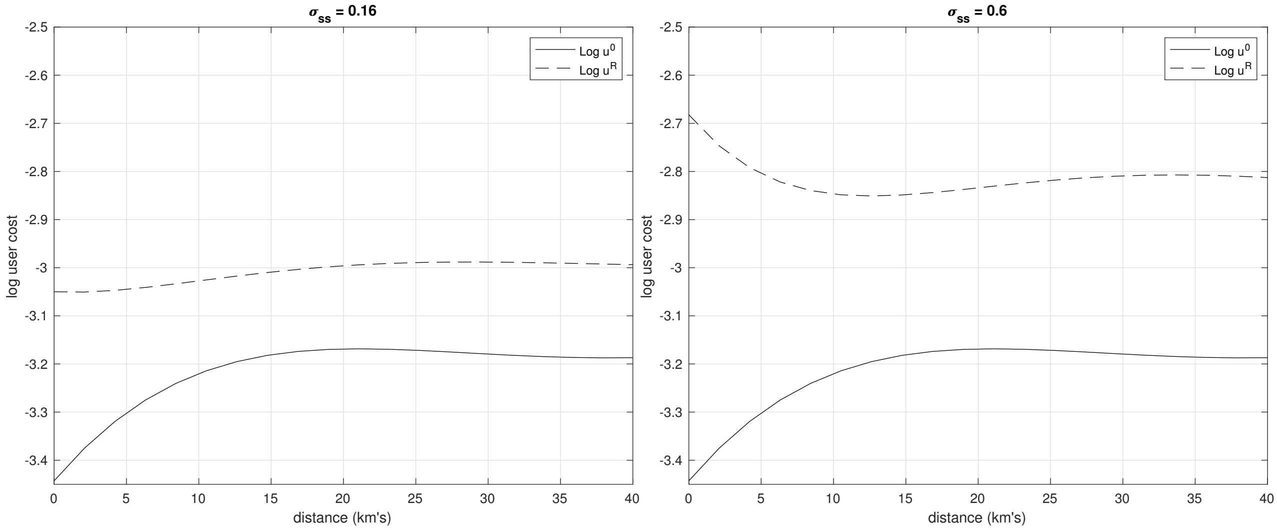


Figure 11: Log user cost vs size

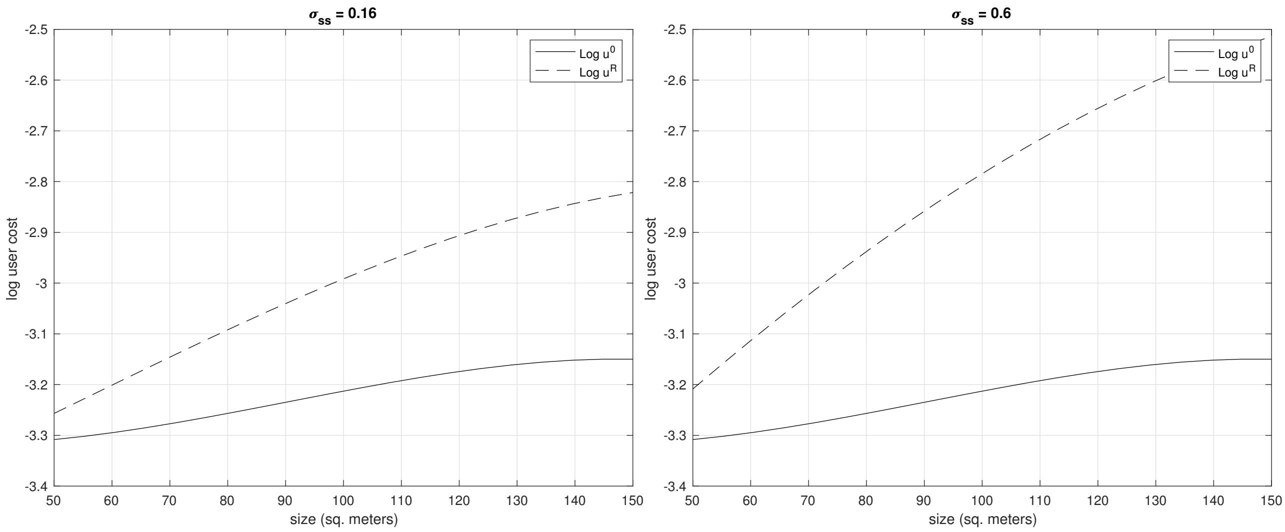


Figure 12: Log user cost vs dwelling type

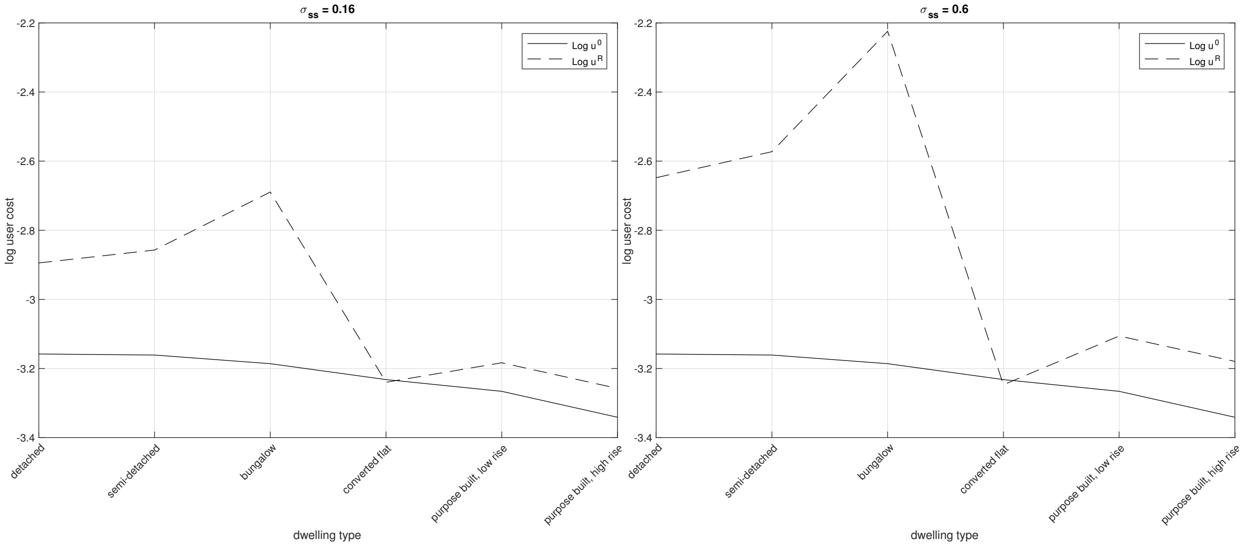


Figure 13: Maintenance cost vs distance

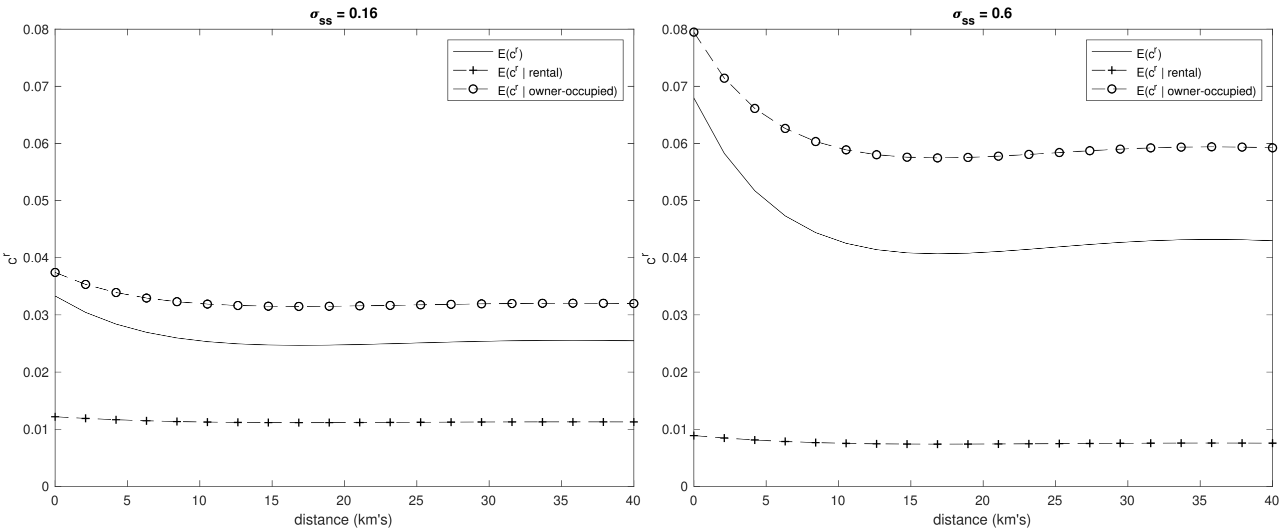


Figure 14: Maintenance cost vs size

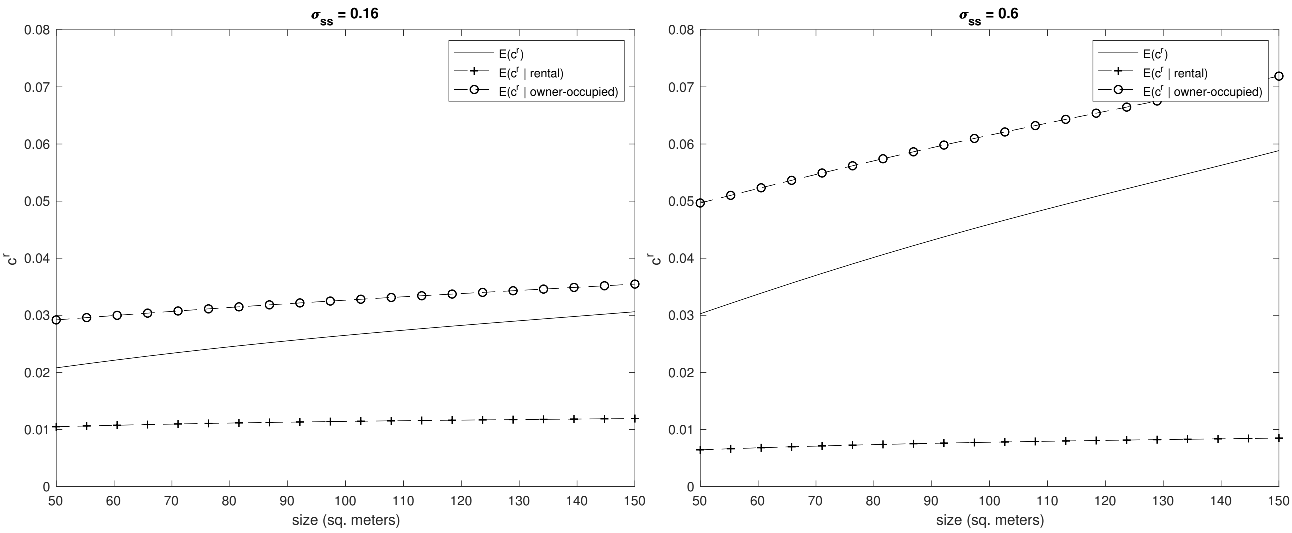


Figure 15: Maintenance cost vs dwelling type

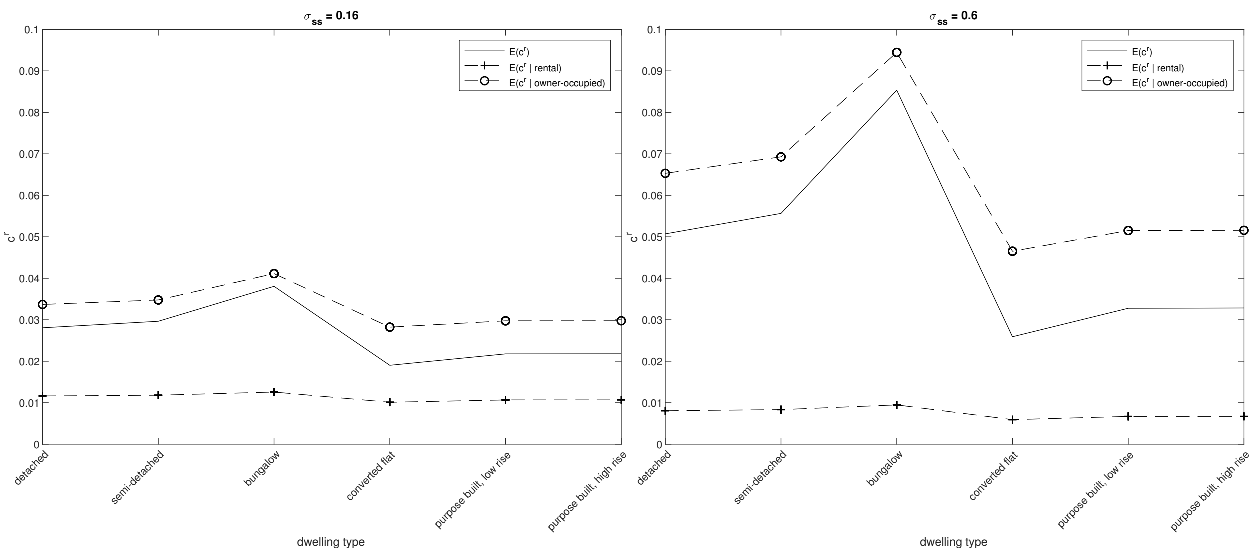


Table 1: Market shares: Greater London and England (%)

Region	EHS Wave	Owner-occupied	Private rented	Social housing
London	2011	66.6	15.7	17.7
	2012	65.3	17.0	17.8
	2013	63.1	18.7	18.1
	2014	62.4	19.5	18.2
England	2011	67.9	14.3	17.8
	2012	67.0	15.1	17.9
	2013	65.3	16.4	18.3
	2014	65.0	17.1	17.9

Note: Market shares are computed using sampling weights for each wave. London refers to the Greater London sample area. The 2011 wave uses data from April 2008 - March 2010. The 2012 wave uses data from April 2009 - March 2011. The 2013 wave uses data from April 2010 - March 2012. The 2014 wave uses data from April 2011 - March 2013. Social housing includes Local Authority provided housing and housing provided by Registered Social Landlords.

Table 2: Market share by distance: Greater London 2011 wave (%)

Distance	Owner-occupied	Private rented	Social housing
less than 10 km	37.9	23.7	38.4
10 - 20 km	61.6	19.8	18.6
20 - 30 km	69.8	13.5	16.8
30 - 50 km	71.4	13.1	15.5
more than 50 km	72.9	13.4	13.7

Note: Market shares are calculated using data and sampling weights from the 2011 wave of the EHS. The 2011 wave uses data from April 2008 - March 2010. Social housing includes Local Authority provided housing and housing provided by Registered Social Landlords.

Table 3: Market share by dwelling size: Greater London 2011 wave (%)

Dwelling size	Owner-occupied	Private rented	Social housing
less than 50 sq. m.	33.1	27.4	39.5
50 - 60 sq. m.	47.5	25.4	27.2
60 - 80 sq. m	60.3	17.1	22.6
80 - 100 sq. m.	74.6	12.6	12.8
more than 100 sq. m.	90.1	7.24	2.63

Note: Market shares are calculated using data and sampling weights from the 2011 wave of the EHS. The 2011 wave uses data from April 2008 - March 2010. Social housing includes Local Authority provided housing and housing provided by Registered Social Landlords.

Table 4: Market share by dwelling type: Greater London 2011 wave (%)

Dwelling Type	Owner-occupied	Private rented	Social housing
semi detached	73.9	13.0	13.7
detached	94.4	5.0	0.40
bungalow	76.8	5.0	18.3
converted flat	39.3	48.5	15.2
low rise	32.2	26.7	38.4
high rise	20.7	19.7	48.1

Note: Market shares are calculated using data and sampling weights from the 2011 wave of the EHS. The 2011 wave uses data from April 2008 - March 2010. The semi-detached category includes “End Terrace" and “Mid Terrace". Social housing includes Local Authority provided housing and housing provided by Registered Social Landlords.

Table 5: Summary statistics by sector

	Rental	Owner-occupied	All
no parking access	0.3115	0.1174	0.1544
street parking	0.2202	0.1130	0.1335
off-street parking	0.4683	0.7696	0.7121
no litter	0.6906	0.8212	0.7963
minor litter	0.2957	0.1704	0.1943
major litter	0.0137	0.0084	0.0094
rear plot depth (sq. meters)	8.1709	15.4083	14.0287
SAP05/09[15]	54.8144	51.8487	52.4140
cavity, insulation	0.1547	0.3136	0.2833
cavity, no insulation	0.3786	0.3512	0.3564
other insulation	0.4667	0.3352	0.3603
heating age 0-3 years	0.2283	0.2451	0.2419
heating age 3-12 years	0.4207	0.4036	0.4069
heating age 12+ years	0.3510	0.3513	0.3512
double glazed (80%+)	0.6708	0.7958	0.7720
0-1 bedrooms	0.2099	0.0410	0.0731
2 bedrooms	0.4011	0.2238	0.2576
3 bedrooms	0.2862	0.4730	0.4374
4 bedrooms	0.0821	0.2057	0.1822
5+ bedrooms	0.0207	0.0565	0.0497
1 bathroom	0.8573	0.7048	0.7339
2 bathrooms	0.1287	0.2458	0.2235
3+ bathrooms	0.0140	0.0494	0.0426
0 living rooms	0.0137	0.0007	0.0032
1 living room	0.8107	0.5277	0.5816
Observations	1,016	3,043	4,059

[15] Each property has an energy efficiency rating calculated by the surveyor using the Standard Assessment Procedure (SAP05 or SAP09). This rating is based on an estimate of each dwelling's energy cost per square meter. It takes account of the cost of space and water heating, ventilation, and lighting. Higher ratings are for more energy efficient properties.

Table 5: Summary statistics by sector

	Rental	Owner-occupied	All
2+ living rooms	0.1756	0.4716	0.4152
big kitchen	0.8406	0.9560	0.9340
fireplaces	0.2817	0.4823	0.4440
attic	0.0444	0.1170	0.1031
balcony	0.0303	0.0270	0.0295
basement	0.0230	0.0199	0.0205
self-reported value		0.8266	
freehold	.	0.8823	
market rent[16]	0.9290		
furnished	0.2335		
partly furnished	0.2169		
unfurnished	0.5496		
log value	.	12.4151	.
log rent	7.8565		.
log value (std. dev.)		0.5009	.
log rent (std. dev.)	0.5419		.
Observations	1,016	3,043	4,059

[16]Self-reported to be let at market rate.

Table 6: Estimation results - baseline model

	Owner-occupied sector Log-value	Rental sector Log-rent	Selection equation
SAP05	-0.002	-0.001	-0.007
	(0.001)	(0.002)	(0.002)
cavity, insulation	0.053	0.110	0.377
	(0.023)	(0.060)	(0.093)
cavity, no insulation	0.040	-0.037	0.020
	(0.027)	(0.077)	(0.115)
heating age: 3-12 years	-0.031	-0.087	-0.173
	(0.015)	(0.041)	(0.069)
heating age: 12+ years	-0.064	-0.115	-0.226
	(0.017)	(0.046)	(0.072)
double glazed (80%+)	-0.039	0.052	0.184
	(0.017)	(0.038)	(0.068)
street parking	-0.018	-0.058	0.018
	(0.026)	(0.049)	(0.087)
off-street parking	0.120	0.112	0.344
	(0.024)	(0.046)	(0.078)
minor litter	-0.092	-0.149	-0.258
	(0.016)	(0.037)	(0.062)
major litter	-0.138	-0.128	-0.146
	(0.054)	(0.129)	(0.218)
rear plot depth	0.004	0.008	0.013
	(0.001)	(0.002)	(0.003)
self-reported value	-0.094		
	(0.015)		
market rent		0.654	
		(0.081)	
partly furnished		0.007	
		(0.032)	
unfurnished		-0.037	
		(0.027)	
$\rho_{is} = corr(\varepsilon_i, \varepsilon_s)$	0.667	0.951	
	(0.083)	(0.012)	
σ_{ii}	0.090	0.386	
	(0.0088)	(0.0525)	
N	4059	4059	
log likelihood	-3.52e+06	-3.26e+06	
R^2	0.650	0.433	

Note: The table displays weighted (using EHS sampling weights) maximum likelihood estimates of Type II Tobit model parameter values estimated using data from the 2011 wave of the EHS. The model includes the variables listed in the table, dummy variables for quarter and dwelling age, and nonparametric functions of size, distance from London, and direction. For the non-parametric functions we use Chebyshev polynomials in distance and in size and Fourier series in angular direction. The number of terms in the series were chosen to minimise the Bayes Information Criteria (BIC). The selected model includes 3rd order polynomials in size (square meters), 7th order polynomials in distance (kilometers), and Fourier series up to order 3. Polynomial and Fourier series coefficients and selected other variables are omitted from the table for concision. Column 1 displays owner-occupied sector results for the model of log value. Column 2 displays rental sector results for the model of log rent. Column 3 displays results from the selection equation. Standard errors are in parentheses.

Table 7: Robustness of selection results

	Model 1	Model 2	Model 3	Model 4	Model 5
$\rho_{os} = corr(\eta_o, \eta_s)$	0.667	0.691	0.703	0.697	0.701
	(0.083)	(0.068)	(0.067)	(0.066)	(0.065)
$\rho_{rs} = corr(\eta_r, \eta_s)$	0.951	0.951	0.950	0.951	0.950
	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)
σ_{oo}	0.090	0.083	0.083	0.080	0.076
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
σ_{rr}	0.386	0.337	0.324	0.324	0.324
	(0.0525)	(0.042)	(0.040)	(0.040)	(0.040)

Note: The table displays weighted (using EHS sampling weights) selected results from 5 model specifications. Each successive model encompasses the previous model. Model 1 is the baseline model results presented above. Model 2 adds indicators for 2 bedrooms, 3 bedrooms, 4 bedrooms, 5 or more bedrooms, 2 bathrooms, 3 or more bathrooms, 1 living room, 2 or more living rooms, 1 or more big kitchens, fireplaces, an attic, a balcony and a basement. Model 3 adds the principal components of property quality, as described in the text. Model 4 adds the first set of principal components for neighborhood quality. Model 5 adds the second set of neighborhood quality principal components.

Table 8: Distribution of predicted probability of being owner-occupied

	Model 1	Model 2	Model 3	Model 4	Model 5
Rental properties					
p1	0.201	0.127	0.111	0.096	0.091
p10	0.354	0.298	0.3	0.294	0.294
p25	0.463	0.451	0.453	0.455	0.443
p50	0.655	0.638	0.63	0.619	0.619
p75	0.813	0.804	0.802	0.802	0.803
p90	0.905	0.906	0.903	0.906	0.904
p99	0.979	0.973	0.979	0.979	0.976
Owner-occupied properties					
p1	0.333	0.339	0.356	0.343	0.34
p10	0.627	0.631	0.628	0.628	0.631
p25	0.794	0.792	0.793	0.795	0.794
p50	0.893	0.901	0.903	0.903	0.901
p75	0.949	0.957	0.957	0.957	0.958
p90	0.976	0.98	0.981	0.981	0.983
p99	0.996	0.996	0.996	0.996	0.998

Note: For each sector and model, the table shows percentiles of the distribution of predicted ownership probabilities.