WORKING PAPERS

N° 1596

November 2024

"Housing Search and Liquidity in Spatial Equilibrium"

Tuuli Vanhapelto and Thierry Magnac

Toulouse School of Economics

Housing Search and Liquidity in Spatial Equilibrium*

Tuuli Vanhapelto† Thierry Magnac‡

November 26, 2024

Abstract

Local housing markets differ in their liquidity, the ease of transacting. Transacting is often easier in urban rather than rural locations, for example. To rationalize these liquidity differences, we set up a model of housing search in the cross-section of multiple interconnected local markets. Markets vary in structural characteristics, leading some to be in higher demand than others, which in turn affects equilibrium liquidity across local markets. Taking the model to data in Finland, we find that the housing market consists of very heterogeneous segments, and especially the value of housing services and the efficiency of the meeting technology matter for the cross-sectional variation in liquidity. Accounting for equilibrium buyer sorting is important: characteristics like the value of housing services affect liquidity both directly but also by attracting more prospective buyers into the market.

*This article was a chapter in Tuuli Vanhapelto's doctoral dissertation in 2023, and we thank the committee members Fabrice Collard, Essi Eerola, Christian Hellwig, Natalia Khorunzhina and Etienne Wasmer for thoughtful comments. We also thank Marko Terviö and Georgios Angelis, and presentation audiences at TSE workshops, the 2020 Meeting of the Finnish Economic Society, the 2021 and 2024 ECHOPPE meetings and the 2024 European Meeting of the Urban Economics Association for helpful feedback and discussions. We thank Etuovi.com and the Finnish Federation of Real Estate Agency for providing access to the data. Tuuli Vanhapelto gratefully ackowledges funding from Yrjö Jahnsson Foundation, Jenny and Antti Wihuri Foundation and Nordea Pankin Säätiö. Thierry Magnac gratefully acknowledges funding support from the Agence Nationale de la Recherche - ECHOPPE grant ANR-17-CE41-0008-01 and ANR-17-EURE-0010 (Investissements d'Avenir programs). The usual disclaimer applies.

†Adam Smith Business School, University of Glasgow, tuuli.vanhapelto@glasgow.ac.uk

‡Toulouse School of Economics, Université Toulouse Capitole, thierry.magnac@tse-fr.eu

1 Introduction

Housing liquidity refers to the ease with which apartments or houses can be bought and sold. Housing is known to be relatively illiquid, and transacting involves a costly search process. This has important macroeconomic implications, since a substantial share of household wealth is held in housing, and holdings of illiquid assets are known to shape marginal propensities to consume (Kaplan & Violante, 2014). Housing liquidity also plays a critical role in household welfare. When housing transactions are less costly, households can more easily adjust their housing consumption in response to life events. This can potentially even translate to greater labor market mobility and higher employment (Head & Lloyd-Ellis, 2012; Karahan & Rhee, 2019).

While it is well established that housing liquidity fluctuates with the business cycle (Krainer, 2001), it is also known to vary across locations (Genesove & Han, 2012; Piazzesi et al., 2020). Selling residential property can be significantly more challenging in remote locations compared to urban areas, for example. While better liquidity in some locations signals higher demand, the mechanisms that explain how the characteristics of local housing housing markets affect liquidity are not yet well understood. This motivates the research presented in this paper, where we explore the cross-sectional differences in housing liquidity using a structural housing search model where prices, sale times, and market tightness are jointly determined in equilibrium across various markets.

Our first contribution is the measurement of local market tightness (i.e. ratio of searchers to sellers) using data from a real estate listing website on the seller-side of the market (e.g. listing durations and prices) as well as on the searcher-side of the market (e.g. subscriptions for email notifications about new listings). We use data from Finland where the housing market is very heterogenous by location. We consider local markets, or segments, defined by geography as well as by housing unit size (distinguishing between small and large units), implying 66 local markets in total. Across these segments, average listing prices range from around 50 000 euros to more than 400 000 euros and average listing times range from some weeks to more than six months. Our measure of market tightness ranges between approximately 0.5 in remote areas and 2.5 in large urban areas, and it has a high predictive power when trying to understand heterogeneity in housing liquidity across local markets. Variation in our measure of tightness explains alone more than 60% of the cross-sectional variation in housing liquidity as measured by sale times.

In consequence, understanding how local market tightness is determined in equilibrium with multiple local markets is crucial for explaining variations in liquidity. As our second contribution, we set up a model of housing search in the cross-section of many interconnected housing market segments in which households choose where to search. Thus, the model integrates a housing search model within a spatial equilibrium framework, as used in urban and spatial economics. Segments are heterogenous in multiple dimensions. We focus in particular on three structural parameters characterizing each market segment: two parameters describing the mean and the dispersion of the value of housing services as well as one parameter indexing matching efficiency. Searching households trade off shorter search times for better matches and sellers trade off shorter sale times for higher prices. We propose a highly tractable method for solving the equilibrium in all markets and for verifying its uniqueness, even with a potentially large number of interconnected markets.

Endogenous segment tightness, or popularity, stemming from searchers' location choice has important implications for liquidity. Firstly, the interconnectedness of local markets can amplify the direct effects of structural characteristics on liquidity. For example, better matching efficiency in a given local market has not only a direct effect on sale times through higher meeting frequency for a fixed tightness but also will increase the popularity of this segment among prospective buyers, which will further increase tightness and decrease seller sale times. Therefore, the general equilibrium effects of parameters on market outcomes through searcher sorting can be larger than what is implied by the direct effects alone. Secondly, when markets are connected, then the observed outcomes of a given segment are affected not only by the structural primitives of that segment but also those of all other segments via searcher sorting. In this case, data from a given segment is not only reflective of the characteristics of that segment but of the entire economy, which must be accounted for when trying to empirically understand the heterogeneity in liquidity.

As our third contribution, we follow an identification strategy that consists in recovering segment-specific parameter vectors describing the distribution of housing quality and the meeting technology from observed data on prices, sale times and market tightness. We adopt the attractive empirical strategy of the quantitative spatial literature of inverting the link between structural parameters and data. Specifically, we invert equations that predict prices, sale times, and value of searching as functions of observed prices, expected sale times and the numbers of active buyers and sellers in each segment. Absent mobility frictions, the value of searching is equal in all segments although it is not directly observable. It is therefore calibrated to match the meeting frequency in one specific segment. It is straightforward to report sensitivity analysis for all estimates as a function of this single underidentified parameter.

Our segment-specific estimates are consistent with the high degree of heterogeneity across segments in equilibrium outcomes. For example, in the most liquid market, the estimated seller match probability (frequency at which sellers meet buyers) is almost 10-fold more than in the least liquid markets. The estimated average monthly value of housing services ranges from approximately 500 to 2,000 euros. While prices in the most expensive segment are around 8 times higher than in the least expensive segment, the value of housing services is only four times higher, as prices are affected by other market characteristics too.

Next, we ask which characteristics of different segments contribute to higher demand, higher local market tightness and eventually higher liquidity. This exercise bridges together the earlier empirical and theoretical research on housing liquidity. Previous research had demonstrated that factors such as higher population and higher incomes, among others, contribute to improved local housing liquidity (Genesove & Han, 2012; Jiang et al., 2024). We provide a structural interpretation for these findings. For example, the relationship between higher population and improved liquidity could operate through different channels: Markets with high population could have a more efficient matching process (for example due to thick-market effects), more valuable housing services (and therefore a higher opportunity cost of not being matched) or a more heterogenous housing stock (and therefore higher returns to search). Our empirical analysis allows to disentangle between these determinants and assess their relative importance in local housing liquidity. We find that, in particular, locations with high population have an efficient meeting technology, and locations with high incomes also have a high value of housing services.

Finally, we use the model and the estimated parameters to conduct two counterfactual exercises. The first counterfactual exercise helps us understand which characteristics of local markets matter the most for the variation in housing liquidity. We shut down different dimensions of parameter heterogeneity across local markets one-by-one. We find that, in particular, heterogeneity in the mean value of housing services and heterogeneity in meeting technology parameters matter for the cross-sectional heterogeneity in liquidity. The second counterfactual illustrates the importance of considering the spillovers between local markets when analysing liquidity. We study what would happen in response to a 10% increase in the value of housing services in urban areas, reflecting for example a shift in the geographic or amenity preferences of the population, while the value of housing services is held constant in rural areas. The increase in the value of housing services in urban areas hurts rural market segments because of changes in relative attractiveness, and rural market segments would experience modest negative effects on equilibrium prices and increases in sale times. The welfare losses through these spillovers are however relatively small in absolute terms - rural homeowners suffer mainly relative to their urban counterparts.

Related literature The literature on housing search is vast, and Han & Strange (2015) provide a review. The majority of the empirical work on housing liquidity focuses on liquidity at the macro level, on the strong procyclicality of the housing market or on endogenous momentum (see, for example, Novy-Marx 2009; Carrillo 2012; Carrillo & Pope 2012; Diaz & Jerez 2013; Head et al. 2014; Ngai & Tenreyro 2014; Eerola & Määttänen 2018; Smith 2020; Anenberg & Bayer 2020; Garriga & Hedlund 2020; Moen et al. 2021; Ngai & Sheedy 2024; Badarinza et al. 2024). These papers share with ours the thematic interest on liquidity, but we are interested in the cross-sectional variation instead of the time-series variation. Only a handful of earlier papers study housing search in the cross-section of multiple market segments (see Head & Lloyd-Ellis (2012), Williams (2018) and Bruneel-Zupanc et al. (2022)). For our analysis, we take as a starting point the model by Krainer (2001), consider a version of the model with multiple market segments that are heterogenous in multiple dimensions, and incorporate the idea of a spatial equilibrium to model the equilibrium in the cross-section of segments.

In terms of the research question, our paper is closest to the small literature on the determinants of housing liquidity in the cross-section of markets. Genesove & Han (2012) provide some of the first evidence on the role of buyer search in shaping housing liquidity, demonstrating how market characteristics such as population growth and income contribute to improved liquidity. Building on this, we offer a more structural interpretation of the data. Our findings complement their work, as our structural parameters appear to capture heterogeneity in the same key characteristics of local markets. Jiang et al. (2024) also provide evidence on how market characteristics affect liquidity, including as measures of liquidity not only time-on-market but also price dispersion. The implications of real estate illiquidity on return heterogeneity and price dispersion are studied in Giacoletti (2021) and Sagi (2021). Other work studying empirically the cross-sectional variation in housing market liquidity include Famiglietti et al. (2020), Amaral et al. (2024) and Carrillo & Williams (2019). We contribute to this strand of literature by documenting new empirical evidence about housing liquidity and search on an online platform as well as by setting up a model where market segments are interconnected and liquidity is determined in the equilibrium across all markets.

Regarding data, model and empirical approach, our paper is closest related to three other papers that also use online search data to study housing markets. Piazzesi et al. (2020), as well as Gargano et al. (2023) who use a similar approach to them, study household search behavior with online data including information exactly where a given household searches. In these approaches, households' choices on where to search are exogenous and observed. We contribute to these papers by modeling households' choices of search location as endogenous, which allows considering counterfactuals where search locations may also change. Kaas et al. (2024) incorporate dynamics to a spatial search model in order to decompose house price changes across local markets to variation in supply factors, demand factors and surplus sharing factors. Our approach is different and complementary in that in our model, value functions such as the value of selling are endogenous, which allows us to study how they respond to changes in primitives. Moreover, we complement these three papers by studying a different question as we emphasize heterogeneity in housing liquidity across local markets.

2 Data and empirical evidence on housing liquidity

Our empirical application uses data from Finland, where there is significant variation in housing market liquidity across the metropolitan areas located mainly in the south of the country and the rural areas in the east and the north. This section presents the main data sources and then documents observations about the variation of equilibrium housing market liquidity in the cross-section of market segments that motivate our analysis.

2.1 Data

A search market clears partially through prices and partially through the time on the market, so we need data on both. While data on transaction prices and quantities is typically available from administrative sources (tax records), data on sale times is not, so we rely on non-administrative data. We use data on housing market liquidity from two sources.

Listing and transaction datasets Etuovi.com is an online house listings website where individuals and real estate agencies can post listings (advertisements) of houses for sale. Etuovi.com is the biggest such website in Finland as measured by the number of individual listings.[1] Etuovi.com has provided us information on all the listings posted on their website between January 2017 and May 2019, but we will restrict our sample to listings posted in 2017 and 2018. For each listing, we observe characteristics like the last listing price before unlisting and the duration of the listing period, as well as apartment characteristics such as the number of rooms.

Most data sources on housing contain information primarily on the seller side of the market, but in the data provided to us by Etuovi.com, we also observe some information related to

[1] Etuovi.com is Finland's leading property listings website as measured by the number of listings as of 6.3.2023. Etuovi.com reached 1.3 million visitors on week 21 in 2020, *Markkinapulssi tutkii - kyselyn tulokset (vko 21/2020).*

the searcher side. At the level of each listing, we observe three variables that are informative about buyer-side search activity: the number of clicks that each listing gathers, the number of times sellers are contacted by buyers via the platform, and the number of email notifications sent via the website to subscribed searchers for each new listing. When constructing our measure of market tightness, we will focus on the email notifications as we think of them as the least noisy signal of buyer activity (clicks can be very noisy for example in high-end segments, and contacts can be uninformative if in some segments buyers contact sellers directly rather than via the platform). The three are, however, highly correlated at the level of market segments with each bilateral correlation of 0.9 and above.

Secondly, we also use data from the Finnish Federation of Real Estate Agency (KVKL Hintaseurantapalvelu)[2], who provided us with a dataset on all transactions intermediated by the member agencies of the organization since the early 2000s. While the listing data contains information on ask prices, and there is no certainty about whether a transaction eventually took place, the transaction data contains information on actual transactions. This also means that we know the final transaction price instead of the ask price. In this dataset as well, we observe a detailed set of unit characteristics, and in particular information about compulsory maintenance costs of the units.

A more detailed description of the data and the sample selection as well as summary statistics for both the listings data and the transaction data are available in Appendix A.1.

Market-level measures As we are interested in describing liquidity at the level of local markets, we have to decide how to stratify the market into local markets or segments. The division could be done based on any observable characteristics of the units, but to remain parsimonious, we focus on two dimensions, geography and size. Starting with geography, we treat the 15 largest cities as regions on their own. We divide the rest of Finland into 18 groups using administrative regions (*maakunnat*) in the 2018 classification. The geographic classification is described in more detail in Appendix A.2. Moreover, we distinguish between smaller apartments (2 rooms or less) and larger apartments (3 or more rooms), since smaller apartments often have shorter sale times. As we have 33 geographic regions and 2 size groups, we end up with 66 market segments in total.

For the purposes of our empirical exercise, we need to measure prices (p_m), average sale times (expected seller time-on-market, denoted $E(TOM)_m$) and the numbers of searchers and sellers, ($n_{searchers,m},\ n_{sellers,m}$), at the segment level (subscript m). In our main analysis,

[2]KVKL Hintaseurantapalvelu, www.hintaseurantapalvelu.fi, Kiinteistönvälitysalan Keskusliitto Ry

we measure p_m by the mean listing price in the segment in the listings data, and $E(TOM)_m$ by the mean listing time in the segment. We also verify that our results are robust to using transaction times and prices (as opposed to listing times and prices). Our preferred method for the measurement of the numbers of active sellers and searchers, implying then values for market tightness, are described below.

Market characteristics We also study the associations of our parameter estimates with municipality characteristics. These are obtained from two different Statistics Finland open access databases: Municipality Key Figures database and the Income Distribution database, for details see Appendix A.3.

2.2 Measuring market tightness

A key ingredient for our analysis is to measure market tightness, the ratio of searchers (buyers) to sellers in a given local market ($\theta_m = \frac{n_{searchers,m}}{n_{sellers,m}}$), using the listings data. To do so, we construct measures of the number of active searchers and sellers in each market segment.

Starting with the number of active sellers, we consider the number of active online listings in a given time period. Since the number of active sellers depends both on the flow of new sellers and on how long they stay active, we compute the number of active sellers in a given time period in segment m as

$$n_{sellers,m} = n_{new\ listings,\,m} \cdot E(TOM)_m \quad (1)$$

where the number of active sellers is given by the product of the flow of new sellers per period and the expected duration of a listing. Both the number of new listings per period and the expected duration of a listing are estimated by sample averages in the listing dataset.

To measure the number of buyers that are actively searching in a given segment, we use the information from the listing platform on the email notifications sent to subscribers. The email notifications operate as follows. Individuals can subscribe for email notifications about apartments responding to pre-determined criteria that they define. Search criteria for these email notifications include standard housing search criteria; For example, a household may choose to be notified about all 2-bedroom units in a given municipality. Subscribed searchers then get a daily or a weekly email containing information on all new listings matching their search criteria. We observe how many searchers received a notification about each new listing in the first week after the listing is posted. The total number of email alerts received per listing is obviously very high as individuals do receive emails about several listings - many people subscribe to such email alerts without necessarily having a serious intention of purchasing a home, with the intention of following how the market evolves. However, our measurement builds on the idea that across segments, variation in the number of emails sent is informative about the variation in the number of searchers, even if the levels might not be informative.

Consider first the total number of notifications sent by the platform in a given period in a given market. This corresponds to the number of new listings in the market, $n_{new\ listings,\ m}$, multiplied by the number of email notifications sent out per each listing in that market, $n_{notifications\ sent\ per\ new\ listing,m}$ (the latter is estimated by the sample average). For example, if there are 10 new listings in a segment in a period and for each new listing, 120 are searchers notified, then the number of notifications sent is equal to 1200. To link this to the number of searching households, we assume that each household only searches in one segment at a time (this will also be true in our model), and all households searching in a given segment want to receive as many notifications. If this is the case, then by simple accounting it has to be that the number of emails received by searchers is given by the number of searchers, $n_{searcher,\ m}$ times the number of notifications per searcher, $n_{notifications\ received\ per\ searcher,m}$. Because the number of notifications sent must be equal to the number of notification received, we have that

$$\underbrace{n_{notifications\ sent\ per\ new\ listing,m} \cdot n_{new\ listing,m}}_{\text{total notifications sent in segment m in given period}} = \underbrace{n_{searcher,m} \cdot n_{notifications\ received\ per\ searcher,m}}_{\text{total notifications received in segment m in given period}} \tag{2}$$

$$\Rightarrow n_{searcher,m} = \frac{n_{notifications\ sent\ per\ new\ listing,m} \cdot n_{new\ listing,m}}{n_{notifications\ received\ per\ searcher,m}} \tag{3}$$

The quantities on the right-hand side of this equation are observed with the exception of $n_{notifications\ received\ per\ searcher,m}$, the number of email notifications that buyers want to receive in different segments, which is unobserved.

We proceed by making an assumption on how many emails searchers want to receive in different segments. One alternative would be to assume that no matter in which market segment a household searches, they will always choose to receive a constant number of new notifications per period (for example, if there were two markets, ”urban” and ”rural”, then households searching in either would be receiving as many emails per week). We think, however, that it is more realistic to assume that in more busy markets, searching households are following the market more intensively. To capture this, we assume that the number of emails that a searching household is willing to receive in segment m is proportional to the inverse seller time-on-market (as we view a shorter seller time-on-market as indicative of a more busy market):

$$n_{notifications\ received\ per\ searcher,m} = \omega\ E(TOM)_m^{-1} \tag{4}$$

We set ω in order to make sure that the aggregate number of selling households in a period equals the aggregate number of searching households, implying $\omega \approx 175$. This means that for example for a market segment where the seller time-on-market is 17 weeks (the average across market segments in our sample), buyers would behave by setting their search criteria on the website so as to receive email information about approximately 10 new listings each week.

Local market tightness for market m follows then directly from setting $\theta_m = \frac{n_{searchers,m}}{n_{sellers,m}}$. An attractive implication of the parameterization in equation (4) is that the implied tightness measures are proportional to the observed number of email notifications only;[3]

$$\frac{\theta_m}{\theta_{m'}} = \frac{n_{notifications\ sent\ per\ new\ listing,m}}{n_{notifications\ sent\ per\ new\ listing,m'}}$$

This means that the resulting variation in the tightness measure across markets reflects only the variation in the observed number of email notifications across markets - not, for example, variation in listing times across markets.

In summary, we have first used the observed quantities directly to produce an estimate about the number of active sellers in a given segment in a given time period. After that, we have used the observed data about how many email notifications are sent out per a new listing, together with a behavioral assumption about how many notifications a searching household wants to receive, to produce an estimate for how many households are searching in a given segment in a given time period. We next summarize the variation in the resulting tighntess measure, which ranges approximately between .5 (two times more sellers than buyers) and 2.5 (five times more buyers than sellers). Appendix A.2 summarizes our tightness measure in more detail.

[3] This can be seen by reorganising equations (1) and (3) as follows:

$$\theta_m = \frac{n_{searcher,m}}{n_{seller,m}} = \frac{n_{notifications\ sent\ per\ new\ listing,m} \cdot E(TOM)_m^{-1}}{n_{notifications\ received\ per\ searcher,m}} = \frac{1}{\omega}\, n_{notifications\ sent\ per\ new\ listing,m}$$

2.3 Empirical Evidence on Housing Liquidity

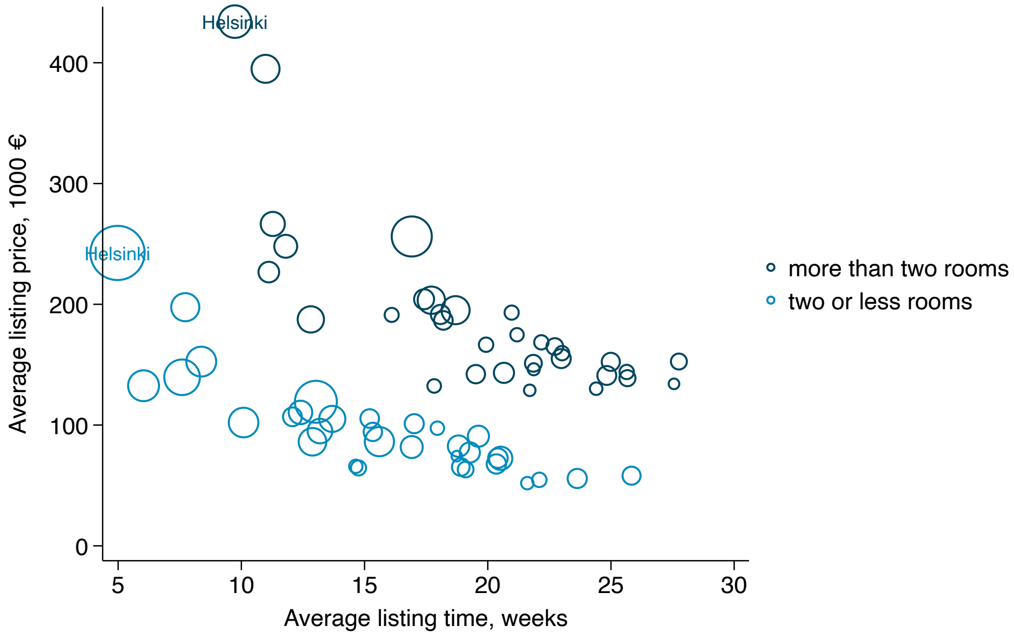
Figure 1 presents the empirical evidence that is the starting point of our analysis. It depicts the variation in three key outcomes at the level of local markets: Prices, listing times and market tightness. The purpose of our modeling exercise that will follow is to understand how these three variables are jointly determined in equilibrium in the cross-section of segments.

Negative correlation between prices and sale times Figure 3a summarizes the cross-sectional correlation between mean listing prices and mean listing times. The first notable feature in the data is that there is substantial heterogeneity in both. Prices in the most expensive segments are more than 8 times higher than prices in the least expensive segments. The longest listing durations are almost 6 times longer than the shortest ones. A second notable feature is that there is a clear negative correlation between listing prices and listing times: The unweighted correlation coefficient is -0.83 for small and -0.77 for large units. The negative correlation suggests that, across markets, some are better off, resulting high prices and short listing times, while others are more disadvantaged, with lower prices and longer listing times. As documented earlier, the time series depicts similar variation: In economic booms, prices are high and sale times are short, and in downturns, prices drop and liquidity declines (Krainer, 2001).

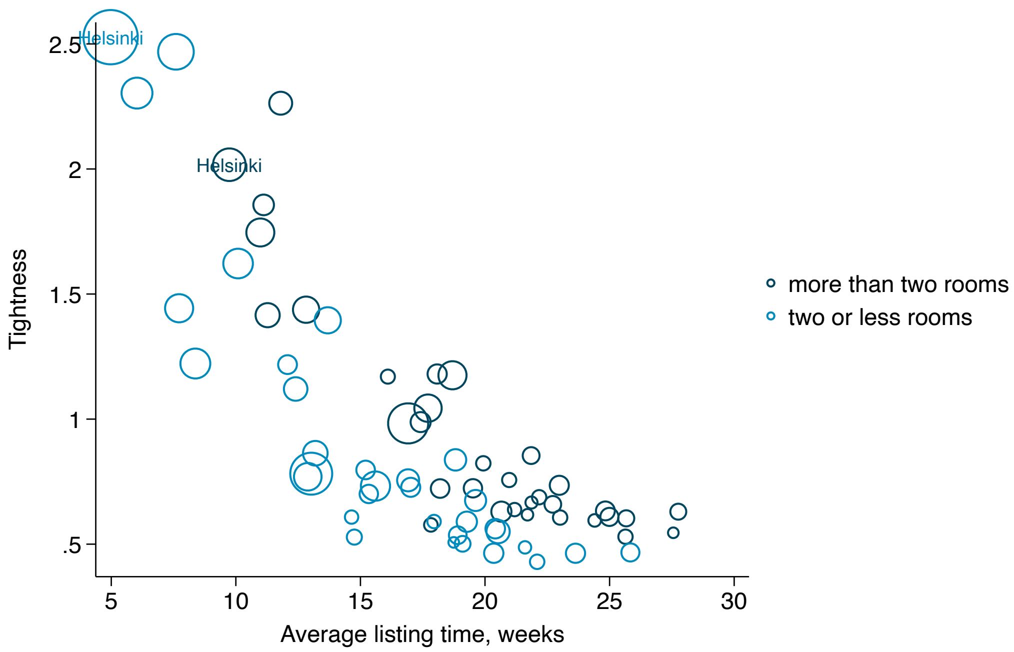
The negative association between prices and times-on-market across market segments obviously need not be causal since it results from an equilibrium. Standard search theory suggests that there is a tradeoff between *longer* sale times or *lower* sales prices. Moreover, the negative association between prices and sale times is not true for all dimension by which the housing market is segmented, such as unit size. In our data, within geographic markets, there is an *opposite* relationship between mean prices and mean sale times across apartments of different sizes: Smaller apartments have lower prices but also shorter sale times than large ones in the same location (in Figure 3a, this is illustrated for the city of Helsinki). In other words, there is a negative relationship between prices and sale times across locations but for a fixed unit size, but on the other hand, within locations but across unit sizes, the relationship is positive. This highlights that to realistically capture the cross-sectional heterogeneity in the data by a model, markets need to differ from one another in more than one dimension of heterogeneity.

Tightness and sale times Figure 3b describes the cross-sectional relationship between our estimate for market tightness and listing times. By construction, the tightness measure ranges from below 1 to above 1 (as we require that the aggregate steady state number of searchers equals the aggregate number of sellers). There is a clear negative association between tightness and listing times (the unweighted correlation coefficient is -0.79 when pooling small and large units together, and in a regression of listing times on tightness, R^2 is above 0.62). This suggests that our measure of tightness can alone explain a lion's share in the variation of listing times. Moreover, unlike for prices, the negative correlation between tightness and listing times is true both within locations across unit sizes and across locations for a fixed unit size.

Since buyer search activity, as measured by the tightness, seems to be a key variable for understanding the variation in liquidity, we next proceed to set up a model where tightness, prices and sale times are jointly determined in equilibrium as functions of the structural characteristics of different markets.



(a) Prices and listing times.



(b) Tightness and listing times.

Figure 1: Prices, sale times and tightness at the level of local markets.

Notes. The figures are based on listing data from Etuovi.com. A unit of observation is a local market as defined in Section 2.1. The size of the circle is proportional to the number of listings.

3 Model

To understand how sale times are associated with prices in the cross-section of market segments, we model a housing market that can be partitioned into distinct but interconnected local markets (also called ”segments” for simplicity). Within a local market, the model structure is inspired by Krainer (2001), but we allow households to choose where they search, generating linkages between different market segments. Prices, expected sale times and market tightness are jointly determined in the equilibrium across all segments.

All proofs for this section are presented in Appendix B.

3.1 Environment

There are N_{houses} houses and $N_{households}$ households in the economy. The housing market is partitioned into M subsegments and each house in the economy can be assigned to one of them, so each segment has $n_{houses,m}$ houses. A household that is matched to a house in market segment m receives a housing dividend, an instantaneous utility of housing consumption every period, given by

$$d_{im} = x_m + \sigma_m \varepsilon_i$$

where x_m is the systemic component in market m, representing the value housing services that all households agree on, and ε_i is a mean-zero idiosyncratic match-specific component for the house-household pair i, so $\sigma_m \varepsilon_i$ captures households' idiosyncratic preferences for house-specific attributes.

Trades take place because each period, a household might receive an exogenous moving shock and become mismatched with their current house. If a household receives a moving shock, which happens with probability $1 - \pi$ (where π is the probability that a match persists from a time period to the next one), they stop receiving flow utility from their current house. Once a household receives a moving shock, they immediately put their house up for sale in the market segment m where it is located, and start searching for a new house. Selling and buying are separate actions and can happen in any order. When a searching household meets a seller and visits a house, an ε_i is drawn from a distribution characterized by a cumulative distribution function $F(\varepsilon)$. Because ε represents independent draws for quality, we assume that $\varepsilon \sim N(0, 1)$. If the searching household chooses to buy in the current period, they start receiving housing dividend $x_m + \sigma_m \varepsilon_i$ in the next period. If a buyer chooses not to buy in the current period, they continue searching in the next period.

Houses are tied to the segment m in which they are located. Sellers face a flow cost c_m at every period they are trying to sell their unit. Sellers wait for potential buyers to arrive to them, and this happens with probability $\delta_m(\theta_m)$ in each period. This meeting probability is a function of market tightness, $\theta_m = \frac{n_{searchers,m}}{n_{sellers,m}}$. If a seller meets with a potential buyer, the buyer draws ε_i. Seller does not observe ε_i, but observes all other information, and conditional on meeting a buyer, makes the buyer a take-it-or-leave-it offer at price p.

The value of having a house on the market for sale in segment m writes

$$q_m = c_m + \beta\big\{(1-\delta_m(\theta_m))q_m + \delta_m(\theta_m)\max_p \big[\mu(p)p + (1-\mu(p))q_m)\big]\big\} \tag{5}$$

where c_m is the flow cost of selling, β is the discount factor, $\delta_m(\theta_m)$ is the probability of meeting with a buyer, and p is the take-it-or-leave it price that the seller sets conditional on having met a buyer. $\mu(p)$ is the belief that the seller has about the probability of the transaction happening, given the price that they have set, and it reflects the tradeoff that sellers face - conditional on having met a buyer, setting a higher price implies a lower probability of transacting.[4]

The seller sets the price optimally to satisfy

$$p_m = \arg\max_p \Big\{\mu(p)p + [1-\mu(p)]q_m\Big\} \tag{6}$$

taking as given the value of selling in the segment. Therefore, the optimal pricing rule p_m must satisfy the necessary first-order condition,

$$\frac{\partial\mu(p)}{\partial p}p\bigg|_{p=p_m} + \mu(p)\bigg|_{p=p_m} - q_m\frac{\partial\mu(p)}{\partial p}\bigg|_{p=p_m} = 0 \tag{7}$$

The equilibrium price which satisfies this first-order condition also satisfies the sufficient second-order condition (see Appendix B). The seller's optimal decision rule is hence to sell for the first household willing to pay the price p_m.

Searching households, or buyers, are mobile. When a household has become mismatched with their former house, they choose in which out of the M segments to search for a new house. After they have decided where to search, matching is random, and buyers in segment m meet with sellers with probability $\lambda_m(\theta_m)$. The value of searching for a new house in

[4]Implicitly, there could be market-makers that would purchase houses from sellers for the price q_m as soon as sellers become mismatched and that would then take care of selling the houses.

segment m is given by

$$s_m = u_m + \beta\big[\lambda_m(\theta_m)E_\varepsilon \max[v_m(\varepsilon_i) - p_m, s_m] + (1 - \lambda_m(\theta_m))s_m\big] \tag{8}$$

where u_m is the flow utility that households receive while they search in segment m (representing the flow utility while renting). If a searching household meets with a seller, they draw ε from $F(\varepsilon)$. If they choose to buy the unit, they make the transfer p_m to the seller and start receiving housing dividend $x_m + \sigma_m\varepsilon_i$ in the next period. If a buyer chooses not to buy in the current period, they continue searching in the next period.

Moreover, because buyers are free to choose where they search, in equilibrum they must be equally well off in all locations. If not, some searchers would switch to market segments offering higher value of search, congesting that market until utilities would be equalized everywhere. Thus, the equilibrium value of search cannot depend on the location, and is denoted by $\overline{s}$.

$v_m(\varepsilon_i)$ is the value of having a match with dividend $x_m + \sigma_m\varepsilon_i$. This consists of the flow utility of a match, $x_m + \sigma_m\varepsilon_i$, and the dynamic value of either remaining matched to that house or receiving a moving shock:

$$v_m(\varepsilon) = x_m + \sigma_m\varepsilon + \beta\big[\pi v_m(\varepsilon) + (1 - \pi)(q_m + \overline{s})\big]. \tag{9}$$

where with probability π the match persists next period and the agent receives v again. On the other hand, with probability $1 - \pi$ the match is broken, in which case the agent no longer receives housing utility but gains the value of selling a house q_m and the value of the search option $\overline{s}$.

$v_m(\varepsilon)$ is increasing and linear in ε, so there exists a reservation value $\tilde{\varepsilon}$ such that the buyer is indifferent between buying the house they have visited in the curret period or searching for another period,

$$v(\tilde{\varepsilon}_m) - p = \overline{s}, \tag{10}$$

$$\tilde{\varepsilon}_m(p) = \frac{1 - \beta\pi}{\sigma_m}p - \frac{1}{\sigma_m}x_m - \frac{(1 - \pi)\beta}{\sigma_m}q_m + \frac{1 - \beta}{\sigma_m}\overline{s}. \tag{11}$$

The last equation is derived from equation (9) and equation (10). The buyer's optimal decision rule is to purchase the first house for which they draw a match value $\varepsilon_i > \tilde{\varepsilon}_m$. The probability that the searcher trades, conditional on having met a buyer, is the probability of drawing a value above the threshold $\tilde{\varepsilon}_m$,

$$1 - F_\varepsilon(\tilde{\varepsilon}_m) \tag{12}$$

Finally, we assume that in equilibrium, sellers have rational expectations: Their belief about selling in a period conditional on the price they set, $\mu(p)$, will be correct so that this probability will indeed be the probability that the buyer will purchase the house,

$$\mu(p_m) = 1 - F_\varepsilon(\tilde{\varepsilon}_m) \tag{13}$$

This gives us the derivative of μ that we need in equation (7):

$$\frac{\partial \mu(p)}{\partial p} = -f(\tilde{\varepsilon}(p)) * \frac{\partial \tilde{\varepsilon}(p)}{\partial p} = -\frac{1-\beta\pi}{\sigma_m} f(\tilde{\varepsilon}(p)) \tag{14}$$

3.2 Equilibrium

Flow equations The number of households in the economy, $N_{households}$, as is the number of houses in each segment, $n_{houses,m}$, are fixed. Houses are in each period either occupied by matched households or held by sellers trying to sell them,

$$n_{houses,m} = n_{occupied,m} + n_{unoccupied,m}$$

Households in segment m are either matched with houses or searching for a new house,

$$n_{households,m} = n_{matched\ households,m} + n_{searchers,m}.$$

Note that it may be that $N_{houses} \neq N_{households}$ or that $n_{houses,m} \neq n_{households,m}$ as there is an implicit rental sector which can accommodate the unmatched households.

Congestion in the intersection of market segments For the value of search to balance between markets, there needs to be a congestion mechanism which ensures that not all households will want to search in the same segment. This is achieved by allowing meeting probabilities δ_m and λ_m (and therefore also equilibrium prices) to depend on market tightness in segment m, the ratio of buyers to sellers *in the segment*. When households are allowed to choose where to search, market tightness in each segment becomes endogenous not only through search times but also the choice of where to search. Searchers choose where to search by trading off expected value of a match, net of the price, relative to the match-finding rate, and this will determine the equilibrium tightness in all segments.

Notice that when sellers make pricing decisions and buyers make purchase decisions, they *do not* internalize the effects that these decisions have on local market tightness. Since pricing and purchasing decisions are made *after* the period's meetings have realized, sellers will not set prices so as to attract more buyers into the segment. Sellers only trade off the probability of selling with a lower price given a level of market tightness.

Assumption 3.1. *(Congestion mechanism.) The meeting probabilities δ_m and λ_m are known functions of the market tightness θ_m and a meeting technology parameter α_m, such that: $\frac{\partial \delta}{\partial \theta} > 0$, $\frac{\partial \lambda}{\partial \theta} < 0$, and $\lim_{\theta \to 0} \delta(\theta) = 0$, $\lim_{\theta \to 0} \lambda(\theta) = 1$, $\lim_{\theta \to +\infty} \delta(\theta) = 1$ and $\lim_{\theta \to +\infty} \lambda(\theta) = 0$, and for fixed θ_m, both probabilities are monotone in α_m.*

For our numerical and empirical exercises, we use a functional form similar to Díaz et al. (2024), where the number of meetings in a segment is given by

$$n_{meetings,m} = \left(n_{sellers,m}^{-\alpha_m} + n_{searchers,m}^{-\alpha_m}\right)^{-\frac{1}{\alpha_m}}$$

where $\alpha_m > 0$ is a matching technology parameter. This implies that the rate at which sellers meet buyers and the rate at which buyers meet sellers are given by

$$\begin{aligned}
\delta(\theta_m; \alpha_m) &= \frac{n_{meetings,m}}{n_{sellers,m}} = (1 + \theta_m^{-\alpha_m})^{-\frac{1}{\alpha_m}} \\
\lambda(\theta_m; \alpha_m) &= \frac{n_{meetings,m}}{n_{searchers,m}} = (1 + \theta_m^{\alpha_m})^{-\frac{1}{\alpha_m}} = \frac{1}{\theta_m} \delta(\theta_m; \alpha_m)
\end{aligned}$$

These meeting probabilities satisfy the conditions required in Assumption 3.1. Holding θ_m fixed, both meeting probabilities increase with the meeting technology parameter α_m.

Equilibrium in the intersection of market segments Under assumption 3.1, meeting probabilities λ_m and δ_m are functions of an exogenous parameter α_m and the market tightness in segment, θ_m. Each market segment is now characterized by exogenous parameters $x_m, u_m, c_m, \sigma_m, \alpha_m$, and $n_{houses,m}$. In equilibrium, θ_m adjusts such that each market segment offers the same utility to searchers. The endogenous elements to be determined in equilibrium are $(p_m, q_m, s_m, \tilde{\varepsilon}_m, \mu_m, n_{sellers,m}, n_{searchers,m})$ $\forall m$.

Definition 3.1. *(Stationary Spatial Equilibrium.) Given market-invariant parameters β and π, a vector of segment characteristics $(x_m, u_m, c_m, \sigma_m, \alpha_m),\ \forall\ m$, the number of houses in each segment $n_{houses,m}$ $\forall m$, the aggregate number of households in the economy $N_{households}$, and a functional form governing $\delta_m(\theta_m;\alpha_m)$ and $\lambda_m(\theta_m;\alpha_m)$, the stationary spatial equilibrium of the economy is a vector $(p_m, q_m, s_m, \tilde{\varepsilon}_m, \mu_m, n_{sellers,m}, n_{searchers,m})$ $\forall m$ such that:*

1. *In each segment, the price p, the value of selling q, the value of searching s, the threshold dividend $\tilde{\varepsilon}$, and the sellers' belief on the probability of selling μ, are such that they satisfy equations (5), (7), (8), (10) and (13).*

2. *The value of searching is equal across all markets and denoted by $\bar{s}$: $s_m = \bar{s}\ \forall m$.*

3. *In each market segment, the number of houses sold in a period equals the number of houses bought,*

$$\delta_m(\theta_m;\alpha_m)\mu_m n_{sellers,m} = \lambda_m(\theta_m;\alpha_m)(1 - F(\tilde{\varepsilon}_m))n_{searchers,m}$$
$$\Rightarrow$$
$$\frac{\delta_m}{\lambda_m} = \frac{n_{searchers,m}}{n_{sellers,m}} = \theta_m$$

4. *In each segment, the flow of households from the matched state into the selling state in a period equals the flow of households away from the selling state,*

$$(1-\pi)\cdot n_{matched\ households,m} = \delta_m \cdot \mu_m \cdot n_{sellers,m}$$

5. *The total number of matched and unmatched houses and households are consistent with the aggregate constant quantities,*

$$n_{houses,m} = n_{matched\ households,m} + n_{sellers,m} \quad \forall m$$
$$n_{households,m} = n_{matched\ households,m} + n_{searchers,m} \quad \forall m$$
$$\sum_{m\in M} n_{households,m} = N_{households}.$$

The equilibrium is symmetric in the sense that all sellers in market m charge the same price and all buyers in market m adopt the same purchasing rule.

Consider next the solution to the equilibrium. We show next the structure of the equilibrium by which Definition 3.1 can be tranformed into a substantially simplified structure.

Lemma 3.1. *Under assumption 3.1, the spatial equilibrium of the economy is characterised by $(2M+1)$ equations in $(2M+1)$ unknowns ($\theta_m \forall m$, $\tilde{\varepsilon}_m \forall m$, $\bar{s}$):*

$$\frac{\beta}{1-\beta\pi}(\gamma(\tilde{\varepsilon}_m,\theta_m)+\lambda(\theta_m)z(\tilde{\varepsilon}_m)+a_m)=\tilde{\varepsilon}_m \quad \forall m, \tag{15}$$

$$\sigma_m\lambda(\theta_m)z(\tilde{\varepsilon}_m)=\frac{(1-\beta)(1-\beta\pi)}{\beta}\left[\overline{s}-\frac{1}{1-\beta}u_m\right] \quad \forall m, \tag{16}$$

$$\sum_{m=1}^{M}\frac{1-\theta_m}{1+\frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi}}n_{houses,m}=N_{households}-N_{houses} \tag{17}$$

where

$$z(\tilde{\varepsilon}_m)=E_\varepsilon((\varepsilon-\tilde{\varepsilon}_m)\mathbf{1}_{\{\varepsilon>\tilde{\varepsilon}_m)\}})=f(\tilde{\varepsilon}_m)-\tilde{\varepsilon}_m(1-F(\tilde{\varepsilon}_m)) \tag{18}$$

$$\gamma(\tilde{\varepsilon}_m,\theta_m)=\frac{1-\beta\pi}{\beta}\frac{1-F(\tilde{\varepsilon}_m)}{f}(\tilde{\varepsilon}_m)+\delta_m(\theta_m)\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}, \tag{19}$$

$$a_m=-\frac{1-\beta\pi}{\beta\sigma_m}(x_m+u_m) \tag{20}$$

Proof. The proof is presented in Appendix B. □

Equation (15) corresponds to part 1 of Definition (3.1), and is related to market-clearing within a single segment. Equation (16) corresponds to part 2 of Definition (3.1), requiring that the value of search does not depend on m. Equation (17) corrresponds to parts 3-5 of Definition (3.1) and implies aggregate market clearing in the sense that the numbers of searchers and sellers in the equilibrium must be consistent with the exogenous aggregate numbers of houses and households.

Proposition 3.1. *(Spatial Equilibrium Existence and Uniqueness.) Given a value of search* $\overline{s}$*, which satisfies* $\overline{s}>0$ *and a feasibility constraint* $\overline{s}\leq\overline{s}_{max}$*, under assumption 3.1, a solution in* $(\theta_m,\tilde{\varepsilon}_m)$ *to equations 15 and 16 exists and is unique. A solution will not exist if* $\overline{s}>\overline{s}_{max}$*.*

Proof. The proof and the characterization of $\overline{s}_{max}$ are presented in Appendix B. □

Proposition 3.1 tells us that given an equilibrium value of search, $\overline{s}$, there is a unique possible combination of a market tightness (given by θ_m) and a solution to the within-segment equilibrium (determined by $\tilde{\varepsilon}_m$) that satisfy the equilibrium equations (15) and (16) . Thus, given an equilibrium value of $\overline{s}$, the equilibrium is unique. Only values of $\overline{s}$ up to an upper bound $\overline{s}_{max}$ that depends on the structural parameters may be considered. For example, if the values of the structural parameters are not very favorable to searchers, then a very high value of $\overline{s}$ is not feasible as an equilibrium outcome. The condition for existence is presented in Appendix B.

Aggregate market clearing Consider next the general equilibrium (or aggregate market clearing) in the sense that the endogenous numbers of searchers and sellers are consistent with the exogenous numbers of houses and households. From Proposition 3.1, the solution in θ_m, $\tilde{\varepsilon}_m$ is unique given a value of search $\overline{s}$. However the proposition does not make the case that there cannot be multiple different equilibria for different feasible values of $\overline{s}$.

However, it is possible to use equation (17) to verify that only one $\overline{s}$ satisfies the aggregate market clearing in the equilibrium definition, which then guarantees the uniqueness of the general equilibrium. Equation (17) establishes a link between the aggregate numbers of houses and households, which are exogenous, and the equilibrium value of search. For given values of N_{houses} and $N_{households}$, Equation (17) simply requires the equilibrium value of search must be such that numbers of searchers and sellers which are implied by the solution characterized by Proposition 3.1 are consistent with N_{houses} and $N_{households}$.

When solving the equilibrium, it is therefore possible to proceed by making a grid on the feasible range of possible values of $\overline{s}$ and solving the implied equilibrium from equations (15) and (16) for each gridpoint. By Proposition 3.1, the solution for a fixed $\overline{s}$ is unique. One can then choose the value of $\overline{s}$ among the gridpoints such that equation (17) holds. Although we have no proof of uniqueness of the solution to this equation, uniqueness can be verified on a case-by-case basis via this tractable procedure. Furthermore, we have never found multiple solutions to this equation. In practice, it turns out that the left-hand side of equation (17) is always increasing in $\overline{s}$, and the right-hand side is fixed. Thus this equation pins down a unique value of $\overline{s}$ on the grid and therefore pins down a unique general equilibrium.

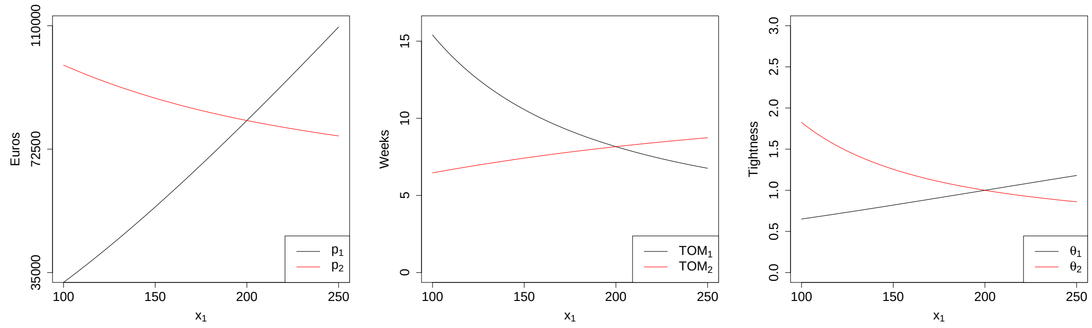
The algorithm we used to solve the model is described in Appendix C. An attractive feature of the equilibrium characterisation and the solution algorithm is that the endogenous quantities in segment m depend on the parameters of other market segments only via the equilibrium value of search $\overline{s}$. This makes the solution very quick to derive when compared to an a strategy where equilibrium in a given market would depend directly on the parameters of all markets. With the strategy that we propose, we only need to consider market clearing equations for one location at a time, after which we verify aggregate market clearing.

3.3 Comparative Statics

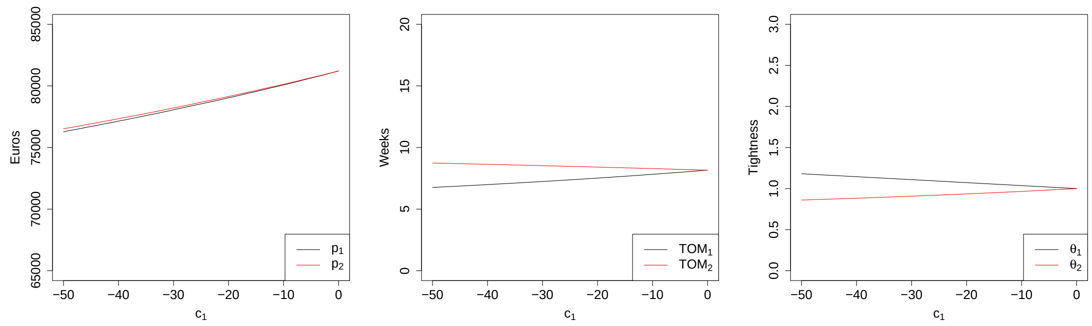
This section reports simple numerical comparative statics from the model to illustrate the equilibrium interconnection between different market segments. To do so, we report equilibrium outcomes from a 2-location quantification of the model.

In the example, the economy consists of 2 cities, $m \in [1, 2]$ which are identical in the baseline. We set $x_1 = x_2 = 200$, $u_1 = u_2 = 0$, $c_1 = c_2 = 0$, $\sigma_1 = \sigma_2 = 2$, $\alpha_1 = \alpha_2 = 0.5$, and $\beta = 0.999$, $\pi = 0.9985$. Both cities have an equal number of houses and the aggregate number of houses equals the aggregate number of households. We then vary one dimension of heterogeneity in city 1 at a time. The characteristics that vary is x_1 in Figure 2a, c_1 in Figure 2b, σ_1 in Figure 2c and α_1 in Figure 2d. Each figure has subpanels which report the equilibrium prices (left panel), seller expected time-on-market (middle panel) and tightness (right panel). Through searcher mobility, changes in the characteristics of one location can also affect the outcomes of the other location. Note that curves are intersecting at the baseline when cities are identical.

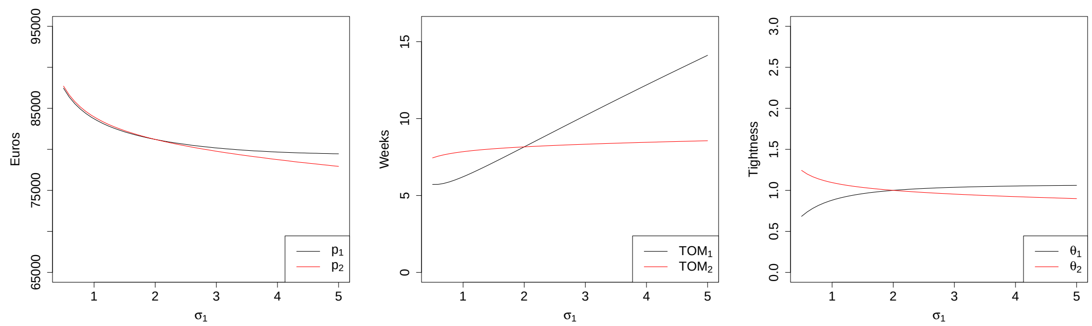
For instance, Figure 2a illustrates how prices, sale times and market tightness vary in each of the two cities as functions of housing quality in city 1, x_1. Changes in the housing quality in city 1 clearly have an effect on prices and sale times in city 1: More valuable housing services translate into higher demand, which is then reflected as both higher prices and shorter sale times (or better liquidity for sellers). Moreover, the value of housing services in city 1 also affects the equilibrium outcomes in city 2. In particular, higher housing quality in city 1 translates *all else equal* to longer sale times (and lower liquidity) in city 2. Figures 2b, 2c and 2d document similar variation in outcomes when c_1, σ_1 and α_1 vary. Qualitatively, prices are most sensitive to x, but listing times (and thus liquidity) are also sensitive to σ and α.



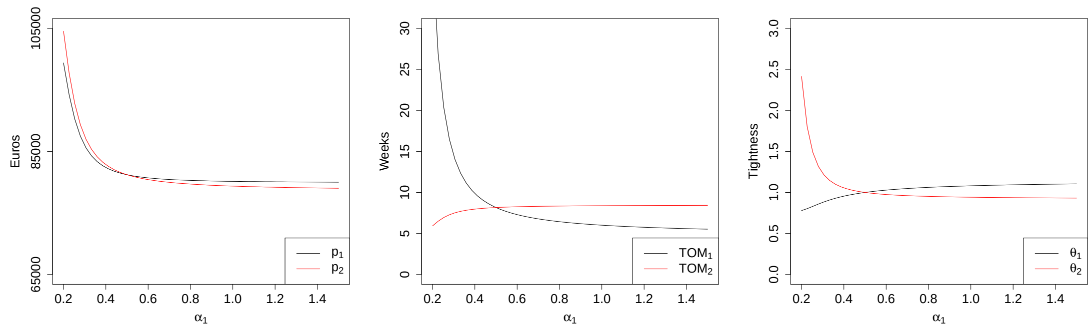
(a) x_1, the average housing quality in city 1, varies.



(b) c_1, the holding cost in city 1, varies.



(c) σ_1, the dispersion of idiosyncratic housing valuations in city 1, varies.



(d) α_1, matching efficiency in city 1, varies.

Figure 2: Prices, sale times and tightness as functions of a varying characteristics in city 1.

Notes. As a baseline, we set $\beta = 0.999$, $\pi = 0.9985$, $u_1 = u_2 = 0$, $c_1 = c_2 = 0$, $\sigma_1 = \sigma_2 = 2$, $\alpha_1 = \alpha_2 = 0.5$, and $x_1 = x_2 = 200$, and then vary one characteristic at a time for city 1.

4 Identification and Estimation

The challenge for identification and estimation in spatial models or multi-segment models in general is that the parameter space can be very large. For example, in our case, each segment is characterised by a vector $x_m, u_m, c_m, \sigma_m, \alpha_m, n_{houses,m}$, making the parameter vector $6 \times M$-dimensional. Therefore, GMM-based methods using grid-search over the parameter space are generally speaking unfeasible. We follow the spatial economics literature and take the model to data by *inverting* model counterparts.

We treat as observed four market characteristics for each market segment: The price p_m, the expected seller time on the market $E(TOM)_m$, and the number of active sellers $n_{sellers,m}$ and buyers $n_{searchers,m}$. We show that under certain assumptions, these segment-level observables map to segment-level parameters, allowing us to recover a parameter vector describing each segment.

All proofs for the propositions of this section are listed in Appendix D.

Calibrated parameters We start by fixing exogenously some of the model parameters. We set the discount factor at $\beta = 0.999$, consistent with an annual discount factor of 0.95, if we consider a model time period to be a week. We also fix the match persistence rate at $\pi = 0.9985$, consistent with a household moving on average every 14 years. This is set so that the aggregate number of houses implied by the model is approximately consistent with the number of owner-occupied houses in Finland.[5]

Next, we calibrate u_m and c_m, the flow utilities (or costs) while searching for a new house and while selling. We set $u_m = 0 \ \forall m$. Implicitly, renting is competitively priced such that rent equals the flow utility of living in a rental unit and therefore the net utility of renting is zero. Thus, even if locations might differ in the value of housing services that they provide to searching (renting) households, absentee landlords are able to extract that surplus fully from households in the form of rent.

To set c_m, we abstract away from costs related directly to selling units, such as the disutility from orgainsing property viewings, and focus on unit maintenance costs that a seller has to pay even when the unit is unoccupied. To estimate these costs, we use the information available in the transaction dataset (KVKL) on building maintenance charges. In Finland, the owners of units in multi-unit buildings have to pay monthly maintenance charges (*hoito-*

[5]The model-consistent number of houses is given below by Equation 21. The data together with this choice of π imply $N_{houses} \approx 1,500,000$. According to Statistics Finland, there were approximately 1,700,000 owner-occupied houses (or owner-occupier households) in Finland in 2018. We intentionally target a slightly lower value, since our measure for the number of sellers is constructed using the listing data and therefore might underestimate the real number of sellers.

vastike) to building cooperatives. The building cooperative then uses the income to pay for expenses such as unit heating, building garbage removal, building management fees, et cetera. The owner of a unit must pay these fees even if the unit is unoccupied, and thus they do reflect at least to an extent the costs of holding a vacant unit for sale. We set c_m as the average maintenance charge in market m. The downside of using maintenance charges to estimate c_m is that they are only observed for units in multi-family housing (in single-family housing, each owner makes their own arrangements for maintenance, and the costs are not reported in sales or listings data). Thus, we assume that the cost of holding a single-family unit vacant is not too different from the cost for units in multi-family housing (realistically, this is likely underestimating the maintenance costs of single-family housing). The calibrated values for c_m are approximately proportional to prices and the average annualized maintenance cost is on average 2% of the average listing prices. Appendix Table A2 summarizes these costs.

Market size One of the primitives characterizing each market is market size as measured by the number of houses in each market; $n_{houses,m}$. To recover them, we use flow equations and information on the number of sellers in each segment, which we treat as observed (for the details on the measurement, please see Section 2.2). Moreover, when the number of houses are recovered, then the aggregate number of households, $N_{households}$, follows directly from information on the aggregate numbers of houses, searchers and sellers. (Note that we have constructed our measure of $n_{searchers,m}$ in a way that guarantees that $N_{searchers} = N_{sellers}$, which using Appendix Equation 128 implies also that $N_{houses} = N_{households}$.)

Proposition 4.1. *(Number of houses and households.) The segment size, as measured by $n_{houses,m}$, can be recovered from observed values of $n_{sellers,m}$ and $E(TOM)_m$:*

$$n_{houses,m} = n_{sellers,m} \cdot \left[1 + \frac{1}{(1-\pi)E(TOM)_m}\right] \tag{21}$$

where the right-hand side quantities are observed. The aggregate number of houses, N_{houses}, then follows immediately. The aggregate number of households, $N_{households}$, follows from N_{houses}, $N_{searchers}$ and $N_{sellers}$.

Proof. Please see Appendix Section B.1.3. The number of sellers can be mapped to the model-consistent market sizes (number of houses in each segment) through the flow equation 120. $N_{households}$ is directly implied by equation 128 if N_{houses}, $N_{searchers}$ and $N_{sellers}$ are known. □

Identification of parameters given $\bar{s}$ and structural constraints Next, we show how to recover x_m, σ_m, α_m from data on p_m, $E(TOM)_m$ and $\theta_m = \frac{n_{searchers,m}}{n_{sellers,m}}$, given the value of search $\bar{s}$. Given $\bar{s}$, we have $3 \times M$ equations for $3 \times M$ unknowns, and the order condition for identification is satisfied. We next explore if other constraints limit the degree of identification. This turns out to be the case: The model structure imposes additional requirements on parameters, in particular from requiring that all model-implied probabilities are bounded between 0 and 1. The implied structural constraints are developed in Appendix D.

Proposition 4.2. *(Structural constraints.) Solutions to the model exists and respect that probabilities are bounded between 0 and 1 if and only if*

$$E(TOM)_m \theta_m \geq 1. \tag{22}$$

Furthermore, solutions $\tilde{\varepsilon}_m$ are well defined if and only if $\bar{s} \geq 0$ respects the following bound:

$$\frac{(1-\beta)(1-\beta\pi)}{\beta}\bar{s} \leq \min_m(\varpi_m \psi(\bar{\varepsilon}_m)), \tag{23}$$

in which variables ϖ_m and $\bar{\varepsilon}_m$ as well as function $\psi(.)$ are defined in the proof.

Proof. See Appendix D. □

Under these conditions, we can now properly invert the model and recover segment-specific parameters.

Proposition 4.3. *(Model inversion.) If the data-generating process is the one described in Definition 3.1 and verifies Proposition 4.2 and if $\bar{s}$ is given and verifies the bound written in Proposition 4.2, then parameters describing each market segment, (x_m, σ_m, α_m) can be point identified as nonlinear transformations of data on p_m, $E(TOM_m)$, θ_m.*

The proposition states that fixing the value of search $\bar{s}$ within proper bounds, it is possible to back out $3 \times M$ parameters. This is straightforward using model counterparts of the observed quantities. In other words, Proposition 4.3 allows us to back out *exactly* the parameters that would generate the observed data if the data generating process was the model of section 3.

Calibrating $\bar{s}$ in the identified interval Our model inversion argument in Proposition 4.3 takes the value of $\bar{s}$ as given, but the structural constraints in Proposition 4.2 imply only a partially identified interval of feasible values for $\bar{s}$ instead of a point-identified value. Therefore, to implement the model inversion in practice, one needs to calibrate the value of search within the partially identified set. Then, using Proposition 4.3, the remaining parameters are well-identified. Since there is only one quantity which needs to be calibrated within the identified interval, and this calibration implies the full remaining parameter vector, it is straightforward to then characterize the sensitivity of the results to the calibrated quantity.

There are multiple possible methods for calibrating $\overline{s}$ within the partially identified interval: one alternative is to calibrate $\overline{s}$ directly. Another alternative is to exogenously calibrate a single parameter in one market segment. For example, fixing σ to a constant in one market segment would recover point identification for $\overline{s}$. An alternative, which we follow here, is to fix the meeting probability δ to an exogenously set constant in a single segment, which also recovers point identification for $\overline{s}$.

Proposition 4.4. *(Fixing $\overline{s}$.) If the data-generating process is the one described in Definition 3.1, then calibrating δ_0 for a single segment 0 together with observed data p_0, $E(TOM)_0$ and θ_0 implies an equilibrium value of search, $\overline{s}$.*

Proof. See Appendix D. □

Fixing δ_0 to an exogenously set constant implies a value of search for that segment, s_0. By Definition 3.1, this has to be the equilibrium value of search everywhere, so $\overline{s}$ is recovered.

Using constraint (23) in Proposition 4.2 and data on p_m, $E(TOM)_m$ and θ_m $\forall m$ implies an upper bound for the identified interval of $\overline{s}$ given by 60,215 euros. The market segment where the constraint of equation 23 is binding is the market for small housing units in the third-largest metropolitan area in Finland (city of Turku). In this segment, selling units is relatively easy: The seller time-on-market $E(TOM)$ is the second-lowest in our sample data, market tightness θ is high, but prices are low compared to locations such as the capital Helsinki. In our baseline calibration, we set $\delta_0 = 0.5$ for that market. This is to reflect that on this market, it is relatively easy for sellers to meet prospective buyers: By assumption, a seller meets a new prospective buyer in expectation every two weeks. This implies a value of search equal to $\overline{s} = 54,429$ euros. We consider alternative calibrations for $\overline{s}$ in Appendix E.1 and show that the results are not very sensitive to the choice of δ_0.

Summary We have shown that all the model primitives needed to compute the equilibrium, $(x_m, \sigma_m, \alpha_m, n_{houses,m})$ can be recovered from data on p_m, $E(TOM)_m$ and $n_{sellers,m}$, $n_{searchers,m}$ and calibrated parameters β, π, u_m $\forall m$, c_m $\forall m$ and δ_0.

5 Results

This section starts by summarizing our parameter estimates, proceeds with analysing how they are associated with market characteristics, then reports some measures for the welfare costs of illiquidity and finally reports results from some counterfactual experiments.

5.1 Parameters

The housing market of the entire country is divided into 66 distinct but interconnected segments as described in Section 2.1. For each of these segments we are able to identify the structural parameters of interest $(x_m, \sigma_m, \alpha_m)$ as well as the equilibrium meeting probabilities (δ_m, λ_m) as nonlinear transforms of the original data, as summarized in Section 4. In our baseline calibration, we have set $\delta_0 = 0.5$ for one segment, implying $\overline{s} = 54,429$.

Table 1 summarizes our baseline results for the segment-specific parameters. The common component of the housing dividend, x_m, which is reported on a monetary scale and should be interpreted as the flow utility from housing as euros in a week, takes an average value of approximately 200 euros. This sounds to us like a reasonable number, as it would be consistent with a monthly housing utility in the order of magnitude of some 800 euros.[6] The highest values of x are unsurprisingly found for large apartments in the capital city Helsinki, where the quantity corresponds to a monthly value of housing services of approximately 2000 euros. The parameter σ, which summarizes the dispersion of the idiosyncratic component in housing dividends, is quite small relative to x, with an average value of 1.33. Small values of σ are consistent with the idea that households do agree on a large share of the value of housing characteristics, and preference heterogeneity accounts for only a small proportion of flow utilities.

Both meeting probabilities λ (e.g. for sellers) and δ (for searchers) are mostly quite small, leading to relatively large trading probabilities conditional on meeting (μ). These endogenous trading probabilities differ across markets - in places where the meeting probabilities are high (mostly small apartments in large towns), the probabilities of transaction conditional on meeting are low (there is a negative correlation between μ and δ of -0.9). The estimates for α, the matching technology parameter, are not very heterogenous by segment in levels. However, even small differences in α can generate important differences in meeting probabilities (for example, the exercise in Figure 2d illustrates how sensitive seller time-on-market can be with

[6]Recall that this number should be interpreted as the utility from living in owner-occupied housing, when living in the rental sector is normalized to give zero net utility to households - if renting and owning are not very different in terms of utility, we would expect to find quantities in the same order of magnitude as rents.

	Mean	p25	p50	p75
x	203.0	144.9	196.3	238.1
σ	1.33	0.98	1.12	1.53
α	0.28	0.24	0.27	0.28
δ	0.08	0.05	0.06	0.09
λ	0.09	0.07	0.08	0.10
Observations	66			

Table 1: Summary statistics for the parameters obtained as nonlinear transforms of the original data.

Notes. x, σ and α are model primitives; δ and λ are equilibrium objects, as they also depend on the equilibrium value of market tightness θ.

respect to α). As we will document below, despite α having a relatively small range, it plays a significant role in driving differences in liquidity.

We report parameter estimates implied by alternative calibrations of δ_0 in Appendix Table A3. While, of course, the exact numbers are sensitive to the calibration of δ_0 and the implied value of $\overline{s}$, we note that qualitatively our results are surprisingly little affected by the calibration. Our baseline calibration had fixed $\delta_0 = 0.5$ for a given location, but when we consider alternatives ranging from $\delta_0 = 0.3$ to $\delta_0 = 0.7$, the distributions of the parameter estimates move very little. For example, the mean values of the parameters are at most modestly affected by the different possible calibrations. To verify that our results are also robust to using data on transaction prices and times (as opposed to listing times and prices), Appendix Table A4 reports parameter results when the inversion is implemented using transaction prices and times. The results are overall similar to our baseline findings.

5.2 Association with market characteristics

Next we use linear regressions to study how our structural parameter estimates are associated with different market-level characteristics: population, income and population growth. These variables are chosen to follow Genesove & Han (2012) who study the associations of these variables with buyer and seller times on market as well as market tightness. They show, for example, that higher population at the local market level is associated with shorter seller sale times. We are interested interpreting these findings through the lens of our model: *through which structural parameter* would higher population affect liquidity. For example, locations with higher population could have better liquidity because they have an efficient meeting technology (high α) or high housing quality (high x). However, instead of giving these regressions a structural interpretation, we view this as exercise as a means to understand in a reduced-form sense what kind of variation the structural parameters are capturing.

To have an interpretation for the relative magnitudes of the coefficients, we standardise both dependent and independent variables to have a mean zero and a variance of 1 in the sample of market segments. Since we defined a segment by the size of the unit and by the geography, but the market characteristics are observed only at the geography level, we include an indicator that takes value 1 if the observation reflects the small apartments in the location.

Tables 2 and 3 describe these regressions for the housing quality parameter x as well as for the meeting technology parameter α, which illustrates how easy it is for sellers to meet potential buyers and vice versa. For the remaining parameters (σ_m as well as the meeting probabilities δ_m and λ_m), similar regression results are displayed in Appendix E.2. For transparency, Appendix Figure A2 also plots some scatterplots of the raw data to illustrate that the significance commented in the text is also visible in the raw data.

As one can expect, in Table 2, a larger value of α (associated with faster meetings) is positively correlated with the average municipality size in the segment. This is possibly related to thick-market effects, which are not explicitly modeled. Furthermore, although in the third column, population growth (average municipality income) is positively (negatively) associated with α, the coefficients are not significant, which is consistent with α indeed capturing mainly factors which affect the matching efficiency. Across the different columns, the coefficient for ”small” is positive, indicating that meeting buyers and sellers of small apartment is generally speaking easier than meeting buyers and sellers of large apartments. This is consistent with smaller matching frictions for smaller units.

On the other hand, market size, population growth and average incomes are all associated with higher housing quality parameter x, as illustrated in table 3. In particular, higher average

	(1)	(2)	(3)	(4)
	α	α	α	α
Municipality size	0.403			0.244
	(0.137)			(0.0980)
Population growth		0.426		0.355
		(0.139)		(0.200)
Household income			0.0646	-0.211
			(0.0788)	(0.154)
small	0.889	0.889	0.889	0.889
	(0.200)	(0.197)	(0.223)	(0.191)
Observations	66	66	66	66

Table 2: Associations of the meeting technology parameter with market segment characteristics.

Notes. The table documents coefficients from an unweighted linear regression of the outcome on indicated dependent variables. Municipality size refers to average municipality population in the market segment, apart from single-municipality market segments where it is the actual population in 2018. Population growth refers to total region population change from 2016 to 2018 relative to 2016 population. Household income refers to average household disposable income in the region in 2018. Both the independent and the three continuous dependent variables are standardised to have mean 0 and variance of 1. "small" is an indicator variable which takes value 1 for the market segments of apartments of two rooms or less. Heteroscedasticity-robust standard errors are in parentheses.

incomes are associated with higher x. This is consistent with the idea that the value of housing services is higher in locations with higher incomes (where, often, also rents would be higher, for example). The value of housing services is also larger on average in large apartments, as we would expect, as indicated by the negative coefficient of "small".

	(1) x	(2) x	(3) x	(4) x
Municipality size	0.639			0.427
	(0.0852)			(0.111)
Population growth		0.564		0.132
		(0.0876)		(0.0700)
Household income			0.523	0.275
			(0.104)	(0.0716)
small	-1.300	-1.300	-1.300	-1.300
	(0.101)	(0.126)	(0.136)	(0.0720)
Observations	66	66	66	66

Table 3: Associations of the housing quality parameter with some market segment characteristics.

Notes. Notes. The table documents coefficients from an unweighted linear regression of the outcome on indicated dependent variables. Municipality size refers to average municipality population in the market segment, apart from single-municipality market segments where it is the actual population in 2018. Population growth refers to total region population change from 2016 to 2018 relative to 2016 population. Household income refers to average household disposable income in the region in 2018. Both the independent and the three continuous dependent variables are standardised to have mean 0 and variance of 1. "small" is an indicator variable which takes value 1 for the market segments of apartments of two rooms or less. Heteroscedasticity-robust standard errors are in parentheses.

5.3 Costs of illiquidity

The model equips us with different ways to measure the costs of illiquidity. The first measure follows Krainer (2001): If housing is a very liquid asset, then succesfully selling a unit should not imply a significant change to the seller's welfare relative to the value of having a unit to sell. This implies that on a very liquid market, the value of having a unit for sale, q_m, should be close to the sales price, p_m. Our first measure of illiquidity is therefore

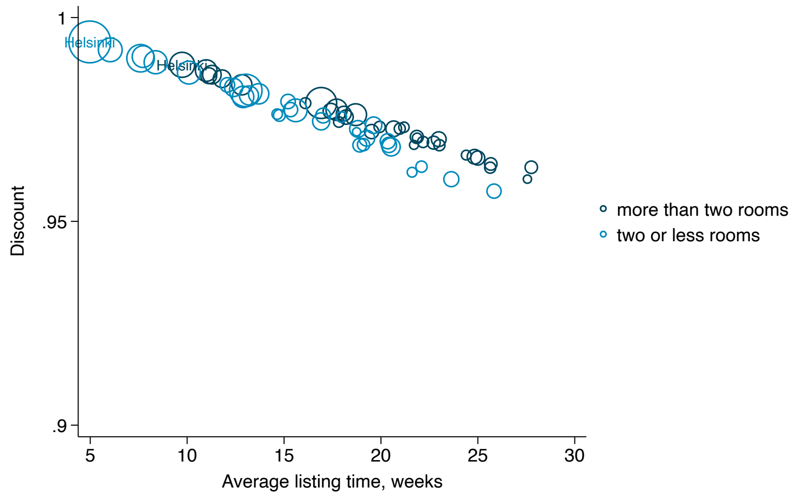
$$\text{Discount 1} = \frac{q_m}{p_m}. \tag{24}$$

The first measure does not take into account the fact that housing illiquidity affects welfare not only through higher costs of selling but also through a costly search for a new unit. To account for this, we also consider the disutility when a match is severed. Again, on a very liquid market, both selling the current home and finding a new one should happen easily, and so the value one gets when becoming mismatched, $q_m + \bar{s}$, should not be too far from the net present value of holding the current match forever. To measure the value of holding the current match forever, we consider the match value of someone with the lowest possible equilibrium housing dividend, given by $x_m + \sigma_m \tilde{\varepsilon}_m$ (implying that the discount is a lower bound for real costs). The second measure for the costs of illiquidity is

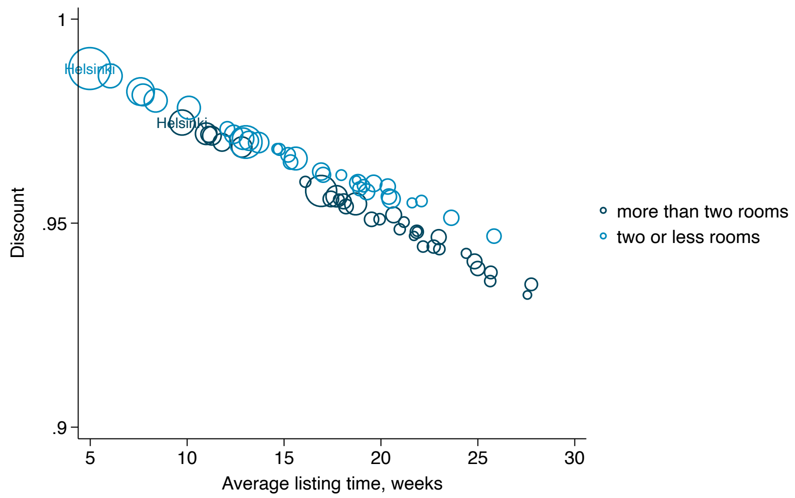
$$\text{Discount 2} = \frac{q_m + \bar{s}}{\frac{1}{1-\beta}\left\{x_m + \sigma_m \tilde{\varepsilon}_m\right\}}. \tag{25}$$

Figure 3 summarizes the two different measures of illiquidity in every segment as a function of the average listing times. The complement (to 1) of Discount 1, which measures the gains from a succesful transaction, ranges from 0.5% to 4.3% across local market segments. The complement (to 1) of Discount 2, which measures the disutility of becoming mismatched relative to the value of holding current match forever, ranges from 1.2% to 6.8%. Both are, almost by construction, strongly associated with sale times.[7] These are, at least for the less liquid locations, substantial welfare costs.

[7] Jiang et al. (2024) note that sale times alone can be a weak measure of liquidity if sales times are short because holding costs (c_m) are high: in this case, transacting happens quickly not because transacting is particularly easy but because not transacting is very costly. However, since transacting costs are modest and fairly homogenous across segments, it seems that in our context, using time-on-market as a proxy for illiquidity is relatively inconsequential.



(a) Discount 1.



(b) Discount 2.

Figure 3: Illiquidity discounts from the model and listing times in the data.

Notes. The illiquidity discount measures are given by equations 24 and 25. Each symbol represents a different local market. Circle size indicates the model-consistent number of houses in each segment.

5.4 Counterfactual experiments

Counterfactual I The aim of our first counterfactual exercise is to understand which characteristics of markets are the most important drivers of the heterogeneity in liquidity. To do so, we proceed by setting different parameters (x, c, σ, α) to estimated medians, one at a time, and after that recomputing the equilibrium and measuring the heterogeneity in sale times as a measure of illiquidity.

The resulting variation in liquidity is summarized in Table 4. The first line indicates the distribution of sale times in the data (these are exactly matched in the baseline estimation). The last line indicates equilibrium sale times when all parameters are constants across markets - by construction, there is no heterogeneity in sale times left in this scenario. Line 2 shows the case where c is set to the same value in all markets. Since heterogeneity in c contributes very little to differences in liquidity, on each of the following lines, c is held at median and, and the remaining parameters (x, σ, α) are varied one at a time. Lines 3-8 indicate these intermediate cases where we shut down heterogeneity in two or three parameters at a time.

The first observation from Table 4 is that on rows 3-5, substantial heterogeneity in liquidity remains when the heterogeneity in the different parameters (x, σ, α) is shut down one-by-one. This indicates that none of the three parameters *alone* drive the variation in liquidity across markets. The variation in α is highly impactful, even though the parameters were not particularly heterogeneous to begin with. Zooming in on rows 6 and 9 we see that after shutting down heterogeneity in both x and in α, only little heterogeneity remains in expected sale times. We interpret this as evidence that in our calibration of the model, housing quality (x) and matching technology efficiency (α) are together the most important drivers of liquidity.

Counterfactual II In a second counterfactual, we highlight the importance of accounting for spillovers between market segments. To illustrate the linkages between segments, we consider what would happen if *some segments,* but not all, were to face a demand shock. We consider an increase in the demand for apartments in urban areas, measured by a 10 % increase in x, absent changes in the valuation of housing services in rural areas. This could reflect for example a shift in preferences towards consumption amenities provided by urban markets. We consider urban areas as the 15 large cities which form their own markets, as listed in Appendix A.

Table 5 reports changes in equilibrium outcomes from our second counterfactual experiment, relative to the baseline equilibrium (data). As we would expect, prices increase and sale times decrease in urban areas following the demand shock. Across different urban areas, the equilibrium changes are quite heterogenous: Prices are not inceasing by 10% in all urban segments, but instead they increase by 11-17% depending on the segment. Simultaneously some of the demand shock translates to improved liquidity in urban areas. These changes illustrate how endogenous buyer sorting can reinforce the intial effect of the quality change. The spillovers on rural markets, where there were no exogenous parameter changes, indicate how they became less attractive in relative terms. Rural prices adjust to the new equilibrium by declining as θ decreases. Thus, lower relative demand in rural areas is reflected in lower prices and worse liquidity, although the changes are modest in size. We interpret this as rural sellers suffering but mainly relative to their urban counterparts.

	E(TOM)				
Scenario	p25	p50	p75	Mean	SD
Baseline (data)	13.3	18.4	21.7	17.7	5.4
Set c to median	13.0	18.2	21.2	17.5	5.5
Set x, c to medians	13.1	16.0	20.1	16.5	5.2
Set α, c to medians	15.4	16.9	19.6	17.3	3.2
Set σ, c to medians	12.2	16.5	19.7	16.0	5.1
Set x, α, c to medians	14.9	16.2	17.0	16.2	1.4
Set x, σ, c to medians	12.3	14.4	19.5	15.2	5.2
Set α, σ, c to medians	14.5	15.4	17.1	15.7	1.8
Set x, α, σ, c to medians	14.8	14.8	14.8	14.8	0.0

Table 4: Counterfactual I: Dispersion in E(TOM) across local markets in different scenarios after shutting down parameter heterogeneity across markets.

Notes. In counterfactual I, parameters are set to medians, as indicated one or multiple at a time, and the equilibrium is re-computed. The relevant medians are reported in Table 1 and Appendix Table A2. The table documents the resulting variation in $E(TOM)$ across market segments.

	Urban market segments			Rural market segments		
	mean	min	max	mean	min	max
Price change, %	13.4	11.1	17.3	-0.9	-1.7	-0.4
Sales time change, %	-2.9	-6.8	-1.0	1.4	1.1	1.5
θ change, %	3.6	1.4	7.6	-2.7	-2.8	-2.4

Table 5: Counterfactual II: Changes in the counterfactual equilibrium compared to the baseline.

Notes. In counterfactual II, urban markets experience a 10% increase in mean housing quality x. Rural markets don't experience changes in structural parameters. The table documents resulting changes in equilibrium outcomes relative to the baseline equilibrium.

6 Conclusions

In this paper, we develop a tractable model of housing search and liquidity in the cross-section of interconnected market segments. The model is tailored to match key empirical observations related to the variation of housing liquidity in the cross-section of market segments. Housing market liquidity can be high if housing quality is high (so that the opportunity cost of not being matched to a house is high), if there is little idiosyncratic variation in preferences for housing (so that most buyers do not want to search for a long time before purchasing), if the matching technology is efficient (meeting probabilities are high), or if the market segment is endogenously popular (so that there are relatively many potential buyers). Different characteristics of markets affect the local equilibrium directly, by affecting the behavior of agents in a given segment, but also indirectly by governing where households choose to search and changing local market tightness.

The model gives us a structured way of interpreting the outcome data on prices, sale times and tightness, since we can partially identify the parameters of the model as nonlinear transforms of observable data. In our empirical application, we study the housing market in Finland. We show that the mean value of housing services and the efficiency of matching are key parameters governing the variation in liquidity across markets. Markets where housing services are valuable are also markets where incomes are high, and markets where matching is efficient are ones where population is high. These findings allow us to re-interpret some of the earlier findings on housing liquidity. We then show that in our context, the heterogeneity in the welfare costs of housing illiquidity are substantial and are in the order of magnitude of multiple percentage points of sale prices.

One limitation of our setup is that we build on a setup where households are fully mobile across locations. Thus, the framework can be seen as suitable for the analysis of long-run steady states, and is limited in understanding the role of different frictions such as for example the role of moving costs in the propagation of shocks. Recent work from Kaas et al. (2024) makes important steps towards thinking about nonstationary dynamics in spatial search. Finally, our setup improves on the earlier work on search in the cross-section of market segments by endogenizing households' choice of where to search, but this comes with some important limitations. For example, as highlighted in the work of Piazzesi et al. (2020), in reality some households may search across multiple segments, and considering too little or too much spatial aggregation is not empirically inconsequential.

References

Amaral, F., Toth, M., & Zdrzale, J. (2024). Urban spatial distribution of housing liquidity. *Unpublished manuscript*.

Anenberg, E., & Bayer, P. (2020). Endogenous sources of volatility in housing markets: The joint buyer–seller problem. *International Economic Review*, *61*(3), 1195–1228.

Badarinza, C., Ramadorai, T., Siljander, J., & Tripathy, J. (2024). Behavioral lock-in: Aggregate implications of reference dependence in the housing market. *Bank of England Working Paper*.

Bruneel-Zupanc, C., Chapelle, G., Eyméoud, J.-B., & Wasmer, E. (2022). Housing prices propagation: A theory of spatial interactions. *Unpublished manuscript*.

Carrillo, P. E. (2012). An empirical stationary equilibrium search model of the housing market. *International Economic Review*, *53*(1), 203–234.

Carrillo, P. E., & Pope, J. C. (2012). Are homes hot or cold potatoes? The distribution of marketing time in the housing market. *Regional Science and Urban Economics*, *42*(1-2), 189–197.

Carrillo, P. E., & Williams, B. (2019). The repeat time-on-the-market index. *Journal of Urban Economics*, *112*, 33–49.

Diaz, A., & Jerez, B. (2013). House prices, sales, and time on the market: A search-theoretic framework. *International Economic Review*, *54*(3), 837–872.

Díaz, A., Jerez, B., & Rincón-Zapatero, J. P. (2024). Housing prices and credit constraints in competitive search. *The Economic Journal*, *134*(657), 220–270.

Eerola, E., & Määttänen, N. (2018). Borrowing constraints and housing market liquidity. *Review of Economic Dynamics*, *27*, 184 - 204.

Famiglietti, M., Garriga, C., & Hedlund, A. (2020). The geography of housing market liquidity during the great recession. *Federal Reserve Bank of St. Louis Review*, *102*(1), 51–77.

Gargano, A., Giacoletti, M., & Jarnecic, E. (2023). Local experiences, search, and spillovers in the housing market. *The Journal of Finance*, *78*(2), 1015–1053.

Garriga, C., & Hedlund, A. (2020). Mortgage debt, consumption, and illiquid housing markets in the great recession. *American Economic Review*, *110*(6), 1603–1634.

Genesove, D., & Han, L. (2012). Search and matching in the housing market. *Journal of Urban Economics*, *72*(1), 31–45.

Giacoletti, M. (2021). Idiosyncratic risk in housing markets. *The Review of Financial Studies*, *34*(8), 3695–3741.

Han, L., & Strange, W. C. (2015). The microstructure of housing markets: Search, bargaining, and brokerage. In *Handbook of regional and urban economics* (Vol. 5, pp. 813–886). Elsevier.

Head, A., & Lloyd-Ellis, H. (2012). Housing liquidity, mobility, and the labour market. *Review of Economic Studies*, *79*(4), 1559–1589.

Head, A., Lloyd-Ellis, H., & Sun, H. (2014, April). Search, liquidity, and the dynamics of house prices and construction. *American Economic Review*, *104*(4), 1172-1210.

Jiang, E. X., Kotova, N., & Zhang, A. L. (2024). Liquidity in residential real estate markets. *Unpublished manuscript*.

Kaas, L., Kocharkov, G., & Syrichas, N. (2024). Understanding spatial house price dynamics in a housing boom. *Unpublished manuscript*.

Kaplan, G., & Violante, G. L. (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica*, *82*(4), 1199–1239.

Karahan, F., & Rhee, S. (2019). Geographic reallocation and unemployment during the great recession: The role of the housing bust. *Journal of Economic Dynamics and Control*, *100*, 47–69.

Krainer, J. (2001). A theory of liquidity in residential real estate markets. *Journal of Urban Economics*, *49*(1), 32 - 53.

Moen, E. R., Nenov, P. T., & Sniekers, F. (2021). Buying first or selling first in housing markets. *Journal of the European Economic Association*, *19*(1), 38–81.

Ngai, R., & Sheedy, K. (2024). The ins and outs of selling houses: Understanding housing market volatility. *International Economic Review*.

Ngai, R., & Tenreyro, S. (2014). Hot and cold seasons in the housing market. *American Economic Review*, *104*(12), 3991–4026.

Novy-Marx, R. (2009). Hot and cold markets. *Real Estate Economics*, *37*(1), 1–22.

Piazzesi, M., Schneider, M., & Stroebel, J. (2020). Segmented housing search. *American Economic Review*, *110*(3), 720–759.

Sagi, J. S. (2021). Asset-level risk and return in real estate investments. *The Review of Financial Studies*, *34*(8), 3647–3694.

Smith, E. (2020). High and low activity spells in housing markets. *Review of Economic Dynamics*, *36*, 1–28.

Williams, J. (2018). Housing markets with endogenous search: Theory and implications. *Journal of Urban Economics*, *105*, 107–120.

Appendices

A Data appendix

A.1 Listing and transaction data

Etuovi.com have provided us with microdata on the listings which are published between January 2017 and May 2019. Our data is drawn from the database on 16 June 2019. To avoid issues related to flow and stock sampling, we restrict the sample to the subset of the listings data that have been published in January 2017 - December 2018. For the houses that had been listed before 31 December 2018 but not unlisted before June 2019, we do not observe the actual listing time, only a lower bound of it. However, as this is only a small proportion of our sample, we abstract away from the related censoring issue, and use the observed lower bound as an estimate for the sale times of the censored units.

The transaction dataset is provided to us by The Finnish Federation of Real Estate Agency (KVKL). While we observe transactions intermediated by the member agencies over a longer time period than for the listings data, we focus on transactions that took place in 2017 or 2018 to be consistent with the listings data. Although the first dataset contains information on listings (which might be different from eventual transactions) and the second one on transactions, both datasets indicate qualitatively similar findings about the cross-sectional variation in housing market liquidity (at the segment level, the correlation of prices and sale times in both datasets is above 0.9).

For both datasets, we exclude new apartments, as the sales strategies of real estate developers could differ from those of households. We also remove outliers in terms of the price, apartment size, and sale time, and observations with missing price or sale time information. Throughout, we exclude Aland islands from the analysis. To measure prices, we consider the price net of any housing cooperative debt whenever this information is available.

Table A1 provides summary statistics for the microdata from Etuovi.com and KVKL. The listing dataset contains more than 229 000 observations and the transaction dataset more than 117 000 observations.

	Listing data			Transaction data		
	mean	p50	sd	mean	p50	sd
Time on the market (days)	111	77	121	92	57	109
Price (1000 euros)	189	154	150	178	150	128
N	229462			117206		

Table A1: Summary statistics for listing and transaction microdata.

Notes. Listing data from Etuovi.com and transaction data from KVKL. In the listings data, the price refers to the listing price and the time on the market to the time for which the listing was online. In the transaction data, the price refers to the final transaction price and the time on the market refers to the time between the sales start date and the transaction date. For both datasets, prices are indicated net of any cooperative debt when applicable. p50 refers to the median.

References:

Listing data: Etuovi.com, Alma Mediapartners Oy

Transaction data: KVKL Hintaseurantapalvelu, www.hintaseurantapalvelu.fi, Kiinteistönvälitysalan Keskusliitto Ry

A.2 Summary statistics of market segment-level data

The main empirical analysis is conducted in a cross-section of 66 market segments, where the segmentation is based on geography and apartment size. For the geographic divisions, wee treat the 15 largest cities as regions on their own. These cities are Helsinki, Espoo, Tampere, Vantaa, Oulu, Turku, Jyväskylä, Kuopio, Lahti, Pori, Kouvola, Joensuu, Lappeenranta, Hämeenlinna and Vaasa. We divide the rest of Finland to 18 groups using administrative regions (*maakunnat*) in the 2018 classification (we exclude Aland islands from the analysis). Figure A1 summarizes the geographic divisions. Further, within each geographic unit, small and large apartments are treated separately.

Table A2 summarizes listing data for each apartment size category separately. As illustrated in section 2, sale times tend to be longer for larger apartments. Market tightness, that is the number of buyers divided by the number of sellers, in particular, has an interquartile range of .5 to 1.1 in small apartments and .6 to 1.2 in large apartments.

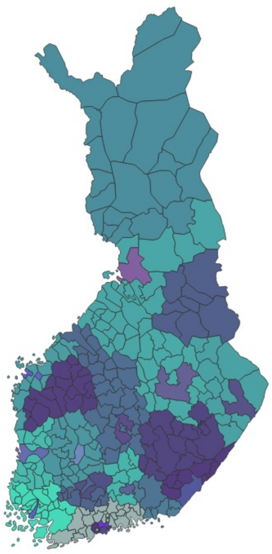


Figure A1: Geographic segmentation of markets.

Notes. Different colors indicate different local markets, and don't have an ordinal interpretation.

	Small apartments					Large apartments				
	min	p25	p50	p75	max	min	p25	p50	p75	max
Mean price (1000 euros)	52	68	86	105	243	129	144	167	195	434
Mean time (days)	35	90	109	135	181	68	122	145	161	194
Mean emails	75	94	127	196	442	93	110	127	205	396
Mean maint. cost (euros/month)	135	160	167	177	200	199	230	249	269	349
Tightness	0.4	0.5	0.7	1.1	2.5	0.5	0.6	0.7	1.2	2.3
N	33					33				

Table A2: Summary statistics for market-segment-level averages in the data.

Notes. Listing data from Etuovi.com for all other variables and from KVKL for maintenance costs. p50 refers to the median, and p25 and p75 to the 25th and 75th percentiles. Mean emails refers to the number of email alerts sent per listing in the first week after the listing was posted. Tightness refers to our measure of market tightness, see section 2.2.

A.3 Other data sources

Municipality characteristics We use data on regional characteristics in the exercise where we compare segment-level parameters to observable segment-level characteristics. The information on population and population growth is obtained from Statistics Finland Municipal Key Figures. The information on municipality-level incomes is obtained via the Statistics Finland Income Distribution statistics (in part administrative, in part survey data).

References:

Official Statistics of Finland: 118w – Number, income and income structure of household-dwelling units by municipality, 1995-2021, referred 3.3.2023.

Statistics Finland, Municipal Key Figures 1987-2018 with 2019 regional classifications, referred 3.3.2023.

Maps The shapefile for the map of Finland is obtained from Statistics Finland.

References:

Municipality area boundaries, Statistics Finland, obtained 17.6.2022. The material can also be downloaded from Statistics Finland's interface service with the licence CC BY 4.0.

https://tilastokeskuskartta.swgis.fi/#

B Proofs

B.1 Model Solution (Proof of Lemma 3.1)

B.1.1 Part 1: Equation 15

To simplify expression, in this Appendix section we denote $\delta(\theta_m; \alpha_m)$ simply by δ_m and respectively for λ_m. However δ_m and λ_m are functions of θ_m throughout.

Price setting Start from the sellers' price setting equation, 6. When setting the price, seller considers only the effect of p on $\mu(p)$, taking as given the value of selling, q_m. Denoting by p the decision variable of the seller and by p^*_m the price which solves

$$p^*_m = \arg\max_p \mu(p)p + (1 - \mu(p))q_m \tag{26}$$

The necessary FOC gives

$$\frac{\partial \mu(p)}{\partial p} \cdot p\bigg|_{p=p^*_m} + \mu(p)\bigg|_{p=p^*_m} - q_m \cdot \frac{\partial \mu(p)}{\partial p}\bigg|_{p=p^*_m} = 0 \tag{27}$$

implying

$$p^*_m = q_m - \frac{\mu(p^*_m)}{\frac{\partial \mu}{\partial p}(p^*_m)} \tag{28}$$

Denote the equilibrium price which will be charged by all sellers in m by $p_m = p^*_m$.

Value of selling The equilibrium value of selling is given by equation 5,

$$q_m = c_m + \beta\big\{(1 - \delta_m)q_m + \delta_m \max_p \big[\mu(p)p + (1 - \mu(p))q_m)\big]\big\} \tag{29}$$

Substituting in $p_m = p_m^*$, we get

$$q_m = c_m + \beta\big\{(1-\delta_m)q_m + \delta_m\big[\mu(p_m)p_m + (1-\mu(p_m))q_m)\big]\big\} \tag{30}$$

$$= c_m + \beta\big\{q_m - \delta_m q_m + \delta_m\mu(p_m)p_m + \delta_m q_m - \delta_m\mu(p_m)q_m\big\} \tag{31}$$

$$= c_m + \beta\big\{q_m + \delta_m\mu(p_m)p_m - \delta_m\mu(p_m)q_m\big\} \tag{32}$$

$$= c_m + \beta(1-\delta_m\mu(p_m))q_m + \beta\delta_m\mu(p_m)p_m \tag{33}$$

$$\big[1-\beta+\beta\delta_m\mu(p_m)\big]q_m = c_m + \beta\delta_m\mu(p_m)p_m \tag{34}$$

$$q_m = \frac{1}{1-\beta+\beta\delta_m\mu(p_m)}\big[c_m + \beta\delta_m\mu(p_m)p_m\big] \tag{35}$$

Equating this with $q_m = p_m + \frac{\mu(p_m)}{\frac{\partial\mu}{\partial p}(p_m)}$ from the FOC, we find

$$p_m + \frac{\mu(p_m)}{\frac{\partial\mu}{\partial p}(p_m)} = \frac{1}{1-\beta+\beta\delta_m\mu(p_m)}\big[c_m + \beta\delta_m\mu(p_m)p_m\big] \tag{36}$$

$$\big[1-\beta+\beta\delta_m\mu(p_m)\big]\big[p_m + \frac{\mu(p_m)}{\frac{\partial\mu}{\partial p}(p_m)}\big] = c_m + \beta\delta_m\mu(p_m)p_m \tag{37}$$

$$(1-\beta)p_m + \big[1-\beta+\beta\delta_m\mu(p_m)\big]\frac{\mu(p_m)}{\frac{\partial\mu}{\partial p}(p_m)} = c_m \tag{38}$$

$$p_m = \frac{1}{1-\beta}c_m - \frac{1-\beta+\beta\delta_m\mu(p_m)}{1-\beta}\frac{\mu(p_m)}{\frac{\partial\mu}{\partial p}(p_m)} \tag{39}$$

We will shortly substitute for μ and $\frac{\partial\mu}{\partial p}$.

Match values The value of a match v given an idiosyncratic dividend ε can be rewritten as

$$v(\varepsilon) = x_m + \sigma_m\varepsilon + \beta\big[\pi\, v(\varepsilon) + (1-\pi)\,\big(q_m + \overline{s}\big)\big] \tag{40}$$

$$(1-\beta\pi)\, v(\varepsilon_i) = x_m + \sigma_m\varepsilon_i + \beta(1-\pi)\,\big(q_m + \overline{s}\big) \tag{41}$$

$$v(\varepsilon) = \frac{1}{1-\beta\pi}\big[x_m + \sigma_m\varepsilon + \beta(1-\pi)\,\big(q_m + \overline{s}\big)\big] \tag{42}$$

Buyer indifference The buyer decides on a threshold $\tilde{\varepsilon}$ such that they will purchase the unit if $\varepsilon_i \geq \tilde{\varepsilon}_m$. When deciding on the threshold, they only consider the effect that the threshold has on their utility, taking as given the values of selling q_m and searching $\overline{s}$. The threshold value of the dividend such that the household is indifferent between searching in the next period or purchasing in this period is, given that the seller is charging price p,

$$v(\tilde{\varepsilon}) - p = \overline{s} \tag{43}$$

$$\frac{1}{1-\beta\pi}\left[x_m + \sigma_m\tilde{\varepsilon} + \beta(1-\pi)\left(q_m + \overline{s}\right)\right] - p = \overline{s} \tag{44}$$

$$x_m + \sigma_m\tilde{\varepsilon} + \beta(1-\pi)\left(q_m + \overline{s}\right) = (1-\beta\pi)[\overline{s} + p] \tag{45}$$

$$\sigma_m\tilde{\varepsilon} = (1-\beta\pi)[\overline{s} + p] - x_m - \beta(1-\pi)\left(q_m + \overline{s}\right) \tag{46}$$

$$\sigma_m\tilde{\varepsilon} = (1-\beta\pi)p - x_m - \beta(1-\pi)q_m + \left[(1-\beta\pi)\overline{s} - \beta(1-\pi)\overline{s}\right] \tag{47}$$

$$\sigma_m\tilde{\varepsilon} = (1-\beta\pi)p - x_m - \beta(1-\pi)q_m + (1-\beta)\,\overline{s} \tag{48}$$

Probability of selling From the seller's price setting view, they consider the effect that the price they set, p, has on their probability of transacting, $\mu(p)$, taking as given the values q_m and $\overline{s}$:

$$\mu(p) = \mathbb{P}(\sigma_m\varepsilon > \sigma_m\tilde{\varepsilon}(p)) = \mathbb{P}(\varepsilon > \tilde{\varepsilon}(p)) = 1 - F(\tilde{\varepsilon}(p)) \tag{49}$$

where

$$\tilde{\varepsilon}(p) = \frac{1-\beta\pi}{\sigma_m}p - \frac{1}{\sigma_m}x_m - \frac{(1-\pi)\beta}{\sigma_m}q_m + \frac{1-\beta}{\sigma_m}\overline{s} \tag{50}$$

So

$$\frac{\partial\tilde{\varepsilon}(p)}{\partial p} = \frac{1-\beta\pi}{\sigma_m} \tag{51}$$

Since $\varepsilon \sim N(0,1)$,

$$\frac{\partial\mu(p)}{\partial p} = -f(\tilde{\varepsilon}(p)) * \frac{\partial\tilde{\varepsilon}}{\partial p} = -f(\tilde{\varepsilon}(p))\frac{1-\beta\pi}{\sigma_m} \tag{52}$$

Prices Substituting $\mu(p_m) = 1 - F(\tilde{\varepsilon}_m)$ and $\frac{\partial\mu(p)}{\partial p} = -\frac{1-\beta\pi}{\sigma_m}f(\tilde{\varepsilon}_m)$ into the pricing equation, we find the equilibrium value of p, which is the price that all sellers in m will set,

$$p_m = \frac{1}{1-\beta}c_m - \frac{1-\beta+\beta\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\beta}\frac{1-F(\tilde{\varepsilon}_m)}{-\frac{1-\beta\pi}{\sigma_m}f(\tilde{\varepsilon}_m)} \tag{53}$$

$$= \frac{1}{1-\beta}c_m + \sigma_m\frac{1}{(1-\beta)(1-\beta\pi)}\big(1-\beta+\beta\delta_m(1-F(\tilde{\varepsilon}_m))\big)\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} \tag{54}$$

$$= \frac{1}{1-\beta}c_m + \frac{\sigma_m}{(1-\beta\pi)}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \frac{\sigma_m\beta\delta_m}{(1-\beta)(1-\beta\pi)}\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)} \tag{55}$$

and using p_m to find q_m,

$$q_m = p_m + \frac{\mu(p_m)}{\frac{\partial\mu}{\partial p}(p_m)} \tag{56}$$

$$= p_m + \frac{1-F(\tilde{\varepsilon}_m)}{-\frac{1-\beta\pi}{\sigma_m}f(\tilde{\varepsilon}_m)} \tag{57}$$

$$= p_m - \frac{\sigma_m}{1-\beta\pi}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} \tag{58}$$

$$= \frac{1}{1-\beta}c_m + \frac{\sigma_m\beta\delta_m}{(1-\beta)(1-\beta\pi)}\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)} \tag{59}$$

Value of searching From the definition of the value of search in segment m, using that in equilibrium, s_n=s_m=$\overline{s}$ $\forall\ n, m$, we can write

$$s_m = u_m + \beta\left(\lambda_m E\max(v_m(\varepsilon_i) - p_m,\ s_m) + (1-\lambda_m)\ s_m\right), \tag{60}$$

$$= u_m + \beta\lambda_m E\max(v_m(\varepsilon_i) - p_m - s_m, 0) + \beta s_m \tag{61}$$

$$= \frac{1}{1-\beta}\big[u_m + \beta\lambda_m E\max(v_m(\varepsilon_i) - p_m - s_m, 0)\big] \tag{62}$$

$$\overline{s} = \frac{1}{1-\beta}\big[u_m + \beta\lambda_m E\max(v_m(\varepsilon_i) - p_m - \overline{s}, 0)\big] \tag{63}$$

$$\tag{64}$$

Noting that

$$
\begin{aligned}
& v_m(\varepsilon_i) - p_m - \overline{s} && (65)\\
&= \frac{1}{1-\beta\pi}\big[x_m + \sigma_m\varepsilon_i + \beta(1-\pi)\ \big(q_m + \overline{s}\big)\big] - p_m - \overline{s} && (66)\\
&= \frac{1}{1-\beta\pi}\sigma_m\varepsilon_i + \frac{1}{1-\beta\pi}\big[x_m + \beta(1-\pi)\ q_m\big] - p_m - [1 - \frac{1}{1-\beta\pi}\beta(1-\pi)]\overline{s} && (67)\\
&= \frac{1}{1-\beta\pi}\sigma_m\varepsilon_i + \underbrace{\frac{1}{1-\beta\pi}\big[x_m + \beta(1-\pi)\ q_m\big] - p_m - \frac{1-\beta}{1-\beta\pi}\overline{s}}_{\frac{1}{1-\beta\pi}\sigma_m\tilde{\varepsilon}_m} && (68)\\
&= \frac{1}{1-\beta\pi}\sigma_m\big(\varepsilon_i - \tilde{\varepsilon}_m\big) && (69)
\end{aligned}
$$

We can rewrite

$$
\begin{aligned}
\mathbb{E}_\varepsilon \max\Big[v_m(\varepsilon_i) - p_m - \overline{s}, 0\Big] &= \mathbb{E}_\varepsilon \max\Big[\frac{1}{1-\beta\pi}\sigma_m(\varepsilon_i - \tilde{\varepsilon}_m), 0\Big] && (70)\\
&= \frac{1}{1-\beta\pi}\sigma_m\mathbb{E}_\varepsilon\Big((\varepsilon - \tilde{\varepsilon}_m)\mathbb{1}_{\varepsilon\geq\tilde{\varepsilon}_m}\Big) && (71)
\end{aligned}
$$

And so s rewrites

$$
s_m = \frac{1}{1-\beta}\big[u_m + \frac{\beta\lambda_m}{1-\beta\pi}\sigma_m\mathbb{E}_\varepsilon\Big((\varepsilon - \tilde{\varepsilon}_m)\mathbb{1}_{\varepsilon\geq\tilde{\varepsilon}_m}\Big)\big] \quad (72)
$$

and $\mathbb{E}_\varepsilon\Big((\varepsilon - \tilde{\varepsilon})\mathbb{1}_{\varepsilon\geq\tilde{\varepsilon}}\Big) = f(\tilde{\varepsilon}) - \tilde{\varepsilon}\big(1 - F(\tilde{\varepsilon})\big)$ since

$$
\begin{aligned}
\mathbb{E}_\varepsilon\big((\varepsilon - \tilde{\varepsilon})\mathbb{1}_{\varepsilon\geq\tilde{\varepsilon}}\big) &= \mathbb{E}_\varepsilon\big(\varepsilon\ \mathbb{1}_{\varepsilon\geq\tilde{\varepsilon}}\big) - \mathbb{E}_\varepsilon\big(\tilde{\varepsilon}\ \mathbb{1}_{\varepsilon\geq\tilde{\varepsilon}}\big) && (73)\\
&= \mathbb{E}_\varepsilon(\varepsilon|\varepsilon > \tilde{\varepsilon})\mathbb{P}(\mathbb{1}_{\varepsilon\geq\tilde{\varepsilon}}) - \mathbb{E}_\varepsilon(\tilde{\varepsilon})\mathbb{P}(\mathbb{1}_{\varepsilon\geq\tilde{\varepsilon}}) && (74)\\
&= \big(\frac{f(\tilde{\varepsilon})}{1 - F(\tilde{\varepsilon})}\big)(1 - F(\tilde{\varepsilon})) - \tilde{\varepsilon}\big(1 - F(\tilde{\varepsilon})\big) && (75)\\
&= f(\tilde{\varepsilon}) - \tilde{\varepsilon}\big(1 - F(\tilde{\varepsilon})\big) && (76)
\end{aligned}
$$

where the second-to-last line uses that for a truncated standard normal distribution, $\mathbb{E}(x|x > a) = \phi(a)/(1 - \Phi(a))$ where $\phi(\cdot)$ denotes the probability density function and Φ the cumulative density function.

Denote

$$z(\tilde{\varepsilon}_m) = \mathbb{E}_\varepsilon\big((\varepsilon - \tilde{\varepsilon}_m)\mathbb{1}_{\varepsilon \geq \tilde{\varepsilon}_m}\big) = f(\tilde{\varepsilon}_m) - \tilde{\varepsilon}_m\big(1 - F(\tilde{\varepsilon}_m)\big) \tag{77}$$

Solution The above equations show that equilibrium values p_m, q_m, s_m and $\mu(p_m)$ can be expressed as functions of the (unknown) equilibrium elements $\tilde{\varepsilon}_m$, $\delta_m(\theta_m)$ and $\lambda_m(\theta_m)$, where δ_m and λ_m are known functions of θ_m.

It remains to find equilibrium values of $\tilde{\varepsilon}_m$ and θ_m, which must satisfy the buyer indifference equation.

Rewriting the buyer indifference equation gives

$$v(\tilde{\varepsilon}_m) - p_m = \overline{s} \tag{78}$$

$$\iff \sigma_m\tilde{\varepsilon}_m = (1-\beta\pi)p_m - x_m - \beta(1-\pi)q_m + (1-\beta)\overline{s} \tag{79}$$

Notice first that

$$(1-\beta\pi)p_m - \beta(1-\pi)q_m = (1-\beta\pi)p_m - \beta(1-\pi)\big(p_m + \frac{\mu(p_m)}{\frac{\partial\mu}{\partial p}(p_m)}\big) \tag{80}$$

$$= (1-\beta)p_m - \beta(1-\pi)\frac{\mu(p_m)}{\frac{\partial\mu}{\partial p}(p_m)} \tag{81}$$

$$= (1-\beta)p_m + \beta(1-\pi)\frac{\sigma_m}{1-\beta\pi}\frac{(1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} \tag{82}$$

Substituting this and the expressions for p_m and $s_m = \overline{s}$:

$$\sigma_m\tilde{\varepsilon}_m = -x_m + (1-\beta)\overline{s} + (1-\beta)p_m + \beta(1-\pi)\frac{\sigma_m}{1-\beta\pi}\frac{(1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} \tag{83}$$

$$= -x_m + \big[u_m + \frac{\beta\lambda_m}{1-\beta\pi}\sigma_m z(\tilde{\varepsilon}_m)\big] \tag{84}$$

$$+ c_m + \frac{\sigma_m(1-\beta)}{(1-\beta\pi)}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \frac{\sigma_m\beta\delta_m}{(1-\beta\pi)}\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)} \tag{85}$$

$$+ \beta(1-\pi)\frac{\sigma_m}{1-\beta\pi}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} \tag{86}$$

$$= -x_m + u_m + c_m + \frac{\beta\sigma_m}{1-\beta\pi}\lambda_m z(\tilde{\varepsilon}_m) + \sigma_m\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \frac{\sigma_m\beta}{(1-\beta\pi)}\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}\delta_m \tag{87}$$

$$\tilde{\varepsilon}_m = \frac{-x_m + \ u_m + c_m}{\sigma_m} + \frac{1 - F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \frac{\beta}{1 - \beta\pi}\Big(\lambda_m(\theta_m) z(\tilde{\varepsilon}_m) + \delta_m(\theta_m)\frac{(1 - F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}\Big) \tag{88}$$

Finally, introducing some additional notation, we can express this as

$$\tilde{\varepsilon}_m = \frac{\beta}{1 - \beta\pi}\Big[\gamma(\tilde{\varepsilon}_m, \theta_m) + \lambda_m(\theta_m) z(\tilde{\varepsilon}_m) + a_m\Big] \tag{89}$$

where

$$z(\tilde{\varepsilon}_m) = \mathbb{E}_\varepsilon\big((\varepsilon - \tilde{\varepsilon}_m)\mathbb{1}_{\varepsilon \geq \tilde{\varepsilon}_m}\big) = f(\tilde{\varepsilon}_m) - \tilde{\varepsilon}_m\big(1 - F(\tilde{\varepsilon}_m)\big) \tag{90}$$

$$\gamma(\tilde{\varepsilon}_m, \theta_m) = \frac{1 - \beta\pi}{\beta}\frac{1 - F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \delta_m(\theta_m)\frac{(1 - F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}, \tag{91}$$

$$a_m = -\frac{1 - \beta\pi}{\beta\sigma_m}(x_m - u_m - c_m). \tag{92}$$

Thus, we have rewritten equations 5, 7, 8, 10 and 13 concisely in a single equation, corresponding to equation 15 in the main text, in $2 \times M$ unknowns, θ_m and $\tilde{\varepsilon}_m$ $\forall\, m.$ If there is a soluton to equation 15, then we can use equations 49 , 53, 56, 72 and to find endogenous values p_m, s_m, q_m as well as μ_m.

Verifying the second-order condition The equilibrium pricing rule must satisfy the necessary FOC

$$\frac{\partial \mu(p)}{\partial p} p\bigg|_{p=p(z_m)} + \mu(p)\bigg|_{p=p(z_m)} - q(x)\frac{\partial \mu(p)}{\partial p}\bigg|_{p=p(z_m)} = 0 \tag{93}$$

$$\iff \tag{94}$$

$$\frac{\partial \mu(p)}{\partial p}\bigg|_{p=p(z_m)} * \Big(p(z_m) - q(x)\Big) + \mu(p)\bigg|_{p=p(z_m)} = 0 \tag{95}$$

The second order condition for the pricing rule to be a maximum writes

$$\frac{\partial^2 \mu(p)}{\partial p^2} p\Big|_{p=p(z_m)} + \frac{\partial \mu(p)}{\partial p}\Big|_{p=p(z_m)} + \frac{\partial \mu(p)}{\partial p}\Big|_{p=p(z_m)} - q(x)\frac{\partial^2 \mu(p)}{\partial p^2}\Big|_{p=p(z_m)} < 0 \tag{96}$$

$$\iff \tag{97}$$

$$\frac{\partial^2 \mu(p)}{\partial p^2}\Big|_{p=p(z_m)} \underbrace{\Big[p(z_m) - q(x)\Big]}_{\text{in equilibrium } -\frac{\mu(p)}{\partial \mu(p)/\partial(p)}|_{p=p(z_m)}} + 2\frac{\partial \mu(p)}{\partial p}\Big|_{p=p(z_m)} < 0 \tag{98}$$

We can use the fact that in equilibrium, the following relations hold:

$$\frac{\partial \tilde{\varepsilon}(p)}{\partial p} = \frac{1-\beta\pi}{\sigma_m} \tag{99}$$

$$\mu(p) = 1 - F(\tilde{\varepsilon}) \tag{100}$$

$$\frac{\partial \mu(p)}{\partial p} = -\frac{1-\beta\pi}{\sigma_m} f_\varepsilon(\tilde{\varepsilon}(p)) \qquad \neq 0 \text{ if } f(\tilde{\varepsilon}(p)) \neq 0 \tag{101}$$

$$\frac{\partial^2 \mu(p)}{\partial p^2} = -\Big(\frac{1-\beta\pi}{\sigma_m}\Big)^2 f'_\varepsilon(\tilde{\varepsilon}(p)) \tag{102}$$

And $\varepsilon \sim \mathcal{N}(0,1)$, we have

$$f(\tilde{\varepsilon}) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{\tilde{\varepsilon}^2}{2}\Big) \tag{103}$$

$$f'(\tilde{\varepsilon}) = \frac{\partial f_\varepsilon(\tilde{\varepsilon})}{\partial \tilde{\varepsilon}} = -\tilde{\varepsilon} f(\tilde{\varepsilon}) = -\tilde{\varepsilon}\frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{\tilde{\varepsilon}^2}{2}\Big) \tag{104}$$

We may therefore replace in the second order condition to find:

$$\frac{\partial^2 \mu(p)}{\partial p^2}\Big|_{p=p(z_m)} \Big(-\frac{\mu(p)}{\partial \mu(p)/\partial(p)}|_{p=p(z_m)}\Big) + 2\frac{\partial \mu(p)}{\partial p}\Big|_{p=p(z_m)} \tag{105}$$

$$= -\Big(\frac{1-\beta\pi}{\sigma_m}\Big)^2 f'(\tilde{\varepsilon}(p))\Big(-\frac{1-F(\tilde{\varepsilon}(p))}{-\frac{1-\beta\pi}{\sigma_m} f_\varepsilon(\tilde{\varepsilon}(p))}\Big) - 2\frac{1-\beta\pi}{\sigma_m} f(\tilde{\varepsilon}(p)) \tag{106}$$

$$= \tilde{\varepsilon}(p)\Big(\frac{1-\beta\pi}{\sigma_m}\Big)^2 f(\tilde{\varepsilon}(p))\Big(\frac{1-F(\tilde{\varepsilon}(p))}{\frac{1-\beta\pi}{\sigma_m} f_\varepsilon(\tilde{\varepsilon}(p))}\Big) - 2\frac{1-\beta\pi}{\sigma_m} f(\tilde{\varepsilon}(p)) \tag{107}$$

$$= \tilde{\varepsilon}(p)\frac{1-\beta\pi}{\sigma_m}\big(1 - F(\tilde{\varepsilon}(p))\big) - 2\frac{1-\beta\pi}{\sigma_m} f(\tilde{\varepsilon}(p)) \tag{108}$$

$$= \frac{1-\beta\pi}{\beta\sigma_m}\big[\tilde{\varepsilon}(p)\big(1 - F(\tilde{\varepsilon}(p))\big) - 2f(\tilde{\varepsilon}(p))\big] \tag{109}$$

And this is less than zero whenever

$$\tilde{\varepsilon}(p)\big(1 - F(\tilde{\varepsilon}(p))\big) - 2f(\tilde{\varepsilon}(p)) < 0 \tag{110}$$

$$\tilde{\varepsilon}(p)\big(1 - F(\tilde{\varepsilon}(p))\big) < 2f(\tilde{\varepsilon}(p)) \tag{111}$$

$$\tilde{\varepsilon}(p) < 2\frac{f(\tilde{\varepsilon}(p))}{1 - F(\tilde{\varepsilon}(p))} \tag{112}$$

For $\varepsilon \sim N(0,1)$, we know that it is true that $\varepsilon < \frac{f(\varepsilon)}{1-F(\varepsilon)}$ $\forall \varepsilon$. This is because

$$\varepsilon < \frac{f(\varepsilon)}{1 - F(\varepsilon)} \tag{113}$$

$$f(\varepsilon) - \varepsilon * (1 - F(\varepsilon)) > 0 \tag{114}$$

$$z(\varepsilon) > 0 \tag{115}$$

and we know that z is a declining function in ε between ∞ and 0 (see Appendix section B.2.1).

Since the right-hand side of Equation 113 is positive, it then follows that $\tilde{\varepsilon}(p) < 2\frac{f(\tilde{\varepsilon}(p))}{1-F(\tilde{\varepsilon}(p))}$ also always holds. Thus, the second-order condition is always satisfied.

B.1.2 Part 2: Equation 16

This part of the proof argues that equation 8, together with the equilibrium conditions in Definition 3.1, implies equation 16. Start from equation 72:

$$s_m(\tilde{\varepsilon}_m, \theta_m) = \frac{1}{1-\beta}\big[u_m + \frac{\beta}{1-\beta\pi}\sigma_m \lambda_m(\theta_m)\mathbb{E}_\varepsilon\Big((\varepsilon - \tilde{\varepsilon}_m)1_{\varepsilon \geq \tilde{\varepsilon}}\Big)\big] \tag{116}$$

Moreover, the following relationship holds in equilibrium:

$$s_m(\tilde{\varepsilon}_m, \theta_m) = \overline{s} \tag{117}$$

Without loss of generality, when u does not depend on m, for example if $u_m = 0\ \forall m$, this becomes

$$\sigma_m \lambda_m(\theta_m) z(\tilde{\varepsilon}_m) = \frac{(1-\beta)(1-\beta\pi)}{\beta}\big[\overline{s} - \frac{1}{1-\beta}u\big] = \overline{z} \tag{118}$$

(If u_m depends on m, we can carry on indexing $\overline{z}_m$.) $\overline{z}$ (or $\overline{s}$) is an equilibrium quantity which depends on all exogenous parameters in all markets.

B.1.3 Part 3: Equation 17 (Houses, sellers and buyers)

N_{houses} indicates the aggregate number of houses and $N_{households}$ the aggregate number of households, which are both fixed. Since there is an implicit rental sector, the case where $N_{households} \neq N_{houses}$ can also be accommodated. Let $n_{houses} = (n_{houses,1}, ..., n_{houses,M})$ be the vector of the number of houses in each segment. We assume it to be fixed. The vector of households in each segment $n_{households} = (n_{households,1}, ..., n_{households,M})$ is endogenous.

Housing flows There are "occupied" houses (with an owner who likes the house) and unoccupied houses (houses for sale, waiting for a buyer):

$$n_{houses,m} = n_{occupied,m} + n_{unoccupied,m}$$

With probability $1-\pi$ a house moves from $n_{occupied,m}$ to $n_{unoccupied,m}$. With probability $\delta_m(1-F(\tilde{\varepsilon}_m))$ a house is bought and moves from $n_{unoccupied,m}$ to $n_{occupied,m}$.

We thus have at the stationary equilibrium:

$$\begin{aligned} n_{occupied,m} &= \pi n_{occupied,m} + \delta_m(1-F(\tilde{\varepsilon}_m))n_{unoccupied,m}, \\ n_{unoccupied,m} &= (1-\delta_m(1-F(\tilde{\varepsilon}_m))n_{unoccupied,m} + (1-\pi)n_{occupied,m} \end{aligned}$$

which implies that:

$$n_{occupied,m} = \frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi} n_{unoccupied,m} \tag{119}$$

and:

$$n_{houses,m} = n_{occupied,m} + n_{unoccupied,m} = (1 + \frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi})n_{unoccupied,m}.$$

In consequence:

$$n_{unoccupied,m} = \frac{1}{1+\frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi}} n_{houses,m}; \quad n_{occupied,m} = \frac{\frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi}}{1+\frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi}} n_{houses,m}. \tag{120}$$

Household flows There are matched households (happy owners) and searching households (renters, waiting for a seller):

$$n_{households,m} = n_{matched\ households,m} + n_{searchers,m}$$

With probability π a household in $n_{matched\ household,m}$ remains in that state, and with the complement probability $1-\pi$ becomes a searcher in (some) segment. With probability $\lambda_m(1-F(\tilde{\varepsilon}_m))$ a searching household buys and moves from $n_{searcher,m}$ to $n_{matched\ household,m}$. As all matched households live in occupied units, we note $n_{matched\ household,m} = n_{occupied,m}$.

We thus have at the stationary equilibrium:

$$n_{occupied,m} = \pi n_{occupied,m} + \lambda_m(1-F(\tilde{\varepsilon}_m))n_{searcher,m},$$

which implies that:

$$n_{occupied,m} = \frac{\lambda_m(1-F(\tilde{\varepsilon}_m))}{1-\pi} n_{searcher,m}. \tag{121}$$

In consequence, with together with equation 120:

$$n_{searcher,m} = \frac{1-\pi}{\lambda_m(1-F(\tilde{\varepsilon}_m))} n_{occupied,m} = \frac{\frac{1-\pi}{\lambda_m(1-F(\tilde{\varepsilon}_m))}\frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi}}{1+\frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi}} n_{houses,m} \tag{122}$$

$$= \frac{\frac{\delta_m}{\lambda_m}}{1+\frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi}} n_{houses,m} \tag{123}$$

Market tightness Using that the number of matched houses and matched households are equal, equations (119) and (121) yield:

$$n_{occupied,m} = \frac{\delta_m(1-F(\tilde{\varepsilon}_m))}{1-\pi} n_{unoccupied,m} = \frac{\lambda_m(1-F(\tilde{\varepsilon}_m))}{1-\pi} n_{searcher,m}$$

so that:

$$\theta_m = \frac{\delta_m}{\lambda_m} = \frac{n_{searcher,m}}{n_{unoccupied,m}}, \tag{124}$$

the number of searchers divided by the number of unmatched houses. This can be related to exogenous $n_{houses,m}$ using equations (123) and (120).

Aggregate quantities The number of houses in each segment, n_{houses}, is fixed by assumption, implying also that the aggregate number of houses N_{houses} is fixed. The aggregate number of households, $N_{households}$, is fixed.

Denote any possible difference between these quantities by a constant ν:

$$\nu = N_{households} - N_{houses}. \tag{125}$$

The aggregate quantities can also be expressed as

$$\sum_{m \in M} n_{houses_m} = \sum_{m \in M} n_{occupied,m} + \sum_{m \in M} n_{unoccupied,m} = N_{houses} \tag{126}$$

$$\sum_{m \in M} n_{households,m} = \sum_{m \in M} n_{matched\ households,m} + \sum_{m \in M} n_{searchers,m} = N_{households} \tag{127}$$

Because $n_{occupied,m} = n_{matched\ households,m}$, this implies that

$$\sum_{m \in M} n_{searchers,m} - \sum_{m \in M} n_{unoccupied,m} = \nu. \tag{128}$$

The general equilibrium of the spatial economy is such that the difference between the aggregate number of searchers and aggregate number of sellers is given by ν, an exogenous constant. We can further express this as

$$\nu = \sum_{m \in M} n_{searchers,m} - \sum_{m \in M} n_{unoccupied,m} \tag{129}$$

$$= \sum_{m \in M} \Big((\theta_m - 1) n_{unoccupied,m} \Big) \tag{130}$$

$$= \sum_{m \in M} \Big(\frac{\theta_m - 1}{1 + \frac{\delta_m (1 - F(\tilde{\varepsilon}_m))}{1 - \pi}} n_{houses,m} \Big) \tag{131}$$

B.2 Proof of Proposition 3.1

B.2.1 Partial Derivatives and Limits of z and γ

Consider first z:

$$z(\tilde{\varepsilon}_m) = \mathbb{E}_\varepsilon\big((\varepsilon - \tilde{\varepsilon}_m)\mathbb{1}_{\varepsilon \geq \tilde{\varepsilon}_m}\big) = f(\tilde{\varepsilon}_m) - \tilde{\varepsilon}_m\big(1 - F(\tilde{\varepsilon}_m)\big) \tag{132}$$

Derivative w.r.t $\tilde{\varepsilon}$ Using that $\varepsilon \sim N(0,1)$, so that $f'(\varepsilon) = -\varepsilon f(\varepsilon)$, the derivative of z w.r.t. $\tilde{\varepsilon}_m$ is given by

$$\frac{\partial z}{\partial \tilde{\varepsilon}_m} = -\tilde{\varepsilon}_m f(\tilde{\varepsilon}_m) - 1 + F(\tilde{\varepsilon}_m) + \tilde{\varepsilon}_m * f(\tilde{\varepsilon}_m) \tag{133}$$

$$= -(1 - F(\tilde{\varepsilon}_m)) \tag{134}$$

which is always negative.

Limit of z when $\tilde{\varepsilon}_m \to +\infty$. Let us look at the limits when $\tilde{\varepsilon} \to +\infty$.

Look first at intermediate objects. Note that $\lim_{\tilde{\varepsilon}\to+\infty} 1 - F(\tilde{\varepsilon}) = 0$, $\lim_{\tilde{\varepsilon}\to+\infty} (1 - F(\tilde{\varepsilon}))^2 = 0$, $\lim_{\tilde{\varepsilon}\to+\infty} f(\tilde{\varepsilon}) = 0$, so we can apply l'Hopital's rule. Applying l'Hopital's rule:

$$\lim_{\tilde{\varepsilon}\to+\infty} \frac{1 - F(\tilde{\varepsilon})}{f(\tilde{\varepsilon})} = \lim_{\tilde{\varepsilon}\to+\infty} \frac{-f(\tilde{\varepsilon})}{f'(\tilde{\varepsilon})} = \lim_{\tilde{\varepsilon}\to+\infty} \frac{-f(\tilde{\varepsilon})}{-\tilde{\varepsilon} f(\tilde{\varepsilon})} = \lim_{\tilde{\varepsilon}\to+\infty} \frac{1}{\tilde{\varepsilon}} = 0$$

$$\lim_{\tilde{\varepsilon}\to+\infty} \frac{(1 - F(\tilde{\varepsilon}))^2}{f(\tilde{\varepsilon})} = \lim_{\tilde{\varepsilon}\to+\infty} \frac{2(1 - F(\tilde{\varepsilon}))(-f(\tilde{\varepsilon}))}{f'(\tilde{\varepsilon})} = \lim_{\tilde{\varepsilon}\to+\infty} \frac{-2(1 - F(\tilde{\varepsilon}))f(\tilde{\varepsilon})}{-\tilde{\varepsilon} f(\tilde{\varepsilon})}$$
$$= \lim_{\tilde{\varepsilon}\to+\infty} \frac{2(1 - F(\tilde{\varepsilon}))}{\tilde{\varepsilon}} = 0$$

Next, to find the limit of z, we can use the following trick: $x(1 - F(x))$ can be rewritten as $\frac{x}{\frac{1}{1-F(x)}} = \frac{x}{(1 - F(x))^{-1}}$. Now, both the numerator and the denominator go to $+\infty$ as

$x \to +\infty$, so we may apply the l'Hopital's rule:

$$\begin{aligned}\lim_{\tilde{\varepsilon}_m \to +\infty} \tilde{\varepsilon}_m(1 - F(\tilde{\varepsilon}_m)) = \lim_{\tilde{\varepsilon}_m \to \infty} \frac{\tilde{\varepsilon}_m}{(1 - F(\tilde{\varepsilon}_m))^{-1}} &= \lim_{\tilde{\varepsilon}_m \to \infty} \frac{1}{-1 * (1 - F(\tilde{\varepsilon}_m))^{-2} * -f(\tilde{\varepsilon}_m)} \\ = \lim_{\tilde{\varepsilon}_m \to \infty} \frac{(1 - F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)} = 0\end{aligned}$$

Thus,

$$\lim_{\tilde{\varepsilon}_m \to +\infty} z(\tilde{\varepsilon}_m) = \lim_{\tilde{\varepsilon}_m \to +\infty} \Big[f(\tilde{\varepsilon}_m) - \tilde{\varepsilon}_m(1 - F(\tilde{\varepsilon}_m)) \Big] \tag{135}$$

$$= 0 \tag{136}$$

Limit of z when $\tilde{\varepsilon}_m \to -\infty$. Start by using the same trick as above: Rewrite

$$-\tilde{\varepsilon}_m F(\tilde{\varepsilon}_m) = \frac{-\tilde{\varepsilon}_m}{\frac{1}{F(\tilde{\varepsilon}_m)}} = \frac{-\tilde{\varepsilon}_m}{F(\tilde{\varepsilon}_m)^{-1}}$$

and now $-\tilde{\varepsilon}_m \to \infty$ as well as $F(\tilde{\varepsilon}_m)^{-1} \to \infty$ as $x \to -\infty$. Therefore we may apply l'Hopital's rule to $\frac{-\tilde{\varepsilon}_m}{F(\tilde{\varepsilon}_m)^{-1}}$:

$$\lim_{\tilde{\varepsilon}_m \to -\infty} \frac{-\tilde{\varepsilon}_m}{F(\tilde{\varepsilon}_m)^{-1}} = \lim_{\tilde{\varepsilon}_m \to -\infty} \frac{-1}{-1F(\tilde{\varepsilon}_m)^{-2} f(\tilde{\varepsilon}_m)} = \lim_{\tilde{\varepsilon}_m \to -\infty} \frac{F(\tilde{\varepsilon}_m)^2}{f(\tilde{\varepsilon}_m)} = 0$$

Which shows that $\tilde{\varepsilon} F(\tilde{\varepsilon}) \to 0$ as $\tilde{\varepsilon} \to -\infty$.

Thus,

$$\lim_{\tilde{\varepsilon}_m \to -\infty} z(\tilde{\varepsilon}_m) = \lim_{\tilde{\varepsilon}_m \to -\infty} \Big[f(\tilde{\varepsilon}_m) - \tilde{\varepsilon}_m(1 - F(\tilde{\varepsilon}_m)) \Big] \tag{137}$$

$$= \lim_{\tilde{\varepsilon}_m \to -\infty} \Big[f(\tilde{\varepsilon}_m) - \tilde{\varepsilon}_m 1 + \tilde{\varepsilon}_m F(\tilde{\varepsilon}_m)) \Big] \tag{138}$$

$$= \infty \tag{139}$$

In summary, $z(\tilde{\varepsilon}_m)$ decreases between $+\infty$ and 0 when $\tilde{\varepsilon}_m$ varies between $-\infty$ and $+\infty$.

Consider next γ.

$$\gamma(\tilde{\varepsilon}_m, \theta_m) = \frac{1-\beta\pi}{\beta}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \delta_m(\theta_m)\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)} \tag{140}$$

Derivative of γ w.r.t $\tilde{\varepsilon}_m$. Consider first $\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)}$.

$$\frac{\partial \frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)}}{\partial \tilde{\varepsilon}_m} = \frac{1}{f(\tilde{\varepsilon}_m)} \cdot (-f(\tilde{\varepsilon}_m)) + (1-F(\tilde{\varepsilon}_m)) \cdot (-1) f(\tilde{\varepsilon}_m)^{-2} \cdot (-\tilde{\varepsilon}_m f(\tilde{\varepsilon}_m)) \tag{141}$$

$$= -1 + \frac{(1-F(\tilde{\varepsilon}_m))}{f(\tilde{\varepsilon}_m)}\tilde{\varepsilon}_m \tag{142}$$

which is always less than zero. This is because $\varepsilon \sim N(0,1)$, we know that it is true that $\varepsilon < \frac{f(\varepsilon)}{1-F(\varepsilon)}\ \forall \varepsilon$, so it is also true that $\varepsilon\frac{1-F(\varepsilon)}{f(\varepsilon)} < 1$.

Consider next $\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}$.

$$\frac{\partial \frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}}{\partial \tilde{\varepsilon}_m} = \frac{1}{f(\tilde{\varepsilon}_m)} 2(1-F(\tilde{\varepsilon}_m)) * (-f(\tilde{\varepsilon}_m)) \tag{143}$$

$$+ (1-F(\tilde{\varepsilon}_m))^2 \cdot (-1) f(\tilde{\varepsilon}_m)^{-2}(-\tilde{\varepsilon}_m f(\tilde{\varepsilon}_m)) \tag{144}$$

$$= -2(1-F(\tilde{\varepsilon}_m)) + \tilde{\varepsilon}_m \frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)} \tag{145}$$

$$= (1-F(\tilde{\varepsilon}_m))\Big\{-2 + \tilde{\varepsilon}_m \frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)}\Big\} \tag{146}$$

which is also always negative.

Now, the partial derivative of γ w.r.t. $\tilde{\varepsilon}_m$, holding constant θ_m, is given by:

$$\frac{\partial \gamma(\tilde{\varepsilon}_m, \theta_m)}{\partial \tilde{\varepsilon}_m} = \frac{1-\beta\pi}{\beta}\frac{\partial \frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)}}{\partial \tilde{\varepsilon}_m} + \delta(\theta_m)\frac{\partial \frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}}{\partial \tilde{\varepsilon}_m} \tag{147}$$

Which is negative since $\frac{1-\beta\pi}{\beta} > 0$, $\delta(\theta) \geq 0$.

Limit of γ when $\tilde{\varepsilon} \to +\infty$. From above arguments it follows directly that

$$\lim_{\tilde{\varepsilon}_m \to +\infty} \gamma(\tilde{\varepsilon}_m, \theta_m) = 0$$

Limit of γ when $\tilde{\varepsilon} \to -\infty$. Start by rewriting γ as

$$\gamma(\tilde{\varepsilon}_m, \theta_m) = \frac{1-\beta\pi}{\beta}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \delta_m(\theta_m)\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)} \tag{148}$$

$$= \frac{1-\beta\pi}{\beta}\Big[\frac{1}{f(\tilde{\varepsilon}_m)} - \frac{F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)}\Big] + \delta_m(\theta_m)\Big[\frac{1}{f(\tilde{\varepsilon}_m)} - \frac{2F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \frac{(F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}\Big] \tag{149}$$

Again, we notice that $\lim_{\tilde{\varepsilon}\to-\infty} F(\tilde{\varepsilon}) = 0$, $\lim_{\tilde{\varepsilon}\to-\infty} (F(\tilde{\varepsilon}))^2 = 0$, $\lim_{\tilde{\varepsilon}\to-\infty} f(\tilde{\varepsilon}) = 0$, so we may apply the l'Hopital's rule. Applying it gives

$$\lim_{\tilde{\varepsilon}\to-\infty} \frac{F(\tilde{\varepsilon})}{f(\tilde{\varepsilon})} = \lim_{\tilde{\varepsilon}\to-\infty} \frac{f(\tilde{\varepsilon})}{-\tilde{\varepsilon} f(\tilde{\varepsilon})} = -\lim_{\tilde{\varepsilon}\to-\infty} \frac{1}{\tilde{\varepsilon}} = 0$$

$$\lim_{\tilde{\varepsilon}\to-\infty} \frac{F(\tilde{\varepsilon})^2}{f(\tilde{\varepsilon})} = 0$$

Thus,

$$\lim_{\tilde{\varepsilon}_m\to-\infty} \gamma(\tilde{\varepsilon}_m, \theta_m) = \lim_{\tilde{\varepsilon}_m\to-\infty} \Big\{\frac{1-\beta\pi}{\beta}\Big[\frac{1}{f(\tilde{\varepsilon}_m)}\Big] + \delta_m(\theta_m)\Big[\frac{1}{f(\tilde{\varepsilon}_m)}\Big]\Big\} \tag{150}$$

$$= +\infty \tag{151}$$

In summary, $\gamma(\tilde{\varepsilon}_m, \theta_m)$ decreases between $+\infty$ and 0 when $\tilde{\varepsilon}_m$ varies between $-\infty$ and $+\infty$.

B.2.2 Existence and uniqueness of solution (Proposition3.1)

To simplify notation and without loss of generality, assume we are in the case where $u_m = 0$ and therefore $\bar{z}_m = \bar{z}$ for all m as in Equation 118 (otherwise continue to index $\bar{z}$ with m and treat $\bar{s}$ constant across m). Start from equations 89 and 118:

$$\begin{aligned} \lambda(\theta_m)z(\tilde{\varepsilon}_m) &= \bar{z}/\sigma_m, \\ g(\tilde{\varepsilon}_m, \theta_m) &= \frac{\beta}{1-\beta\pi}(\gamma(\tilde{\varepsilon}_m, \theta_m) + \lambda(\theta_m)z(\tilde{\varepsilon}_m) + a_m) = \tilde{\varepsilon}_m \end{aligned} \tag{152}$$

in which:

$$\begin{aligned} z(\tilde{\varepsilon}_m) &= E_\varepsilon((\varepsilon - \tilde{\varepsilon}_m)\mathbf{1}\{\varepsilon > \tilde{\varepsilon}_m\}), \\ \gamma(\tilde{\varepsilon}_m, \theta_m) &= \frac{1-\beta\pi}{\beta}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \delta(\theta_m)\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)}, \\ a_m &= -\frac{1-\beta\pi}{\beta\sigma_m}\Big[x_m - u_m - c_m\Big]. \end{aligned}$$

And we know from above section that

$$\frac{\partial z}{\partial \tilde{\varepsilon}_m} < 0, \frac{\partial \gamma}{\partial \tilde{\varepsilon}_m} < 0, \tag{153}$$

and that $z(\tilde{\varepsilon}_m)$ and $\gamma(\tilde{\varepsilon}_m, \theta_m)$ are both decreasing between $+\infty$ and 0 when $\tilde{\varepsilon}_m$ varies between $-\infty$ and $+\infty$. By assumption 3.1,

$$\frac{\partial \delta_m}{\partial \theta_m} > 0, \frac{\partial \lambda_m}{\partial \theta_m} < 0, \tag{154}$$

Consider the first equation in system (152). For any $\bar{z} > 0$ and $\theta_m > 0$, as $\bar{z}/(\sigma_m\lambda(\theta_m)) > 0$, and $z(\tilde{\varepsilon}_m)$ is decreasing between $+\infty$ (when $\tilde{\varepsilon}_m \to -\infty$) and 0 (when $\tilde{\varepsilon}_m \to +\infty$), there is a unique solution $\tilde{\varepsilon}_m = \phi_1(\theta_m, \bar{z})$ for any $\theta_m > 0$. Using equations (153) , we obtain that $\frac{\partial \phi_1}{\partial \theta_m} < 0$ and $\frac{\partial \phi_1}{\partial \bar{z}} < 0$. In addition, the limits on the right and left of the domain of $\phi_1(\theta_m, \bar{z})$ as a function of θ_m are obtained using the follwing argument. When $\theta_m = 0$, $\lambda(\theta_m) = 1$ and $\phi_1(\theta_m, \bar{z}) = z^{-1}(\bar{z}/\sigma_m)$ where z^{-1} is the inverse of $z(\tilde{\varepsilon}_m)$. On the other hand, when $\theta_m \to +\infty$, $\lambda(\theta_m) \to 0$ and $z(\tilde{\varepsilon}_m) = \frac{\bar{z}/\sigma_m}{\lambda(\theta_m)} \to +\infty$ for all $\bar{z} > 0$ and thus $\phi_1(\theta_m, \bar{z}) \to -\infty$.

Consider the second equation in (152). Substituting in the first one,

$$\gamma(\tilde{\varepsilon}_m, \theta_m) - \frac{1-\beta\pi}{\beta}\tilde{\varepsilon}_m + a_m = -\frac{\bar{z}}{\sigma_m}.$$

For any $\bar{z} > 0$ and $\theta_m > 0$, as the LHS is decreasing in $\tilde{\varepsilon}_m$ and varying between $+\infty$ (when $\tilde{\varepsilon}_m \to -\infty$) and $-\infty$ (when $\tilde{\varepsilon}_m \to +\infty$), there is a unique solution $\tilde{\varepsilon}_m = \phi_2(\theta_m, \bar{z})$. It is direct to show from equations (153) that $\frac{\partial\phi_2}{\partial\theta_m} > 0$ (since $\frac{\partial\delta}{\partial\theta_m} > 0$) and $\frac{\partial\phi_2}{\partial\bar{z}} > 0$ (since $\frac{\partial\gamma}{\partial\tilde{\varepsilon}_m} < 0$). In addition the limits on the right and left of $\phi_2(\theta_m, \bar{z})$ as a function of θ_m can be obtained easily.

When $\theta_m = 0$, $\delta(\theta_m) = 0$ and $\phi_2(\theta_m, \bar{z})$ is equal to the unique solution $\phi_2(0, \bar{z})$:

$$\frac{1-\beta\pi}{\beta}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} - \frac{1-\beta\pi}{\beta}\tilde{\varepsilon}_m + a_m = -\frac{\bar{z}}{\sigma_m},$$

On the other hand, when $\theta_m \to +\infty$, $\delta(\theta_m) \to 1$ and $\phi_2(\theta_m, \bar{z})$ is the unique solution $\phi_2(+\infty, \bar{z})$ of

$$\frac{1-\beta\pi}{\beta}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)} - \frac{1-\beta\pi}{\beta}\tilde{\varepsilon}_m + a_m = -\frac{\bar{z}}{\sigma_m}.$$

In consequence, the solution continuously increases between $\phi_2(0, \bar{z})$ and $\phi_2(\infty, \bar{z})$. As $\phi_1(\theta_m, \bar{z})$ decreases continuously between $z^{-1}(\bar{z}/\sigma_m)$ and $-\infty$, the condition for existence is that

$$z^{-1}(\bar{z}/\sigma_m) \geq \phi_2(0, \bar{z}).$$

Note that the LHS decreases with $\bar{z}$ while the RHS is increasing in $\bar{z}$. Moreover, the LHS is tending to $+\infty$ when $\bar{z} \to 0$ and tending to $-\infty$ when $\bar{z} \to +\infty$. In addition, the RHS is tending to a constant when $\bar{z} \to 0$ and tending to $+\infty$ when $\bar{z} \to +\infty$. Thus there are values of $\bar{z}$ such that the solution does not exist.

If the solution exists, denote the full solution as the zero in θ_m of function:

$$\psi(\theta_m, \bar{z}) = \phi_1(\theta_m, \bar{z}) - \phi_2(\theta_m, \bar{z}).$$

As $\frac{\partial\phi_1}{\partial\theta_m} < 0$ and $\frac{\partial\phi_2}{\partial\theta_m} > 0$, then $\frac{\partial\psi}{\partial\theta_m} < 0$ and the solution is unique. Furthermore, as $\frac{\partial\phi_1}{\partial\bar{z}} < 0$ and $\frac{\partial\phi_2}{\partial\bar{z}} > 0$, we have $\frac{\partial\psi}{\partial\bar{z}} < 0$. The solution is differentiable in $\bar{z}$ and $\frac{\partial\theta}{\partial\bar{z}} < 0$.

C Solution Algorithm

Notation The equilibrium is characterized by equations 89 and 118:

$$
\begin{aligned}
s(\tilde{\varepsilon}_m, \theta_m) &= \frac{1}{1-\beta}\frac{\beta}{1-\beta\pi}\sigma_m \lambda(\theta_m) z(\tilde{\varepsilon}_m) = \overline{s}, \\
g(\tilde{\varepsilon}_m, \theta_m) &= \frac{\beta}{1-\beta\pi}(\gamma(\tilde{\varepsilon}_m, \theta_m) + \lambda(\theta_m) z(\tilde{\varepsilon}_m) + a_m) = \tilde{\varepsilon}_m.
\end{aligned}
$$

where:

$$
\begin{aligned}
z(\tilde{\varepsilon}_m) &= E_\varepsilon((\varepsilon - \tilde{\varepsilon}_m)\mathbf{1}\{\varepsilon > \tilde{\varepsilon}_m\}), \\
\gamma(\tilde{\varepsilon}_m, \theta_m) &= \frac{1-\beta\pi}{\beta}\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} + \delta(\theta_m)\frac{(1-F(\tilde{\varepsilon}_m))^2}{f(\tilde{\varepsilon}_m)},
\end{aligned}
$$

as well as the aggregate market clearing given by equation (17). Moreover, we may write the first equation as :

$$
\frac{(1-\beta)(1-\beta\pi)}{\beta} s(\tilde{\varepsilon}_m, \theta_m) = \sigma_m \lambda(\theta_m) z(\tilde{\varepsilon}_m) = \overline{z} \quad \Rightarrow \quad \lambda(\theta_m) = \frac{\overline{z}}{z(\tilde{\varepsilon}_m)\sigma_m}
$$

(where we are assuming that we are in the case of equation 118 where u_m does not depend on m and so $\overline{z}$ does not depend on m either).

Solution algorithm First use the parameters of each location together with Appendix section B.2.2 to determine the highest feasible value for $\overline{s}$ (or $\overline{z}$) in each location such that the model equations 89 and 118 can have a solution (as discussed in section B.2.2, there is no solution to the relevant system of equations for very high values of $\overline{z}$). In the spatial equilibrium, the highest possible value of $\bar{z}$ such that the model equations can have a solution in all locations must be the smallest of these. Next, make a grid over the possible values of $\bar{z}$ (consider $\bar{z}$ strictly positive and up to this highest possible value).

For each point $\overline{z}_i$ on the grid, for each location m, find the values of θ_m and $\tilde{\varepsilon}_m$ that are consistent with $\overline{z}_i$. Note that this can be done location-by-location, as all dependencies across locations go through $\bar{z}$.

1. Take an initial guess for the vector of θ's (for example, $\theta^{(0)} = (1, 1, ..., 1)$)
2. Use $g(\tilde{\varepsilon}_m, \theta_m) = \tilde{\varepsilon}_m$ to solve $\tilde{\varepsilon}_m^{(0)} = \tilde{\varepsilon}_m(\theta_m^{(0)})\ \forall m$.

   There is a unique solution for a fixed θ since $\frac{\partial \gamma}{\partial \tilde{\varepsilon}} < 0$ and $\frac{\partial z}{\partial \tilde{\varepsilon}} < 0$ both decrease between

$+\infty$ and 0, so $g(\tilde{\varepsilon}_m, \theta_m) - \tilde{\varepsilon}_m = 0$ decreases in $\tilde{\varepsilon}$ between $+\infty$ and $-\infty$.

Compute implied $z_m^{(0)} = z(\tilde{\varepsilon}_m^{(0)})$.

3. [a] Consider first the case where $0 < \frac{\overline{z}}{z_m^{(0)}\sigma_m} < 1$.

   We know that

$$\lambda(z_m^{(0)}, \overline{z}, \sigma_m) = \frac{\overline{z}}{z_m^{(0)}\sigma_m}$$

   and

$$\lambda(\theta_m)$$

   is a monotone function of θ giving values between 0 and 1 so we can invert the relationship

$$\lambda(\theta_m) = \frac{\overline{z}_i}{z_m^{(0)}\sigma_m}$$

   and this inversion will imply a value θ.

   Inverting $\lambda(\theta)$ w.r.t. θ and equating given $\overline{z}$ and given $z_m^{(0)}$, implies

$$\theta_m(\overline{z}_i, z_m^{(0)}) = \lambda^{-1}(\overline{z}_i, z_m^{(0)})$$

   Update the guess for θ using this final expression,

$$\theta_m^{(1)} = \theta_m(\overline{z}_i, z_m^{(0)})$$

   [b] Consider next the case where $\frac{\overline{z}_i}{z_m^{(0)}\sigma_m} > 1$, and therefore the expression for λ cannot be inverted. If $\lambda(\theta_m^{(0)}) \cdot z_m^{(0)} > \frac{\overline{z}}{\sigma}$, update the guess for θ upwards. If $\lambda(\theta_m^{(0)}) \cdot z_m^{(0)} < \frac{\overline{z}}{\sigma}$, update the guess for θ downwards.

   [c] Consider finally the case where $\frac{\overline{z}}{z_m^{(0)}\sigma_m} \approx 0$. Update the guess for θ upwards.

4. Go back to step 2. Repeat until convergence.
5. If it would happen that the algorithm did not converge using steps 3a-3c for some location m, it is possible also to do a grid for θ_m and search for the gridpoint on which equations 89 and 118 are satisfied. The solution exists (as we are considering the feasible range of $\bar{z}$) and is unique (by Proposition 3.1). However, it is more efficient to use the iterative procedure of steps 3a-3c to see if the algorith converges before doing a grid search.

Now we have computed the $\theta_m(\overline{z}_i)$, $\tilde{\varepsilon}_m(\overline{z}_i)$ for all m that are consistent with each possible value of $\overline{z}_i$ on the grid.

Use the equation

$$\sum_{m=1}^{M} \frac{1-\theta_m(\overline{z}_i)}{1+\frac{\delta(\theta_m(\overline{z}_i))\cdot(1-F(\tilde{\varepsilon}_m(\overline{z}_i)))}{1-\pi}} n_{houses,m} = \nu.$$

to find the correct $\overline{z}_i$. Verify that only 1 value of $\overline{z}_i$ on the grid satisfies the equality (the left-hand side of the summation is increasing in $\overline{z}$).

D Identification

For each m, we observe $p_m, E(TOM)_m$ and θ_m. π and β are fixed. u_m is set to 0 by assumption and c_m to pre-specified values.

The question of identification is to assess whether primitives x_m, σ_m and α_m are identified using that for any m, we have the set of equilibrium equations:

$$\frac{1-F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} = (p_m - \frac{c_m}{1-\beta})\frac{1}{\sigma_m}\frac{(1-\beta\pi)(1-\beta)}{1-\beta+\beta E(TOM)_m^{-1}} \tag{155}$$

$$s_m = \frac{1}{1-\beta}\big[\frac{\beta\lambda_m}{1-\beta\pi}\sigma_m \mathbb{E}_\varepsilon\big((\varepsilon-\tilde{\varepsilon}_m)\mathbf{1}_{\varepsilon\geq\tilde{\varepsilon}_m}\big)\big] \tag{156}$$

in which:

$$\lambda_m = \frac{\delta_m}{\theta_m} \text{ and } \delta_m = \frac{1}{E(TOM)_m(1-F(\tilde{\varepsilon}_m))}. \tag{157}$$

We first investigate the structural constraints derived from this setting.

D.1 Structural constraints

Fix σ_m to an arbitrary value. Use Equation (155) and derive $\tilde{\varepsilon}_m$. We also have:

$$\delta_m = \frac{1}{E(TOM)_m(1-F(\tilde{\varepsilon}_m))}$$

Because δ_m, μ_m, λ_m and $\delta_m\mu_m$, $\lambda_m\mu_m$ are probabilities, and $\lambda_m = \frac{\delta_m}{\theta_m}$, we must impose that

$$\mu(\tilde{\varepsilon}_m) < 1 \tag{158}$$

$$\delta_m < 1 \tag{159}$$

$$\lambda_m < 1 \iff \delta_m < \theta_m \tag{160}$$

$$\mu(\tilde{\varepsilon}_m)\delta_m < 1 \tag{161}$$

$$\mu(\tilde{\varepsilon}_m)\lambda_m < 1 \iff \mu(\tilde{\varepsilon}_m)\delta_m < \theta_m \tag{162}$$

The first equation in the system is always satisfied as long as $\tilde{\varepsilon}_m > -\infty$. For the second equation to be satisfied, we can only consider $\tilde{\epsilon}_m < \bar{\epsilon}_{1,m}$ where $\bar{\epsilon}_{1,m}$ is given by

$$\bar{\epsilon}_{1,m} = F^{-1}(1 - \frac{1}{E(TOM)_m})$$

For the 3rd equation to be satisfied, we can only consider $\tilde{\epsilon}_m < \bar{\epsilon}_{2,m}$ where $\bar{\epsilon}_{2,m}$ is given by

$$\bar{\epsilon}_{2,m} = F^{-1}(1 - \frac{1}{E(TOM)_m \cdot \theta_m})$$

This condition is only binding when $\theta < 1$. The 4th equation is satisfied whenever the 1st and the 2nd equations are. The 5th equation is satisfied whenever

$$\frac{1}{E(TOM)} < \theta \tag{163}$$

This is a condition that only concerns the data and is independent of the choice of σ_m. This is a second restriction on the data, together with the requirement that $E(TOM)_m > 1$ (as selling in the model always takes at least one period).

Concisely, $\tilde{\varepsilon}_m$ exists if and only if $E(TOM)_m \min(1, \theta_m) > 1$ and this constraint does not depend on σ_m. Moreover, values of $\tilde{\epsilon}_m$ are possible only if they satisfy

$$\tilde{\epsilon}_m < \bar{\epsilon}_m = \min\left[F^{-1}(1 - \frac{1}{E(TOM)_m}),\ F^{-1}(1 - \frac{1}{E(TOM)_m \cdot \theta_m})\right] \tag{164}$$

All solutions $\tilde{\varepsilon}_m < \bar{\varepsilon}_m$ can be considered.

Consequences for σ_m Retruning to the value of σ_m, Equation (155) can be written as:

$$\frac{1 - F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)}\sigma_m = (p_m - \frac{c_m}{1-\beta})\frac{(1-\beta\pi)(1-\beta)}{1 - \beta + \beta E(TOM)_m^{-1}} \tag{165}$$

in which the RHS is a function of β, π and observables p_m and $E(TOM)_m$. Taking derivatives:

$$d\frac{1 - F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)}\sigma_m + \frac{1 - F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)}d\sigma_m = 0.$$

which proves that σ_m is a function of $\tilde{\varepsilon}_m$ written as $\sigma_m(\tilde{\varepsilon}_m)$. Use:

$$d\frac{1-F}{f} = \frac{-f^2 + (1-F)\tilde{\varepsilon}_m f}{f^2}d\tilde{\varepsilon}_m = -\frac{z(\tilde{\varepsilon}_m)}{f}d\tilde{\varepsilon}_m,$$

in which $z(\tilde{\varepsilon}_m) = E_\varepsilon((\varepsilon - \tilde{\varepsilon}_m)\mathbf{1}\{\varepsilon > \tilde{\varepsilon}_m\}) = f(\tilde{\varepsilon}_m) - \tilde{\varepsilon}_m(1 - F(\tilde{\varepsilon}_m))$. We thus get:

$$-\sigma_m \frac{z(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} d\tilde{\varepsilon}_m + \frac{1 - F(\tilde{\varepsilon}_m)}{f(\tilde{\varepsilon}_m)} d\sigma_m = 0,$$

or:

$$\frac{1}{\sigma_m}\frac{d\sigma_m}{d\tilde{\varepsilon}_m} = \frac{z(\tilde{\varepsilon}_m)}{1 - F(\tilde{\varepsilon}_m)} > 0.$$

Furthermore, $\lim_{\tilde{\varepsilon}_m \to -\infty} \sigma_m(\tilde{\varepsilon}_m) = 0$, $\lim_{\tilde{\varepsilon}_m \to \infty} \sigma_m(\tilde{\varepsilon}_m) = +\infty$.

This means that there exists $\bar{\sigma}_m = \sigma_m(\bar{\varepsilon}_m)$ and all solutions $\sigma_m < \bar{\sigma}_m$ are admissible. We can write:

$$\bar{\sigma}_m = (p_m - \frac{c_m}{1-\beta}) \frac{(1-\beta\pi)(1-\beta)}{1 - \beta + \beta E(TOM)_m^{-1}} \frac{f(\bar{\varepsilon}_m)}{1 - F(\bar{\varepsilon}_m)}$$

in which $\bar{\varepsilon}_m$ is defined in equation (164).

The value of search Rewrite equation (156) uisng the expression for $z(\tilde{\varepsilon}_m)$ as:

$$\frac{(1-\beta)(1-\beta\pi)}{\beta} s_m = \lambda_m \sigma_m z(\tilde{\varepsilon}_m). \tag{166}$$

and use equation (157) to replace λ_m as a function of $\tilde{\varepsilon}_m$. From equation (165), derive that the value of search is proportional to:

$$\begin{aligned} \lambda_m \sigma_m z(\tilde{\varepsilon}_m) &= \frac{1}{E(TOM)_m (1 - F(\tilde{\varepsilon}_m))\theta_m} (p_m - \frac{c_m}{1-\beta}) \frac{(1-\beta\pi)(1-\beta)}{1 - \beta + \beta E(TOM)_m^{-1}} \frac{f(\tilde{\varepsilon}_m)}{1 - F(\tilde{\varepsilon}_m)} z(\tilde{\varepsilon}_m) \\ &= \varpi_m \frac{f(\tilde{\varepsilon}_m)}{(1 - F(\tilde{\varepsilon}_m))^2} z(\tilde{\varepsilon}_m), \end{aligned} \tag{167}$$

in which ϖ_m can be derived from data p_m, c_m, θ_m, and $E(TOM)_m$:

$$\varpi_m = (p_m - \frac{c_m}{1-\beta}) \frac{(1-\beta\pi)(1-\beta)}{E(TOM)_m \theta_m (1 - \beta + \beta E(TOM)_m^{-1})}.$$

Define further

$$\psi(\tilde{\varepsilon}_m) = \frac{f(\tilde{\varepsilon}_m)}{(1 - F(\tilde{\varepsilon}_m))^2} z(\tilde{\varepsilon}_m) = \frac{f(\tilde{\varepsilon}_m)}{(1 - F(\tilde{\varepsilon}_m))^2} (f(\tilde{\varepsilon}_m) - \tilde{\varepsilon}_m(1 - F(\tilde{\varepsilon}_m)),$$

by replacing $z(\tilde{\varepsilon}_m)$ by its value. We thus get:

$$\psi(\tilde{\varepsilon}_m) = \frac{f(\tilde{\varepsilon}_m)}{(1 - F(\tilde{\varepsilon}_m))}(\frac{f(\tilde{\varepsilon}_m)}{(1 - F(\tilde{\varepsilon}_m))} - \tilde{\varepsilon}_m) = h_m(h_m - \tilde{\varepsilon}_m),$$

in which the hazard rate $h_m = \frac{f(\tilde{\varepsilon}_m)}{(1-F(\tilde{\varepsilon}_m))}$.

It is now easy to show that :

$$\frac{d}{d\tilde{\varepsilon}_m}(\frac{f}{1-F}) = \frac{-\varepsilon f(1-F) + f^2}{(1-F)^2} = \psi > 0,$$

and that

$$\frac{d^2}{d(\tilde{\varepsilon}_m)^2}(\frac{f}{1-F}) = \psi' > 0,$$

since the hazard rate for a normal distribution e.g. $h = \frac{f}{1-F}$ is convex (e.g. see https://math.stackexchange.com/questions/1349555/standard-normal-distribution-hazard-rate).

The range of admissible values We now impose that the value of search given by (167) is the same in all segments and for instance equal to constant K:

$$\varpi_m \psi(\tilde{\varepsilon}_m) = K. \tag{168}$$

We now have to impose that $\tilde{\varepsilon}_m \leq \bar{\varepsilon}_m$ for all m and this limits the range of possible values of K, since $\psi' > 0$. Specifically, the admissible range for K is given by:

$$K < \min_m(\varpi_m \psi(\bar{\varepsilon}_m)),$$

in which $\bar{\varepsilon}_m$ is defined in equation (164). This delivers the maximum value for $\bar{s} \geq 0$ which is given by:

$$\frac{(1-\beta)(1-\beta\pi)}{\beta}\bar{s} < \min_m(\varpi_m \psi(\bar{\varepsilon}_m)). \tag{169}$$

The reciprocal is straightforward and developed in the next section.

D.2 Inverting the model, given $\overline{s}$

Suppose from now on that $\bar{s}$ is arbirtrarily fixed within bounds (169). Then, the following quantities are fixed or can be estimated from data: p_m, $E(TOM)_m$, θ_m and $\bar{s}$ and we want to recover x_m, σ_m and α_m.

Identification proceeds in four steps:

1. Use value $K = \frac{\beta}{(1-\beta)(1-\beta\pi)}\bar{s}$ in equation (168) and derive $\tilde{\varepsilon}_m = \psi^{-1}(K/\varpi_m)$ for all m which is well defined and identified.
2. Use equation (157) to write that $\delta_m = \frac{1}{E(TOM)_m(1-F(\tilde{\varepsilon}_m))}$ and $\lambda_m = \frac{\delta_m}{\theta_m}$ to identify δ_m and λ_m.
3. Then use Equation (155) to identify σ_m.
4. Finish by observing that α_m will follow immediately from δ_m or λ_m.

D.3 Calibrating $\overline{s}$

Since the value of $\bar{s}$ is unobserved, one parameter in the system needs to be calibrated or fixed in order to recover a calibrated value for $\bar{s}$. There are different possible strategies for calibrating $\bar{s}$. We proceed by setting δ to a constant in a given market segment.

Consider a market indexed by 0. Use

$$E(TOM)_0 = \frac{1}{(1-F(\tilde{\varepsilon}_0))\delta_0}$$

to recover

$$\tilde{\varepsilon}_0 = F^{-1}(1 - \frac{1}{E(TOM)_0\delta_0})$$

Now, using $\tilde{\epsilon}_0$ as well as p_0 and $E(TOM)_0$, Equation 155 pins down σ_0. $\tilde{\epsilon}_0$, $E(TOM)_0$ and θ_0 pin down λ_0. Then, using λ_0, σ_0 and $\tilde{\epsilon}_0$, Equation 156 pins down s_0. Finally, by Definition 3.1, $\overline{s} = s_0$.

E Additional Tables and Figures

E.1 Robustness

	$\delta_0 = 0.3$				$\delta_0 = 0.4$				$\delta_0 = 0.6$				$\delta_0 = 0.7$			
	Mean	p25	p50	p75	Mean	p25	p50	p75	Mean	p25	p50	p75	Mean	p25	p50	p75
x	194.2	136.2	188.6	229.2	199.9	141.9	193.8	235.0	204.9	146.9	197.8	240.0	206.3	148.2	199.2	241.4
σ	1.02	0.78	0.89	1.17	1.22	0.91	1.03	1.40	1.41	1.03	1.17	1.60	1.47	1.07	1.21	1.65
α	0.27	0.24	0.26	0.28	0.27	0.24	0.26	0.28	0.28	0.25	0.27	0.28	0.29	0.25	0.27	0.28
δ	0.07	0.05	0.06	0.08	0.08	0.05	0.06	0.09	0.09	0.05	0.06	0.09	0.09	0.05	0.06	0.09
λ	0.08	0.07	0.08	0.10	0.09	0.07	0.08	0.10	0.09	0.07	0.08	0.10	0.09	0.07	0.08	0.10
Observations	66				66				66				66			
$\overline{s}$	45819.9				51426.5				56354.4				57717.5			

Table A3: Alternative Calibrations for δ_0, implied values of $\overline{s}$ and implied parameter values.

	Mean	p25	p50	p75
x	190.4	143.4	183.6	219.6
σ	1.19	0.81	0.94	1.27
α	0.30	0.26	0.28	0.31
δ	0.09	0.06	0.07	0.09
λ	0.10	0.08	0.10	0.13
Observations	66			

Table A4: Parameter estimates using transaction prices and times.

Notes. This table documents parameter estimates from an analysis which is similar to our baseline model inversion (see Table 1), but instead of measuring p_m by the average listing price in the segment and $E(TOM)_m$ by the average listing time in the segment, p_m is measured using the average transaction price in the segment and $E(TOM)_m$ is measured using the average sale time in the segment, both in the transactions dataset (KVKL). Other variables and calibrations correspond to out baseline model inversion.

E.2 Additional tables and figures related to parameters and market characteristics

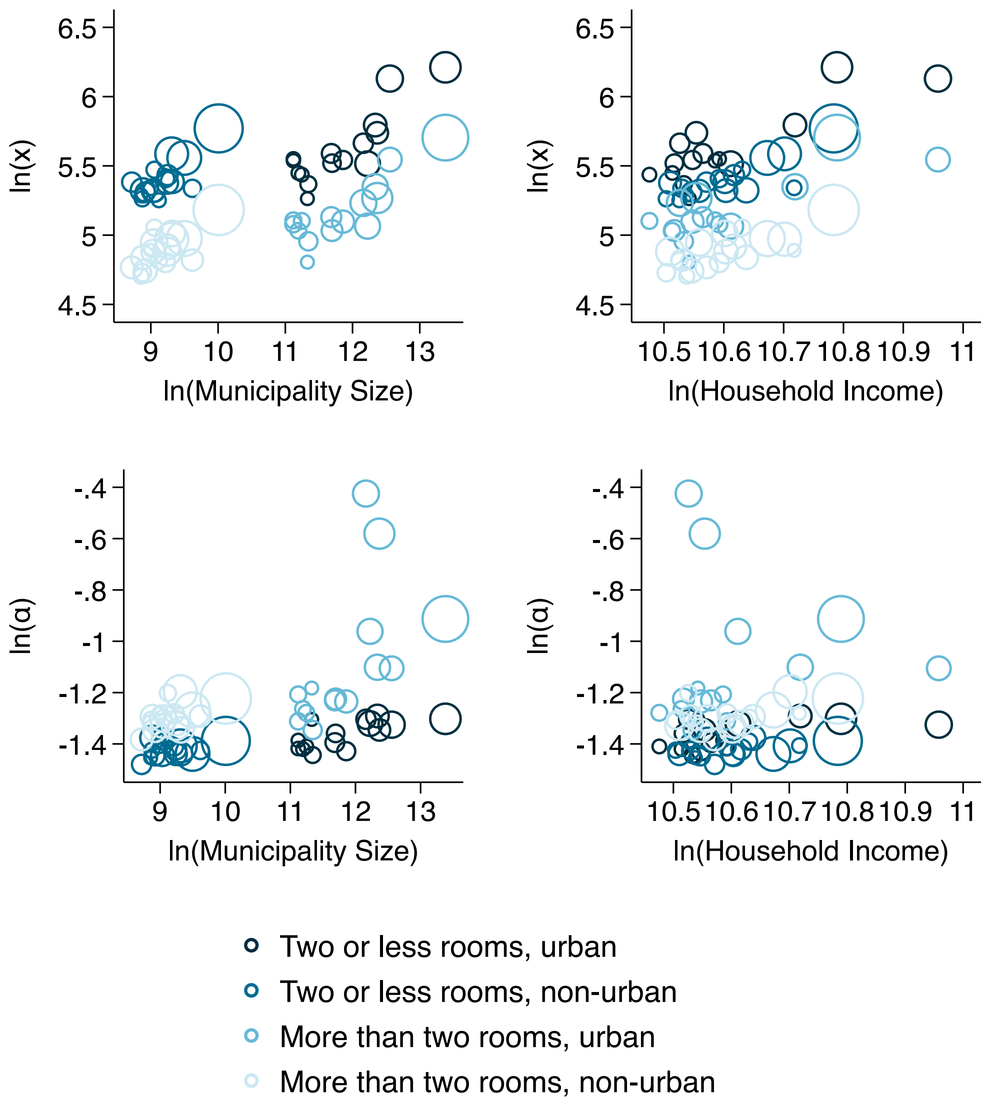


Figure A2: Scatterplots summarizing the relationship between two of the model parameters (x and α) and segment characteristics.

Notes. Municipality size refers to average municipality population in the market segment, apart from single-municipality market segments where it is the actual population in 2018. Household income refers to average household disposable income in the region in 2018. Both x- and y-scales are in natural logarithms. Urban segments refer to the 15 locations where the segment is a given city. Non-urban segments refer to the 18 locations where the remaining areas (outside of the 15 largest cities) are classified into grous based on administrative regions. Circle size indicates the model-consistent number of houses in each segment.

	(1) σ	(2) σ	(3) σ	(4) σ
Municipality size	0.430			0.127
	(0.178)			(0.0734)
Population growth		0.580		0.697
		(0.161)		(0.192)
Household income			-0.0457	-0.434
			(0.0788)	(0.124)
small	-0.0270	-0.0270	-0.0270	-0.0270
	(0.226)	(0.204)	(0.250)	(0.183)
Observations	66	66	66	66

Table A5: Associations of the housing quality parameter with some market segment characteristics.

Notes. The table documents coefficients from an unweighted linear regression of the outcome on indicated dependent variables. Municipality size refers to average municipality population in the market segment, apart from single-municipality market segments where it is the actual population in 2018. Population growth refers to total region population change from 2016 to 2018 relative to 2016 population. Household income refers to average household disposable income in the region in 2018. Both the independent and the three continuous dependent variables are standardised to have mean 0 and variance of 1. ”small” is an indicator variable which takes value 1 for the market segments of apartments of two rooms or less. Heteroscedasticity-robust standard errors are in parentheses.

	(1) δ	(2) δ	(3) δ	(4) δ
Municipality size	0.525			0.333
	(0.148)			(0.123)
Population growth		0.538		0.408
		(0.142)		(0.203)
Household income			0.120	-0.219
			(0.0904)	(0.150)
small	0.612	0.612	0.612	0.612
	(0.198)	(0.196)	(0.236)	(0.186)
Observations	66	66	66	66

Table A6: Associations of the housing quality parameter with some market segment characteristics.

Notes. The table documents coefficients from an unweighted linear regression of the outcome on indicated dependent variables. Municipality size refers to average municipality population in the market segment, apart from single-municipality market segments where it is the actual population in 2018. Population growth refers to total region population change from 2016 to 2018 relative to 2016 population. Household income refers to average household disposable income in the region in 2018. Both the independent and the three continuous dependent variables are standardised to have mean 0 and variance of 1. ”small” is an indicator variable which takes value 1 for the market segments of apartments of two rooms or less. Heteroscedasticity-robust standard errors are in parentheses.

	(1) λ	(2) λ	(3) λ	(4) λ
Municipality size	0.0351			0.0705
	(0.0809)			(0.0800)
Population growth		0.0161		0.0244
		(0.110)		(0.168)
Household income			-0.0793	-0.121
			(0.0548)	(0.126)
small	1.429	1.429	1.429	1.429
	(0.173)	(0.174)	(0.172)	(0.174)
Observations	66	66	66	66

Table A7: Associations of the housing quality parameter with some market segment characteristics.

Notes. The table documents coefficients from an unweighted linear regression of the outcome on indicated dependent variables. Municipality size refers to average municipality population in the market segment, apart from single-municipality market segments where it is the actual population in 2018. Population growth refers to total region population change from 2016 to 2018 relative to 2016 population. Household income refers to average household disposable income in the region in 2018. Both the independent and the three continuous dependent variables are standardised to have mean 0 and variance of 1. ”small” is an indicator variable which takes value 1 for the market segments of apartments of two rooms or less. Heteroscedasticity-robust standard errors are in parentheses.