Reference Dependence in the Housing Market*

Steffen Andersen Cristian Badarinza Lu Liu
Julie Marx and Tarun Ramadorai†

May 13, 2020

Abstract

We model listing decisions in the housing market, and structurally estimate household preference and constraint parameters using comprehensive Danish data. Sellers optimize expected utility from property sales, subject to down-payment constraints, and internalize the effect of their choices on final sale prices and time-on-the-market. The data exhibit variation in the listing price-gains relationship with "demand concavity;" bunching in the sales distribution; and a rising listing propensity with gains. A new fact is that gains and down-payment constraints have interactive effects on listing prices. We find reference-dependence around the nominal purchase price and modest loss-aversion, but our canonical model cannot fully explain the new facts.

*We thank Jan David Bakker, Nick Barberis, Richard Blundell, Pedro Bordalo, John Campbell, João Cocco, Joshua Coval, Stefano DellaVigna, Andreas Fuster, Nicola Gennaioli, Arpit Gupta, Adam Guren, Chris Hansman, Ralph Koijen, Ulrike Malmendier, Atif Mian, Karthik Muralidharan, Tomek Piskorski, Andrei Shleifer, Jeremy Stein, Ansgar Walther, Joshua White, Toni Whited, and seminar participants at the FCA-Imperial Conference on Household Finance, CBS Doctoral Workshop, Bank of England, King's College Conference on Financial Markets, SITE (Financial Regulation), CEPR Household Finance conference, UC Berkeley, Rice University, UT Dallas, Southern Methodist University, University of Chicago, and University of Michigan for useful comments.

†Andersen: Copenhagen Business School, Email: san.fi@cbs.dk. Badarinza: National University of Singapore, Email: cristian.badarinza@nus.edu.sg. Liu: Imperial College London, Email: l.liu16@imperial.ac.uk. Marx: Copenhagen Business School, Email: jma.fi@cbs.dk. Ramadorai (Corresponding author): Imperial College London, Tanaka Building, South Kensington Campus, London SW7 2AZ, and CEPR. Tel.: +44 207 594 99 10, Email: t.ramadorai@imperial.ac.uk.

1 Introduction

Housing is typically the largest household asset, and mortgages, typically the largest liability (Campbell, 2006, Badarinza et al. 2016). Given these high stakes, household decisions in the housing market have been a rich source of field evidence on economic agents' preferences, beliefs, and constraints. A prominent example is the highly-cited finding that listing prices for houses rise sharply when their sellers face nominal losses relative to the initial purchase price—prima facie field evidence of reference-dependent loss-aversion (Kahneman and Tversky, 1989, Köszegi and Rabin, 2006, 2007).[1]

The controlled environment of the lab allows for precise investigation that is uncontaminated by the frictions and distractions that agents face in the field. To reclaim some of this precision in field settings, DellaVigna (2009, 2018) calls for a *structural behavioral* approach to more carefully and realistically map reduced-form facts to deep structural parameters. We therefore set up a new model of household selling decisions which incorporates realistic housing market frictions. We structurally estimate the parameters of the model using a large and granular administrative dataset which tracks the entire stock of Danish housing, and the universe of Danish listings and housing transactions between 2009 and 2016, matched to household demographic characteristics and financial information. These rich data also yield several new facts about household decisions that we cannot match using canonical model features, making them targets for future theoretical work.

In our model, sellers face both an extensive margin decision of whether or not to list, and an intensive margin choice of the listing price. Sellers receive expected utility, modelled à la Köszegi and Rabin (2006), both from the final sale price of the property as well as (potentially asymmetrically) from any gains or losses relative to a fixed reference price, which we simply set to the nominal purchase price of the property. Sellers enjoy

[1] Genesove and Mayer (2001) present early and carefully documented evidence from the Boston housing market. More recent evidence includes Ferreira et al. 2010, Anenberg, 2011, Schulhofer-Wohl, 2012, Hong et al. 2016, and Bracke and Tenreyro 2018.

additional "gains from trade" when listings successfully convert to final sales, receive an outside option utility value otherwise, and face down-payment constraints à la Stein (1995). When making their decisions, sellers take as given the impacts of the chosen listing price on both the probability of sale and the final sale price, i.e., they are (realistically) constrained by the impacts of their listing decisions on final outcomes, given the behavior of demand in the housing market. The model yields a number of important insights, which together suggest a more structured investigation of the moments in the data. While we describe specific predictions more fully in the paper, we summarize a few important ones here, as well as how they map to the data.

First, a model with "linear reference dependence" (without loss aversion) can be separately identified from one with pure final utility of wealth (i.e., the final sales price) by simply inspecting how listing prices vary with gains. The key is that the reference price alters the wedge between the utility in the event of a sale and the outside option. Intuitively, for a reference-dependent household, gains in the event of a sale are additive to the utility from the final sale price, and provide an additional incentive to lower listing prices to increase the likelihood of a successful transaction, and vice versa for losses.

Second, we confirm in our model that the kink in preferences characterizing loss aversion generates more than one kink in listing prices. Sellers just to the left of the reference point have a strong incentive to increase their listing price, aiming for marginally positive realized gains. But for sellers that face more substantial losses, such "fishing" for higher prices is limited because higher listing prices imply a higher probability that the transaction will not go through.

In the data, the listing price schedule has a characteristic "hockey stick" shape first identified by Genesove and Mayer (2001), rising substantially as expected losses mount, and being virtually flat in gains. Our empirical estimates are similar in magnitude to those estimated in that paper despite the differences in sample period, location, and sample size.[2] At first glance, they are consistent with a significant degree of reference dependence, at least in the loss domain.[3]

Third, the shape of demand is very important for model outcomes. If sale probabilities respond linearly and negatively to higher listing prices, there are material incentives to set low list prices to induce quick sales. However, Guren (2018) shows in the U.S. that housing markets are characterized by "concave demand," i.e., past a certain level, lowering list prices does not boost sale probabilities, but simply negatively impacts realized sale prices; we confirm this finding in the Danish data.[4] The insight that we derive from the model is that seller optimization when demand is nonlinear *can induce kinks in the optimal listing price schedule with no need for kinks in preferences such as loss aversion.* Intuitively, reference-dependent sellers expecting gains set low list prices in a world with linear demand, but in a world with concave demand, there is little reward for lowering list prices past a point.

We validate this insight in the data by comparing regional housing markets in Denmark with varying degrees of demand concavity—and find that the slopes of the listing price responses to gains and losses vary just as the model predicts. We adopt an instrumental variables approach to confirm this fact, using regional variation in the homogeneity of the Danish housing stock. Intuitively, a more homogeneous "cookie-cutter" housing stock makes valuation more transparent, and should therefore increase demand sensitivity to listing prices. In accordance with this intuition, there is visible evidence of more pronounced demand concavity in more homogeneous sub-markets, and we can precisely

[2]In the original Genesove and Mayer sample of Boston condominiums between 1990 and 1997 [N=5,792], list prices rise between 2.5 and 3.5% for every 10% nominal loss faced by the seller. We find rises of 4.4 and 5.3% for the same 10% nominal loss in the Danish data of apartments, row houses, and detached houses between 2009 and 2016 [N=175,646].

[3]As Genesove and Mayer (2001), Clapp et al. (2018), and others note, variation in unobservable quality, as well as potential under- or over-payment at the time of house purchase are important sources of measurement error. We adopt a wide range of strategies to check robustness to this possible confound, including property-specific fixed effects estimation, bounding strategies previously proposed in the literature (Genesove and Mayer, 2001), an instrumental variables approach proposed by Guren (2018), and a Regression Kink Design (Card et al., 2015b).

[4]We also show using these data that there are substantial increases in the *volatility* of time on the market associated with higher listing premia, a new and important observation.

predict a less steeply sloped listing price schedule using this instrument.

Fourth, the model reconfirms an issue recognized in prior work (e.g., Genesove and Mayer, 1997, 2001), that downsizing aversion à la Stein (1995) can be difficult to separate from reference-dependent loss aversion. This is because down-payment constraints on mortgages create an incentive for households to "fish" with higher listing prices, since household leverage magnifies the effect of declines in collateral value, severely compressing the size of houses into which households can move. We find strong evidence of this effect of household leverage on listing prices in the Danish data, but document sufficient variation to separately identify these two effects from one another, i.e., there is a large enough share of both "unconstrained losers" and "constrained winners" in the data, a result of cross-sectional and time-series variation in reference prices, variation in initial leverage, and subsequent (cash-out) refinancing and remortgaging decisions taken by households following initial home purchase.

Fifth, household listing behavior also has material implications for quantities. This means that we can also identify some of the parameters of the model using bunching in the house sales distribution, a strategy recently proposed and implemented for identifying reference dependence in a range of field settings (see, e.g., Kleven, 2016, and Rees-Jones, 2018). In the data, we see clear bunching in the realized sales distribution when gains since purchase are just positive, and "holes" in this distribution just to the left of this point; prima facie evidence of reference-dependent loss-aversion around the nominal purchase price.[5]

Finally, while bunching in the distribution of house sales captures *ex post*-negotiation outcomes, the model shows that *extensive margin* decisions better capture sellers' *ex ante* listing behavior in the extent to which selling propensities vary with expected gains. This can be identified by estimating listing propensities for the entire Danish housing stock of

[5]As we note later in the paper, nonlinearities arising from demand concavity can also generate bunching in the realized sales distribution, which can be hard to distinguish from bunching arising from reference-dependent loss aversion.

over 1.7 million housing units as a function of expected gains.[6] When we undertake this analysis in the data, we find that there is a slight but visible increase in the propensity to list as expected gains rise.

We estimate the seven parameters in our model using seven selected moments from the data, including those described above, using classical minimum distance estimation in our exactly identified system. The resulting parameter point estimates yield $\eta = 0.981$ (*s.e.* 0.312), meaning that gains count about as much as final prices for final utility, and $\lambda = 1.525$ (*s.e.* 0.422), a modest degree of loss aversion, lower than early estimates between 2 and 2.5 (e.g., Kahneman et al. 1990, Tversky and Kahneman, 1992), but closer to those in more recent literature (e.g., Imas et al. 2016 find $\lambda = 1.59$). The role of concave demand is very important for these parameter estimates—in a restricted model in which we assume that demand is (counterfactually) linear, estimated $\eta = 0.750$ (*s.e.* 0.291) and $\lambda = 3.285$ (*s.e.* 0.867). This strongly reinforces a broader message (see, e.g., Blundell, 2017) that realistic frictions need to be incorporated when mapping reduced-form facts to inferences about deeper underlying parameters.[7]

The parameters reveal strong evidence of the down-payment channel originally identified by Stein (1995), and we confirm this evidence using separate estimates of financial constraints from the distribution of household net financial wealth in Denmark. The parameter estimates also show that there are significant "gains from trade" from successful house listings, and substantial (psychological and transactions) costs associated with listing, as we later discuss.

The model does a good job of matching the selected moments with plausible preference parameters. However as an out-of-sample exercise, we conduct a broader evaluation of

[6]Another critical issue here is that modeling the extensive margin decision is a good way to explicitly account for any selection effects that may drive patterns of observed *intensive margin* listing premia in the data. We discuss this in detail in the model section below.

[7]This also highlights that frictions in matching in the housing market are another important part of the explanation for the positive correlation between volume and price observed in housing markets, an original motivation for the mechanisms identified by both Stein (1995) and Genesove and Mayer (2001) which our model also incorporates.

how the model matches the *entire surface* of the listing premium along the home equity and gains dimensions. A novel pattern that we find in the data is that home equity and expected losses have *interactive* effects on listing prices in this market. The canonical model that we set up is unable to match these new facts, which we view as targets for future theoretical work.

To be more specific, when home equity levels are low, i.e., when down-payment constraints are tighter, households set high listing prices that vary little around the nominal loss reference point. In contrast, households that are relatively unconstrained set listing prices that are significantly steeper in expected losses. Households' listing price responses to down-payment constraints are also modified by their interaction with nominal losses. Mortgage issuance by banks in Denmark is limited to an LTV of no greater than 80%,[8] and for households facing nominal losses since purchase, listing prices rise visibly as home equity falls below this down-payment constraint threshold of 20%. But for households expecting nominal gains, there is a strong upward shift in this constraint threshold (i.e., to values above 20%) in the *level* of home equity at which they raise listing prices. We discuss these findings and conjecture mechanisms to explain them towards the end of the paper; we view them as important targets for future theoretical work that is beyond the scope of our analysis in this paper.

The paper is organized as follows. Section 2 introduces the model of household listing behavior. Section 3 discusses the construction of our merged data set, and provides descriptive statistics about these data. Section 4 estimates the moments that we use for structural estimation and uncovers new facts about the behavior of listing prices and listing decisions. Section 5 describes our structural estimation procedure, and reports our parameter estimates. Section 6 describes validation exercises including our instrumental variables analysis, and highlights areas where the model falls short in explaining features of the data. Section 7 concludes.

[8]We later describe the precise institutional features of the Danish setting, which permits additional non-mortgage borrowing at substantially higher rates for higher LTV mortgages.

2 A Model of Household Listing Behavior

We set up a model in which a household (the "seller"), optimally decides on a listing price (the "intensive margin"), as well as whether or not to list a house (the "extensive margin"). The model flexibly embeds different preferences and constraints that have commonly been used to explain patterns in listing behavior. We then structurally estimate the model to recover the preference and constraint parameters that best match the data.

2.1 Setup

The market consists of a continuum of sellers and buyers of residential property. For simplicity, we consider a model with two periods. In period 0, a subset of property owners receive a "moving" shock $\theta \sim \text{Uniform}(\theta_{\min}, \theta_{\max})$, and decide whether or not to put a property up for sale on the market, and the optimal listing price in case of listing. We model θ as a "gain from trade" (Stein, 1995), i.e., a boost to lifetime utility associated with successfully moving, which captures a variety of reasons for moving, including labor market moves to opportunity, or the desire to upsize arising from a newly expanded family.

In period 1, buyers visit properties that are up for sale. If the resulting negotiations succeed, the property is transferred to the buyer for a market-clearing price. If negotiations fail, the seller stays in the property, with a constant level of utility $\underline{u}$.

Our goal is to explain the structural relationship between listing decisions and seller preferences and constraints. To sharpen this focus, we model buyer decisions and equilibrium negotiation outcomes in reduced-form, and focus on recovering seller "policy functions" from this setup. In particular, let α denote the probability that a willing buyer will be found, and P the final sale price resulting from the negotiation between buyer and seller. Building on Guren (2018), we note that sufficient statistic formulas (Chetty, 2009) for equilibrium outcomes are mappings between sale probabilities $\alpha(\ell)$, final sale prices

$P(\ell)$, and the listing premium $\ell = L - \widehat{P}$,[9] i.e. the difference between the listing price L set by the seller, and some measure of the "expected" property value $\widehat{P}$.[10] In particular, the realized premium of the final sales price $P(\ell)$ over the expected property value $\widehat{P}$, and the probability of a quick sale $\alpha(\ell)$ arise from the seller's listing behavior, and the subsequent negotiation process with the buyer.

A typical seller's decision in period 0 can be written as:

$$\max_{s\in\{0,1\}} \left\{ (s) \max_{\ell} \underbrace{\left[\alpha(\ell)\left(U\left(P(\ell),\cdot\right)+\theta\right)+(1-\alpha(\ell))\underline{u}-\varphi\right]}_{EU(\ell)} + (1-s)\underline{u} \right\} \tag{1}$$

For a property that is listed, we distinguish between two possible outcomes in period 1, conditional on the level of the listing premium. With probability $\alpha(\ell)$ the negotiation succeeds, and the seller receives utility from selling the property for an equilibrium price $P(\ell)$. With probability $1-\alpha(\ell)$ the listing fails, in which case the seller falls back to their outside option level of utility $\underline{u}$. In addition, owners who decide to list incur a one-time cost φ, which is sunk at the point of listing—all utility costs associated with listing (e.g., psychological "hassle factors", search, listing and transaction fees) are captured by this single parameter.

The seller decides on the extensive margin of whether ($s = 1$) or not ($s = 0$) to list, and a listing premium ℓ to maximize expected utility from final sale of the property. When so doing, the seller takes $\alpha(\ell)$ and $P(\ell)$, i.e., the "demand" functions, as given;[11]

[9] In the model solution and calibration exercise, we normalize $\widehat{P}$ to 1. All model quantities can therefore be thought of as being expressed in logs, consistent with the definitions of gains/losses and home equity used in our empirical work.

[10] Guren (2018) assumes that the buyer's expected or "reference" value is given by the average listing price in a given zip code and year. This allows for more flexibility, allowing listing prices to systematically deviate from hedonic/fundamental property values across time and locations. We begin with a simpler benchmark, setting $\widehat{P}$ to the fundamental value of the house in the interests of internal consistency of the model. As we show later, this distinction does not play a significant role in our empirical work, as Denmark has a relatively homogenous and liquid housing market, and we show that the listing premium based on hedonic prices more strongly predicts a decrease in the probability of sale than the alternative based on average listing prices in a direct comparison in the online appendix.

[11] As we describe later, we flexibly allow for the seller to perceive $\alpha(\ell)$ differently from the (ex-post) estimated mapping function in the data by adding a parameter δ to the model, i.e., the seller essentially

we estimate these functions in the data as a reduced-form for equilibrium outcomes in the negotiation process in period 1, which the seller internalizes when optimizing utility. This assumption simplifies the model, and allows us to more closely focus on our goal, namely, the underlying parameters of seller utility and constraints.

The functions $\alpha(\ell)$ and $P(\ell)$ restrict the seller's action space, and capture the basic tradeoff that sellers face: a larger ℓ can lead to a higher ultimate transaction price, while at the same time it decreases the probability that a willing buyer will be found within a reasonable time frame.[12] These points capture the link between listing premia, final realized sales premia, and time-on-the-market or TOM originally detected by Genesove and Mayer (2001). In the remainder of the paper, we refer to these two functions $\alpha(\ell)$ and $P(\ell)$ collectively as *concave demand*, following Guren (2018), who documents using U.S. data that above average list prices increase time-on-the market or TOM (i.e., they reduce the probability of final sale), while below average list prices reduce seller revenue with little effect on TOM. We find essentially the same pattern in the Danish data.

We turn next to describing the components of $U(P(\ell), \cdot) = u(P(\ell), \cdot) - \kappa(P(\ell), \cdot)$, which allows us to nest a range of preferences $u(P(\ell), \cdot)$, including reference-dependent loss-aversion à la Kahneman and Tversky (1979) and Koszegi and Rabin (2006, 2007), as well as down-payment constraints à la Stein (1995) $\kappa(P(\ell), \cdot)$.

maximizes subject to an $(\alpha(\ell) + \delta)$ probability.

[12]In our estimation, we define a *period* as equal to six months. In this case, the function $\alpha(\ell)$ captures the probability that the transaction goes through within a time frame of six months after the initial listing.

2.2 Reference Dependent Loss Aversion

We adopt a standard formulation of reference-dependent loss averse preferences as in Section IV of Koszegi and Rabin (2006), writing $u(P(\ell), \cdot)$ as:

$$u(P(\ell), R) = \begin{cases} P(\ell) + \lambda\eta G(\ell), \text{ if } G(\ell) < 0 \\ P(\ell) + \eta G(\ell), \text{ if } G(\ell) \geq 0 \end{cases}. \tag{2}$$

In equation (2), we simply set the reference price level to R. Realized gains $G(\ell)$ relative to this reference level are given by $G(\ell) = P(\ell) - R$. Throughout the paper, we simply assume that R is fixed, and in our empirical application, we simply set R at the original nominal purchase price of the property.[13]

The parameter η captures the degree of reference dependence. Sellers derive utility both from the terminal value of wealth (i.e. the final price realized from the sale of the house), as well as from the realized gain relative to the reference price R.

The parameter $\lambda > 1$ governs the degree of loss aversion. This specification of the problem assumes that utility is piecewise linear in nominal gains and losses relative to the reference point, with a kink at zero, and has been used widely both in the lab (e.g., Ericson and Fuster, 2011), and in the field (e.g., Anagol et al., 2018).

This utility specification nests several possibilities depending on parameter values. When $\eta = 0$, we recover final utility of wealth $u(P(\ell), R) = P(\ell)$. When $\eta > 0$ and $\lambda = 1$ we are in a "linear reference-dependence" framework, i.e., $u(P(\ell), R) = P(\ell) + \eta G(\ell)$.[14] We discuss the implications of these different parameter values on optimal listing premia ℓ^* from the model later in this section.

[13] While this is a restrictive assumption, we find strong evidence to suggest the importance of this particular specification of the reference point in our empirical work. We follow Blundell (2017), trading off a more detailed description of the decision-making problem in the field against stronger assumptions that permit measurement of important underlying parameters.

[14] While at first glance this model might seem to simply be a scaled version of $u(P(\ell), R) = P(\ell)$, as we later discuss, the presence of the outside option $\underline{u}$ means that these models generate very different predictions from one another.

2.3 Down-payment Constraints

We now turn to describing $\kappa(P(\ell), \cdot)$. Let M denote the level of the household's outstanding mortgage, and γ the required down-payment on a *new* mortgage origination. For a given price level $P(\ell)$, the "realized" home equity position of the household is $H(\ell) = P(\ell) - M$. Under the assumption that H is put towards the downpayment on the next home, we can distinguish between constrained (i.e., downsizing averse) households for which $H(\ell) < \gamma$, and unconstrained households for which $H(\ell) \geq \gamma$.

In the face of binding down-payment constraints, only unconstrained sellers can move to another property of the same or greater value. However, there are several ways in which households could relax these constraints despite legal restrictions on LTV at mortgage initiation (which, as we discuss later, are strictly set at 20% in Denmark). The first, obvious way is for the household to downsize to a less expensive home than $P(\ell)$, or indeed, to move to the rental market, which might incur a utility cost. The second is that households can engage in non-mortgage borrowing to fill the gap $\gamma - H(\ell)$. A common approach in Denmark is to borrow from a bank or occasionally from the seller of the property to bridge funding gaps between 80% and 95% loan-to-value (LTV).[15] A third (typically unobservable) possibility is that households can bring additional funds to the table by liquidating other assets.[16] We therefore assume that violating the down-payment constraint does not lead the seller to withdraw the sale offer. Rather, the seller incurs a monetary penalty of μ for every unit by which realized home equity drops below the

[15] Danish households can borrow using "Pantebreve" or "debt letters" to bridge funding gaps above LTV of 80%. Over the sample period, this was possible at spreads of between 200 and 500 bp over the mortgage rate. For reference, see categories *DNRNURI* and *DNRNUPI* in the Danmarks Nationalbank's statistical data bank.

[16] In Stein (1995), M represents the outstanding mortgage debt net of any liquid assets that the household can put towards the downpayment. The granular data that we can employ allow us to measure the net financial assets that households can bring to the table to supplement realized home equity. We later verify using these data that our inferences are sensible when considering these additional funds.

constraint threshold:[17]

$$\kappa(P(\ell)) = \begin{cases} \mu(\gamma - H(\ell)), \text{ if } H(\ell) < \gamma \\ 0, \text{ if } H(\ell) \geq \gamma \end{cases}. \tag{3}$$

2.4 State Variables

In the model, besides the moving shock θ, seller decisions are conditional on two state variables, namely, the reference point R, and the size of the outstanding mortgage M. To more directly map these to quantities that we estimate in the data, and to make our setup more directly comparable to extant empirical and theoretical literature, we normalize these two variables by the property's fundamental (in our empirical application, hedonic) value, and employ the seller's expected or "potential" gains $\widehat{G} = \widehat{P} - R$, and similarly, "potential" home equity $\widehat{H} = \widehat{P} - M$ as state variables instead. Realized gains $G(\ell)$ and home equity $H(\ell)$ are their "potential" levels plus a markup/premium which is the outcome of the seller's listing behavior and the negotiation process in the market, i.e., mediated by $\alpha(\ell)$ and $P(\ell)$.

We next discuss selected predictions of this model to build intuition, and to help choose key moments of the data with which we estimate the model's structural parameters.

2.5 Model Predictions

2.5.1 Preference Parameters and Optimal Listing Premia

The first-order condition determining the optimal ℓ^* balances the marginal utility benefit (MB) of a higher premium, conditional on a successful sale, against the marginal cost (MC) of an increased chance of the transaction failing, and the consequent fall to the

[17] i.e.,

$$U(P(\ell)) = \begin{cases} u(P(\ell) - \mu(\gamma - H(\ell)), \text{ if } H(\ell) < \gamma \\ u(P(\ell)), \text{ if } H(\ell) \geq \gamma \end{cases}.$$

outside option utility level:

$$MB = \alpha(\ell)U'(P(\ell)) = \alpha'(\ell)[\underline{u} - \theta - U(P(\ell))] = MC. \tag{4}$$

We now outline how equation (4) is affected by different parameter values in equation (2). (For the purposes of this discussion, we assume that $\mu = 0$ in equation (3) to focus on preferences.) If utility stems purely from the terminal house price (i.e. $\eta = 0$), ℓ^* will depend on the hedonic value of the property $\widehat{P}$ and the size of the moving shock θ, but will be invariant to the reference level R. However, the case of linear reference dependence ($\eta > 0$, $\lambda = 1$) is different. While R does not affect MB in this case, R does affect MC by altering the distance between $\underline{u}$ and the achievable utility level in the event of a successful transaction. Intuitively, if the household can realize a gain from a successful transaction (i.e., when R is sufficiently low), the utility from sale will be higher. The resulting ℓ^* will be lower, as the household seeks to increase the probability of a successful transaction. The opposite is true when the household faces a loss (i.e., when R is sufficiently high), which results in a higher ℓ^*.[18]

In the case of reference-dependent loss-aversion ($\eta > 0$ and $\lambda > 1$), the kink in the piecewise linear utility function leads to a piecewise linear pattern in ℓ^* at $\widehat{G} = 0$. We illustrate the intuition behind this result in Figure 1, for a simplified case in which demand is linear.

Taking as given the trade-off they face between prices and the probability of a successful sale, sellers assign an optimal listing premium for each level of potential gains $\widehat{G}$. The listing premium they set determines the gains that they ultimately realize. Importantly, there exists a level of potential gains $\widehat{G}_0$ which maps to a realized gain of *exactly*

[18]Note that it is critical to assume that households do not receive utility from simply living in a house that has appreciated relative to their reference point R, i.e. they do not enjoy utility from "paper" gains until they are realized. If this condition does not hold, the model is degenerate in that R is irrelevant both for the choice of the listing premium (intensive margin) and the decision to list (extensive margin). We demonstrate this result analytically in the online appendix.

zero. Sellers with potential gains below $\widehat{G}_0$ will want to avoid realizing a loss, meaning that they slightly adjust their optimal listing premium upwards. This adjustment works well (i.e., results in an acceptable probability of a failed sale) until a lower limit $\widehat{G}_1$ is reached. Beyond this point, it is sub-optimal to aim for a realized gain of zero (despite the pain associated with a loss), so the seller has no choice but to accept the loss. In the online appendix, we work through this intuition analytically. Some broad implications include that the slope of ℓ^* along the $\widehat{G}$ dimension helps to identify the extent of reference dependence η, and slope differences below and around $\widehat{G} = 0$ help to identify λ.

2.5.2 Concave Demand

The demand functions $\alpha(\ell)$ and $P(\ell)$ are a critical determinant of listing behavior and the expected shape of ℓ^* in this model. To build intuition, consider the effect of linear demand functions $\alpha(\ell) = \alpha_0 - \alpha_1 \ell$ and $P(\ell) = \widehat{P} + \beta_0 + \beta_1 \ell$ in a linear reference-dependent utility model (with $\eta = 1$, $\lambda = 1$) without financial constraints ($\mu = 0$). In this case:

$$\ell^* = \frac{1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0 + \theta}{\beta_1} - \frac{1}{\beta_1}\widehat{G}\right). \tag{5}$$

Equation (5) shows that when the probability of sale is less sensitive to ℓ (i.e., when α_1 is lower), in accordance with intuition, ℓ^* is higher.[19]

This intuition carries over to a case in which $\alpha(\ell)$ has the concave shape first identified by Guren (2018), and has important and interesting implications for the observed shape of ℓ^* as $\widehat{G}$ varies. Figure 2 graphically illustrates this mechanism, allowing variation in $\alpha(\ell)$ around $\underline{\ell} = 0$, i.e., the point at which $L = \widehat{P}$.

In Panel A of Figure 2, consider a seller facing gains $\widehat{G} > 0$. From equation (4), low listing premia are optimal for this seller, to avoid hitting the lower outside option $\underline{u}$ associated with a failed listing. However, when the seller is faced with concave demand

[19] We derive the equation explicitly in the online appendix.

(i.e., the flattening out in $\alpha(\ell)$ when $\ell < \underline{\ell}$, continuing to assume that $P(\ell) = \beta_0 + \beta_1\ell$), lowering ℓ below $\underline{\ell}$ does not boost the sale probability $\alpha(\ell)$, but it does continue to negatively impact the realized sale price $P(\ell)$, meaning that ℓ^* "flattens out" in this world.

The tradeoff faced by sellers facing losses $\widehat{G} < 0$ is different—they set a higher listing premium to offset these losses conditional on a willing buyer being found. This is traded off against a lower probability that such a buyer will be found at all. A stronger demand response to ℓ makes this type of "fishing" more costly. The bottom panel of Figure 3 illustrates this force—a steeper buyer response to differences between ℓ and $\underline{\ell}$ (i.e., increased demand concavity) in turn leads to an attenuation of the slope of ℓ^*. In the extreme case in which concave demand has an infinite slope around some level of the listing premium, rational sellers' ℓ^* collapses to a constant—which would be observationally equivalent to the case in which sellers are not reference dependent at all ($\eta = 0$).

The main predictions from the model in this case are that first, the optimal ℓ^* in a linear reference dependent model ($\eta > 0$, $\lambda = 1$) in the presence of concave demand exhibits a flatter slope as gains rise, relative to the case of linear demand. The resulting ℓ^* profile along the dimension of realized gains G shows a characteristic "hockey stick" shape of the type detected by Genesove and Mayer (2001), but for very different reasons. Second, the model predicts a tight link between the shape of $\alpha(\ell)$ and the slope of ℓ^*. We use this insight and cross-sectional variation in the concavity of demand across different segments of the Danish market to aid structural parameter identification. For example, if $\eta = 0$, in this model, demand concavity does not affect the slope of the ℓ^* profile along the G dimension. In contrast, a high η leads to a high "pass-through" of demand concavity into optimal listing premia.

Finally, we add a parameter δ to the model, which allows us to estimate the extent to which sellers perceive the constraints imposed on their behavior by demand concavity estimated in the data. In essence, as we describe later, this addition means that sellers perceive the slope to the right of $\underline{\ell}$ as the slope estimated in the data plus a factor δ. Sellers in our model are explicitly allowed to have reference-dependent preferences—which provides an incentive for them to set listing premia that are far away from $\underline{\ell}$. In Guren's (2018) model, demand concavity provides a strong incentive for sellers to stay close to the "average" listing premium, a force he refers to as strategic complementarity. We can view δ as one way to assess the relative strength of these forces, namely, reference dependence and strategic complementarity—a high (negative) level of δ would be associated with sellers perceiving high costs for deviating from $\underline{\ell}$, and vice versa.[20]

2.5.3 Bunching in the Distribution of House Sales

Household listing behavior also has material implications for quantities. Non-linearities in ℓ^* along the G dimension will also be reflected in the degree to which transactions complete, as well as on the final price at which these transactions occur.

In a model with $\eta > 0$, sellers with $G < 0$ choose relatively higher ℓ^*. This has two consequences. First, the likelihood that willing buyers will be found is lower, given $\alpha(\ell)$, and the consequent likelihood of observing such transactions is lower. Second, if the transactions do go through, the associated realized gains on these transactions will be higher, shifting mass in the final sales distribution towards sales with realized $G > 0$. This effect is especially pronounced if sellers are loss averse, i.e., when $\lambda > 1$, in which case we should observe a jump in the final distribution of house sales precisely at $G = 0$ with greater mass in this distribution just to the right of this point and less mass just to the left of this point. The online appendix presents a figure which illustrates this force.

It is also worth noting that concave demand generates a non-linear listing premium profile for the reasons documented above. This has similar effects on ℓ^*, as this force pushes sellers not to lower listing premia below a point. This asymmetry also generates bunching towards positive values of realized gains, which will depend on the level of $\underline{\ell}$.

[20]As we show later, high listing premia are also associated greater volatility in TOM, meaning that δ can also be perceived as a way to assess sellers' degree of risk aversion in this domain.

A subtle point here is that any change in the precise specification of the reference point R in the presence of loss aversion will change the location at which bunching is observed. Indeed, heterogeneity in reference points will make it hard to observe the precise location of bunching. To complicate matters further, variations in the level of $\underline{\ell}$ are a confound, rendering it difficult to distinguish models with heterogeneous reference points from models with spatial or temporal variation in $\underline{\ell}$, the point at which demand concavity kicks in. We avoid this complexity in our setup by simply taking the stance that R is the nominal purchase price of the property and evaluating the extent to which we see bunching given this assumption. As we will later see, this is not an outlandish assumption—we observe significant evidence in the data of bunching at this point, suggesting its relevance to sellers.

2.5.4 Extensive Margin

The model also has predictions for listing decisions. Any force that induces a wedge between the expected utility from a successful listing and the outside option $\underline{u}$ can lead the seller to decide that listing is sub-optimal. While sellers will respond to such a wedge by adjusting their decisions along the *intensive margin*, they may also find that beyond a certain level of the listing premium, the likelihood of the transaction going through is so low that listing is sub-optimal to begin with. In particular, the model predicts that sellers with lower expected gains and lower expected home equity are less likely to find it optimal to list, which delivers clear predictions on the extensive margin.

This is important, because it allows us to exploit the observed listing behaviour of sellers with respect to potential gains as an additional moment to inform the structural estimation of the degree of reference dependence. While bunching in the distribution of realized house sales described in the previous section is useful, it only captures *ex post*-negotiation outcomes. The extensive margin decisions better capture sellers' *ex ante* listing behaviour in response to their potential gains and home equity position.

Another critical issue here is that modeling the extensive margin decision is a good way to explicitly account for any selection effects that may drive patterns of observed *intensive margin* listing premia in the data, an issue that the prior literature (e.g., Genesove and Mayer, 1997, 2001, Anenberg, 2011, Guren, 2018) has been unable to control for as a result of data limitations. For example, if sellers that decide not to list are more conservative (i.e., they set lower listing premia), and those who decide to list are more aggressive (i.e., they set higher listing premia) the resulting selection effect would lead to a higher observed non-linearity in listing premia around reference points that would bias parameter estimates and inferences conducted only using the intensive margin.

The moving shock θ is a key determinant of such selection effects. For any given level of the moving shock, the listing decision is still a simple binary choice, but once we model the entire distribution of shocks, we can directly account for the heterogeneity of listing decisions and calculate average listing premia in the population. By construction, these average listing premia incorporate the endogenous first-stage selection effects and can be immediately mapped onto the data.

2.5.5 Additional Effects of the Extensive Margin on the Intensive Margin

There are more subtle implications of the model which link the extensive and the intensive margin. The parameter θ alters the distance between the outside option and the utility from a successful listing, which pushes the seller to set higher average ℓ. However, this force can move ℓ into regions of concave demand in which the response of buyers is more (or less) pronounced, i.e., the probability of a successful sale is more (or less) responsive to the chosen listing premium because of the nonlinearity in $\alpha(\ell)$. Different levels of θ, therefore, are associated with different magnitudes of the seller's reactions around the relevant preference and constraint thresholds (i.e., R and M) in gains and home equity. Thus, variation in θ will interact with different levels of R and M in a manner which can smooth and blur the kinks in the ℓ^* profile discussed above—both in the model and in the data. The online appendix illustrates this force with a specific example.

We turn next to describing the data and our estimation of key moments as a precursor to structural estimation of the parameters.

3 Data and Descriptive Statistics

Our data span all transactions and electronic listings (which comprise the overwhelming majority of listings) of owner-occupied real estate in Denmark between 2009 and 2016. In addition to listing information, we also acquire information on property sales dates and sales prices, the previous purchase price of the sold or listed property, hedonic characteristics of the property, and a range of demographic characteristics of the households engaging in these listings and transactions, including variables that accurately capture households' financial position at each point in time. In our empirical work, we also combine the data in the Danish housing register with the listings data to assess the determinants of the extensive margin listing decision for all properties in Denmark over the sample period. That is, we can also assess the fraction of the total housing *stock* that is listed, conditional on functions of the hedonic value such as potential gains relative to the original purchase price, or the owner's potential level of home equity.

We link administrative data from various sources; all data other than the listings data are made available to us by Statistics Denmark. We briefly describe these data below; the online appendix contains detailed information about the sources, construction, and matching involved in assembling the dataset.

3.1 Data Sources

3.1.1 Property Transactions and Other Property Data

We acquire comprehensive administrative data on registered properties, property transactions, property ownership, and hedonic characteristics of properties from the registers of the Danish Tax and Customs Administration (SKAT) and the Danish housing register (Bygnings-og Boligregister, BBR). These data are available from 1992 to 2016. In our hedonic model, described later, we also include the (predetermined at the point of inclusion in the model) biennial property-tax-assessment value of the property that is provided by SKAT, which assesses property values every second year.[21]

Loss aversion and down-payment constraints were originally proposed as explanations for the puzzling *aggregate* correlation between house prices and measures of housing liquidity, such as the number of transactions, or the time that the average house spends on the market. In the online appendix, we show the price-volume correlation in Denmark over a broader period containing our sample period. The plot looks very similar to the broad patterns observed in the US.

3.1.2 Property Listings Data

Property listings are provided to us by RealView (http://realview.dk/en/), who attempt to comprehensively capture all electronic listings of owner-occupied housing in Denmark. We link these transactions to the cleaned/filtered sale transactions in the official property registers, 76.56 percent of transactions have associated listing data.[22] For each property listing, we know the address, listing date, listing price, size and time of any adjustments to the listing price, changes in the broker associated with the property, and the sale or retraction date for the property.

[21] As we describe later, this is the same practice followed by Genesove and Mayer (1997, 2001); it does not greatly affect the fit of the hedonic model, and barely affects our substantive inferences when we remove this variable.

[22] We more closely investigate the roughly 25% of transactions that do not have an associated electronic listing. 10% of the transactions can be explained by the different (more imprecise) recording of addresses in the listing data relative to the registered transactions data. The remaining 15% of unmatched transactions can be explained by: (a) off-market transactions (i.e., direct private transfers between friends and family, or between unconnected households); and (b) broker errors in reporting non-publicly announced listings ("skuffesager") to boligsiden.dk. We find that on average, unmatched transactions are more expensive than matched transactions. Sellers of more expensive houses tend to prefer the skuffesalg option for both privacy and security reasons.

3.1.3 Mortgage Data

To establish the level of the owner's home equity in each property at each date, we obtain data on the mortgage attached to each property from the Danish central bank (Danmarks Nationalbank), which collects these data from mortgage banks. The data are available annually for each owner from 2009 to 2016, cover all mortgage banks and all mortgages in Denmark and contain information on the mortgage principal, outstanding mortgage balance each year, the loan-to-value ratio, and the mortgage interest rate. If several mortgages are outstanding for the same property, we simply sum them, and calculate a weighted average interest rate and loan-to-value ratio for the property and mortgage in question.

3.1.4 Owner/Seller Demographics

We source demographic data on individuals and households from the official Danish Civil Registration System (CPR Registeret). In addition to each individual's personal identification number (CPR), gender, age, and marital history, the records also contain a family identification number that links members of the same household. This means that we can aggregate individual data on wealth and income to the household level.[23] We also calculate a measure of households' education using the average length of years spent in education across all adults in the household. These data come from the education records of the Danish Ministry of Education. We source individual income and wealth data from the official records at SKAT, which hold detailed information by CPR numbers for the entire Danish population.

[23]Households consist of one or two adults and any children below the age of 25 living at the same address.

3.1.5 Final Merged Data

We only keep transactions for which we can measure both nominal losses and home equity. We have transactions data available from 1992 to the present, meaning that we can only measure the purchase price of properties that were bought during or after 1992. Moreover, the mortgage data run from 2009 to 2016. In addition, the sample is restricted to properties for which we know both the ID of the owner, as well as that of the owner's household, in order to match with demographic information. For listings that end in a final sale, we also drop within-household transactions and transactions that Statistics Denmark flag as anomalous or unusual. We restrict our analysis to residential households, in our main analysis dropping listings from households that own more than three properties in total, as they are more likely property investors than owner-occupiers.[24]

In the online appendix, we describe the data construction filters and their effects on our final sample in more detail. Once all filters are applied, the sample comprises 217,028 listings of Danish owner-occupied housing in the period between 2009 and 2016, for both sold (70.6%) and retracted (29.4%) properties, matched to mortgages and other household financial and demographic information. These listings correspond to a total of 193,850 unique households, and 181,020 unique properties. Most households that we observe in the data sell one house during the sample period, but roughly 9% of households sell two houses over the sample period, and roughly 1.5 percent of households sell three or more houses.

In our main analysis we study the 175,646 households that have a mortgage, but we return to the 41,382 households with no mortgage as a robustness check. In addition, we use the entire housing stock (13,305,501 observations of 1,736,172 unique properties) to understand the extensive margin decision of whether to list the properties for sale.

[24]Genesove and Mayer (2001) separately estimate loss aversion for these groups of homeowners and speculators. We simply drop the speculators in this analysis, choosing to focus our parameter estimation in this paper on the homeowners.

3.2 Hedonic Pricing Model

To calculate potential gains and potential home equity, we require a measure of expected price $\widehat{P}$ for each property-year in the data. To do so, we estimate a standard hedonic pricing model on our sample of sold listings and use this model to predict prices for the entire sample of listed properties, including those that are not sold.[25]

The hedonic model predicts the log of the sale price P_{it} of all sold properties i in each year t, using a set of property characteristics:

$$\ln(P_{it}) = \xi + \xi_t + \xi_m + \xi_{tm} + \beta_{ft}\mathbb{1}_{i=f}\mathbb{1}_{t=\tau} + \boldsymbol{\beta}\mathbf{X_{it}} + \boldsymbol{\beta}_{\mathbf{fx}}\mathbb{1}_{i=f}\mathbf{X_{it}} + \Phi(v_{it}) + \mathbb{1}_{i=f}\Phi(v_{it}) + \varepsilon_{it}, \tag{6}$$

where $\mathbf{X_{it}}$ is a vector of property characteristics, namely ln(lot size), ln(interior size), number of rooms, bathrooms, and showers, a dummy variable for whether the property was unoccupied at the time of sale or retraction, the age of the building, dummy variables for whether the property is located in a rural area or a historic area, and ln(distance of the property to the nearest major city).

ξ is a constant, ξ_t are year fixed effects, ξ_m are fixed effects for different municipalities (there are 98 municipalities in Denmark), ξ_{tm} are year cross municipality fixed effects, and $\mathbb{1}_{i=f}$ is an indicator variable for whether the property is an apartment (denoted by f for flat) rather than a house.[26] $\Phi(v_{it})$ is a third-order polynomial of the previous-year tax assessor valuation of the property.[27] We interact the apartment dummy with time

[25]Later in the paper, we also assess the extent to which gains, losses and home equity determine the *decision* to list. We estimate a separate hedonic model for all properties, including all unlisted properties, in order to conduct these additional tests.

[26]In the online appendix, we also include cohort effects ξ_c in the hedonic regression, and continue to find robust evidence of all patterns uncovered in our empirical analysis, showing that intra-cohort variation in gains and losses is also associated with changes in listing premia.

[27]Genesove and Mayer (1997, 2001) also consider tax assessment data in their hedonic model. Importantly, the tax assessment valuation is carried out before the time of the transaction, in some cases even many years before. Until 2013, the tax authority re-evaluates properties every second year. The assessment, which is valid from January 1st each year, is established on October 1st of the prior year. In the years between assessments, the valuation is adjusted by including local-area price changes. This

dummies, as well as with the hedonic characteristics and the tax valuation polynomial, to allow for a different relationship between hedonics and apartment prices.

When we estimate the model, the R^2 statistic equals 0.86 in the full sample.[28] The large sample size allows us to include many fixed effects into the model, helping to deliver a better-fitting model. This helps to ameliorate concerns of noise or unobserved quality in the measure $\widehat{P}$, an important concern when estimating the effects of both loss aversion and home equity (e.g., Genesove and Mayer, 1997, 2001, Anenberg, 2011, Clapp, et al., 2018). We also adopt a number of alternative approaches to deal with the important issue of unobserved quality and its effects on our inferences, as we describe later.

3.3 Gains, Losses, and Home Equity

Armed with the hedonic pricing model, we estimate $\widehat{G}$ and $\widehat{H}$ in percentage terms, i.e., $\widehat{G} = \widehat{\ln P} - \ln R$, where R is set to the nominal purchase price of the property, and $\widehat{H} = \widehat{\ln P} - \ln M$, where M is reported by the household's mortgage bank each year. The online appendix plots the distributions of $\widehat{G}$ and $\widehat{H}$ in the data, both winsorized at the 1% and 99% percentile points (there are several large values given the substantial time elapsed since the purchase of some properties in the data).[29] Mean $\widehat{G}$ in the data is 33% and median $\widehat{G}$ is 26%, and 25% (75%) of property-years have $\widehat{G} < 0$ ($\widehat{G} \geq 0$). Mean $\widehat{H}$ is 26%, median $\widehat{H}$ is 24%, and 75% (25%) of property-years have $\widehat{H} < 0$ ($\widehat{H} \geq 0$). Modal $\widehat{H}$ is close to 21%, which is to be expected, as Denmark has a constraint on the issuance of mortgages—the Danish Mortgage Act specifies that LTV at issuance by mortgage banks

adjustment has been frozen since 2013, recording such price changes as of 2011. Only in the case of significant value-enhancing adjustments to a house or apartment would a re-assessment have taken place thereafter–and once again, is pre-determined at the point of property sale.

[28]The online appendix contains several details about the hedonic model and estimates. We also estimate the model in levels rather than logs, with an R^2 of 0.88. Moreover, the R^2 when we eliminate the tax assessor valuation from the hedonic characteristics is 0.73.To check the robustness of our results to the specification of the hedonic model, we also amend it in various ways as outlined in the appendix. Our results are qualitatively, and for the most part, quantitatively unaffected by these amendments.

[29]$\widehat{H}$ is also winsorized at 100% to restrict it to this level for tiny mortgages, given the log difference approach that we use to compute it.

is restricted to be 80% or lower, though, as mentioned earlier, households can engage in non-mortgage borrowing to fill the gap at substantially higher rates. This constraint does not change over our sample period.[30]

It is potentially a significant challenge to estimate the independent effects of down-payment constraints[31] and gains on households' listing decisions, since $\widehat{G}$ and $\widehat{H}$ are jointly dependent on $\widehat{\ln P}$. However, there are multiple other factors that influence this correlation, including the LTV ratio at origination (i.e., variation in initial downpayments), and households' post-initial-issuance remortgaging decisions including "cash-out" refinancing. In the online appendix, we plot the joint distribution of these two variables, and show that in addition to unconstrained winners ($\widehat{H} \geq 20\%$ and $\widehat{G} \geq 0$) and constrained losers ($\widehat{H} < 20\%$ and $\widehat{G} < 0$) which comprise 68.4% of the sample, we have both constrained winners (24.9% of the sample, $\widehat{H} < 20\%$ and $\widehat{G} \geq 0$) and the final category of unconstrained losers (6.7% of the sample, $\widehat{H} \geq 20\%$ and $\widehat{G} < 0$).[32] While this is reassuring, it could well be the case that this variation is confined to one particular part of the sample period, i.e., driven by time-variation in Danish house prices. To check this, in the online appendix we also plot the shares of seller groups in the data across each of the years in our sample. The figure shows that aggregate price variation does shift the relative shares in each group across years, with price rises increasing the fraction of unconstrained winners relative to losing and constrained groups. However, the relative shares still look

[30] The online appendix documents the changes in the Act over our 2009 to 2016 sample period. While the constraint does not move during this period, there are a few changes in the wording of the act, a change in the maximum maturity of certain categories of loans in February 2012 from 35 to 40 years, and the revision of certain stipulations on the issuance of bonds backed by mortgage loans. None of these materially affect our inferences.

[31] The notion of constraints applies only if households are reluctant to downsizing. In the online appendix, we show, using a subsample of 14,939 households for which we can find two subsequent housing transactions and mortgage down-payment data, that there is a high correlation between the current house value, and the price of the next home that these households purchase, and that the price of the next home almost always lies above the price of the current house.

[32] $M > R$ is frequently observed in the data (44.8% of observations). This is primarily because of households' subsequent actions to remortgage to higher levels than their mortgage at issuance. This generally arises from "cash-out" refinancing, but could also arise from disadvantageous subsequent refinancing by homeowners, or fluctuations in adjustable rate mortgage payments causing households to increase mortgage principal to reduce monthly payment volatility.

stable over the sample period, alleviating concerns that different groups simply come from different time periods, i.e., identification of any effects is likely to arise mainly from the cross-section rather than the time-series. We also verify that the inclusion of cohort and cohort-cross-municipality fixed effects in the hedonic model does not affect our inferences materially.

3.4 Listing Premia, Realized Premia, Time-on-the Market, and Probability of Sale

Our measure of the listing premium in the data is $\hat{\ell} = \ln L - \widehat{\ln P}$, where L is the reported initial listing price observed in the data.[33] The online appendix plots the distribution of $\hat{\ell}$. Mean (median) $\hat{\ell}$ is 13.4% (11.8%), and $\hat{\ell} > 0$ ($\hat{\ell} < 0$) for 75% (25%) of the sample.

The online appendix also shows the distribution of TOM in the data. We winsorize this distribution at 200 weeks, viewing properties that spend roughly 4 years on the market as essentially retracted. Mean (median) TOM in the data is 36 weeks (25 weeks). This is higher than the value of roughly 17 weeks reported in Genesove and Han (2012).

We next inspect the inputs to the function $\alpha(\hat{\ell})$ in the data. The left plot in Panel A of Figure 3 shows how TOM relates to $\hat{\ell}$ in the data using a simple binned scatter plot. When $\hat{\ell}$ is below 0, TOM barely varies with $\hat{\ell}$; however, TOM moves roughly linearly with $\hat{\ell}$ when $\hat{\ell}$ is positive and moderately high. Interestingly, we also observe that the relationship between $\hat{\ell}$ and TOM flattens out as $\hat{\ell}$ rises to very high values above 40%. This behavior is mirrored in the right-hand side of Panel A of Figure 3, which shows the share of seller *retracted* listings, which also rises with $\hat{\ell}$. Here we also see more "concavity" as $\hat{\ell}$ drops below zero, in that the retraction rate rises the farther $\hat{\ell}$ falls below zero.

The left plot in Panel B of Figure 3 simply converts the two plots in Panel A into a

[33] We confirm, estimating Genesove and Mayer's (2001) specifications on our data (see online appendix), that the coefficient on $\widehat{\ln P}$ in our data using ther regression, controlling for a range of other determinants, is close to 1. We discuss below how our results are also robust to using the alternative approach of Genesove and Mayer (2001), and discuss identification and measurement concerns in greater detail there as well.

single number, which is the probability of house sale within six months, i.e., $\alpha(\hat{\ell})$ (we pick six months to match the median 25 week TOM observed in the sample), which we plot on the y-axis as a function of $\hat{\ell}$ on the x-axis. To smooth the average point estimate at each level of the listing premium, we use a generalized logistic function or GLF (Richards, 1959, Zwietering et al., 1990, Mead, 2017) of the form:

$$\alpha(\hat{\ell}) = A + \frac{K - A}{\left(C + Qe^{-B\hat{\ell}}\right)^{1/\nu}}. \tag{7}$$

The solid line corresponds to this set of smoothed point estimates. The GLF is well suited to capture concave demand, in that it is bounded both from above (K) and below (A). Moreover, we are able to easily vary the degree of concavity in a convenient way, through the single parameter B. In our estimation of the parameters, we restrict $A = 0$ to impose that the probability of sale asymptotically converges to 0 for arbitrary high levels of $\hat{\ell}$.

The right-hand plot in Panel B of Figure 3 shows how the markup of $P(\hat{\ell})$ over the hedonic value (i.e. $\ln P(\hat{\ell}) - \widehat{\ln P}$ or the "realized premium" of the final sales price over the hedonic value) varies with $\hat{\ell}$. This realized premium rises virtually one-for-one with $\hat{\ell}$ when $\hat{\ell}$ is low, but flattens out as $\hat{\ell}$ rises. In Denmark, virtually identically to the patterns detected by Guren (2018) in three U.S. markets, low list prices appear to reduce seller revenue with little corresponding decline in TOM, and akin to Genesove and Mayer (2001), who analyze the Boston housing market between 1990 and 1997, the Danish data also reveals that sellers who set high $\hat{\ell}$ suffer longer TOM, but ultimately achieve higher prices (i.e., high realized premia) on these sales. The solid line shows a simple linear fit of this relationship that we use in the model.

As described in the model section, the model parameter δ (which we implement as an addition to B in equation (7)) accounts for the possibility that sellers may perceive demand concavity differently from our measures in the data. Concretely, we measure δ as the relative reduction in the probability of sale for a listing premium of 10%. The right-hand plot in Panel C of Figure 3 shows that the data also exhibit strong pattern of rising *volatility* of TOM as $\hat{\ell}$ rises, a new fact that we uncover here. Among other reasons, this is a force that could potentially deliver a negative δ, i.e., a downward adjustment of probability of sale for any given $\hat{\ell}$, as a result of TOM volatility aversion on the part of sellers (we show the potential effects of δ on perceived $\alpha(\hat{\ell})$ in the left-hand plot in Panel C).[34]

4 Moments in the Data

In this section, we describe important moments of the data. These are a combination of previously established results that we verify in our data, and new facts about the behavior of $\hat{\ell}$ and the extensive margin listing decision. Rather than attempting to use the entire dataset to estimate the parameters of the model, we select key moments that are critical for identification, and use these as the input into structural parameter estimation. One approach to validation that we later adopt is to generate predictions from the model using these estimated parameters, and to evaluate these predictions against the data.

4.1 Listing Premia, Gains/Losses, and Home Equity

Panel A of Figure 4 is a 3-D plot of $\hat{\ell}$ against both $\widehat{G}$ and $\widehat{H}$ in the data, averaged in bins of 3 percentage points. The plot shows that $\hat{\ell}$ declines in both $\widehat{G}$ and $\widehat{H}$, consistent with the patterns previously identified in the literature. Unusually, given the large administrative data that we have access to, the plot captures the variation $\hat{\ell}$ along both dimensions simultaneously, and clearly reveals both *independent* and *interactive* variation along both dimensions. We describe these interactions in more detail towards the end of the paper, and evaluate the extent to which we can match these relationships using the model.

[34]The more steeply sloping line corresponds to the actual δ parameter that we acquire from structural estimation, described later.

4.1.1 Estimation Moments

To make progress at this stage, we discipline the estimation moments, restricting our focus to two cross-sections of the 3-D plot. We first evaluate the relationship between $\hat{\ell}$ and $\widehat{G}$, setting $\widehat{H} = 20\%$, and the relationship between $\hat{\ell}$ and $\widehat{H}$ setting $\widehat{G} = 0\%$. These are shown as dotted lines in Panel A of Figure 4, and Panel B plots these cross-sections. The left-hand plot in Panel B shows that $\hat{\ell}$ increases substantially more with losses ($\widehat{G} < 0$) than it declines with gains ($\widehat{G} > 0$). There is also some visual evidence of a kink in the relationship at $\widehat{G} = 0$. The right-hand plot of Panel B shows a strong negative relationship between $\hat{\ell}$ and $\widehat{H}$, but little evidence of a kink in this plot at the $H = 20\%$ mark.

We estimate three moments for structural parameter estimation in Panel B. The first is the level of $\hat{\ell}$ at $\widehat{G} = 0\%$, which is 10.4%. The second is the slope of $\hat{\ell}$ for $\widehat{G} < 0\%$, which equals -0.492. The third is the slope of $\hat{\ell}$ for $\widehat{H} < 20\%$ which is -0.304.[35]

4.1.2 Unobserved Quality

An important and often-repeated concern in the literature measuring the relationship between $\hat{\ell}$ and $\widehat{G}$ (and indeed, in estimating the function $\alpha(\hat{\ell})$) is that observed non-linearities could simply arise from measurement error in the underlying model for $\widehat{P}$. We discuss this issue in detail in the online appendix, which also shows plots on the robustness of the relationship between $\hat{\ell}$ and $\widehat{G}$, and the shape of $\alpha(\hat{\ell})$ to a range of different models that we use to estimate $\widehat{P}$. The bottom line is that the asymmetries seen in the relationship between $\hat{\ell}$ and $\widehat{G}$, as well as those seen in measured demand concavity are robust to estimating several different models of $\widehat{P}$ proposed in the literature as well as more novel approaches that we adopt to deal with unobserved quality, including the use of property-specific fixed effects to absorb time-invariant unobserved house features; instrumenting variation in $\widehat{P}$ using regional house price indexes following Guren (2018); Genesove and Mayer (2001)'s bounding approach; adding fixed effects, demographics,

[35]We compute standard errors for our structural parameter estimates using a bootstrap procedure.

and interactions to the hedonic model; and the use of the regression kink design approach suggested by Card et al. (2015b) (and implemented e.g., by Landais, 2015, Nielsen et al. 2010, Card et al. 2015a), which relies on quasi-random assignment at thresholds of particular "running variables" that induce kinks in agents' responses.

4.2 Regional Variation in Demand Concavity and Listing Premium Slopes

The model predicts two possible and distinct sources of the differential slopes of ℓ^* across gains and losses. One is that in the presence of loss aversion (i.e., $\lambda > 0$), there are kinks in ℓ^* around $\widehat{G} = 0$, which can be smoothed into a differential slope by variation in θ. The second is buyer sensitivity to ℓ, i.e. the degree of demand concavity $\alpha(\ell)$. The bottom panel of Figure 3 illustrates this second mechanism in the model, which predicts that sellers set a steeper ℓ^* slope when $\widehat{G} < 0$ in markets where $\alpha(\ell)$ demand is *less* steeply sloped and vice versa. This predicts a tight correlation between the slope of $\alpha(\hat{\ell})$ and the slope of $\hat{\ell}$ when $\widehat{G} < 0$.

To check whether this predicted correlation moment is observed in the data, we separately estimate the slope of $\hat{\ell}$ in the domain $\widehat{G} < 0$, as well as separate $\alpha(\ell)$ functions (in particular, the slope of $\alpha(\hat{\ell})$ when $\hat{\ell} \geq 0$) in different municipalities of Denmark, corresponding to different local housing markets.[36]

Figure 5 Panel A shows results when we sort municipalities by their estimated demand concavity (i.e., the slope of $\alpha(\hat{\ell})$ when $\hat{\ell} \geq 0$). The right-hand panel of the plot illustrates that there is indeed substantial variation in demand concavity across municipalities, showing municipalities in the top and bottom 5% of estimated demand concavity. The slope for municipalities with strong demand concavity (top 5%) lies between -1.4 and -1.1, while the slope for municipalities with weak demand concavity (bottom 5%) lies between

[36]Municipalities are a natural local market unit—there are 98 in Denmark, each of around 60,000 inhabitants, and with roughly 1,800 listings on average. We also re-do this exercise using shires, which are a smaller geographical delineation covering 80 listings on average as a cross-check.

-0.1 and -0.3. The left-hand panel of Figure 5 Panel A shows the corresponding figure for the relationship between $\hat{\ell}$ and $\widehat{G}$ for these municipalities. Indeed, as the model predicts, markets with strong demand concavity exhibit a substantially weaker slope of $\hat{\ell}$ in the domain $\widehat{G} < 0$ (-0.1 to -0.4) than markets with weak demand concavity (-0.5 to -0.9).[37]

In the left-hand plot of Panel B of Figure 5, we plot the relationship between the slope of $\hat{\ell}$ in the domain $\widehat{G} < 0$ and demand concavity (i.e., the slope of $\alpha(\hat{\ell})$ when $\hat{\ell} \geq 0$) across all municipalities. As the model predicts, this relationship is tight and negative across municipalities.

We turn these observations into moments for the purposes of estimation. For each municipality, we start with the concavity of demand (calculated as above, i.e., the slope of $\alpha(\hat{\ell})$ when $\hat{\ell} \geq 0$). We then compute the slope of $\hat{\ell}$ with respect to $\widehat{G}$ above and below $\widehat{G} = 0$. The fourth moment is then given by regressing the municipality-average listing premium slope for $\widehat{G} < 0$ on the municipality-specific concavity of demand. We estimate this regression slope to be equal to -0.384 (*s.e.* 0.071). Analogously, to obtain the fifth moment we regress the local (municipality-level) listing premium slope for $\widehat{G} \geq 0$ on the local concavity of demand. The corresponding estimate is equal to -0.098 (*s.e.* 0.043).

Towards the end of the paper, we describe a validation analysis that we undertake to confirm the mechanism in the data using instruments for demand concavity. We now turn to describing the sixth and seventh moments that we estimate.

[37]We also observe important differences between the *levels* of $\alpha(\hat{\ell})$ across these markets i.e., there are both "hot" and "cold" municipalities à la Ngai and Tenreyro (2014); for the purposes of our investigation, we focus on the slope differentials, and to show these, Figure 5 normalizes sub-markets to have the same level of the listing premium. Un-normalized plots in the online appendix reveal that the *level* of $\hat{\ell}$ is lower when the level of $\alpha(\hat{\ell})$ is higher and vice versa; consistent with Ngai and Tenreyro (2014), the levels of $\alpha(\hat{\ell})$ and $P(\hat{\ell})$ are strongly positively correlated across sub-markets.

4.3 Bunching

The right-hand plot of Panel A in Figure 6 documents significant bunching of transactions in the positive domain of realized gains G, with a sharp jump around $G = 0$, and with significant mass apparently extending further into the domain $G > 0$. The "raw" change in sales frequency in the interval $G \in [-3\%, +3\%]$ around $G = 0\%$ is one measure of bunching, and this estimate equals 30%. However, we do not need to rely solely on the distribution of realized transactions. The left-hand plot of Panel A of the figure shows the distribution of potential gains $\widehat{G}$. This is a useful counterfactual for the distribution of realized G, because in the model, when $\eta > 0$, realized gains arise from potential gains which are transformed through the choice of ℓ^* and the associated probability of sale. In contrast, in the case of $\eta = 0$, the model predicts that the distribution of G would simply be a constant linear transformation of the distribution of $\widehat{G}$. The position of the pronounced jump in the distribution precisely at $G = 0\%$ is also a clue that $\lambda > 1$, and offers empirical support for the choice of R as the nominal purchase price (see Kleven, 2016, for a discussion of bunching at reference points).

To calculate our preferred measure of bunching (the sixth moment that we employ), we normalize the sales frequency over realized gains G (right-hand plot) by the sales frequency over $\widehat{G}$ expected gains (left-hand plot), which for the same interval $G \in [-3\%, +3\%]$ around $G = 0\%$ delivers a bunching measure of 29%, strikingly similar to the "raw" estimate. The magnitude is also very close when estimating counterfactuals using a smooth polynomial function (see, e.g., Kleven (2016)), and is robust to conditioning on $\widehat{H} = 20\%$.

4.4 Extensive margin

To estimate our final moment, we compute the fraction of the housing stock in Denmark that lists at each level of $\widehat{G}$, by estimating $\widehat{\ln P}$ for all properties in Denmark for which we have data on the nominal purchase price R (12,565,190 property-years in the data).

Panel B of Figure 6 shows a mild increase in the probability of listing as $\widehat{G}$ increases. This slope is our seventh and final empirical moment for structural parameter estimation. We also note that the unconditional average listing propensity is 3.75% of the housing stock (corresponding to between 2% and 4% of the housing stock listed in each sample year). We not not attempt to explain the average propensity to list as we view this as beyond the scope of this paper, and would require us to take a strong stance on the factors that drive the moving decision, currently summarized by our estimates of θ.

Table 1 summarizes the seven moment estimates that we employ in structural parameter estimation, as well as associated OLS and cluster-robust bootstrap standard errors (clustering individual draws by shire). We move next to describing our structural parameter estimation.

5 Structural Estimation

5.1 Moments in the Model

To match the data moments inside the model, we make a few assumptions. First, we simply use the estimated demand concavity $\alpha(\ell)$ and $P(\ell)$ shown in Panel B of Figure 3 as two of these inputs. Second, we set $\gamma = 20\%$ according to Danish law. Third, we normalize all quantities in the model, setting the property's fundamental value $\widehat{P} = 1$ and we set the outside option $\underline{u} = \widehat{P}$. Fourth, we define the variables $\widehat{G} = \widehat{P} - R$ and $\widehat{H} = \widehat{P} - M$ as the model equivalents of potential gains and home equity in the data.

Next, consider the set of parameters from the model:

$$\mathbf{x} = \Big[\eta, \quad \lambda, \quad \delta, \quad \mu, \quad \theta_{\min}, \quad \theta_{\max}, \quad \varphi\Big]'. \tag{8}$$

To obtain policy functions of state variables and parameters, we solve the model numerically, inputting grids of $\widehat{G}$ and $\widehat{H}$, and yielding:

$$\left[s^*(\widehat{G},\widehat{H},\theta,\mathbf{x}),\ell^*(\widehat{G},\widehat{H},\theta,\mathbf{x})\right] = \arg\max_{s\in\{0,1\}}\left\{(s)\max_{\ell}\left\{EU(\ell,\widehat{G},\widehat{H},\theta,\mathbf{x})\right\}+(1-s)\underline{u}\right\}. \tag{9}$$

We then compute aggregates, i.e., averages in the population of listing probabilities, and average listing premia which account for the extensive margin decision:

$$S^*(\widehat{G},\widehat{H},\mathbf{x}) = \int s^*(\widehat{G},\widehat{H},\theta,\mathbf{x})d\theta, \tag{10}$$

$$\mathscr{L}^*(\widehat{G},\widehat{H},\mathbf{x}) = \int_{s^*=1} \ell^*(\widehat{G},\widehat{H},\theta,\mathbf{x})d\theta. \tag{11}$$

These functions then allow us to compute the set of seven model-implied moments $\mathbf{M_m}(\mathbf{x})^{7\times1}$ corresponding to the moments in the data $\mathbf{M_d}^{7\times1}$ described above.

The first moment is the average listing premium $\mathscr{L}^*(\widehat{G}=0\%,\widehat{H}=20\%,\mathbf{x})$. The second is a slope from regressing $\mathscr{L}^*(\widehat{G},\widehat{H}=20\%,\mathbf{x})$ on the grid of $\widehat{G}$ for $\widehat{G}<0$. The third is a slope from regressing $\mathscr{L}^*(\widehat{G}=0\%,\widehat{H},\mathbf{x})$ on the grid of $\widehat{H}$ for $\widehat{H}<20\%$.

We next propose a simple procedure to approximate the regional correlation moments (i.e., the relationship between variation in demand concavity and the slope of the listing premium) inside the model. Let $\kappa_{\widehat{G}<0}$ be the slope from a regression of $\mathscr{L}^*(\widehat{G},\widehat{H}=20\%,\mathbf{x})$ on the grid of $\widehat{G}$ for $\widehat{G}<0$, and $\kappa_{\widehat{G}\geq0}$ the analogous slope for $\widehat{G}\geq0$ ($\kappa_{\widehat{G}<0}$ and $\kappa_{\widehat{G}\geq0}$ simply capture the slopes of the listing premium above and below potential gains of zero). Now consider a change $\tilde{\delta}$ in demand concavity. We re-compute each of the κ slopes for $\delta-\frac{\tilde{\delta}}{2}$ and $\delta+\frac{\tilde{\delta}}{2}$, which is a first-order approximation of the degree to which a change in concave demand "passes through" to the slopes of $\mathscr{L}^*$ above and below $\widehat{G}=0\%$. The fourth and fifth moments inside the model are then given by $\frac{\kappa^+_{\widehat{G}<0}-\kappa^-_{\widehat{G}<0}}{\tilde{\delta}}$ and $\frac{\kappa^+_{\widehat{G}\geq0}-\kappa^-_{\widehat{G}\geq0}}{\tilde{\delta}}$.

The sixth moment measures bunching of transactions around realized gains of zero. To calculate this measure, we begin with a randomly generated sample of $N=1,000$ draws of $\widehat{G}$ from a uniform distribution with limits $(-50\%,+50\%)$. For each observation in the sample, we obtain the optimal aggregate listing premium $\mathscr{L}^*$ for a level of home equity equal to 20% and the average level of the moving shock, and calculate realized gains as $G = P(\mathscr{L}^*) - R$. In addition, we model the likelihood that the transaction goes through by drawing a random number ϵ from a uniform distribution and including the observation in the final sample of transactions if $\epsilon < \alpha(\mathscr{L}^*)$. The measure of bunching is then given by the relative density of transactions in the positive vs. the negative domain, in the interval [-5%, +5%].[38]

Finally, the seventh moment is given by the slope from a regression of $S^*(\widehat{G}, \widehat{H} = 20\%, \mathbf{x})$ on the grid of $\widehat{G}$, to match the corresponding extensive margin moment in the data.

5.2 Classical Minimum Distance Estimation

From the moments in the data and in the model, we calculate:

$$g(\mathbf{x}) = M_m(\mathbf{x}) - M_d.$$

Since the system is exactly identified, i.e., seven moments and seven parameters, we can estimate the structural parameters $\widehat{\mathbf{x}}$ simply as:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} g(\mathbf{x})' g(\mathbf{x}).$$

The asymptotic variance of the parameters is given by:

$$\overline{avar}(\widehat{\mathbf{x}}) = \left[\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \overline{W} \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}'} \right]^{-1},$$

where we set $\overline{W}$ to the inverse of the normalized covariance matrix of moments $\mathbf{x}$. We

[38] We choose this slightly wider interval than in the data to avoid situations in which our results may be influenced by the grid sizes.

consider both a simple (diagonal) case: $\overline{W}_{ii} = (\sigma_i^2/N_i)^{-1}$, as well as the (shire-clustered) bootstrap full covariance matrix. Finally, we make inferences about the parameter estimates using the asymptotic relationship:

$$\widehat{\mathbf{x}} \to^d N(\mathbf{x}, \overline{avar}(\widehat{\mathbf{x}})).$$

5.3 Parameter Estimates

Table 2 shows the estimated parameters and associated standard errors. The data favor a model of reference dependence with $\eta = 0.981$ with a degree of loss aversion $\lambda = 1.525$. This λ estimate is lower than that commonly considered in the early literature, which lies between 2 and 2.5 (e.g., Kahneman et al. 1990, Tversky and Kahneman, 1992), but is closer to estimates reported in more recent literature (e.g., Imas et al. 2016 finds a value of $\lambda = 1.59$).[39]

The parameter $\mu = 1.035$ best matches the average $\hat{\ell}$ slope with respect to $\widehat{H}$, i.e., there is an 103.5 bp penalty (expressed as a fraction of the mortgage amount) for every percent that H drops below $\gamma = 20\%$. This parameter can be contrasted with an average rate increase of roughly 50 bp on the whole loan if the household were to borrow an additional 10% in the unsecured Danish lending market.[40] The relatively larger number suggests that households in Denmark faced financial constraints preventing them from borrowing. In support of this, we find that the median household in our sample has *negative* net liquid financial wealth of roughly -9%, i.e., their unsecured debt is greater than their liquid financial assets (stocks, bonds, cash) by this amount.

[39] Given how close the estimated η is to 1, we re-estimated a restricted version of the model where $\eta = 1$. Further details are discussed in the online appendix. We obtained similar estimates of $\lambda = 1.58$ (*s.e.* 0.25), $\mu = 1.08$ (*s.e.* 0.19), $\delta = -0.11$ (*s.e.* 0.05), $\theta_{\min} = 0.25$ (*s.e.* 0.20), $\theta_{\max} = 1.10$ (*s.e.* 0.40) and $\varphi = 0.04$ (*s.e.* 0.04).

[40] Households in this market face between 200-500 basis points increases in interest rates for every percentage point of borrowing in this market between 80 and 95 LTV over our sample period. Taking 450 bp as the point estimate within this range, at an 80% LTV an additional ten percent borrowing adds roughly 50 bp to the overall loan.

We find that $\delta = -0.098$, which corresponds to a perceived relative reduction of the probability of sale of 9.8%, for a household listing at $\ell = 10\%$, and that the distribution θ ~Uniform($\theta_{\text{min}}, \theta_{\text{max}}$) has parameters $\theta_{\text{min}} = 0.228$ and $\theta_s = 1.037$. These "moving shocks" correspond to the present discounted value of future benefits from successfully selling and/or moving, and are on the order of 22.8% of the hedonic price for a household at the minimum of the distribution, and approximately equal to the entire hedonic value for a household at the maximum of the distribution. Finally, we find that the estimated "all-in" cost of listing is 3.9% of the hedonic value of the house, though this parameter is not precisely estimated in our setup.

5.4 Sensitivity Analysis

Andrews et al. (2017) argue that in method-of-moments estimation of the type that we use, it is often useful to understand the mapping from moments to estimated parameters. In Figure 7 we propose a simple and less formal application of this idea. Solid red lines (which overlap with the dashed black lines) indicate the level of the moment in the data (which are exactly matched by the indicated moments in the model given that the system is exactly identified). Dotted red lines show the 95% confidence interval in the data for the parameter estimate based on bootstrap standard errors. The horizontal solid lines show how sensitive the moments are to varying each of the parameters, describing how each moment varies when we re-compute the model-implied moments varying each of the structural parameters by two standard deviations. This also provides useful intuition on the sources of identification in the data for each of the model's parameters.

We also evaluate the importance of correctly modelling demand concavity. We do so by adopting a naïve approach to estimation that eschews this important feature and simply assumes that demand is linear. To do so, we preserve the $P(\ell)$ function, but simply estimate a linear $\alpha(\ell)$ function, and re-estimate the parameters (apart from δ) under this assumption. We find that in the case of this restricted model, we estimate $\eta = 0.750$ with a degree of loss aversion $\lambda = 3.285$, a radical departure from the more realistic estimates that we extract when demand is permitted to be concave.

6 Model Fit, Validation, and Open Questions

6.1 Model Fit

The plots in Panel A of Figure 8 compare the model-implied patterns of optimal listing premia with those observed in the data, contrasting the 3-dimensional plots in the data (left-hand plot) and the model (right-hand plot). The plot shows that there is a pronounced increase in $\hat{\ell}$ for $G < 0$, and shows a similar increase in $\hat{\ell}$ when $\widehat{H}$ declines. What is striking about the plot is that it suggests that the position of any reference point is not uniquely determined by $\widehat{G}$ or $\widehat{H}$ alone. As we briefly mentioned earlier, there seems to be considerable variation in the slope of the relationship between $\hat{\ell}$ and both $\widehat{G}$ and $\widehat{H}$ that depends on the level of the *other* variable. Put differently, both in the data and in the model, it appears as if the effects of losses and constraints interact with one another, and that the factors affecting household behavior are neither one nor the other variable in isolation. We explore this issue in greater detail below.

6.2 Interactions

Panel B of Figure 8 shows that the model is unable to capture these interactions. The figure plots selected cross-sections of the listing premium surface in the data, using a smooth function of the bins for ease of visualization as dashed lines, alongside their model equivalents as solid lines.[41] The left-hand plot in Panel B shows that there is a change in the slope of the $\hat{\ell}$ - $\widehat{G}$ relationship as $\widehat{H}$ varies, and the right-hand plot, that there seems to be a change in the inflection point in the $\hat{\ell}$ - $\widehat{H}$ relationship as $\widehat{G}$ varies.

[41]We simply use the GLF function introduced in equation (7) for this purpose. The online appendix shows a plot of the actual bins in the data alongside the model-implied listing premia.

Note that the average level of $\hat{\ell}$ in the data declines substantially as households become less constrained, and increases substantially as households become more constrained—this is simply the unconditional relationship between $\hat{\ell}$ and $\widehat{H}$, seen in a different way in the left-hand plot. What is more interesting is that controlling for this change in level, the *slope* of $\hat{\ell}$ as a function of $\widehat{G}$ is affected by the level of $\widehat{H}$. The important new fact is that down-payment-unconstrained households exhibit seemingly greater levels of reference dependence along the gain/loss dimension, exhibiting a pronounced increase in the slope to the left of $\widehat{G} = 0$. In contrast, down-payment constrained households exhibit a flatter $\hat{\ell}$ along the $\widehat{G}$ dimension. The right-hand plot in Panel B of the figure shows the $\hat{\ell}$ - $\widehat{H}$ relationship, where again, the level differences reflect the $\hat{\ell}$ - $\widehat{G}$ relationship. Another interesting fact emerges—along the $\widehat{H}$ dimension, while the slope around the threshold does not change, *the position of the kink* in $\hat{\ell}$ *increases* with the level of past experienced gains.

These new facts appear to require a more intricate model of preferences and/or constraints than the literature has thus far proposed, which cannot be rationalized by our canonical model, which captures many of the forces thus far proposed in the literature. We briefly speculate on the possible types of models that may rationalize these findings here, with a view towards motivating theoretical work on a broader class of preference and constraint specifications.

One possible rationalization of the variation in the $\hat{\ell}$ - $\widehat{G}$ relationship with $\widehat{H}$ is that the luxury of being unconstrained appears to cause more psychological motivations such as loss aversion to come to the fore. Put differently, unconstrained households seem constrained by their loss aversion à la Genesove and Mayer (2001), while constrained households respond to their real constraints by engaging in "fishing" behavior à la Stein (1995). It may also be that this finding can be rationalized by a more complex specification of reference points such as expectations-dependent reference points (e.g., Köszegi and Rabin, 2006, 2007, and Crawford and Meng, 2011).

Turning to the change in the position of the kink in the $\hat{\ell}$ - $\widehat{H}$ relationship as $\widehat{G}$ varies, it appears as if a household's propensity to engage in "fishing" behavior kicks in at a level of $\widehat{H}$ that is strongly influenced by their expected $\widehat{G}$. One possible rationalization of this is that households facing nominal losses feel constrained at levels of home equity (i.e., $H = 20\%$) that would force them to downsize, while those expecting nominal gains may have in mind a larger "reference" level of housing into which they would like to upsize (or indeed, a larger fraction of home equity in the next house). To achieve this larger reference level of housing, they begin "fishing" at levels of $H > 20\%$ in hopes of achieving the higher downpayment on the new, larger house. To provide suggestive evidence on this story, in the online appendix we focus on a sample of 14,440 households for which we can find two subsequent housing transactions and mortgage downpayment data. For this limited subsample, we show a binned scatter plot of the $\hat{\ell}$ on the subsequently sold listing against the realized downpayment on the subsequent house, controlling for the level of $\widehat{H}$ on the subsequently sold listing. We find evidence that the downpayment on the new house is correlated with $\hat{\ell}$, which, given our evidence of $\widehat{G}$ predicting $\hat{\ell}$, is consistent with the idea that households shifting their reference level of housing on the basis of expected gains.

6.3 Demand Concavity, Housing Stock Homogeneity, and Listing Premia

Earlier, we documented how regional variation in demand concavity correlates with regional variation in the shape of the listing premium schedule. This relationship could be driven by a number of different underlying forces. For instance, demand may respond to primitive drivers of supply rather than the other way around—i.e., some markets may be populated by more loss-averse sellers, and buyer sensitivity to ℓ^* might simply accommodate this regional variation in preferences. Another possibility is that this regional relationship simply captures the different incidence of common shocks to demand and market quality.

Our model is partial equilibrium, and describes a different underlying mechanism for this correlation, namely, that sellers are optimizing in the presence of the constraints imposed by demand concavity. In order to understand whether the left-hand plot of Panel B of Figure 5 is potentially consistent with sellers responding to such incentives, we implement an instrumental variables (IV) approach. Our IV approach is driven by the intuition that the degree of demand concavity is related to the ease of value estimation and hence price comparison for buyers. Intuitively, a more homogeneous "cookie-cutter" housing stock can make valuation more transparent, and should therefore increase buyers' sensitivity to ℓ. That is, this intuition predicts that markets with high homogeneity should exhibit more pronounced demand concavity.

Our main instrument is the share of apartments and row houses listed in a given sub-market. Row houses in Denmark are houses of similar or uniform design joined by common walls, and apartments have less scope for unobserved characteristics such as garden sheds and annexes than regular detached houses.[42] As an alternative, we also use the distance (computed by taking the shire-level distance to the closest of the four cities, averaged over all shires in a given municipality) to the four largest cities in Denmark (Copenhagen, Aarhus, Odense, and Aalborg) as a measure of how rural a given market is, and how far away from cities people live on average. This alternative relies on the possibility that homogeneous housing units are more likely to be built in suburbs or in cities, rather than in the countryside.

In the case of both instruments, the identifying assumption is that these measures of homogeneity of the housing stock only affect the slope of $\hat{\ell}$ with respect to $\widehat{G}$ through their effect on $\alpha(\hat{\ell})$. To account for cross-market differences in model-predicted demand-side factors affecting the slope of $\hat{\ell}$ with respect to $\widehat{G}$ and $\widehat{H}$, we also include specifications

[42] In the online appendix, we show pictures of typical row houses in Denmark.

which control for the average age, education length, financial assets, and income of sellers in a given sub-market.

Figure 5 on the right-hand side of Panel B shows strong evidence of the "first-stage" correlation, i.e., demand concavity on the y-axis against homogeneity measured by the share of apartments and row-houses in a given municipality on the x-axis, with each dot representing a municipality (more negative values of demand concavity mean a sharper slope of $\alpha(\hat{\ell})$ to the right of $\hat{\ell} = 0$). Table 3 reports the results of the more formal IV exercise. Column 1 shows the simple OLS relationship between the slope of $\hat{\ell}$ for $\widehat{G} < 0$ on demand concavity slope (slope of $\alpha(\hat{\ell})$ for $\hat{\ell} \geq 0$) across municipalities,[43] with a baseline level of -0.384. Column 2 uses the apartment-and row-house share as an instrument for demand concavity, and the just identified two-stage least squares (2SLS) specification yields a coefficient estimate of -0.479. With both instruments (i.e., including the distance to the largest cities as well), the overidentified 2SLS specification gives a result of -0.465 without, and -0.468 with controls for average household characteristics in the municipality. The first-stage F-statistics are between 25 and 37, assuaging weak-instrument concerns (Stock and Yogo, 2002) and we cannot reject the null of the Hansen overidentification test of a correctly specified model and exogenous instruments at conventional significance levels.[44] These results appear to validate the mechanism that we propose in the model.

7 Conclusion

Using a model that we structurally estimate on comprehensive Danish housing market data, we acquire new estimates of key behavioral parameters and estimates of household

[43]Municipalities are required to have at least 30 observations where $\widehat{G} < 0$, leaving 95 out of 98 municipalities, but results are robust to keeping all municipalities.

[44]These results are robust to using a logit model, different cutoffs ($\ell \geq 5$, 10, 15%) for the demand concavity estimation, cuts of the data such as excluding the largest cities Copenhagen and Arhus, and regressions at the shire level. These robustness checks are all available in the online appendix.

constraints from an important high-stakes household decision. Our work shows that in this market, households do exhibit reference dependence around the nominal purchase price plus modest loss aversion. However these inferences require an understanding of the institutional setting in the housing market—we document that inferences about preferences can be strongly affected by the correct specification of demand in this search and matching market. We also find evidence of the strong role of down-payment constraints on household decisions, and acquire estimates of the size of gains from trade for successful house sales, as well as new estimates of the all-in costs of house listing.

The model cannot completely match some new facts which we identify in the data, which we view as a new target for behavioral economics theory. Nominal losses and down-payment constraints interact with one another, in the sense that reference dependent behavior is less evident when households are facing more severe constraints, and most pronounced for unconstrained households. We also find that home equity constraints loom larger for households facing nominal losses. However, for households facing nominal gains, we find evidence consistent with an upward shift in the point at which they feel constrained. This could be explained by households resetting their desired size or quality of housing upwards in response to experienced gains.

In micro terms, this interaction between reference dependence and constraints could have implications for the way we model behavior. We tend to assume that agents optimize their (potentially behavioral) preferences subject to constraints, and in numerous models, agents may also wish to impose constraints on themselves to "meta-optimize" (Gul and Pesendorfer, 2001, 2004, Fudenberg and Levine, 2005, Ashraf et al. 2006, DellaVigna and Malmendier 2006). However, if constraints affect the incidence of behavioral biases, or indeed, if being in a zone that is more prone to bias affects the response to constraints, our models must of necessity become more complicated to accommodate such behavior.

Taking a more macro perspective, reference dependence appears important for understanding aggregate housing market dynamics. The housing price-volume correlation tends to fluctuate, and especially during housing market downturns, prices and liquidity can move in lockstep. This has important implications for labor mobility, which responds strongly to housing "lock" (Ferreira et al., 2012). Interaction effects such as the effect of expected losses on the household response to constraints could also help to make sense of the seemingly extreme reactions of housing markets to apparently small changes in underlying prices, and inform mortgage market policy (Campbell, 2012, Piskorski and Seru, 2018).

References

Anagol, S., V. Balasubramaniam, and T. Ramadorai (2018): "Endowment effects in the field: Evidence from India's IPO lotteries," *The Review of Economic Studies*, 85(4), 1971–2004.

Andersen, S., J. Y. Campbell, K. M. Nielsen, and T. Ramadorai (2018): "Inattention and Inertia in Household Finance: Evidence from the Danish Mortgage Market," *Unpublished working paper*.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics*, 132(4), 1553–1592.

Anenberg, E. (2011): "Loss aversion, equity constraints and seller behavior in the real estate market," *Regional Science and Urban Economics*, 41(1), 67–76.

Ashraf, N., D. Karlan, and W. Yin (2006): "Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines," *The Quarterly Journal of Economics*, 121(2), 635–672.

Badarinza, C., J. Y. Campbell, and T. Ramadorai (2016): "International Comparative Household Finance," *Annual Review of Economics*, 8(1).

Blundell, R. (2017): "What have we learned from structural models?," *American Economic Review: Papers & Proceedings*, 107(5), 287–92.

Bracke, P., and S. Tenreyro (2019): "History dependence in the housing market," *Bank of England Working Paper*.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82(6), 2295–2326.

Campbell, J. Y. (2006): "Household finance," *The Journal of Finance*, 61(4), 1553–1604.

——— (2012): "Mortgage market design," *Review of Finance*, 17(1), 1–33.

Card, D., A. Johnston, P. Leung, A. Mas, and Z. Pei (2015a): "The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in Missouri, 2003-2013," *American Economic Review*, 105(5), 126–30.

Card, D., D. S. Lee, Z. Pei, and A. Weber (2015b): "Inference on causal effects in a generalized regression kink design," *Econometrica*, 83(6), 2453–2483.

——— (2017): "Regression kink design: Theory and practice," in *Regression discontinuity designs: Theory and applications*, pp. 341–382. Emerald Publishing Limited.

Chetty, R. (2009): "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods," *Annual Review of Economics*, 1(1), 451–488.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *American Economic Review*, 104(9), 2633–79.

Clapp, J. M., R. Lu-Andrews, and T. Zhou (2018): "Controlling Unobserved Heterogeneity in Repeat Sales Models: Application to the Disposition Effect in Housing," *University of Connecticut School of Business Research Paper*, (18-16).

Crawford, V. P., and J. Meng (2011): "New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income," *American Economic Review*, 101(5), 1912–32.

DellaVigna, S. (2009): "Psychology and economics: Evidence from the field," *Journal of Economic Literature*, 47(2), 315–72.

——— (2018): "Structural behavioral economics," *National Bureau of Economic Research Working Paper.*

DellaVigna, S., and U. Malmendier (2006): "Paying not to go to the gym," *American Economic Review*, 96(3), 694–719.

Ferreira, F., J. Gyourko, and J. Tracy (2010): "Housing busts and household mobility," *Journal of Urban Economics*, 68(1), 34–45.

Ferreira, F., J. Gyourko, J. Tracy, et al. (2012): "Housing busts and household mobility: an update," *Economic Policy Review*, (Nov), 1–15.

Fudenberg, D., and D. K. Levine (2006): "A dual-self model of impulse control," *American Economic Review*, 96(5), 1449–1476.

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

Genesove, D., and C. Mayer (2001): "Loss aversion and seller behavior: Evidence from the housing market," *The Quarterly Journal of Economics*, 116(4), 1233–1260.

Genesove, D., and C. J. Mayer (1997): "Equity and time to sale in the real estate market," *The American Economic Review*, 87(3), 255.

Gul, F., and W. Pesendorfer (2001): "Temptation and self-control," *Econometrica*, 69(6), 1403–1435.

——— (2004): "Self-control and the theory of consumption," *Econometrica*, 72(1), 119–158.

Guren, A. M. (2018): "House price momentum and strategic complementarity," *Journal of Political Economy*, 126(3), 1172–1218.

Hahn, J., P. Todd, and W. Van der Klaauw (2001): "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 69(1), 201–209.

Hong, D., R. Loh, and M. Warachka (2016): "Realization utility and real estate," *Unpublished working paper.*

Imas, A., S. Sadoff, and A. Samek (2016): "Do people anticipate loss aversion?," *Management Science*, 63(5), 1271–1284.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990): "Experimental tests of the endowment effect and the Coase theorem," *Journal of Political Economy*, 98(6), 1325–1348.

Kahneman, D., and A. Tversky (1979): "Prospect theory: An analysis of decision under risk," *Econometrica*, pp. 263–291.

Kleven, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464.

Kőszegi, B., and M. Rabin (2006): "A model of reference-dependent preferences," *The Quarterly Journal of Economics*, 121(4), 1133–1165.

——— (2007): "Reference-dependent risk attitudes," *American Economic Review*, 97(4), 1047–1073.

Landais, C. (2015): "Assessing the welfare effects of unemployment benefits using the regression kink design," *American Economic Journal: Economic Policy*, 7(4), 243–78.

Marzilli Ericson, K. M., and A. Fuster (2011): "Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments," *The Quarterly Journal of Economics*, 126(4), 1879–1907.

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142(2), 698–714.

Mead, R. (2017): *Statistical methods in agriculture and experimental biology*. Chapman and Hall.

Newey, W. K., and D. McFadden (1994): "Handbook of Econometrics," *Elsevier*, 4.

Nielsen, H. S., T. Sørensen, and C. Taber (2010): "Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform," *American Economic Journal: Economic Policy*, 2(2), 185–215.

Piskorski, T., and A. Seru (2018): "Mortgage market design: Lessons from the Great Recession," *Brookings Papers on Economic Activity*, 2018(1), 429–513.

Rees-Jones, A. (2018): "Quantifying loss-averse tax manipulation," *The Review of Economic Studies*, 85(2), 1251–1278.

Richards, F. (1959): "A flexible growth function for empirical use," *Journal of Experimental Botany*, 10(2), 290–301.

Schulhofer-Wohl, S. (2012): "Negative equity does not reduce homeowners' mobility," *Federal Reserve Bank of Minneapolis Quarterly Review*, 35(1), 2–15.

Stein, J. (1995): "Prices and Trading Volume in the Housing Market: A Model With Down Payment Effects," *The Quarterly Journal of Economics*, 110(2), 379–406.

Stock, J. H., and M. Yogo (2002): "Testing for weak instruments in linear IV regression," .

Tversky, A., and D. Kahneman (1991): "Loss aversion in riskless choice: A reference-dependent model," *The Quarterly Journal of Economics*, 106(4), 1039–1061.

Zwietering, M., I. Jongenburger, F. Rombouts, and K. Van't Riet (1990): "Modeling of the bacterial growth curve," *Appl. Environ. Microbiol.*, 56(6), 1875–1881.

Figure 1

Reference dependence and loss aversion

The figure illustrates how each specification of utility function is reflected in the sellers' optimal choice of listing premia. We plot a stylized version of listing premium profiles, for the case in which demand functions $\alpha(\ell)$ and $\beta(\ell)$ are linear and the household is not facing financing constraints. In the online appendix, we describe and solve an analytical version of this model, which illustrates the driving forces of optimal listing premia for different levels of potential gains.

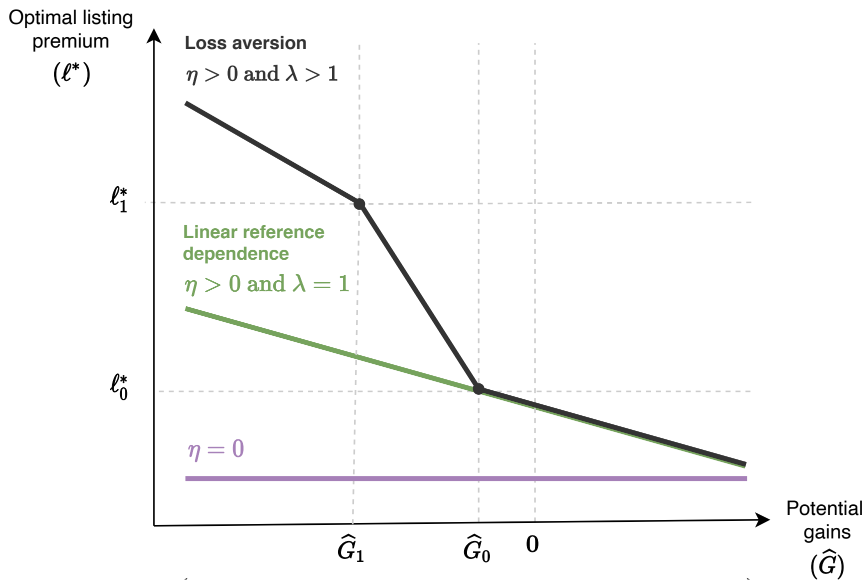
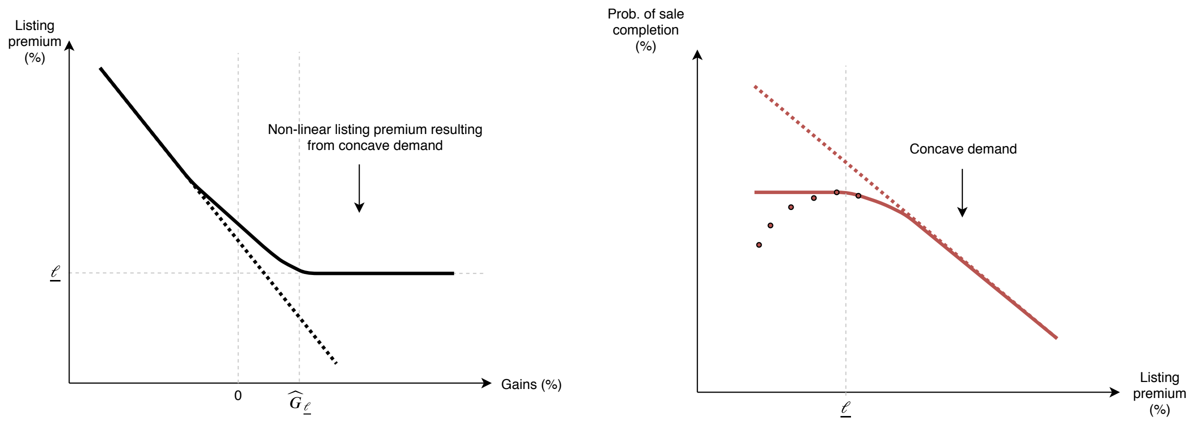


Figure 2

Concave demand

This figure illustrates the link between concave demand and the choice of optimal listing premia. In Panel A we plot a stylized listing profile resulting from a case of pure reference dependence with no loss aversion ($\eta > 0$ and $\lambda = 1$). Since the probability of sale does not respond to listing premia set below a certain level $\underline{\ell}$, it is rational for sellers to not respond to the exact magnitude of the expected gain. In Panel B we show how a steeper slope of demand translates into a general flattening out of the listing premium profile.

Panel A



Panel B

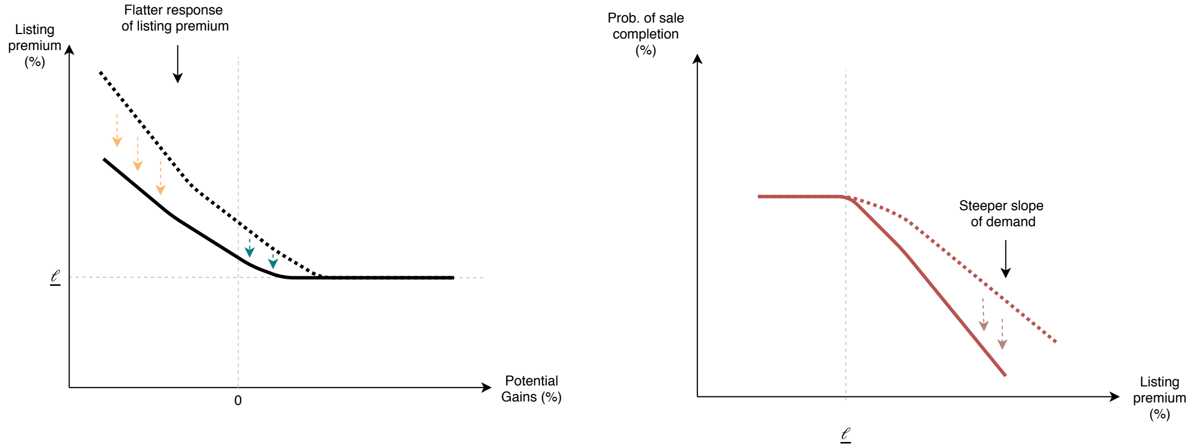
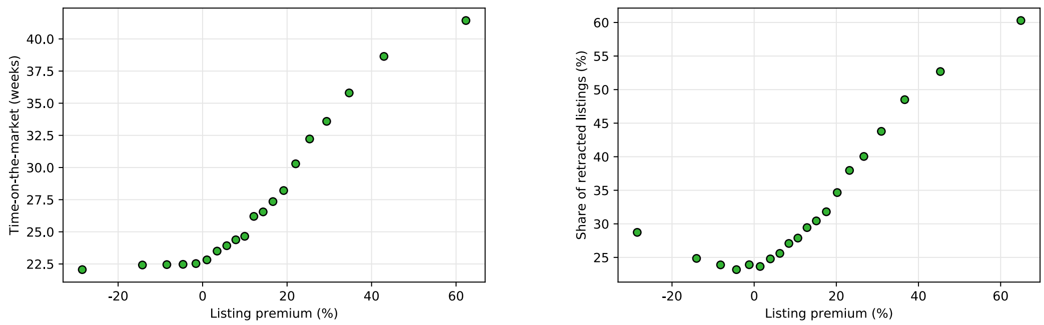


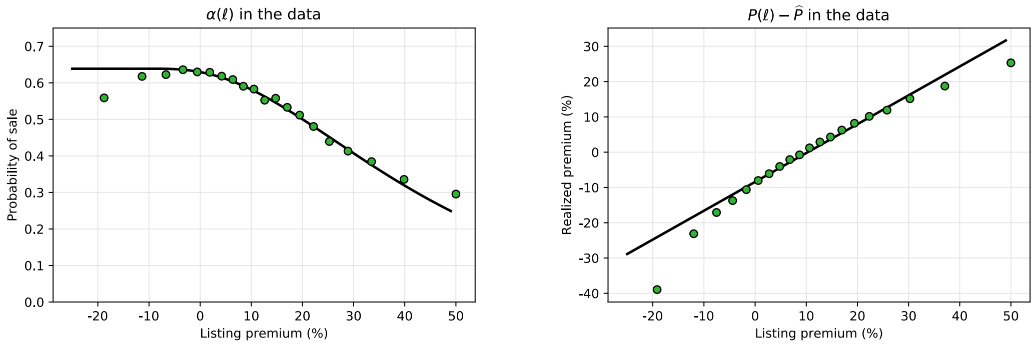
Figure 3
Concave demand in the data

Panel A shows how the time-on-the-market and the listing retraction rate depend on the listing premium. The left-hand side of Panel B shows the average probability of sale within six months $\alpha(\ell)$ across equal-sized bins of the listing premium in the sample. The right-hand side of Panel B shows the average realized premium $P(\ell) - \widehat{P}$ across equal-sized bins of the listing premium. Panel C illustrates the additional scaling factor δ, which is related to the interquartile range of time on the market (in weeks) across equal-spaced bins of the listing premium (connected by a line in the right-hand side figure), as a measure of the uncertainty around the probability of sale, which is increasing in $\ell \geq 0$.

Panel A



Panel B



Panel C

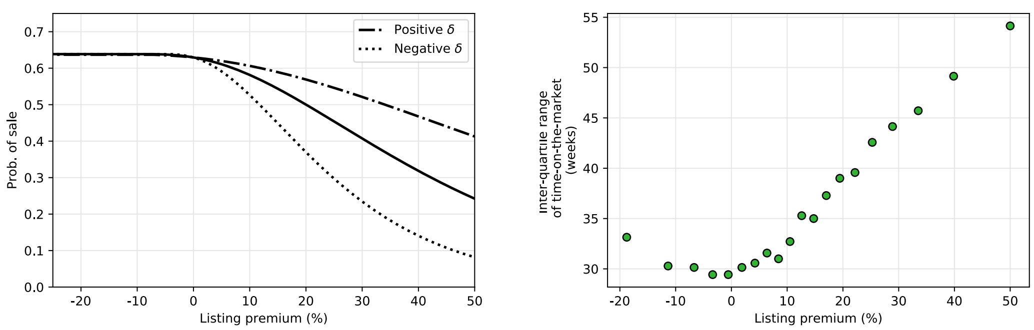
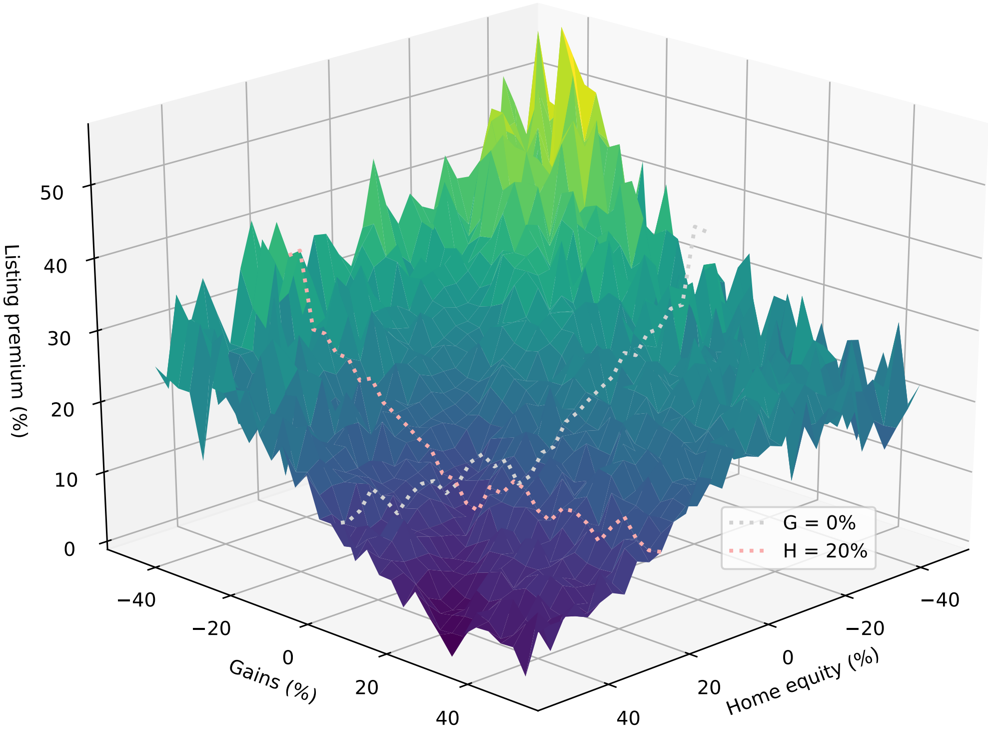


Figure 4

Listing premia, gains, and home equity

Panel A reports binned average values (in 3% steps) for the listing premium (ℓ) along both levels of expected gains and home equity. Panel B shows the underlying binned values for two cross-sections: In the left plot, we condition on a home equity level of 20%, and in the right plot on a level of expected gains of 0%. We use these two representative cross-sections to generate the empirical moments used in structural estimation.

Panel A: Listing premium surface



Panel B: Listing premia moments

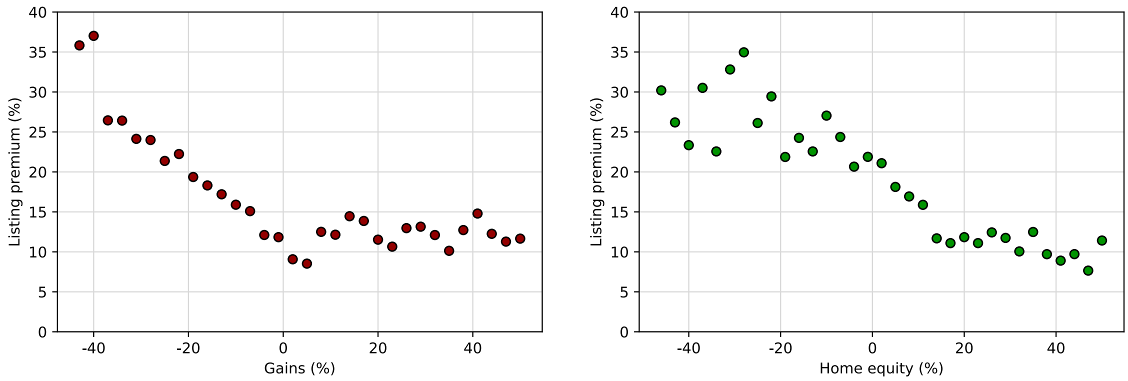
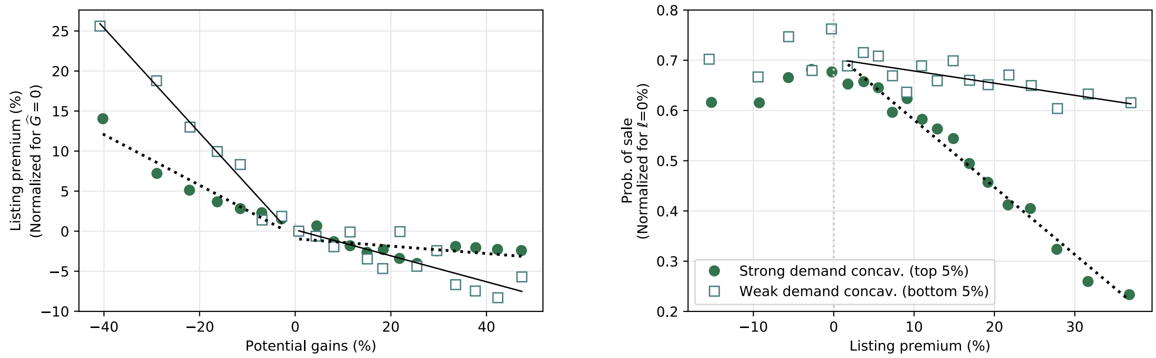


Figure 5

Listing premium-gain slope and demand concavity

Panel A shows the listing premium over gains (left-hand side) and demand concavity (right-hand side) patterns. We sort municipalities by the estimated demand concavity, using municipalities in the top and bottom 5% of observations. Demand concavity is estimated as the slope coefficient of the effect of the listing premium on the probability of sale within six months, for $\ell \in [0, 50]$. The slope of the listing premium over gains is calculated for $\widehat{G} < 0$. For better illustration of the main effect, we adjust the quantities measured to have the same level at $G = 0\%$ and $\ell = 0\%$ respectively. The left-hand side of Panel B shows the correlation between the estimated listing premium slope and demand concavity across municipalities using a binned scatter plot with equal-sized bins. The right-hand side of Panel B shows a binned scatter plot of the correlation between the main instrument, the share of listed apartments and row houses in a given municipality, and demand concavity in a binned scatter plot with equal-sized bins.

Panel A



Panel B

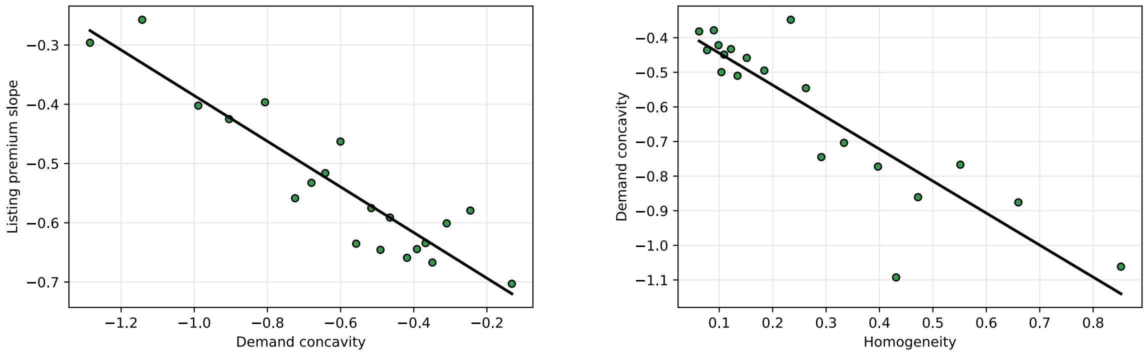
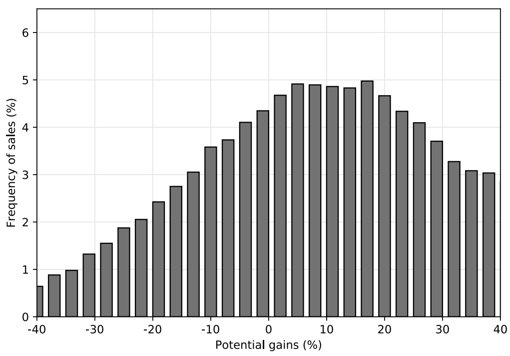


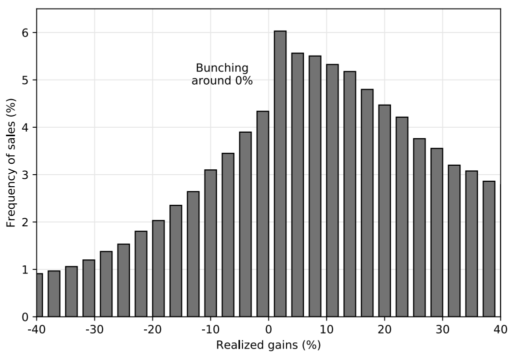
Figure 6

Moments: bunching and extensive margin

Panel A reports the frequency of observed transactions in terms of potential gains (left-hand side plot) and realized gains (right-hand side plot). This serves as the basis for the estimation of bunching, which we use as an empirical moment in our structural estimation exercise. Panel B reports the likelihood of listing with respect to potential gains. We calculate this by calculating the observed number of listings relative to the total stock of properties with potential gains in a given bin.

Panel A





Panel B

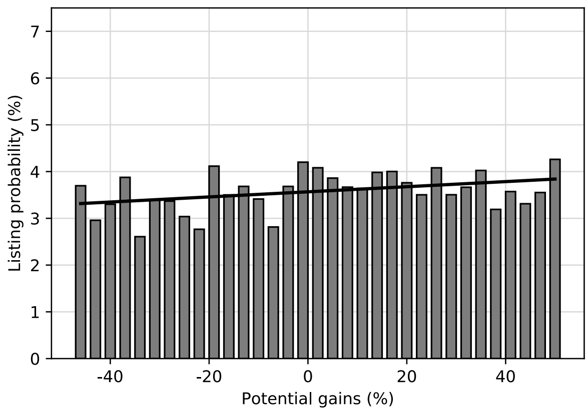


Figure 7

Model sensitivity to parameters

This figure illustrates the mapping from moments to estimated parameters. In the spirit of Andrews et al. (2017), we vary each of the structural parameters by two times their respective estimated standard error and re-compute model-implied moments. Solid red lines indicate the level of the moment in the data. Dotted red lines show the 95% confidence interval in the data based on bootstrap standard errors. The horizontal solid lines show how sensitive the moments are to variation in each of the parameters.

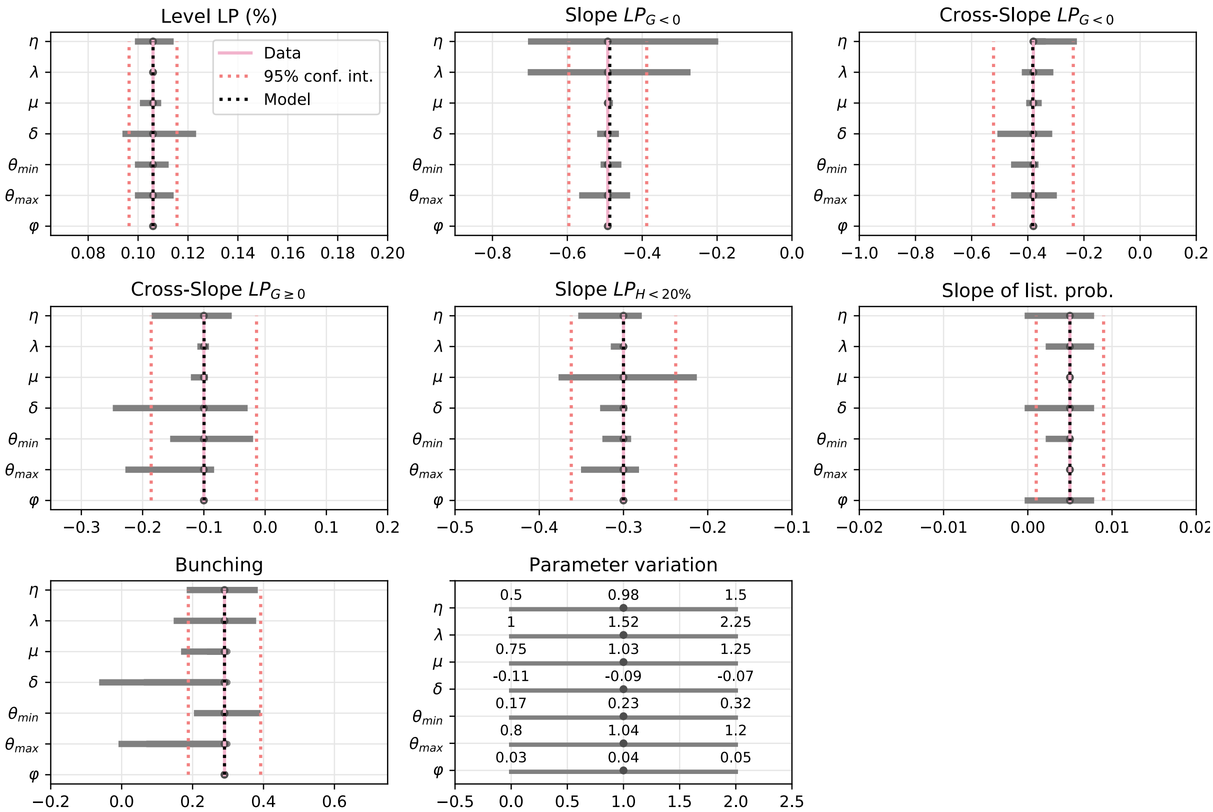
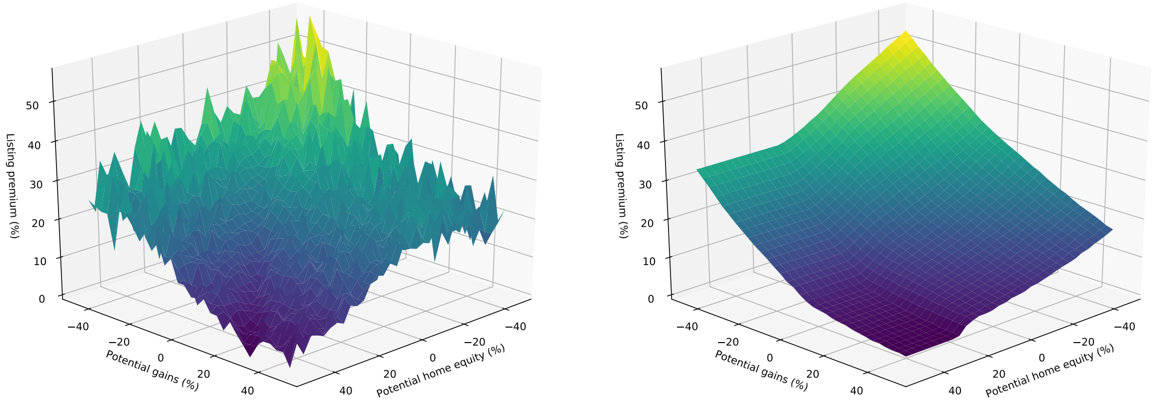


Figure 8

Model fit

Panel A reports listing premia by potential gains and home equity, both in the data and in the model. We use the set of seven estimated parameters to evaluate average quantities in the model, accounting for the extensive margin decision of whether to list the property for sale or not. Panel B illustrates the model fit for conditional listing premia profiles, conditioning on different levels of potential gains and home equity. Dotted lines indicate observations in the data (for which we report fitted values using generalized logistic functions) and solid lines their model-implied counterparts.

Panel A



Panel B

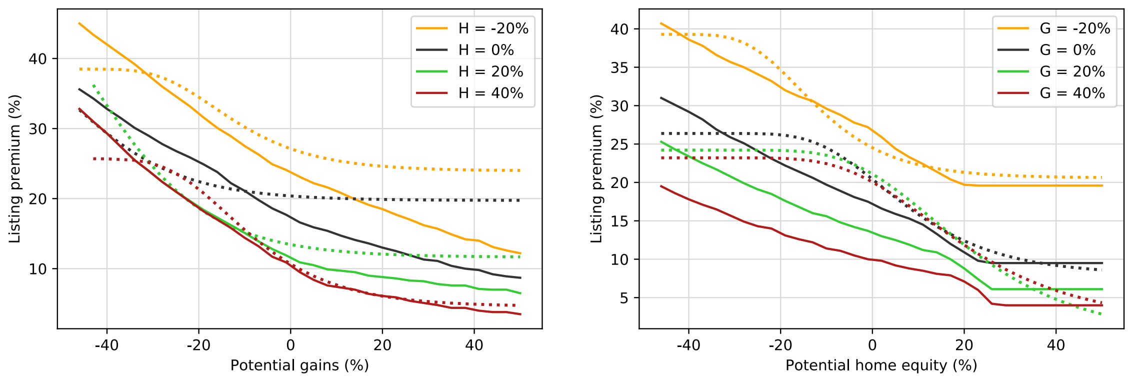


Table 1
Overview of moments and other estimates from the data

The table reports estimated empirical moments in the data. The first two capture the level and the slope of the listing premium with respect to the seller's level of potential gains, for $\widehat{G} > 0\%$, conditional on a home equity level of $\widehat{H} = 20\%$. The third moment is the slope of the listing premium with respect to potential home equity, for $\widehat{H} < 20\%$, conditional on gains of $\widehat{G} = 0\%$. The fourth and fifth moments are obtained as slope coefficients from cross-sectional regressions by municipality. For each municipality, we compute the slope $LP - \widehat{G}$ for $\widehat{G} < 0\%$ and $\widehat{G} \geq 0\%$ respectively, as well as the concavity of demand (i.e. the slope $\alpha - \ell$ for $\ell > 0$). The sixth moment is the slope of the listing probability with respect to the potential gains, conditional on a home equity level of $H = 20\%$. The final moment captures the bunching of transactions around realized gains of 0%, calculated as the relative frequency of transactions in the [0,3%] interval of realized gains, relative to the [-3%,0) interval. In parentheses, we report bootstrap standard errors, clustered at the shire level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

1.	Level of LP for $\widehat{G} = 0\%$	0.104***	(0.005)
2.	Slope LP-$\widehat{G}$ for $\widehat{G} < 0\%$	-0.492***	(0.052)
3.	Slope LP-$\widehat{H}$ for $\widehat{H} < 20\%$	-0.304***	(0.031)
4.	Cross-sectional slope LP-$\widehat{G}$-α for $\widehat{G} < 0\%$	-0.384***	(0.071)
5.	Cross-sectional slope LP-$\widehat{G}$-α for $\widehat{G} \geq 0\%$	-0.098**	(0.043)
6.	Slope of list. prob. by $\widehat{G}$	0.005**	(0.002)
7.	Bunching above $G = 0\%$	0.290***	(0.051)

Table 2
Estimated parameters

The table reports structural parameter estimates obtained through classical minimum distance estimation. We recover concave demand $\alpha(\ell)$ and $P(\ell)$ from the data and set the down-payment constraint $\gamma = 20\%$. In parentheses, we report standard errors based on the estimated bootstrap variance-covariance matrix in the data, clustered at the shire level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

η	=	0.981***	(0.312)
λ	=	1.525***	(0.422)
μ	=	1.035***	(0.140)
δ	=	-0.093***	(0.025)
$\theta_{\min}$	=	0.228	(0.186)
$\theta_{\max}$	=	1.037***	(0.174)
φ	=	0.039	(0.040)

Table 3

Listing premium-slope over gains and demand concavity slope regressions

This table reports regression results for the relationship between the listing premium slope over gains and demand concavity. The dependent variable in all regressions is the slope of the listing premium over $\widehat{G} < 0$ across municipalities.[45] Column 1 reports the baseline correlation with the demand concavity slope across municipalities using OLS. Column 2 reports the 2-stage least squares regression instrumenting demand concavity with the apartment- and row-house share. Columns 3 and 4 report the overidentified 2SLS regression with both instruments, row-house and apartment share and average distance to city, without and with household controls (age, education length, net financial assets and log income), respectively. In parentheses, we report bootstrap standard errors, clustered at the shire level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

	OLS	2SLS		
	(1)	(2)	(3)	(4)
		Single IV	*Overidentified*	
Demand concavity	-0.384***	-0.479***	-0.465***	-0.468*
	(0.071)	(0.121)	(0.083)	(0.267)
Household controls				✓
Observations	95	95	95	95
R^2	0.396			
First-stage F-stat		37.126	36.153	25.376
Hansen J-stat (p-val)			0.222	0.912

Reference Dependence in the Housing Market

Online Appendix

(For online publication)

Steffen Andersen[*] Cristian Badarinza[†] Lu Liu[‡]
Julie Marx[§] Tarun Ramadorai[¶]

May 13, 2020

[*]Copenhagen Business School and CEPR, Email: san.fi@cbs.dk.
[†]National University of Singapore, Email: cristian.badarinza@nus.edu.sg
[‡]Imperial College London, Email: l.liu16@imperial.ac.uk.
[§]Copenhagen Business School, Email: jma.fi@cbs.dk.
[¶]Corresponding author: Imperial College London, Tanaka Building, South Kensington Campus, London SW7 2AZ, and CEPR. Tel.: +44 207 594 99 10. Email: t.ramadorai@imperial.ac.uk.

Contents

1 Further Details on Framework

1.1 Reference Dependence and Loss Aversion

Figure A.1 illustrates the seller's utility function for three cases. The first ($\eta = 0$) corresponds to the utility from terminal value of wealth. The second ($\eta > 0, \lambda = 1$) captures linear reference dependence and the third ($\eta > 0$ and $\lambda > 1$) reference-dependent loss aversion.

1.2 Derivation of $\widehat{G}_0$ and $\widehat{G}_1$

We now derive the potential gain levels $\widehat{G}_0$ and $\widehat{G}_1$ discussed in Figure 1 in the paper, for a simple case where the demand functions are linear: $\alpha(\ell) = \alpha_0 - \alpha_1 \ell$ and $\beta(\ell) = \beta_0 + \beta_1 \ell$.

In this case, expected utility is given by:

$$U^*(\ell|\widehat{G}) = \max_{\ell}(\alpha_0 - \alpha_1 \ell) \left[\underbrace{\widehat{P} + \beta_0 + \beta_1 \ell}_{P(\ell)} + \eta \underbrace{(\widehat{G} + \beta_0 + \beta_1 \ell)}_{G(\ell)} + \theta \right] + (1 - \alpha_0 + \alpha_1 \ell)\widehat{P}. \quad (1)$$

The first-order condition for the choice of ℓ^* is then:

$$\alpha_0(1+\eta)\beta_1 - \alpha_1 \left[\widehat{P} + (1+\eta)\beta_0 + \eta\widehat{G} + \theta - \widehat{P} \right] - 2(1+\eta)\alpha_1\beta_1\ell^* = 0, \quad (2)$$

which implies the optimal solution:

$$\begin{aligned} \ell^*(\widehat{G}) &= \frac{\alpha_0(1+\eta)\beta_1 - \alpha_1 \left[(1+\eta)\beta_0 + \eta\widehat{G} + \theta \right]}{2(1+\eta)\alpha_1\beta_1} \\ &= \frac{1}{2} \left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\eta} - \frac{1}{\beta_1}\frac{\eta}{1+\eta}\widehat{G} \right). \end{aligned} \quad (3)$$

Deeper in the loss domain, we have:

$$\ell^*_\lambda(\widehat{G}) = \frac{1}{2} \left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\lambda\eta} - \frac{1}{\beta_1}\frac{\lambda\eta}{1+\lambda\eta}\widehat{G} \right). \quad (4)$$

Realized gains result from a markup over potential gains, depending on the chosen optimal listing premium:[1]

[1]Note that $G = \widehat{G} + \beta(\ell^*(\widehat{G})) = \beta_0 + \beta_1\widetilde{\gamma}_0 + (1 - \beta_1\widetilde{\gamma}_1)\widehat{G}$ if we define $\ell^*(\widehat{G}) = \widetilde{\gamma}_0 - \widetilde{\gamma}_1\widehat{G}$, and $\ell\lambda^*(\widehat{G}) = \widetilde{\gamma}_{\lambda,0} - \widetilde{\gamma}_{\lambda,1}\widehat{G}$.

$$G(\widehat{G}) = \widehat{G} + \beta(\ell^*(\widehat{G})) \tag{5}$$

Defining $\gamma_0 = \beta_0 + \frac{\beta_1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\eta}\right)$ and $\gamma_1 = 1 - \frac{1}{2}\frac{\eta}{1+\eta}$, we can simplify the expressions for the relationship between realized gains and potential gains:

$$G(\widehat{G}) = \gamma_0 + \gamma_1 \widehat{G} \tag{6}$$

With loss aversion, realized gains are then given by a step function:

$$G(\widehat{G}) = \begin{cases} \gamma_0 + \gamma_1 \widehat{G} & \text{if } \widehat{G} > \widehat{G}_0, \\ 0 & \text{if } \widehat{G} \in [\widehat{G_1}, \widehat{G}_0], \\ \gamma_{\lambda,0} + \gamma_{\lambda,1}\widehat{G} & \text{if } \widehat{G} < \widehat{G}_1. \end{cases} \tag{7}$$

Here, we have:

$$\widehat{G}_0 = -\frac{\gamma_0}{\gamma_1} \text{ and } \widehat{G}_1 = -\frac{\gamma_{\lambda,0}}{\gamma_{\lambda,1}}, \tag{8}$$

with $\gamma_{\lambda,0}$ and $\gamma_{\lambda,1}$ defined analogously to γ_0 and γ_1 above.

1.3 Irrelevance of R with Utility from Passive Gains

We assume that households do not receive utility from simply living in a house that has appreciated relative to their reference point R, i.e. they do not enjoy utility from passive "paper" gains until they are realized. If this condition does not hold, the model is degenerate in that R is irrelevant both for the choice of the listing premium (intensive margin) and the decision to list (extensive margin). Consider the following utility function:

$$\begin{aligned} U =& \alpha(\ell)\left(P(\ell) + \underbrace{P(\ell) - R}_{G(\ell)}\right) + (1-\alpha(\ell))\left(\widehat{P} + \underbrace{\widehat{P} - R}_{\widehat{G}}\right) \\ =& 2\alpha(\ell)P(\ell) + 2(1-\alpha(\ell))\widehat{P} - R. \end{aligned}$$

In this case, R is a simple scaling factor. It does not affect either marginal utility or marginal cost.

1.4 Derivation of Equation (5) in the Paper

To build intuition on how the demand functions $\alpha(\ell)$ and $P(\ell)$ determine listing behavior, consider the case of linear demand functions $\alpha(\ell) = \alpha_0 - \alpha_1 \ell$ and $P(\ell) = \widehat{P} + \beta_0 + \beta_1 \ell$ in a "pure" reference-dependent utility model (with $\eta = 1$, $\lambda = 1$) without financial constraints ($\mu = 0$). In this case, we have:[2]

$$U(\ell) = (\alpha_0 - \alpha_1 \ell)(\widehat{P} + \widehat{G} + \beta_0 + \beta_1 \ell + \theta) + (1 - \alpha_0 + \alpha_1 \ell)\widehat{P}.$$

The first-order condition for the choice of ℓ^* is given by:

$$\alpha_0 \beta_1 - \alpha_1(\widehat{P} + \widehat{G} + \beta_0 + \beta) - 2\alpha_1 \beta_1 \ell^* = 0,$$

which implies the optimal solution:

$$\ell^* = \frac{\alpha_0 \beta_1 - \alpha_1(\widehat{G} + \beta_0 + \theta)}{2\alpha_1 \beta_1} = \frac{1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0 + \theta}{\beta_1} - \frac{1}{\beta_1}\widehat{G}\right).$$

1.5 Bunching

Figure A.2 illustrates the bunching of transactions around realized gains of zero, in the case of loss aversion. This is a direct outcome of the kink in the listing premium profile. Listing premia are sharply increasing around $G = 0$ because sellers avoid realizing losses, and choose listing premia that are just high enough to push them above the threshold.

1.6 Specific Example of Extensive Margin Effects on Intensive Margin

To develop intuition, consider sellers with very high θ, who would naturally choose very low average listing premia. However, because of concave demand, such sellers will converge on the *the same* level ($\underline{\ell}$) of listing premia, the point beyond which there is no further improvement in the probability of sale. Put differently, if θ is sufficiently high, the chosen listing premium is essentially affected only by its effects on the probability of sale, and ℓ^* will barely respond to preferences (potential gains) and constraints (potential home equity). At the opposite end, if θ is very low (i.e., there are only tiny incentives to move), the average listing premium will be so high that responding to potential gains and losses is

[2] In the paper, we denote $\widehat{G} = \widehat{P} - R$ as potential gains and $G(\ell) = P(\ell) - R = \widehat{G} + \beta_0 + \beta_1 \ell$ as realized gains. For consistency in the discussion of model predictions, we did not use the "hat" notation in Equation (5).

either immaterial or too costly. Taken together, if the distribution of θ is such that there is substantial mass in either (or both) of these areas, the average observed listing premium in the market will show no evidence of reference dependence, and appear unaffected by down-payment constraints.[3] More realistically, a more smooth distribution of θ will blur the effects of both reference dependence and constraints on the intensive margin.

2 Detailed Data Description

Our data span all transactions and electronic listings (which comprise the overwhelming majority of listings) of owner-occupied real estate in Denmark between 2009 and 2016. In addition to listing information, we also acquire information on property sales dates and sales prices, the previous purchase price of the sold or listed property, hedonic characteristics of the property, and a range of demographic characteristics of the households engaging in these listings and transactions, including variables that accurately capture households' financial position at each point in time. We link administrative data from various sources; all data other than the listings data are made available to us by Statistics Denmark. We describe the different data sources and dataset construction below.

2.1 Property Transactions and Other Property Data

We acquire administrative data on property transactions, property ownership, and housing characteristics from the registers of the Danish Tax and Customs Administration (SKAT). These data are available from 1992 to 2016. SKAT receives information on property transactions from the Danish Gazette (Statstidende)—legally, registration of any transfer of ownership must be publicly announced in the Danish Gazette, ensuring that these data are comprehensive. Each registered property transaction reports the sale price, the date at which it occurred, and a property identification number.

The Danish housing register (Bygnings-og Boligregister, BBR) contains detailed characteristics on the entire *stock* of Danish houses, such as size, location, and other hedonic characteristics. We link property transactions to these hedonic characteristics using the property identification number. We use these characteristics in a hedonic model to predict property prices, and when doing so, we also include on the right-hand-side the (predetermined at the point of inclusion in the model) biennial property-tax-assessment value

[3] Naturally, these patterns will also strongly be reflected in decisions along the extensive margin. This is a possibility that which we plan to explore in the future (e.g., most intuitively, the majority of low-θ owners may decide not to list), in a setup in which the drivers of the moving decision can be more clearly identified and mapped onto observable household characteristics.

of the property that is provided by SKAT, which assesses property values every second year.[4] SKAT also captures the personal identification number (CPR) of the owner of every property in Denmark. This enables us to identify the property seller, since the seller is the owner at the beginning of the year in which the transaction occurred.

In our empirical work, we combine the data in the housing register with the listings data to assess the determinants of the extensive margin listing decision for all properties in Denmark over the sample period. That is, we can assess the fraction of the total housing stock that is listed, conditional on functions of the hedonic value such as potential gains relative to the original purchase price, or the owner's potential level of home equity.

Loss aversion and down-payment constraints were originally proposed as explanations for the puzzling *aggregate* correlation between house prices and measures of housing liquidity, such as the number of transactions, or the time that the average house spends on the market. In Figure A.3 we show the price-volume correlation in Denmark over a broader period containing our sample period. The plot looks very similar to the broad patterns observed in the US.

2.2 Property Listings Data

Property listings are provided to us by RealView (http://realview.dk/en/), who attempt to comprehensively capture all electronic listings of owner-occupied housing in Denmark. RealView data cover the universe of listings in the portal www.boligsiden.dk, in addition to additional data collected directly from brokers. The data include private (i.e., open to only a selected set of prospective buyers) electronic listings, but do not include off-market property transactions, i.e., direct private transfers between households. Of the total number of cleaned/filtered sale transactions in the official property registers (described below), 76.56 percent have associated listing data.[5] For each property listing, we know the address, listing date, listing price, size and time of any adjustments to the listing price, changes in the broker associated with the property, and the sale or retraction date

[4]As we describe later, this is the same practice followed by Genesove and Mayer (1997, 2001); it does not greatly affect the fit of the hedonic model, and barely affects our substantive inferences when we remove this variable.

[5]We more closely investigate the roughly 25% of transactions that do not have an associated electronic listing. 10% of the transactions can be explained by the different (more imprecise) recording of addresses in the listing data relative to the registered transactions data. The remaining 15% of unmatched transactions can be explained by: (a) off-market transactions (i.e., direct private transfers between friends and family, or between unconnected households); and (b) broker errors in reporting non-publicly announced listings ("skuffesager") to boligsiden.dk. We find that on average, unmatched transactions are more expensive than matched transactions. Sellers of more expensive houses tend to prefer the skuffesalg option for both privacy and security reasons.

for the property. The address of the property is de-identified by Statistics Denmark, and used to link these listings data to administrative property transactions data.

2.3 Mortgage Data

To establish the level of the owner's home equity in each property at each date, we need details of the mortgage attached to each property. We obtain mortgage data from the Danish central bank (Danmarks Nationalbank), which collects these data from mortgage banks through Finance Denmark, the business association for banks, mortgage institutions, asset management, securities trading, and investment funds in Denmark. The data are available annually for each owner from 2009 to 2016, cover all mortgage banks and all mortgages in Denmark and contain information on the mortgage principal, outstanding mortgage balance each year, the loan-to-value ratio, and the mortgage interest rate. The data contain the personal identification number of the borrower as well as the property number of the attached property, allowing us to merge data sets across all sources. If several mortgages are outstanding for the same property, we simply sum them, and calculate a weighted average interest rate and loan-to-value ratio for the property and mortgage in question.

2.4 Owner/Seller Demographics

We source demographic data on individuals and households from the official Danish Civil Registration System (CPR Registeret). In addition to each individual's personal identification number (CPR), gender, age, and marital history, the records also contain a family identification number that links members of the same household. This means that we can aggregate individual data on wealth and income to the household level.[6] We also calculate a measure of households' education using the average length of years spent in education across all adults in the household. These data come from the education records of the Danish Ministry of Education.

Individual income and wealth data also come from the official records at SKAT, which hold detailed information by CPR numbers for the entire Danish population. SKAT receives this information directly from the relevant third-party sources, e.g., employers who supply statements of wages paid to their employees, as well as financial institutions who supply information on their customers' balance sheets. Since these data are used to facilitate taxation at source, they are of high quality.

[6]Households consist of one or two adults and any children below the age of 25 living at the same address.

2.5 Final Merged Data

Our analysis depends on measuring both nominal losses and home equity. This imposes some restrictions on the sample. We have transactions data available from 1992 to the present, meaning that we can only measure the purchase price of properties that were bought during or after 1992. Moreover, the mortgage data run from 2009 to 2016. In addition, the sample is restricted to properties for which we know both the ID of the owner, as well as that of the owner's household, in order to match with demographic information.

For listings that end in a final sale, we drop within-household transactions and transactions that Statistics Denmark flag as anomalous or unusual. We flag (but do not drop) listings by households that do not have a stable structure, that is, we create a dummy for those listings for which the household ceases to exist as a unit in the year following the listing owing to death or divorce. We also flag households with missing education information. We restrict our analysis to residential households, in our main analysis dropping listings from households that own more than three properties in total, as they are more likely property investors than owner-occupiers.[7]

Once these filters are applied, the sample comprises 217,028 listings of Danish owner-occupied housing in the period between 2009 and 2016, for both sold (70.6%) and retracted (29.4%) properties, matched to mortgages and other household financial and demographic information. These listings correspond to a total of 193,850 unique households, and 181,020 unique properties. Most households that we observe in the data sell one house during the sample period, but roughly 9% of households sell two houses over the sample period, and roughly 1.5 percent of households sell three or more houses.

In our main analysis we study the 175,646 households that have a mortgage, but we return to the 41,382 households with no mortgage as a robustness check. In addition, we use the entire housing stock (13,305,501 observations of 1,736,172 unique properties) to understand the extensive margin decision of whether to list the properties for sale.

Table A.1 describes the cleaning and sample selection process from the raw listings data to the final matched data, with $N = 175,646$.

[7]Genesove and Mayer (2001) separately estimate loss aversion for these groups of homeowners and speculators. We simply drop the speculators in this analysis, choosing to focus our parameter estimation in this paper on the homeowners.

3 Summary Statistics

3.1 Liquid Financial Wealth

Figure A.5 Panel A shows the distribution of liquid financial assets in the sample. The wealthiest households in the sample have above 2 million DKK, which is roughly US$ 300,000 in liquid financial assets (cash, stocks, and bonds). The median level of liquid financial assets is 71,000 DKK and the mean in the sample is 247,000 DKK. When we divide *gross* financial assets by mortgage size, we find that households, at the median, could relax their constraints by around 6.25 percent if they were to liquidate all financial asset holdings. However, the right-hand side of the top panel of the figure shows that this would be misleading. Looking at *net* financial assets, once short-term non-mortgage liabilities (mainly unsecured debt) are accounted for, substantially changes this picture. The median level of net financial assets in the sample is -106,000 DKK and the mean is -136,000 DKK, and the picture shows that households' available net financial assets actually effectively *tighten* constraints for around 60 percent of the households in our sample. When we divide *net* financial assets by mortgage size we find, for households with seemingly positive levels of financial assets, that the constraints are in fact tighter by 9.3% at the median. Put differently, if households were to liquidate all financial asset holdings and attempt to repay outstanding unsecured debt, at the median, they would fall short by 9.3%, rather than be able to use liquid financial wealth to augment their down payments. We therefore control for the amount of net financial assets in several of our specifications to ensure that we accurately measure the impact of these constraints on household decisions. This is a significant advance, given the measurement concerns that have affected prior work in this area.

3.2 Age and Education

Given the natural reduction in labor income generating opportunities as households approach retirement, we might also expect that mortgage credit availability reduces as households age. And both age and education have been shown in prior work to affect the incidence of departures from optimal household decision-making (e.g., Agarwal et al., 2009, Andersen, et al., 2018), meaning that we might expect preference-based heterogeneity across households along these dimensions. Figure A.5 Panel B shows the age and education distributions of households in the sample. As expected, home-owning households with mortgages are both older and more educated than the overall distribution of households.

3.3 Gains, Losses and Home Equity – Independent Variation

There are several challenges associated with estimating the independent and joint effects of down-payment constraints and gains on households' listing decisions. One important challenge is that home equity and expected gains/losses are likely to be highly correlated with one another, mainly because of their joint dependence on $\widehat{\ln P}$. Other factors that influence this correlation are the LTV ratio at origination, and households' decisions to remortgage to higher levels or to engage in subsequent "cash-out" refinancing after the initial issuance of the mortgage. A second challenge in cleanly estimating the effects of both constraints and gains on household behavior is that their effects could interact in complex ways. This means that sufficient independent variation is necessary to be able to estimate any interaction effects with reasonable precision.

We therefore document the extent to which there is independent variation in gains and home equity in the data. We first provide a simple classification of the household-years in the data into four groups, based on estimated $\widehat{\ln P}$, the purchase price of the home R, and the mortgage amount M. The groups are:

1. Unconstrained Winners (48.8%): $H \geq 20\%$ and $G \geq 0$.

2. Constrained Winners (26.5%): $H < 20\%$ and $G \geq 0$.[8]

3. Unconstrained Losers (6.2%): $H \geq 20\%$ and $G < 0$

4. Constrained Losers (18.6%): $H < 20\%$ and $G < 0$

The density of the data in each of the four groups is shown in Figure A.7. We show a vertical line at zero gains, and a horizontal line at 20% home equity. Under the assumption that households wish to move into a house of at least the same size as they currently own, and do not possess additional resources that they can bring to bear to augment the down payment, 20% current home equity is the constraint point, rather than zero home equity.

The figure shows that, as expected, there is a high correlation between the extent of home equity constraints and the gains and losses experienced by households. However, in our sample, there is considerable density off the principal diagonal of the plot. While

[8] $M > R$ is frequently observed in the data (44.8% of observations). This is primarily because of households' subsequent actions to remortgage to higher levels than their mortgage at issuance. This generally arises from "cash-out" refinancing, but could also arise from disadvantageous subsequent refinancing by homeowners, or fluctuations in adjustable rate mortgage payments causing households to increase mortgage principal to reduce monthly payment volatility.

this is reassuring, it could well be the case that this variation is confined to one particular part of the sample period, i.e., driven by time-variation in Danish house prices.

To check this, Figure A.8 plots the shares of seller groups in the data across each of the years in our sample. The figure shows that aggregate price variation does shift the relative shares in each group across years, with price rises increasing the fraction of unconstrained winners relative to losing and constrained groups. However, the relative shares still look fairly stable over the sample period, alleviating concerns that different groups simply come from different time periods, i.e., identification of any effects is likely to arise mainly from the cross section rather than the time series.

3.4 Generalized Logistic Functions and Interaction Effects

This rich set of interactions calls for a flexible and parsimonious model capable of capturing the observed shapes of the LP-G and LP-H relationships. To better document the facts about these patterns in the data, we estimate a simple model of reference points, borrowing a function commonly used in the biology literature to model the growth of organisms and populations. This is the generalized logistic function, also known as a Richards curve (Richards, 1959, Zwietering et al., 1990, Mead, 2017):

$$E[LP(V)] = A + \frac{K - A}{(1 + Qe^{-BV})^{1/\nu}}. \tag{9}$$

Here, the parameters A and K control the lower and upper asymptotes of the sigmoid function, and the parameters Q, B and ν control the position of the reference (i.e. inflection) point as well as the slope of the sigmoid curve at the reference point.

Figure A.11 plots the relationships estimated using the model in equation (9). We set V first as gains ($V = G$), and next, as the level of home equity $V = H$. Panel A of the figure has G along the x-axis, and LP along the y-axis. However, we now condition on three levels of H: the blue line shows the LP-G relationship for households with levels of H between 20 and 40% (i.e., effectively unconstrained households), while the red lines show the same relationship when households are increasingly constrained (the dashed red line when H is between -5% and 20%, and the solid line when H is between -15% and -5%).

To better understand these plots, we note that the average level of LP declines substantially as households become less constrained, and increases substantially as households become more constrained—this is simply the unconditional relationship between LP and H, seen in a different way in this plot. (Panel B of the figure shows the level differences that reflect the LP-G relationship, i.e., higher levels of LP for those with high realized losses (in red) relative to those experiencing gains (blue)).

What is more interesting here is that controlling for this change in level, the *slope* of LP as a function of G is also affected by the level of H. The important new fact is that down-payment-unconstrained households exhibit seemingly greater levels of reference dependence along the gain/loss dimension, exhibiting a pronounced increase in the slope to the left of $G = 0$. In contrast, down-payment constrained households exhibit a flatter LP across the G dimension.

The bottom panel shows another interesting fact—along the home equity dimension, while the slope around the threshold does not change, *the position of the kink* in the listing premium *increases* with the level of past experienced gains.

3.5 Conditional Effects on Listing Premia

Of course, these observations could simply be capturing the effect of other potential determinants for which the plots do not control, and indeed, we may be concerned yet again about the independent effects of G and H on LP. To check whether these conditional effects do indeed exist controlling for one another, and for a range of other determinants, and to verify whether they are statistically significant, we estimate the following piecewise-linear specification:

$$\begin{aligned} LP_{it} = \mu_t + \mu_m + \boldsymbol{\xi_0}\mathbf{X_{it}} + \boldsymbol{\xi_1}\mathbf{B_{it}} + \alpha_1 \mathbb{1}_{G_{it}<0} + \alpha_2 \mathbb{1}_{\mathrm{H}_{it}<20\%} \\ + (\beta_0 + \underbrace{\beta_1 \mathbb{1}_{G_{it}<0}}_{\text{Gains}} + \boldsymbol{\beta_2}\mathbf{B_{it}} + \underbrace{\boldsymbol{\beta_3} \mathbb{1}_{G_{it}<0}\mathbf{B_{it}}}_{\text{Conditional effect}})G_{it} \\ + (\gamma_0 + \underbrace{\gamma_1 \mathbb{1}_{\mathrm{H}_{it}<20\%}}_{\text{Down-payment constraint}} + \boldsymbol{\gamma_2}\mathbf{B_{it}} + \underbrace{\boldsymbol{\gamma_3} \mathbb{1}_{\mathrm{H}_{it}<20\%}\mathbf{B_{it}}}_{\text{Conditional effect}})\mathrm{H}_{it} \\ + \varepsilon_{it}. \end{aligned} \tag{10}$$

Equation (10) allows LP to depend (piecewise) linearly on both home equity H and gains G (through β_0, γ_0). We include time (μ_t) and municipality (μ_m) fixed effects, and controls $\mathbf{X_{it}}$ (household age, years of education, and net financial assets). The piecewise linear specification also allows for kinks in the linear relationship at a reference point of 0 for nominal gains, and 20% for home equity through β_1 and γ_1—these coefficients capture the "unconditional" effects of gains and home equity on household behavior. The baseline estimation is reported in Table A.3. To capture the conditional behavior, we bin both home equity and gains (as well as the other conditioning variables) and introduce dummy variables **B** into the regression of the respective other dimension to capture the different *LP*-*G* (and *LP*-*H*) relationships for these groups. We allow for **B** to modify both the unconditional relationship with G and H (β_2, γ_2), as well as any slope differential at the reference points (β_3, γ_3).[9]

Despite the considerable number of parameters in equation (10), the estimates point to interesting conditional variation in the data. The y-axis of Panel A of Figure A.9 shows the point estimate for the slope of the *LP*-*G* relationship for different bins of household covariates shown on the x-axis.

Panel B of Figure A.10 investigates the effect of down-payment constraints, conditioning the *LP*-*H* relationship on the level of household covariates.

4 Hedonic Pricing Model and Alternatives

4.1 Unobserved Quality

An important and often-repeated concern in the literature measuring the relationship between $\hat{\ell}$ and $\widehat{G}$ (and indeed, in estimating the function $\alpha(\hat{\ell})$) is that observed non-linearities could simply arise from measurement error in the underlying model for $\widehat{P}$.

We show that the asymmetries seen in the relationship between $\hat{\ell}$ and $\widehat{G}$, as well as in measured demand concavity are robust to estimating several different models of $\widehat{P}$ proposed in the literature as well as more novel approaches that we adopt to deal with unobserved quality (Figure A.14).

First, we use repeat sales to difference out property-specific fixed effects to absorb time-invariant unobserved house features (e.g., a sea view).[10] Second, we follow Guren (2018), and use more and less granular regional indexes of house price changes as instruments for $\widehat{P}$, as house price changes at the shire level are more plausibly exogenous to individual sellers who can only affect the unobserved quality of their house. Third, we implement the approach proposed by Genesove and Mayer (2001) to establish bounds on the effect of both unobservable property quality, and the possibility that the seller over- or under-

[9]Since we do not want to model any higher-order effects in this context, we exclude the respective gains bins from **B** when interacted linearly with the gains variable, and home equity bins when interacted linearly with the home equity variable. That is, we allow only for "cross-effects" in this specification.

[10]Exploiting the detailed nature of the Danish administrative data and ability to match additional data sources, we are also in the process of acquiring data on tax exemptions for renovation expenses to account for potentially time-variation in unobservable quality that may affect property value. The idea is that households who face a loss do not incur renovation expenses to improve the quality of the house that are systematically larger than those of households who face a gain.

paid at the initial purchase.[11] Fourth, we note that these concerns are potentially more muted in our setting since our hedonic model is estimated with high precision given the larger sample of transactions in our data, and the consequent ability to utilize a range of fixed effects. We additionally augment the model with cohort fixed effects, interactions between hedonics and house size, and controls for detailed demographic characteristics and households' financial circumstances, and continue to find the patterns in $\hat{\ell}$ and $\alpha(\hat{\ell})$ when we make these changes.[12] Fifth, we implement a quasi-experimental approach to establish a significant change in slope in a narrow neighbourhood around $\widehat{G} = 0$ (for $\hat{\ell}$) and in $\hat{\ell} = 0$ (for $\alpha(\hat{\ell})$), while other observable characteristics are visibly smooth around $\widehat{G} = 0$ and $\ell = 0$. This Regression Kink Design (RKD) (suggested by Card et al. 2015b and implemented e.g., by Landais, 2015, Nielsen et al. 2010, Card et al. 2015a) relies on quasi-random assignment at thresholds of particular "running variables" that induce kinks in agents' responses. As long as households can only imperfectly manipulate which side of the threshold they are on, unobserved property quality should not have a significant kink precisely at the threshold and the resulting differences in behavior above and below the threshold can be interpreted as causal. The different approaches are further described below.

4.2 Baseline Hedonic Model

We estimate the expected market price using a hedonic price model on our final sample of traded properties and predict prices for the entire sample of listed properties. The price in logs is estimated using the hedonic model

$$
\begin{aligned}
ln(P_{it}) = {} & \xi + \xi_t + \xi_m + \xi_{tm} + \beta_{ft} \mathbb{1}_{i=f} \mathbb{1}_{t=\tau} \\
& + \boldsymbol{\beta} \mathbf{X_{it}} + \boldsymbol{\beta}_{\mathbf{fx}} \mathbb{1}_{i=f} \mathbf{X_{it}} \\
& + \Phi(v_{it}) + \mathbb{1}_{i=f} \Phi(v_{it}) + \varepsilon_{it}.
\end{aligned}
$$

ξ is a constant, ξ_t are year fixed effects, ξ_m are municipality fixed effects (98 municipalities in total), and ξ_{tm} are municipality-year fixed effects. $\mathbb{1}_{i=f}$ is an indicator variable for whether the property is an apartment (denoted by f for flat) rather than a house. $\mathbf{X_{it}}$ is a vector of the following property characteristics: ln(lot size), ln(interior size), number of rooms, number of bathrooms, number of showers, a dummy variable for whether

[11]Compared to Genesove and Mayer's (2001) estimates of 2.5% to 3.5%, we find a range for listing price increases between 4.4% to 5.3% for every 10% increase in expected loss.

[12]We note that many of these variables have hitherto been considered unobservables in prior literature.

the property was unoccupied at the time of sale or retraction, ln(age of the building), a dummy variable for whether the property is located in a rural area, a dummy for whether the building is registered as historic, ln(distance to nearest of Denmarks four largest cities). $\Phi(v_{it})$ is a third-order polynomial of the previous-year tax assessed valuation of the property. The R^2 of the regression is 0.8638. The model fit is shown in Figure A.6.

4.3 Repeat Sales Estimation

To control for time-invariant unobserved heterogeneity in properties, we apply property fixed effect in a repeat sales sample. Since the hedonic model is based on our final sample of sold listings from 2009-2016, we run the fixed effects model on repeat sales within our final sample, but due to the short window, repeat sales are not as frequent. In order to increase repeat sales sample size, we also estimate the fixed effect model on repeat sales in the entire population of Danish real estate sales from 1992 to 2016. We estimate $\widehat{ln(P_{it})}$ using the model

$$
\begin{aligned}
ln(P_{it}) = & \xi + \xi_t + \xi_m + \xi_{tm} + \xi_p + \beta_{ft}\mathbb{1}_{i=f}\mathbb{1}_{t=\tau} \\
& + \boldsymbol{\beta}\mathbf{Y_{it}} + \boldsymbol{\beta}_{\mathbf{fy}}\mathbb{1}_{i=f}\mathbf{Y_{it}} \\
& + \Phi(v_{it}) + \mathbb{1}_{i=f}\Phi(v_{it}) + \varepsilon_{it}.
\end{aligned}
$$

with ξ_p being property fixed effects and $\mathbf{Y_{it}}$ being a vector of the following (potentially) time-invariant property characteristics: ln(interior size), number of rooms, number of bathrooms, number of showers, a dummy variable for whether the property was unoccupied at the time of sale or retraction, ln(age of the building), a dummy for whether the building is registered as historic. R^2 from estimation of the model is 0.9011.

4.4 Additional Models of $\widehat{P}$

We further include house prices estimated based on a municipality-, and shire-level house price index, respectively, and a model extension using size interactions and cohort (purchase year) fixed effects. An overview of the alternative model specifications is given in Table A.2 and the results are compared in Figures A.12 and A.13.

4.5 Genesove and Mayer (2001) Bounding Approach

We follow Genesove and Mayer (2001) to establish bounds on the relationship between expected gains and list prices given unobserved heterogeneity and variation in over-and under-payment at the previous transaction. In particular, we replicate Table 2 in their paper in Table A.4. As a baseline, comparing column (2) and (1), the effect from a 10% increase in expected losses can be bounded between a 4.4 to 5.3% increase in list prices, compared to their 2.5 lower bound and 3.5% upper bound estimate.

4.6 Regression Kink Design (RKD)

Following Card et al. (2017), we compute the RKD estimate of a given running variable V as follows:

$$\tau = \lim_{v \to \overline{v}_+} \frac{dE[LP_{it}|V_{it} = v]}{dv}\bigg|_{V_{it}=v} - \lim_{v \to \overline{v}_-} \frac{dE[LP_{it}|V_{it} = v]}{dv}\bigg|_{V_{it}=v}, \tag{11}$$

based on the following RKD specification (Landais 2015):

$$E[LP_{it}|V_{it} = v] = \kappa_m + \kappa_t + \boldsymbol{\xi}\mathbf{X_{it}} + \left[\sum_{p=1}^{\overline{p}} \gamma_p(\nu - \overline{\nu})^p + \nu_p(v - \overline{v})^p \mathbb{1}_{V \geq \overline{v}}\right]. \tag{12}$$

$$\text{where } |v - \overline{v}| < b. \tag{13}$$

As before, we include time (κ_t) and municipality (κ_m) fixed effects, and controls $\mathbf{X_{it}}$. These include household characteristics (age, education length, and net financial assets), as well as the previous purchase year, which we include to ensure that households are balanced along the dimension of housing choice, and is predetermined at the point of inclusion in this specification. V is the assignment variable, $\overline{v}$ is the kink threshold, $\mathbb{1}_{V \geq \overline{v}}$ is an indicator whether the experienced property return is above the threshold, and b is the bandwidth size.

To estimate the change in listing premium slope across gains, we choose $V = G$ as the assignment variable, and $\overline{v} = 0$ as the kink point. To estimate the effect of demand concavity, $V = \ell$, with a baseline kink threshold of $\overline{v} = 0\%$. Table A.5 reports results across bandwidths $b \in \{b^*, 15, 20\}$ around each of the running variables. b^* denotes the mean-squared-error optimally chosen bandwidth following Calonico et al (2014) and we use a polynomial order $p = 2$ for gains, and $p = 1$ for demand concavity.[13] Figures A.16

[13]The precision but not the size of the estimate for unconstrained households depends on the use of a local linear compared to a local quadratic function. Hahn et al. (2001) show that the degree of the

to A.18 show further robustness for the RKD using gains.

5 Institutional Background

5.1 Amendments to the Danish Mortgage-Credit Loans and Mortgage-Credit Bonds Act

Changes to the law regulating the loan-to-value ratio of mortgage loans between 2009 to 2016 are listed in Table A.12.

5.2 Foreclosures

Homeowners who cannot pay their mortgage or property tax may benefit from selling their home -– even if they have negative home equity — since they otherwise risk to be declared personally bankrupt by their creditors. If declared personally bankrupt, the property will be sold at a foreclosure auction. Foreclosures in most cases result in prices significantly below market price. Selling in the market will thus potentially allow homeowners to repay a bigger fraction of their debt. Homeowners with negative home equity may even be tempted to set higher listing prices to cover an even higher fraction of the debt. Whether this is optimal is debatable, since setting a higher listing price probably also reduces the probability of selling the property before a foreclosure process could begin.

5.3 The Foreclosure Process in Denmark

A foreclosure takes place if a homeowner repeatedly fails to make mortgage or property tax payments. After the first failed payment, the creditor (the mortgage lender or the tax authorities) first send reminders to the home owners and after approximately six weeks send the case to a debt collection agency. If the home owner after two to three months still fails to pay the creditor, the creditor will go to court (Fogedretten) and initiate a foreclosure. The court calls for a meeting between the owner and the creditor to guide the owner in the foreclosure process. At the meeting the owner and creditor can negotiate a

polynomial is critical in determining the statistical significance of the estimated effects. In particular, the second-order polynomial needed to identify derivative effects leads to an asymptotic variance of the estimate that is larger by a factor of 10 relative to the first-order polynomial. We verify that the qualitative patterns that we detect are broadly unaffected by the use of either polynomial order, but that the standard errors, consistent with Hahn et al. (2001), are substantially higher for the second-order polynomial, reported in Figure A.19.

short extension of four weeks to give the owner a chance to sell the property in the market. If that fails, the court has another four weeks, using a real estate agent, to attempt to sell the property in the market. After the attempts to sell in the market, the creditor will produce a sales presentation for the foreclosures, presenting the property and the extra fees that a buyer has to pay in addition to the bid price. The court sets the foreclosure date and at least two weeks before announces the foreclosure in the Danish Gazette (Statstidende), online, and in relevant newspapers. At the foreclosure auction interested buyers make price bids and highest bid determines the buyer and the price. If the buyer meets some financial requirements, the buyer takes over the property immediately and the owner is forced out. However, the owner can (and often will) ask for a second auction to be set within four weeks from the first. All bids from the first auction are binding in the second, but if a higher bid appears, the new bidder will win the auction.

The entire process from first failed payment to foreclosure typically takes six to nine months. At any point the owner can stop the foreclosure process by selling in the market and repaying the debt.

Selling in the market is preferred to foreclosure since foreclosure prices are significantly lower than market prices. Buyers have few opportunities to assess the house and have to buy the house "as seen" without the opportunity to make any future claims on the seller, making it a risky trade. In addition, buyers have to pay additional fees of more than 0.5 percent of the price.

6 Additional Tables and Figures

Figure A.1
Reference Dependence and Loss Aversion: Utility Functions

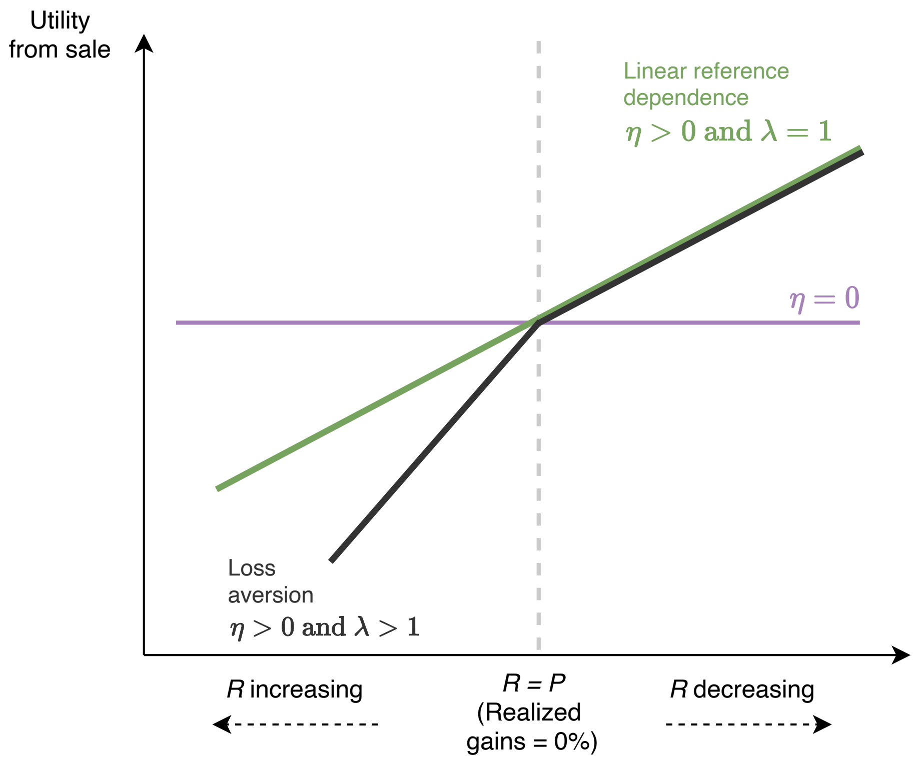


Figure A.2
Bunching

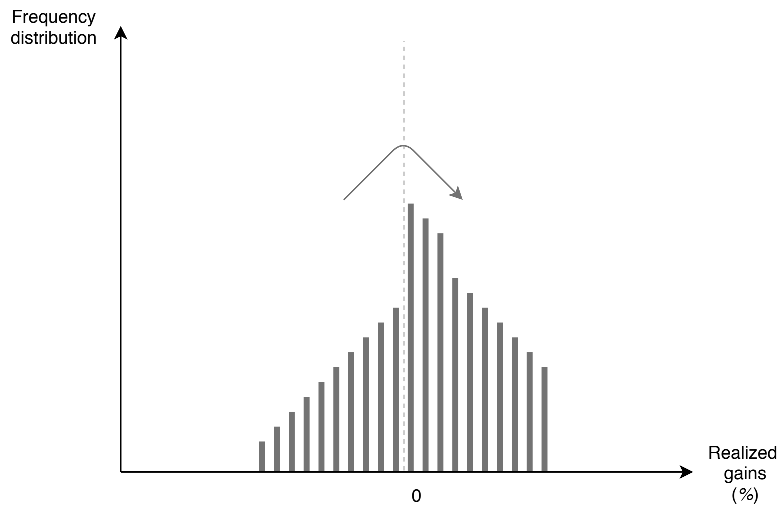


Figure A.3
Price-Volume Correlation

This figure shows quarterly average realized house sales prices (in DKK per square meter) on the right-hand axis, and the number of houses sold in Denmark on the left-hand axis, between 2004Q1 and 2018Q2. The sample period for our analysis covers the years 2009 to 2016. Aggregate housing market statistics are provided by Finans Danmark, the private association of banks and mortgage lenders in Denmark.

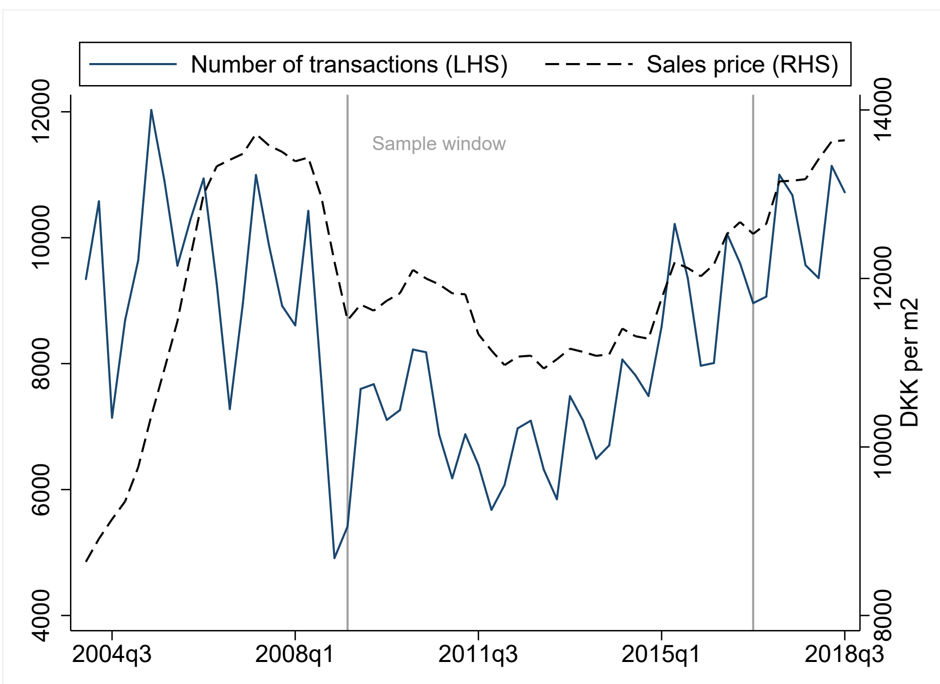
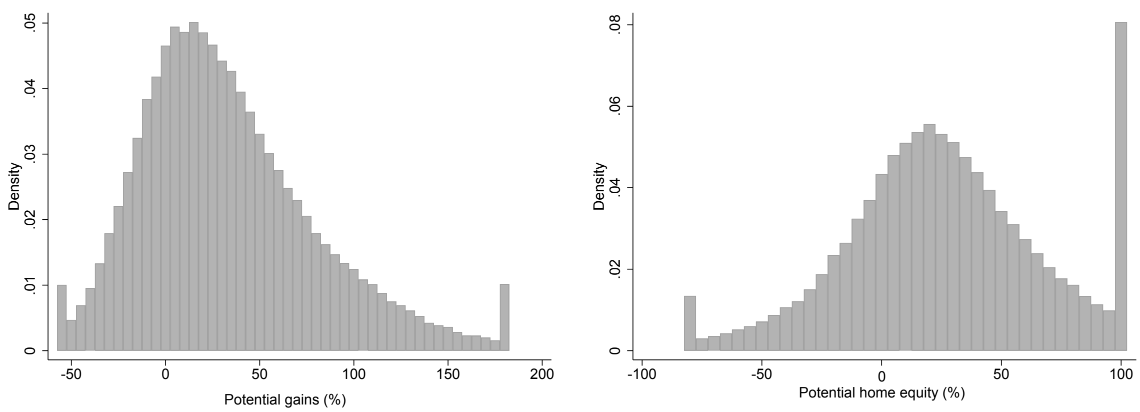


Figure A.4

Summary Statistics: Transaction Characteristics

This figure shows four histograms of main variables of interest. Gain (G) is computed as the log difference between the estimated hedonic price ($\widehat{P}$) and the previous purchase price (R), i.e. $G = \ln \widehat{P} - \ln R$, in percent. Home equity (H) is computed as the log difference between the estimated hedonic price and the current mortgage value (M), i.e. $H = \ln \widehat{P} - \ln M$, in percent. H is truncated at 100 in order to avoid small mortgage balances leading to log differences greater than 100. The listing premium (LP) measures the log difference between the ask price and estimated hedonic price, in percent. All are winsorized at 1 percent in both ends. Time on the market (TOM) measures the time in weeks between when a house is listed and recorded as sold. Each listing spell is restricted to 200 weeks.

Panel A



Panel B

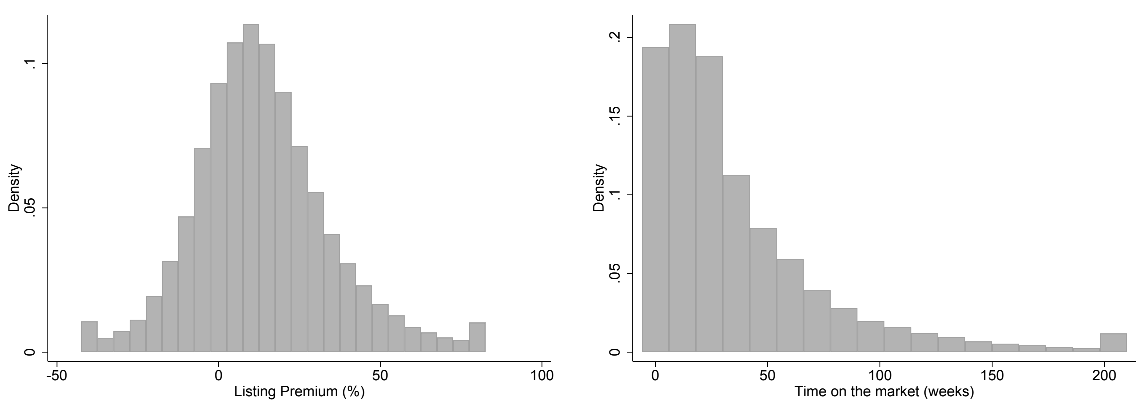
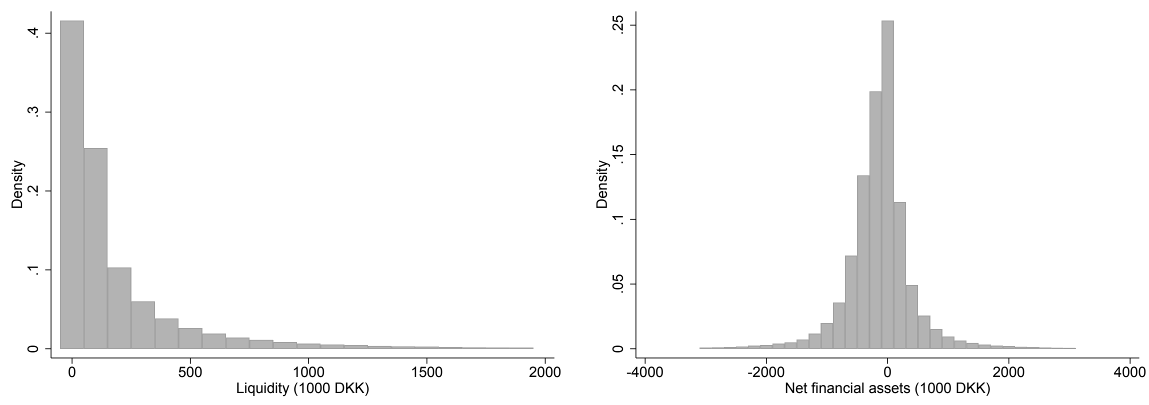


Figure A.5

Summary Statistics: Household Characteristics

This figure shows four histograms of household characteristics. Panel A shows the distribution of available liquid assets. Liquidity is measured as liquid financial wealth (deposit holdings, stocks and bonds). Net financial wealth is measured as liquid financial wealth net of bank debt. Panel B shows household characteristics. Age measures the average age in the household, and education length measures the average length of years spent in education across all adults in the household.

Panel A



Panel B

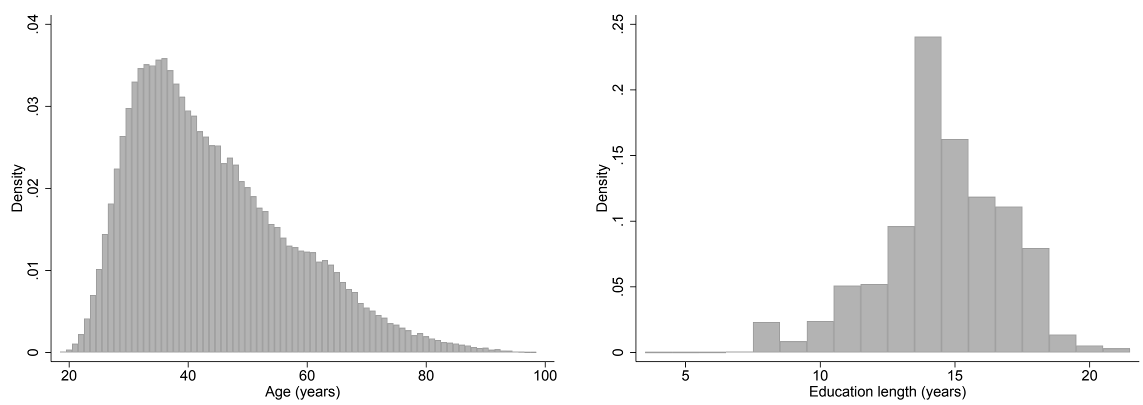


Figure A.6

Actual vs. Predicted Price of Sold Properties

This figure shows a binned scatter plot of the estimated log hedonic price $\ln(P_{it})$ versus the realized log sales price, for the sample of listings that resulted in a sale ($N = 117,408$). The hedonic model is as follows: $\ln(P_{it}) = \xi + \xi_t + \xi_m + \xi_{tm} + \beta_{ft}\mathbb{1}_{i=f}\mathbb{1}_{t=\tau} + \boldsymbol{\beta}\mathbf{X_{it}} + \boldsymbol{\beta_{fx}}\mathbb{1}_{i=f}\mathbf{X_{it}} + \Phi(v_{it}) + \mathbb{1}_{i=f}\Phi(v_{it}) + \varepsilon_{it}$, where $\mathbf{X_{it}}$ is a vector of property characteristics, namely ln(lot size), ln(interior size), number of rooms, number of bathrooms, number of showers, a dummy variable for whether the property was unoccupied at the time of sale or retraction, ln(age of the building), a dummy variable for whether the property is located in a rural area, a dummy for whether the building registered as historic, and ln(distance of the property to the nearest major city). ξ is a constant, ξ_t are year fixed effects, ξ_m are fixed effects for different municipalities (98 municipalities in total), and $\mathbb{1}_{i=f}$ is an indicator variable for whether the property is an apartment (denoted by f for flat) rather than a house. $\Phi(v_{it})$ is a third-order polynomial of the previous-year tax assessor valuation of the property. The R^2 of the regression is 0.86.

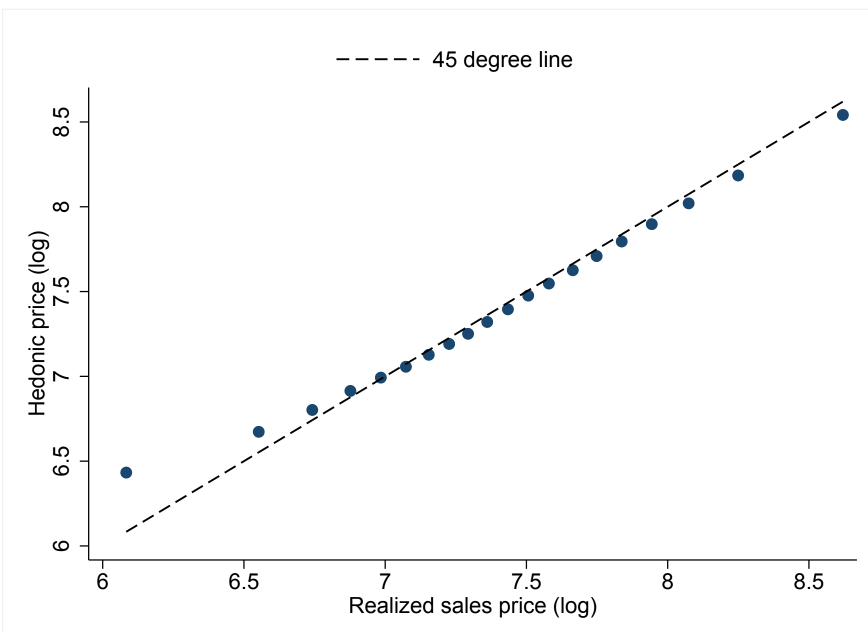


Figure A.7
Gains and Home Equity

This figure plots the joint distribution of the experienced gain and home equity position of households, at the time of listing. The color scheme refers to the relative frequency of observations in gain and home equity bins of 10 percentage points, where each color corresponds to a decile in the joint frequency distribution. The darker shading indicates a higher density of observations. Gain-home equity bins that did not have sufficient observations are shaded in white. The dotted blue lines separate the joint distribution in four groups: (1) Unconstrained Winners (H ≥20% and G ≥0) covering 48.8% of the sample, (2) Constrained Winners (H <20% and G ≥0) with 26.5%, (3) Unconstrained Losers (H ≥20% and G <0) with 6.2%, and (4) Constrained Losers (H <20% and G <0) accounting for 18.6% of the sample.

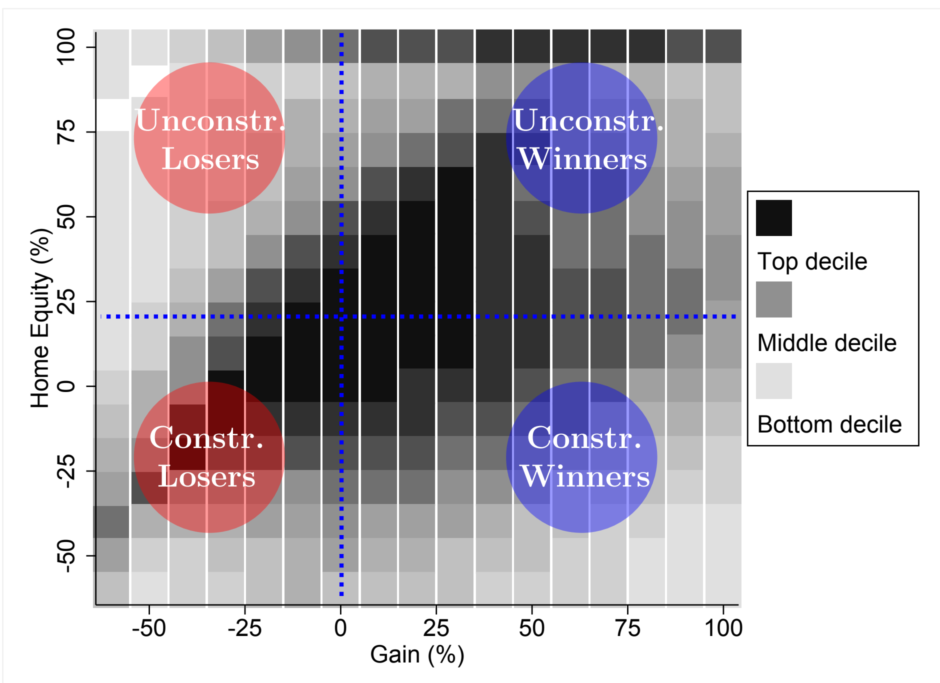


Figure A.8

Seller Groups - Listed (Relative Shares)

This figure shows the relative share of each seller group over time. The four groups are defined as follows: I) Unconstrained Winners ($H \geq 20\%$ and $G \geq 0$), II) Constrained Winners ($H < 20\%$ and $G \geq 0$), III) Unconstrained Losers ($H \geq 20\%$ and $G < 0$), IV) Constrained Losers ($H < 20\%$ and $G < 0$).

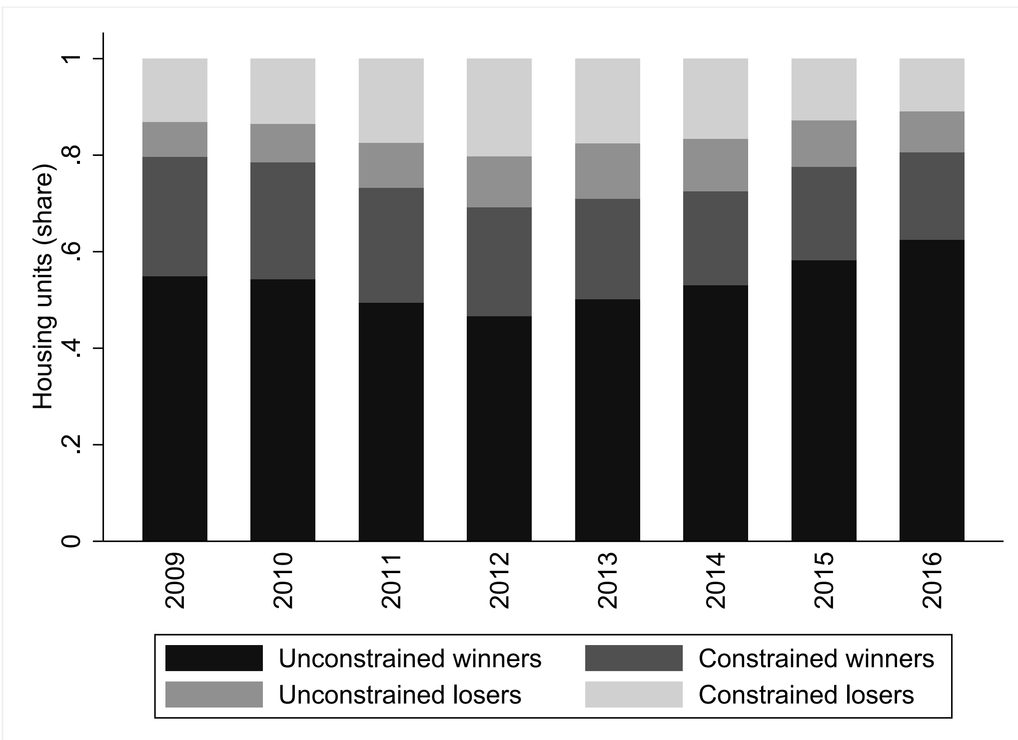


Figure A.9

Loss Aversion: Understanding Heterogeneity

This figure shows the effect of experienced gains on the ask-market-premium (AMP) across quantile bins of covariates (age, education length and net financial wealth). It reports estimated coefficients across different bins of covariates, which corresponds to the slope across the loss domain ($G < 0$), conditional on additional controls for home equity, and time and municipality fixed effects. The sign for $\beta_1 + \beta_3$ is reversed such that an increase in the coefficient can be read as an increase in the effect.

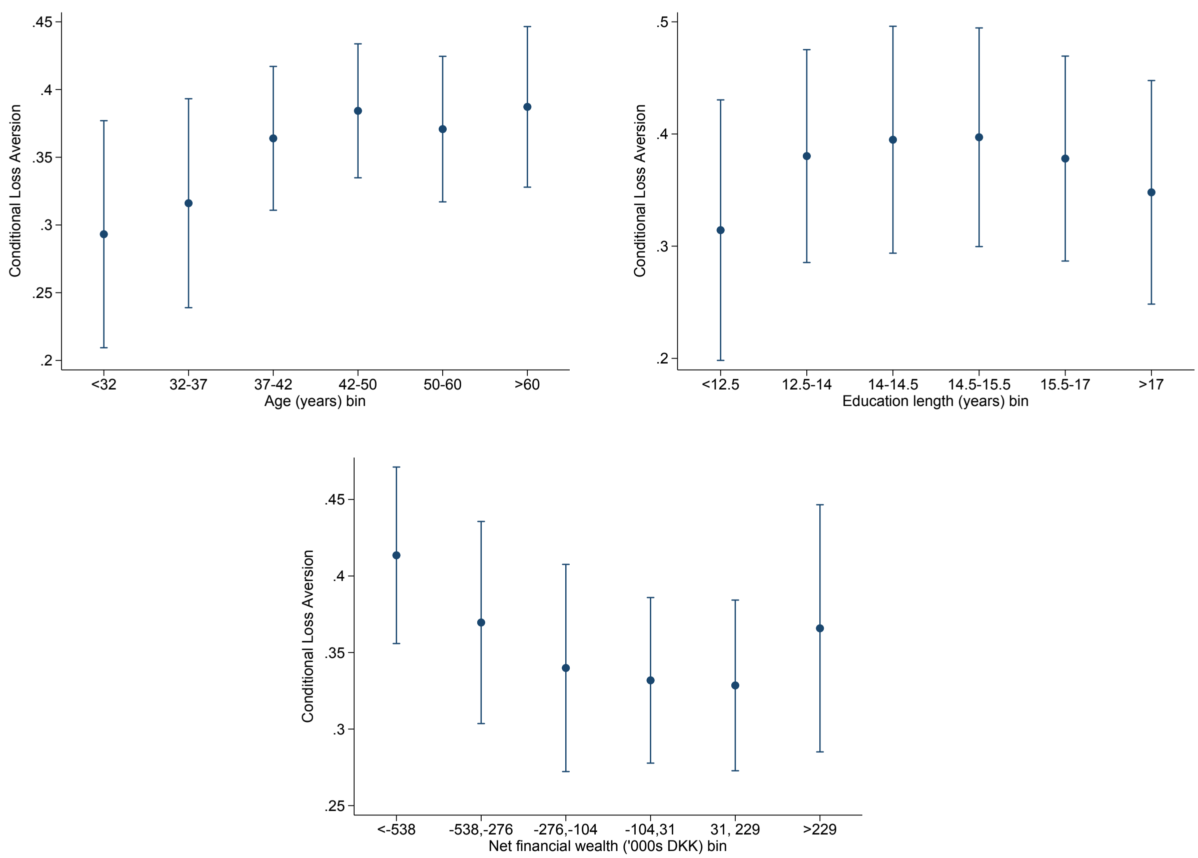


Figure A.10

Down-Payment Constraints: Understanding Heterogeneity

This figure shows the effect of home equity on the ask-market-premium (AMP) across quantile bins of covariates (age, education length, and net financial wealth). It reports the estimated coefficients across different bins of covariates, which corresponds to the slope across the constrained domain ($H < 20\%$), conditional on additional controls for experienced gains, and time and municipality fixed effects. The sign for $\gamma_1 + \gamma_3$ is reversed such that an increase in the coefficient can be read as an increase in the effect.

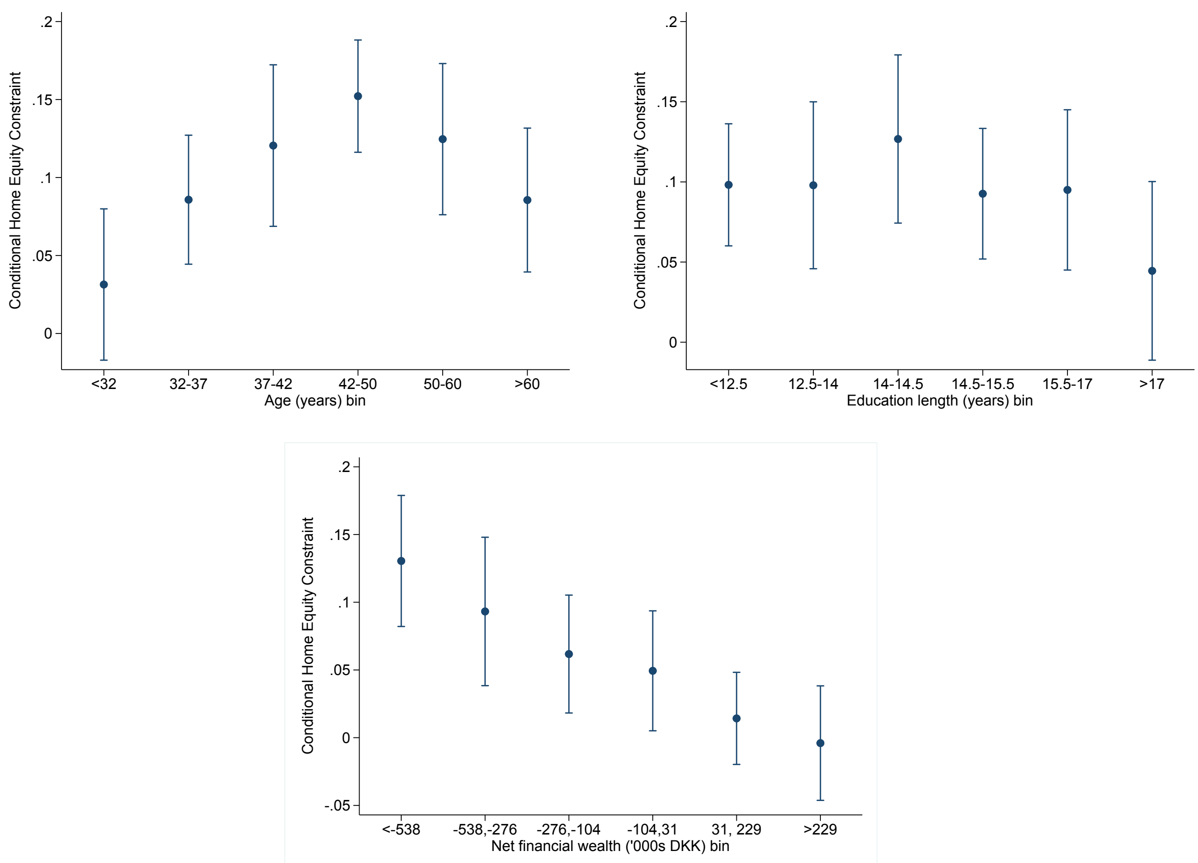
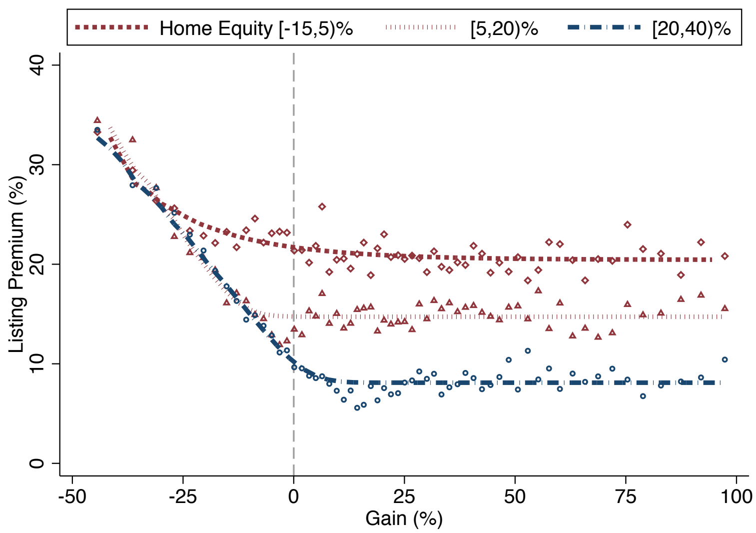


Figure A.11

Estimation of Generalized Logistic Functions (GLF)

This figure shows the effect of experienced gains (Panel A) and home equity (Panel B) on the listing premium. We report estimated relationships which follow a non-linear model specified in the form of a generalized logistics function $E[AMP(V)] = A + \frac{K-A}{(1+Qe^{-BV})^{1/\nu}}$, for which the underlying parameters A, K, Q, B, ν are estimated through a non-linear least squares procedure, and the assignment variables are $V = G$ and $V = H$ respectively. The solid dots indicate bin scatter points, for equally spaced bins of experienced gains and home equity.



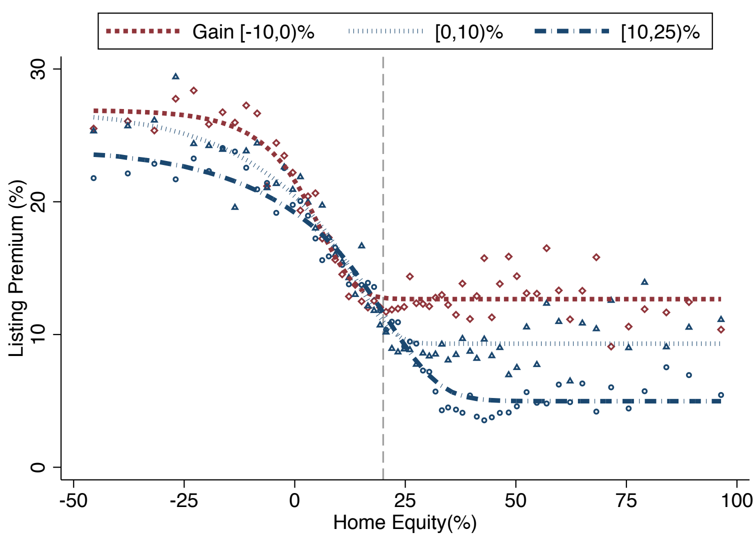


Figure A.12

Coverage of Alternative Models of $\widehat{P}$

This graph shows the number of observations for which we can estimate $\widehat{P}$ for different alternative models. *Hedonic* is a comprehensive hedonic model and our baseline specification. *Ext. hedonic* is an extended version of *Hedonic* which adds purchase year fixed effects and interacts all hedonic controls with three dummies for interior size. *Repeat* adds property fixed effects to *Hedonic* and is therefore restricted to repeated sales within the sample. *Mun. index* is the purchase price adjusted for local, i.e. municipality level, price changes and *Shire index* is the purchase price adjusted for local, shire level, price changes. If not indicated otherwise, models are estimated on the final sample of (repeated) sales from 2009 to 2016. If (full) is indicated, the model is estimated on the full sample of (repeated) sales from 1992 to 2016. *Repeat* > 2 *(full)* is restricted to properties sold at least three times during the full sample period.

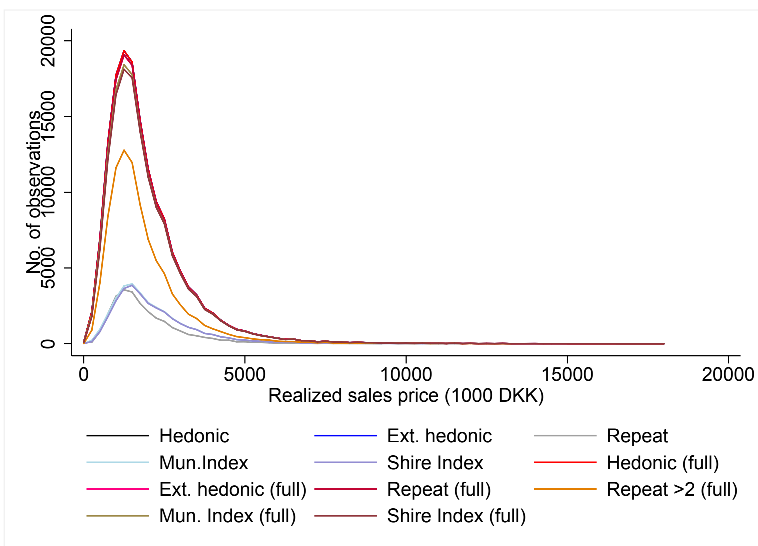
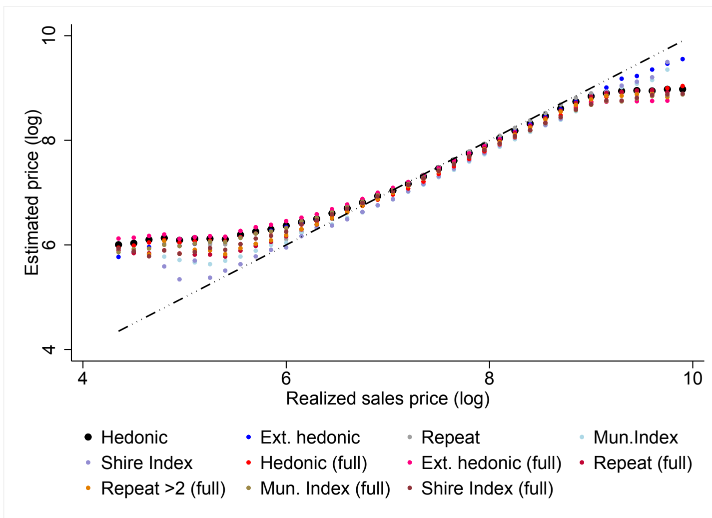


Figure A.13

Estimated vs. Realized ln(price)

This graph compares the model estimated price to the realized sales price in logs. *Hedonic* is a comprehensive hedonic model, and the baseline model for our main analysis. *Ext. hedonic* is an extended version of *Hedonic* which adds purchase year fixed effects and interacts all hedonic controls with three dummies for interior size. *Repeat* adds property fixed effects to *Hedonic* and is therefore restricted to repeated sales within the sample. *Mun. index* is the purchase price adjusted for local, municipality level, price changes and *Shire index* is the purchase price adjusted for local, shire level, price changes. If not indicated otherwise, models are estimated on the final sample of (repeated) sales from 2009 to 2016. If (full) is indicated, the model is estimated on the full sample of (repeated) sales from 1992 to 2016. *Repeat* > 2*(full)* is restricted to properties sold at least three times during the full sample period.

Panel A: All



Panel B: Below 5 mil. DKK

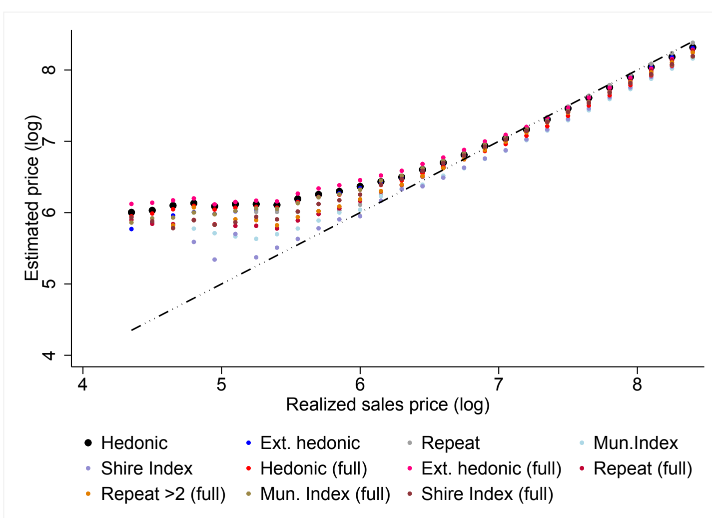
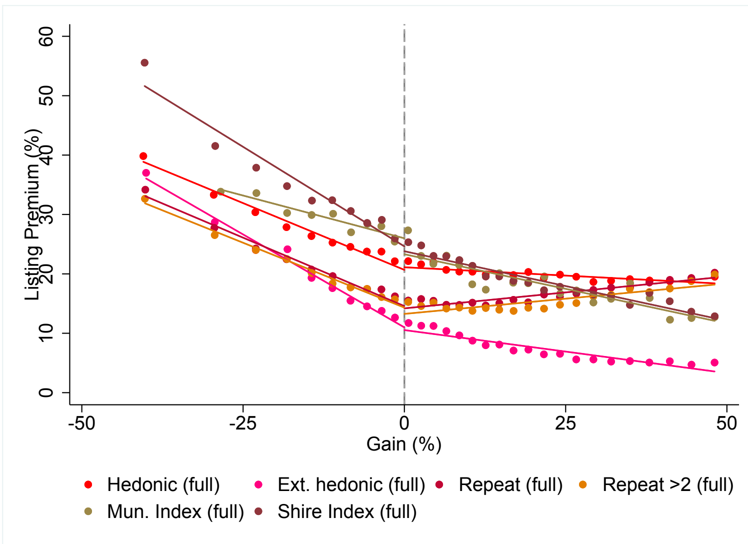


Figure A.14

Robustness to Alternative Models of $\widehat{P}$

These figures show the robustness of our two key empirical shapes to alternative specifications of $\widehat{P}$. Panel A show the listing price-to-gains relationship and Panel B shows demand concavity. *Hedonic (full)* is a comprehensive hedonic model. *Ext. hedonic (full)* is an extended version of *Hedonic (full)* which adds purchase year fixed effects and interacts all hedonic controls with three dummies for interior size. *Repeat (full)* adds property fixed effects to *Hedonic (full)* and is therefore restricted to repeated sales within the sample. *Repeat* > 2*(full)* is restricted to properties sold at least three times during the full sample period. *Mun. index (full)* is simply the purchase price adjusted for local, municipality level, price changes and *Shire index (full)* is the purchase price adjusted for local, shire level, price changes. All are estimated on the full sample of (repeated) sales from 1992 to 2016.

Panel A



Panel B

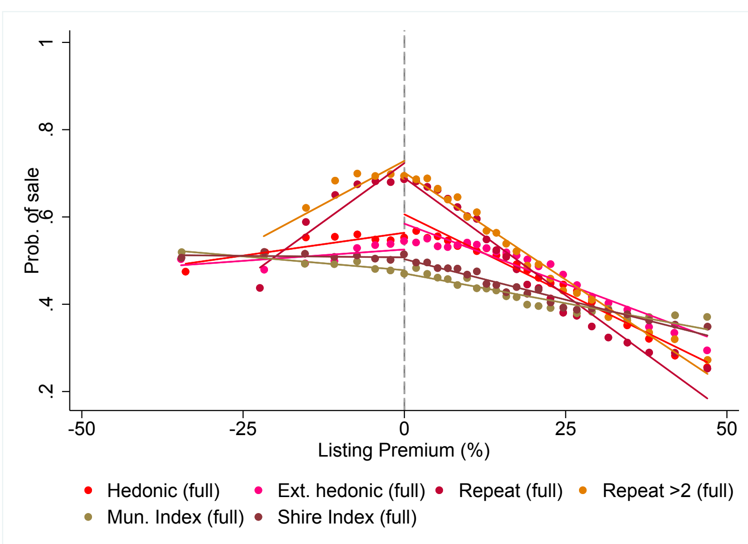


Figure A.15

Residual Listing Premium and Gains and Home Equity

This figure shows the relationship between residual listing premium and gains or home equity, respectively. The residual listing premium is computed with household controls (age, education length, net financial assets) and municipality and year fixed effects partialled out.

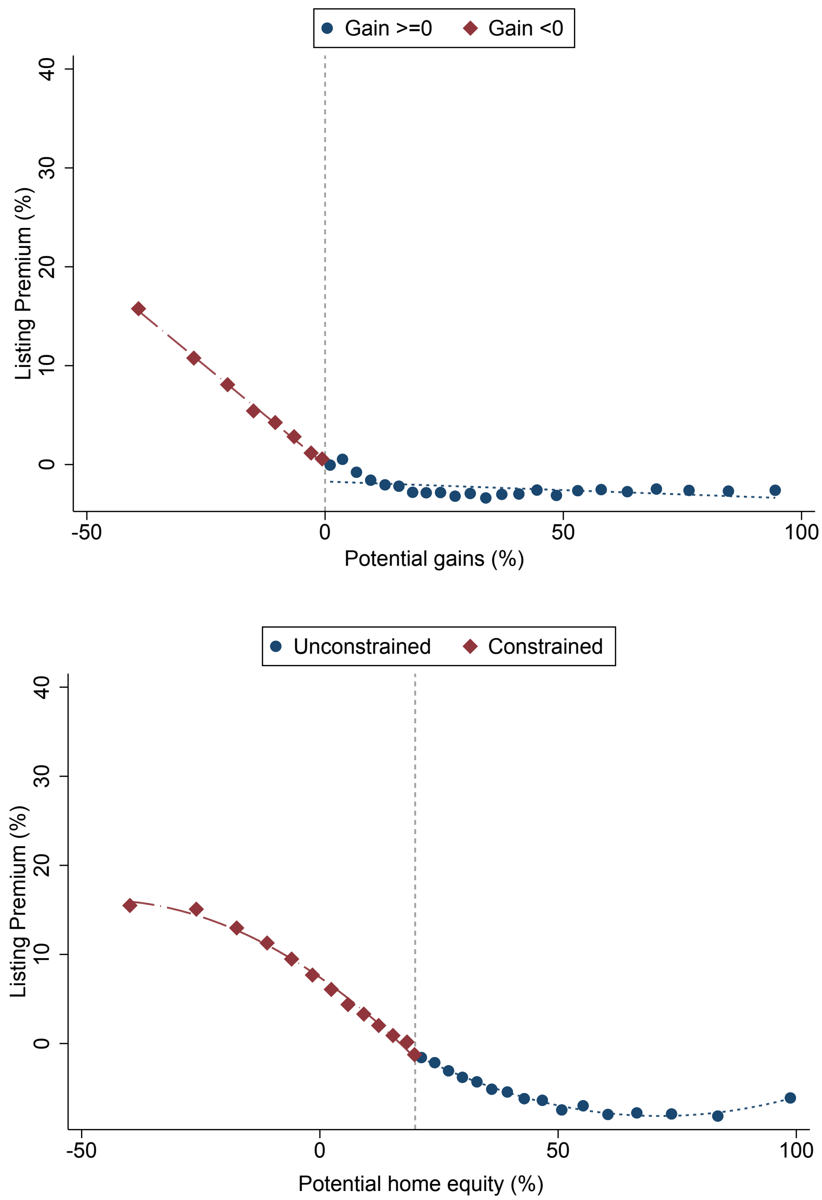


Figure A.16

RKD Validation: Smooth Density of Assignment Variable

This figure shows the number of observations in bins of the assignment variable, gain. Following Landais (2015), the results for the McCrary (2008) test for continuity of the assignment variable and a similar test for the continuity of the derivative are further shown on the figure. We cannot reject the null of continuity of the derivative of the assignment variables at the kink at the 5% significance level.[14]

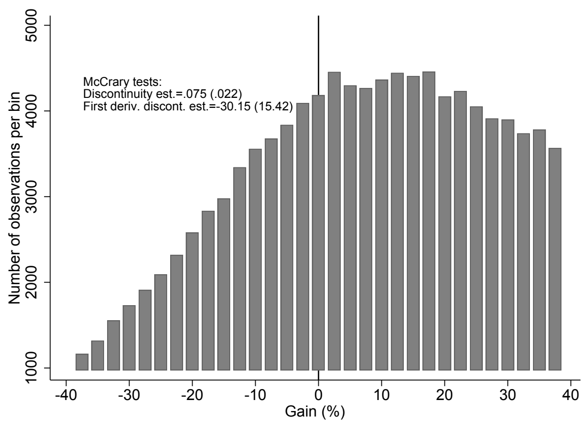


Figure A.17

RKD Validation: Covariates Smooth around Cutoff

This figure shows binned means of covariates (home equity/gain, age, length of education, liquidity, bank debt, financial wealth) over bins of the assignment variable, gain. It provides visual evidence for these covariates evolving smoothly around and not having a kink at the cutoff point.

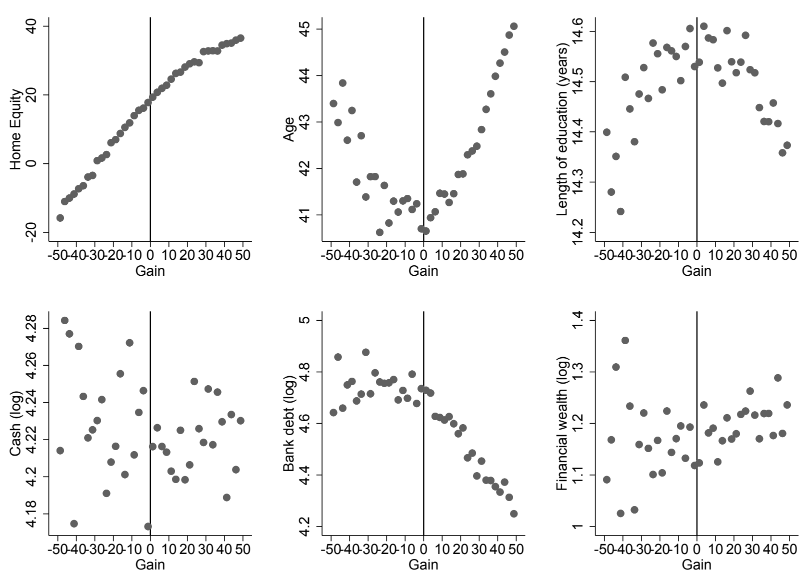


Figure A.18

RKD Robustness: Estimates for Different Bandwidths (Gain)

This figure plots the range of RKD estimates and 95% confidence intervals across bandwidths ranging from 5 to 50, using a local quadratic regression. The optimal bandwidth is indicated based on the MSE-optimal bandwidth selector from Calonico et al. (2014).

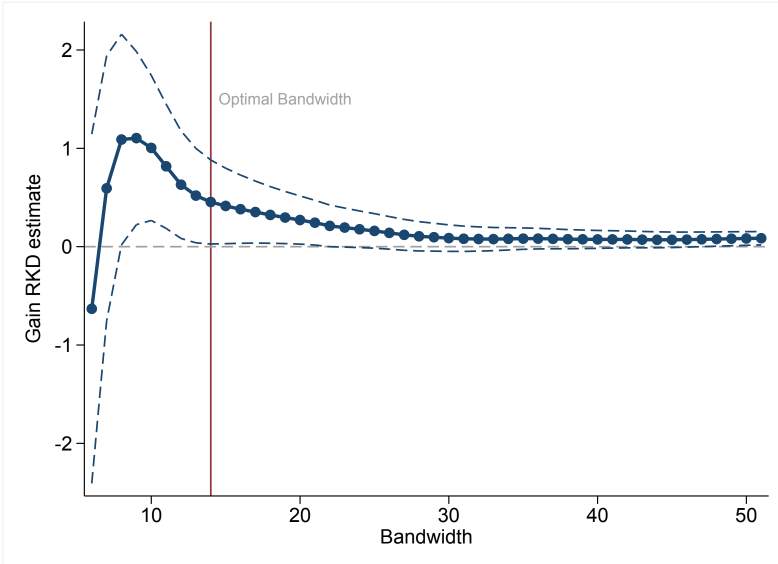


Figure A.19

RKD Estimation: Local Linear vs. Local Quadratic Estimation Results

This figure compares RK estimates using a local linear regression with estimates using a local quadratic regression, across different bandwidths $b \in \{b^*, 10, 20\}$, for gain (G) and probability of sale (P), respectively. b^* refers to the MSE-optimal bandwidth selector from Calonico et al. (2014).

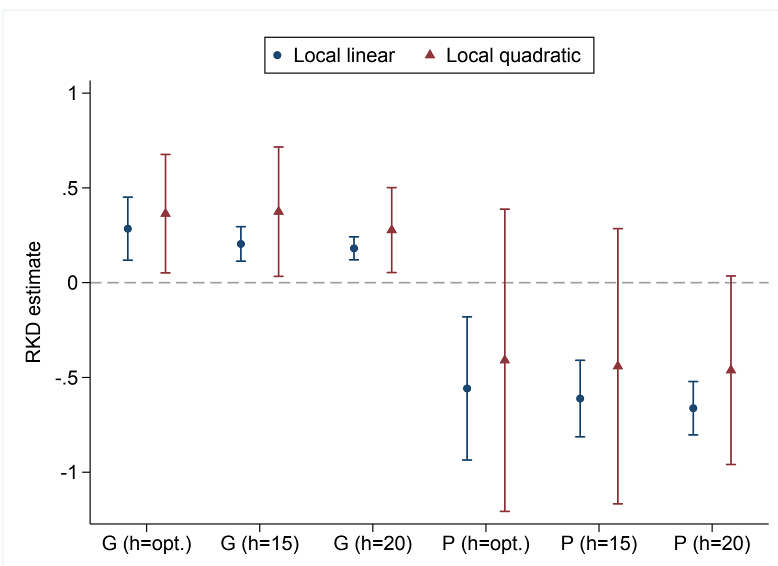


Figure A.20
Non-Mortgage Sample

This figure shows the relationship between listing premium and gains for the sample of households with no mortgage ($N = 41,382$), using a binned scatter plot of equal-sized bins for $\widehat{G} \in [-50, 50]$.

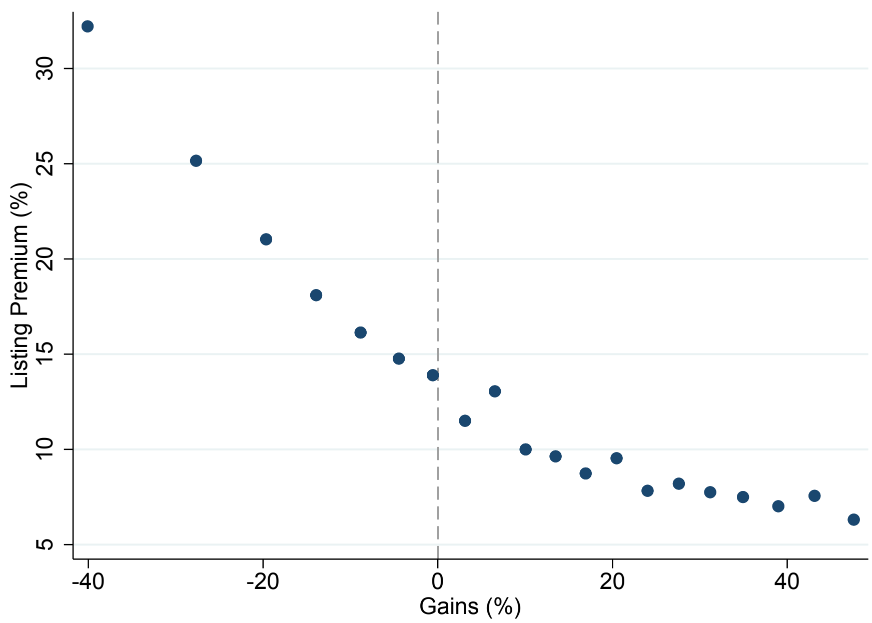


Figure A.21

Correlation between $\alpha(\ell)$ and $P(\ell)$ Levels

This figure shows the correlation between the level of the relationship between probability to sale as a function of the listing premium ($\alpha(\ell)$) on the x-axis and the level of the mapping between listing prices and realized prices ($P(\ell)$) on the y-axis across markets segmented by municipality.

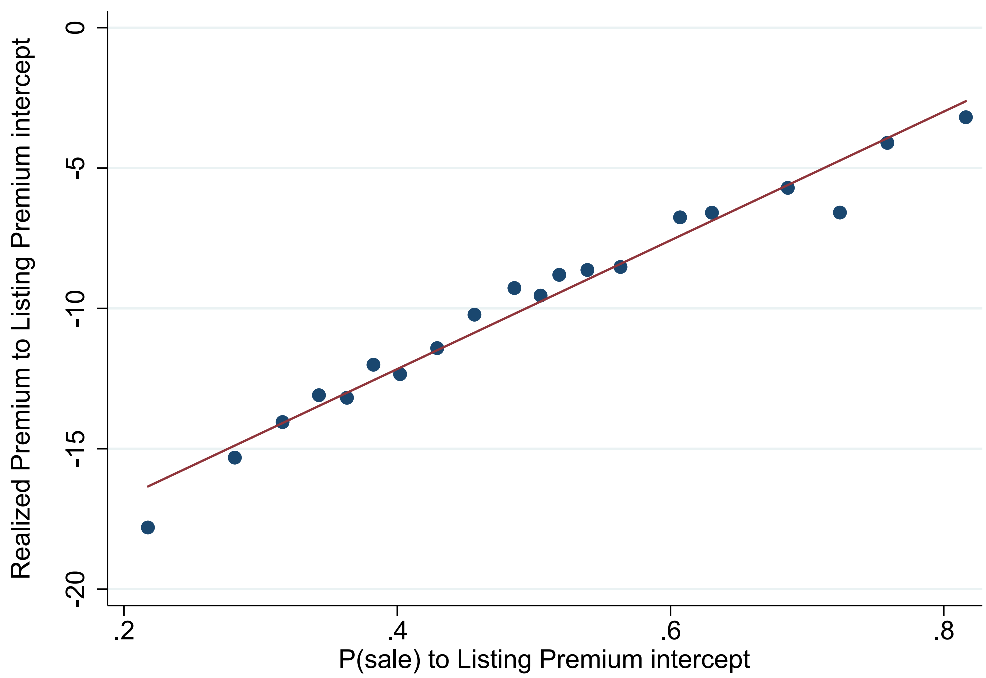


Figure A.22

Listing Premium Predicts Down-Payment

This figure shows a binned scatter plot of the ask-market-premium against the down-payment of a seller's next house, controlling for current home equity (H), based on a sub-sample of the data for which we have information on the next house purchase price and mortgage value ($N = 14,440$).

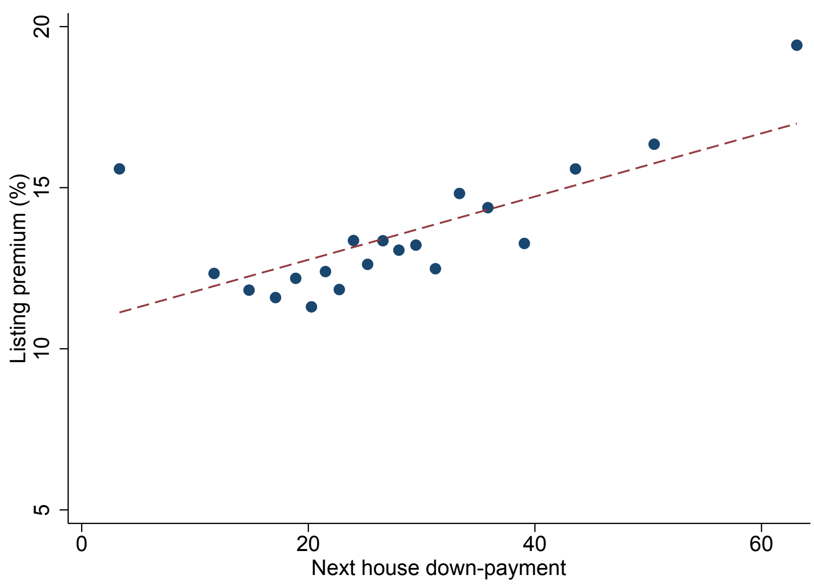


Figure A.23
Current and Next House Price

This figure shows a binned scatter plot of the current home price against the next house price (in 2015 DKK), based on a sub-sample of the data for which we have information on the next house purchase price and mortgage value ($N = 14,440$).

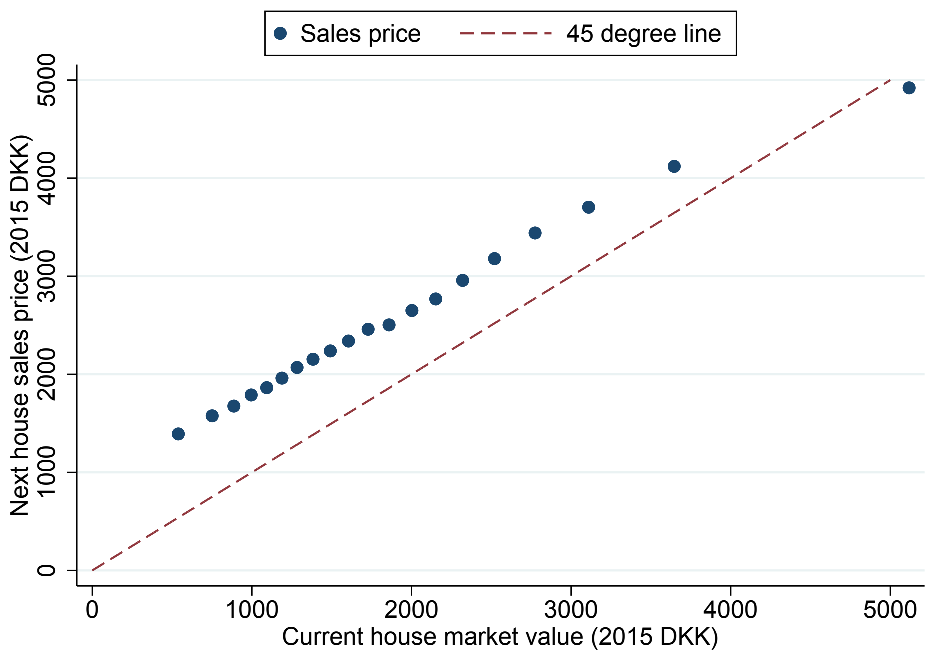


Figure A.24

Understanding the Extensive margin: Home Equity

This figure reports the share of listed houses relative to the stock of all houses, across 5% bins of home equity.

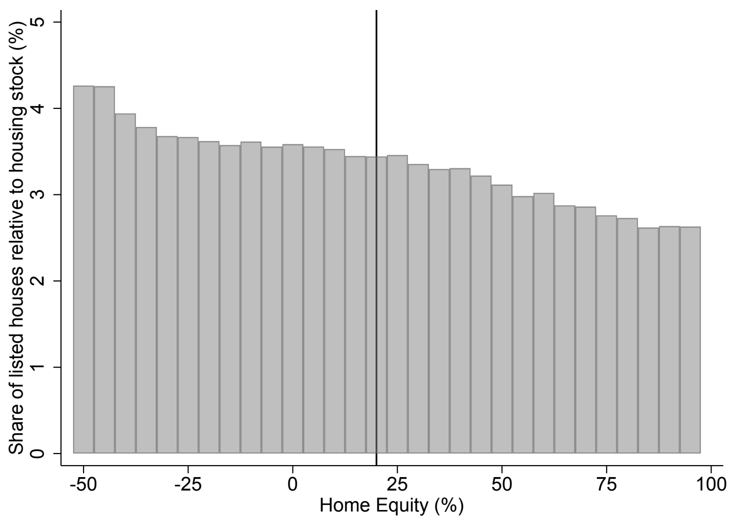


Figure A.25

Illustration of Homogeneity of Housing Stock for IV Estimation

Panel A illustrates what is defined as "row houses" in the Danish building and housing register (Bygnings- og Boligregistret). Each registered property can be looked up on the register via . The right-hand side shows a screenshot of the property outline of a house that is part of a row house unit. On contrast, Panel B shows the property outline of a detached single family house, which has visibly different features from other surrounding houses and is less homogeneous than the row house unit.

Panel A

Panel B

Figure A.26

Listing Premium-Gain Slope and Demand Concavity

This figure the listing premium over gains (left-hand side) and demand concavity (right-hand side) patterns when sorting municipalities by the estimated demand concavity, using municipalities in the top and bottom 5% of observations. Demand concavity is estimated as the slope coefficient of the effect of listing premium on probability of sale within six months, for $\ell \in [0, 50]$. The listing premium over gains slope is the slope coefficient of the effect of expected gains $\widehat{G}$ on listing premia, for $\widehat{G} < 0$.

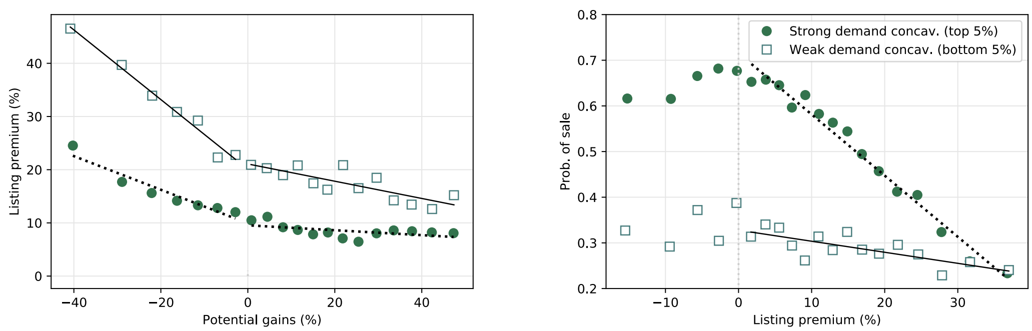


Figure A.27

Model fit

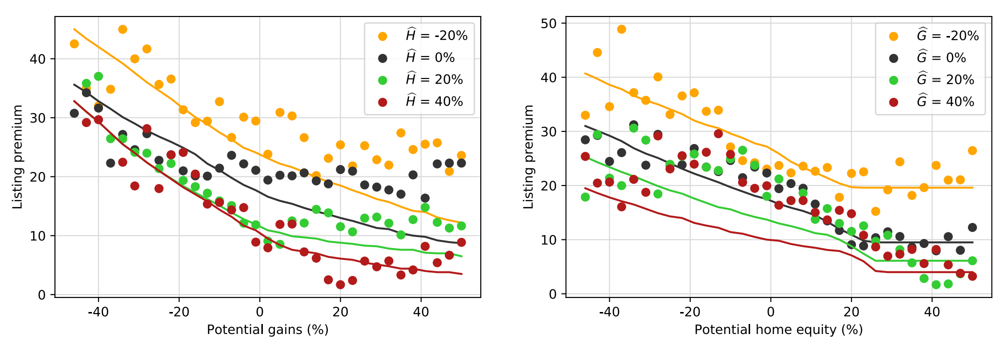


Table A.1

Construction of Main Dataset

This table describes the cleaning and sample selection process from the raw listings data to the final matched data, with $N = 175,646$.

All listings	**614,864**
Unmatched in registers[a]	-107,499
	507,365
Cleaning	
Owner ID not determined[b]	-71,944
Owners not found[c]	-4,026
Error in listing or purchase date[d]	-1,520
No purchase price[e]	-167,970
No ask price	-906
No predicted purchase price	-137
No hedonic price	-5
	260,857
Selecting	
Summerhouses	-23,706
Investors[f]	-20,123
No mortgage	-41,382
Final data	**175,646**

[a] *Reasons could be misreported addresses or not ordinary owner-occupied housing.*
[b] *E.g. properties with several owners from different households.*
[c] *No owner ID in registers*
[d] *Listing date is before purchase date*
[e] *Purchased before 1992*
[f] *Seller owns more than 3 properties*

Table A.2

Overview of Alternative Models for $\widehat{P}$

This table provides an overview of the alternative models for $\widehat{P}$ and the number of observations used for model estimation as well as the resulting number of estimated prices in the final sample. R^2 is from the model estimation of logs.

	Main hedonic	Extended hedonic	Repeat sales	Repeat sales (sold more than twice)	Municipality Price Index	Shire Price Index (2)
			Final sample estimation (2009-2016)			
Timeinvariant property characteristics	✓	✓	✓	-		
Timevariant property characteristics	✓	✓		-		
Property - size interactions		✓		-		
Municipality - sales year fixed effects	✓	✓	✓	-	✓	
Municipality - purchase year fixed effects		✓		-		
Shire - sales year fixed effects				-		✓
Property fixed effects			✓	-		
Final sample size	217,028	216,960	27,311	-	42,373	40,144
Estimation sample	153,256	153,228	25,967	-	153,256	150,268
R^2	0.8606	0.867	0.9709	-	0.3579	0.52
			Full sample estimation (1992-2016)			
Timeinvariant property characteristics	✓	✓	✓	✓		
Timevariant property characteristics	✓	✓				
Property - size interactions		✓				
Municipality - sales year fixed effects	✓	✓	✓	✓	✓	
Municipality - purchase year fixed effects		✓				
Shire - sales year fixed effects				-		✓
Property fixed effects			✓	✓		
Final sample size	217,028	192,719	194,680	114,818	205,966	201,311
Estimation sample	1,842,113	792,599	1,330,034	776,784	1,842,113	1,675,683
R^2	0.8162	0.7768	0.9011	0.882	0.513	0.6139

Table A.3

Loss Aversion and Down-Payment Constraints: Baseline Results

This table reports results for four regressions. Column (4) represents the estimated coefficients from the saturated regression

$$\ell_{it} = \mu_t + \mu_m + \boldsymbol{\xi_0}\mathbf{X_{it}} + \alpha_1 \mathbb{1}_{G_{it}<0} + \alpha_2 \mathbb{1}_{\mathrm{H}_{it}<20\%} + (\beta_0 + \beta_1 \mathbb{1}_{G_{it}<0})G_{it} + (\gamma_0 + \gamma_1 \mathbb{1}_{\mathrm{H}_{it}<20\%})H_{it} + \epsilon_{it},$$

where ℓ_{it} is the listing premium, μ_t and μ_m are year and municipality fixed effects, respectively, and $\mathbb{1}_{G_{it}<0}$ and $\mathbb{1}_{\mathrm{H}_{it}<20\%}$ are indicator functions for households who face an expected gain or home equity lower than 20%, respectively. Column (1) and (2) report results for specifications with only gain or home equity coefficients separately, and column (3) corresponds to column (4) but excludes household controls (age, liquid financial wealth and bank debt). Standard errors are clustered by year and municipality. */**/*** denote $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively.

	(1) LP	(2) LP	(3) LP	(4) LP
α_1	0.795*		-0.181	-0.206
	(0.351)		(0.294)	(0.277)
β_0	-0.041***		-0.014***	-0.018***
	(0.004)		(0.004)	(0.004)
β_1	-0.473***		-0.368***	-0.362***
	(0.035)		(0.030)	(0.031)
α_2		8.679***	6.798***	6.686***
		(0.787)	(0.752)	(0.733)
γ_0		-0.082***	-0.071***	-0.074***
		(0.007)	(0.006)	(0.006)
γ_1		-0.104***	-0.084***	-0.081***
		(0.026)	(0.022)	(0.023)
Household controls				✓
Year FE	✓	✓	✓	✓
Observations	173873	173873	173873	173873
R^2	0.182	0.230	0.266	0.270

Table A.4

Replicating Main Results from Genesove and Mayer (2001)

This table replicates Table 2 from Genesove and Mayer (2001) using our main dataset. The dependent variable is the log ask price. LOSS is the previous log selling price less the expected log selling price, truncated from below at 0, and LOSS (squared) is the term squared. LTV if $\geq$ 80 is the current LTV of the property if the LTV is greater equal to 80 and 0 otherwise. Estimated hedonic house prices are assumed to be additive in baseline value and market index, where baseline value captures the value of hedonic characteristics of the property and the market index reflects time-series variation in aggregate house prices. Residual from last sales price is the pricing error from the previous sale and months since last sale counts the number of months between the previous and current sale.

	(1) Ask (log)	(2) Ask (log)	(3) Ask (log)	(4) Ask (log)	(5) Ask (log)	(6) Ask (log)
LOSS	0.532***	0.444***	0.497***	0.328***	0.552***	0.466***
	(0.015)	(0.015)	(0.026)	(0.026)	(0.015)	(0.015)
LOSS (squared)			0.092*	0.297***		
			(0.053)	(0.051)		
LTV if $\geq$80	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Baseline value	0.988***	0.985***	0.988***	0.985***	0.989***	0.986***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Market index at listing	0.985***	0.982***	0.985***	0.982***		
	(0.003)	(0.003)	(0.003)	(0.003)		
Residual from last sales price		-0.082***		-0.084***		-0.078***
		(0.003)		(0.003)		(0.003)
Months since last sale	-0.000***	-0.000***	-0.000***	-0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.455***	0.486***	0.456***	0.490***	75.543***	75.336***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.218)	(0.217)
Year-Quarter FE					✓	✓
Observations	175646	175646	175646	175646	175646	175646
R^2	0.875	0.876	0.875	0.876	0.878	0.879

Table A.5

Regression Kink Design

The table shows results from sharp RKD tests of loss aversion, using the 0% gain cutoff, and demand concavity, using the 0% listing premium cutoff, for varying bandwidths $b \in \{b^*, 15, 20\}$. b^* refers to the optimally chosen bandwidth using a MSE-optimal bandwidth selector from Calonico et al. (2014). The control variables are year fixed effects, household controls (age, education length and net financial wealth) and year of previous purchase. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

	(1) Gain	(2) Gain	(3) Gain	(4) P(sale)	(5) P(sale)	(6) P(sale)
RK estimate	0.364**	0.375**	0.277**	-0.558***	-0.611***	-0.662***
	(0.159)	(0.174)	(0.114)	(0.193)	(0.103)	(0.072)
Cutoff	0.00	0.00	0.00	0.00	0.00	0.00
Bandwidth	16	15	20	9	15	20
Polynomial order	2	2	2	1	1	1
N below cutoff	43068	43068	43068	42731	42731	42731
N above cutoff	130809	130809	130809	131146	131146	131146

Table A.6

IV Robustness: Shire Level

This table reports regression results for the relationship between the listing premium slope over gains and demand concavity. The dependent variable in all regressions is the slope of the listing premium over $\widehat{G} < 0$, across shires with at least 30 observations. Column 1 reports the baseline correlation with the demand concavity slope across municipalities using OLS. Column 2 reports the 2-stage least squares regression instrumenting demand concavity with the apartment-and row-house share. Column 3 reports the overidentified 2SLS regression with both instruments, row-house and apartment share and average distance to city. Panel B includes household controls (age, education length, net financial assets, and log income). Standard errors are clustered at the municipality-year level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

Panel A

	(1) OLS	(2) 2SLS	(3) 2SLS (overid)
Demand concavity	-0.134***	-0.431***	-0.389***
	(0.027)	(0.122)	(0.114)
Observations	433	433	433
R^2	0.053		
First-stage F-stat	23.991	12.482	11.612
Hansen J-stat (p-val)			0.185

Panel B

	(1) OLS	(2) 2SLS	(3) 2SLS (overid)
Demand concavity	-0.087***	-0.431***	-0.427***
	(0.027)	(0.126)	(0.115)
Household controls	✓	✓	✓
Observations	433	433	433
R^2	0.167		
First-stage F-stat	17.082	13.271	13.767
Hansen J-stat (p-val)			0.936

Table A.7

IV Robustness: Logit Demand Concavity

This table reports regression results for the relationship between the listing premium slope over gains and demand concavity, using a logit specification for demand concavity. The dependent variable in all regressions is the slope of the listing premium over $\widehat{G} < 0$, across municipalities with at least at least 30 observations where $\widehat{G} < 0$ (Panel A), and shires with at least 30 observations, respectively (Panel B). Column 1 reports the baseline correlation with the demand concavity slope across municipalities using OLS. Column 2 reports the 2-stage least squares regression instrumenting demand concavity with the apartment-and row-house share. Column 3 reports the overidentified 2SLS regression with both instruments, row-house and apartment share and average distance to city. Panel B includes household controls (age, education length, net financial assets, and log income). Standard errors are clustered at the municipality-year level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

Panel A

	(1) OLS	(2) 2SLS	(3) 2SLS (overid)
Demand concavity	-0.228***	-0.457***	-0.464***
	(0.053)	(0.131)	(0.123)
Household controls	✓	✓	✓
Observations	95	95	95
R^2	0.607		
First-stage F-stat	27.520	21.835	22.026
Hansen J-stat (p-val)			0.888

Panel B

	(1) OLS	(2) 2SLS	(3) 2SLS (overid)
Demand concavity	-0.060***	-0.377***	-0.383***
	(0.023)	(0.114)	(0.108)
Household controls	✓	✓	✓
Observations	433	433	433
R^2	0.161		
First-stage F-stat	16.330	12.422	12.634
Hansen J-stat (p-val)			0.869

Table A.8

IV Robustness: Excluding Copenhagen and Aarhus

This table reports regression results for the relationship between the listing premium slope over gains and demand concavity, excluding the two largest cities in Denmark, Copenhagen and Aarhus. The dependent variable in all regressions is the slope of the listing premium over $\widehat{G} < 0$, across municipalities with at least at least 30 observations where $\widehat{G} < 0$ (Panel A), and shires with at least 30 observations, respectively (Panel B). Column 1 reports the baseline correlation with the demand concavity slope across municipalities using OLS. Column 2 reports the 2-stage least squares regression instrumenting demand concavity with the apartment-and row-house share. Column 3 reports the overidentified 2SLS regression with both instruments, row-house and apartment share and average distance to city. Panel B includes household controls (age, education length, net financial assets, and log income). Standard errors are clustered at the municipality-year level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

Panel A

	(1) OLS	(2) 2SLS	(3) 2SLS (overid)
Demand concavity	-0.291***	-0.476***	-0.487***
	(0.060)	(0.126)	(0.119)
Household controls	✓	✓	✓
Observations	93	93	93
R^2	0.628		
First-stage F-stat	29.382	24.845	25.041
Hansen J-stat (p-val)			0.792

Panel B

	(1) OLS	(2) 2SLS	(3) 2SLS (overid)
Demand concavity	-0.069**	-0.379***	-0.403***
	(0.030)	(0.118)	(0.115)
Household controls	✓	✓	✓
Observations	364	364	364
R^2	0.182		
First-stage F-stat	15.948	13.444	13.428
Hansen J-stat (p-val)			0.497

Table A.9

Alternative Listing Premia: Premia to Average House Prices

This table reports regression results for the relationship of listing premia (*LP*), the premium to the average house price in a given market (Premium to Avg.), and the *LP* residual when partialling out the effect of Premium to Avg. Premium to Avg. is defined as the ask price less the market average mu_{it}. Column (1) to (4) report results when mu_{it} is defined over municipality-years (with similar results for municipality-year-quarter, not reported here), and column (5)-(8) report results when mu_{it} is defined over shire-years. Standard errors are clustered at the municipality-year level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

	(1) P(sale)	(2) P(sale)	(3) P(sale)	(4) P(sale)	(5) P(sale)	(6) P(sale)	(7) P(sale)	(8) P(sale)
LP	-0.005***		-0.005***		-0.005***		-0.005***	
	(0.000)		(0.000)		(0.000)		(0.000)	
Premium to Avg		-0.001***	0.000***	0.000***		-0.001***	0.000***	0.000***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
LP residual				-0.004***				-0.005***
				(0.000)				(0.000)
Observations	175585	175585	175585	175585	174373	174373	174373	174373
R^2	0.046	0.004	0.046	0.034	0.046	0.003	0.047	0.037

Table A.10

Estimated Parameters (Alternative Identification, $\eta = 1$).

The table reports empirical moments (Panel A) and structural parameter estimates obtained through classical minimum distance estimation (Panel B). We recover concave demand $\alpha(\ell)$ and $P(\ell)$ from the data and set the down-payment constraint $\gamma = 20\%$. In parentheses, we report standard errors based on the estimated bootstrap variance-covariance matrix in the data, clustered at the shire level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

Panel A

Moments

1.	Level of LP for $G = 0\%$	0.104***	(0.005)
2.	Slope LP-G for $G < 0\%$	-0.492***	(0.052)
3.	Slope LP-G for $G \geq 0\%$	0.040*	(0.021)
4.	Slope LP-H for $H < 20\%$	-0.304***	(0.031)
5.	Slope of list. prob. by G	0.005**	(0.002)
6.	Bunching above $G = 0\%$	0.290***	(0.051)

Panel B

Parameter estimates

λ	=	1.58***	(0.25)
μ	=	1.08***	(0.19)
ξ	=	-0.11**	(0.05)
$\theta_{\min}$	=	0.25	(0.20)
$\theta_{\max}$	=	1.10***	(0.40)
φ	=	0.04	(0.04)

Table A.11

Estimated Parameters (Alternative Identification, No Concave Demand).

The table reports structural parameter estimates obtained through classical minimum distance estimation, in a model in which we assume linear demand $\alpha(\ell) = 0.6 - 0.53\ell$ estimated in the data. In this case we need to drop the moments implied by the cross-sectional variation of concave demand, and so we consider just a set of three moments (Level of LP for $G = 0\%$, Slope of LP-G for $\widehat{G} < 0\%$ and bunching above $G = 0\%$), and three parameters (η, λ and $\theta_{\max}$). All other parameters are as in the baseline specification. In parentheses, we report standard errors based on the estimated bootstrap variance-covariance matrix in the data, clustered at the shire level. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively.

η	=	0.750^{***}	(0.291)
λ	=	3.285^{***}	(0.867)
$\theta_{\max}$	=	4.535^{***}	(0.815)

Table A.12

Amendments to the Danish Mortgage-Credit Loans and Mortgage-Credit Bonds Act in the period from 2009 to 2016

May 2009	Allows a bankruptcy estate to make changes to fees in special circumstances
June 2010	Adjustments about bankruptcies
June 2010	Change of wording
December 2010	Change of wording
February 2012	Maximum maturity for loans to public housing, youth housing, and private housing cooperatives is extended from 35 to 40 years
December 2012	Elaboration of the rules on digital communication with the FSA
December 2012	Elaboration on the opportunity for mortgage credit institutions to take up loans to meet their obligation to provide supplementary collateral.
March 2014	Establish the terms under which the mortgage-credit institution can initiate sale of bonds if the term to maturity on a mortgage-credit loan is longer than the term to maturity on the underlying mortgage-credit bonds.
March 2014	Implements EU regulation. Change of wording on the definition of market value.
December 2014	Small additions to the terms under which the mortgage-credit institution can initiate sale of bonds if the term to maturity on a mortgage-credit loan is longer than the term to maturity on the underlying mortgage-credit bonds.
April 2015	Changes to the terms under which the mortgage-credit institution can initiate sale of bonds if the term to maturity on a mortgage-credit loan is longer than the term to maturity on the underlying mortgage-credit bonds.