Reference Dependence in the Housing Market*

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

July 21, 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. Our estimated parameters indicate reference dependence around the nominal purchase price and modest loss aversion. A new and interesting fact that the canonical model cannot match is that gains and down-payment constraints have interactive effects on listing prices.

*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, Henrik Kleven, Ralph Koijen, Ulrike Malmendier, Atif Mian, Karthik Muralidharan, Tomek Piskorski, Claudia Robles-Garcia, 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, BEAM, Cambridge Virtual Real Estate Seminar, Bank of England/Imperial/LSE Household Finance and Housing 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, Gomes et al. 2020). Decisions in the housing market are highly consequential, and are therefore a rich and valuable source of field evidence on households' underlying preferences, beliefs, and constraints. An influential example is the finding that listing prices for houses rise sharply when their sellers face nominal losses relative to the initial purchase price, originally documented by Genesove and Mayer (2001), and reconfirmed and extended in subsequent literature (see, e.g., Engelhardt (2003), Anenberg, 2011, Hong et al. 2019, and Bracke and Tenreyro 2020). This finding has generally been accepted as prima facie evidence of reference-dependent loss aversion (Kahneman and Tversky, 1989, Köszegi and Rabin, 2006, 2007).

Mapping these facts back to underlying preference parameters requires confronting challenges not fully addressed by the extant literature. A rigorous mapping permitting quantitative assessment of parameter magnitudes requires an explicit model of reference-dependent sellers. A plausible model would incorporate additional realistic constraints, such as the fact that optimizing sellers' listing decisions may be disciplined by demand-side responses.[1] Moreover, such a model would predict the behavior of a range of observables in addition to prices—which can be harnessed to accurately pin down parameters. For example, recent work assessing reference dependence in the field extracts information from transactions quantities (see, e.g., Kleven, 2016 and Rees-Jones, 2018), suggesting new moments to match in the residential housing market setting.

In this paper, we develop a new model of house selling decisions incorporating 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,

[1]Recent progress has been made on documenting the shape of housing demand (e.g., Guren, 2018), but it is important to understand how this affects inferences about the relationship between listing prices and sellers' "potential gains".

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 an extensive margin decision of whether to list, as well as an intensive margin choice of the listing price. Sellers maximize expected utility 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. We adopt a standard piecewise linear formulation of reference-dependent utility, characterized by two parameters: η captures how gains are weighed relative to the utility of the final sale price, and λ captures the asymmetric disutility of losses, i.e., conventionally, when $\lambda > 1$, sellers are loss averse. Sellers enjoy additional "gains from trade" from successful sales, receive an outside option utility level otherwise, and face down-payment constraints à la Stein (1995). Sellers take into account how their choices affect outcomes, i.e., the probability of sale as well as the final sale price, given housing demand.

We summarize a few important insights from the model here. When sellers exhibit "linear reference dependence" ($\eta > 0$, i.e., gains and losses matter to sellers, but $\lambda = 1$, i.e., there is no asymmetry between gains and losses), optimal listing premia decline linearly with "potential gains" (the difference between the expected sale price and the reference price) accrued since purchase. Intuitively, such linearly reference-dependent sellers facing losses require a greater final sales price to elevate the total utility received from a successful sale above that of the outside option. This leads them to raise (lower) listing prices in the face of potential losses (gains).[2] In addition, if sellers are loss averse, with $\lambda > 1$, then optimal listing premia slope up more sharply when sellers face potential losses than when they face potential gains, reflecting the asymmetry in underlying preferences.

These predictions on listing premia are mirrored in the behavior of quantities. With

[2]In the trivial case of no reference dependence, i.e., when $\eta = 0$, the model predicts that optimal listing premia are simply flat in potential gains.

linear reference dependence, completed transactions more frequently occur at realized gains (when the final sales price exceeds the reference price) than at realized losses. Put differently, $\eta > 0$ implies a shift of mass to the right in the distribution of transactions along the realized gains dimension, relative to the distribution when $\eta = 0$. With loss aversion, there is, in addition, sharp bunching of transactions precisely at realized gains of zero, and a more pronounced shift of mass of transactions away from realized losses.

Reference dependence and loss aversion also affect the extensive margin. The model predicts that the propensity to list rises in potential gains if $\eta > 0$. When $\lambda > 1$, there is also a pronounced decline over the domain of potential losses. Accounting for the extensive margin decision additionally helps to clean up inferences on the intensive margin, which can otherwise be biased by the drivers of selection into listing.

This discussion suggests that mapping reduced-form facts to underlying preference parameters is straightforward, but several key confounds can interfere. For one, the model reconfirms an issue recognized in prior work (e.g., Genesove and Mayer, 1997, 2001), that downsizing aversion à la Stein (1995) is difficult to separate from loss aversion. Down-payment constraints on mortgages create an incentive for households to "fish" with higher listing prices, since household leverage magnifies declines in collateral value, severely compressing the size of houses into which households can move. This effect of household leverage strongly manifests itself in listing prices in the data, but we document significant independent variation with potential gains, allowing us to cleanly identify loss aversion.

Second, accurate measurement of sellers' "potential gains" is important for our exercise. We confirm that the hedonic model that we employ to predict house prices in our main analysis fits the data with high explanatory power ($R^2 = 0.86$), and that our empirical work is robust to alternative house price prediction approaches. Third, relatedly, as Genesove and Mayer (2001), Clapp et al. (2018), and others note, variation in the unobservable property quality and potential under- or over-payment at the time of property purchase are important sources of measurement error. As we describe later, we adopt a wide range of strategies to check robustness to this possible confound.[3]

Fourth, the shape of demand is very important for model outcomes. If sale probabilities respond linearly and negatively to higher listing prices ("linear demand"), there are material incentives to set low list prices to induce quick sales. However, Guren (2018) shows that U.S. housing markets are characterized by "concave demand," i.e., past a point, lowering list prices does not boost sale probabilities, but does negatively impact realized sale prices; we confirm this finding in the Danish data.[4] The model reveals that this can generate a nonlinear optimal listing price schedule even without any underlying loss aversion. Intuitively, in the face of linear demand, a seller with $\eta > 0$ and $\lambda = 1$ linearly lowers list prices with potential gains, focusing on inducing a swift sale. However, when facing concave demand, lowering list prices past a point is unproductive, leading to an observed "flattening out" in the optimal listing price schedule, which is then nonlinear even though $\lambda = 1$. A related and important observation from the model is that sharp demand responses to raising listing prices are associated with weaker listing price responses to losses, and vice versa.

Keeping these potential confounds in mind, we outline the main facts in the data. First, the listing price schedule has the characteristic "hockey stick" shape first identified by Genesove and Mayer (2001), rising substantially as expected losses mount, and virtually flat in gains. Our estimates are similar in magnitude to those in that paper despite the differences in location, sample period, and sample size.[5] Second, listing premia vary considerably across regional housing markets in Denmark which exhibit varying degrees

[3]This includes estimation with property-specific fixed effects, applying bounding strategies previously proposed in the literature (Genesove and Mayer, 2001), utilizing an instrumental variables approach proposed by Guren (2018), and employing 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.

[5]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 5.4% for the same 10% nominal loss in the Danish data of apartments, row houses, and detached houses between 2009 and 2016 [N=173,065].

of demand concavity. This variation is consistent with the model: steep listing premia responses to losses are observed in markets with weaker demand concavity, and vice versa. These regional moments provide additional discipline to our structural estimation exercise and help account for the demand-concavity confound. Third, we see sharp bunching in the sales distribution at realized gains of zero, and a significant shift in mass in the distribution of sales towards realized gains and away from realized losses. Fourth, we estimate listing propensities for the entire Danish housing stock of over 5.5 million housing units as a function of potential gains. There is a visible increase in the propensity to list houses on the market as potential gains rise, and the slope appears more pronounced over the potential loss domain than the potential gain domain.

Taken together, these facts appear consistent with underlying preferences that are both reference dependent and loss averse around the original nominal purchase price of the house. To more rigorously map these facts back to the model, we structurally estimate seven model parameters using seven selected moments from the data (including those described above) using classical minimum distance estimation in this exactly identified system. The resulting point estimates yield $\eta = 0.948$ (*s.e.* 0.344), meaning that gains count about as much as final prices for final utility, and $\lambda = 1.576$ (*s.e.* 0.570), 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 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).[6] This strongly reinforces a broader message (see, e.g., Blundell, 2017) that realistic frictions need to be incorporated when mapping reduced-form facts from the data to inferences about deeper underlying parameters, strengthening the case for applying a structural

[6] 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)—both of which our model incorporates.

behavioral approach (DellaVigna (2009, 2018)) to field evidence. Finally, the estimated parameters also reveal strong evidence of the down-payment channel originally identified by Stein (1995), reveal significant "gains from trade" from successful house listings, and highlight that there are substantial (psychological and transactions) costs associated with listing.

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 how the model matches the entire surface of the listing premium along the home equity and gains dimensions. A novel pattern that we uncover, and that our model cannot match, is that home equity and expected losses have interactive effects on listing prices in this market. 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%,[7] 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 potential targets for future theoretical work.

The paper is organized as follows. Section 2 introduces the model of household listing behavior. Section 3 discusses the construction of our merged dataset, and provides descriptive statistics about these data. Section 4 introduces the moments that we use for structural estimation and uncovers new facts about the behavior of listing prices and

[7]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.

listing decisions. Section 5 describes our structural estimation procedure, and reports parameter estimates. Section 6 describes validation exercises, and highlights areas where the model falls short in explaining features of the data. Section 7 concludes.

2 A Model of Household Listing Behavior

We develop 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 framework can flexibly embed different preferences and constraints that have commonly been used to explain patterns in listing behavior. In this section we describe the main features and specific predictions of the model, which we later structurally estimate to recover key preference and constraint parameters from the data.

2.1 General Framework

The market consists of a continuum of sellers and buyers of residential property. There are two periods in the model: in period 0, some fraction of property owners receive a shock $\theta \sim \text{Uniform}(\theta_{\min}, \theta_{\max})$, and decide (i) whether or not to put their property up for sale, and (ii) the optimal listing price in case of listing. This "moving shock" θ can be thought of as a "gain from trade" (Stein, 1995), i.e., a boost to lifetime utility which sellers receive in the event of successfully selling and 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 final sale price. If negotiations fail, the seller stays in the property, and a receives a constant level of utility $\underline{u}$.

We seek to uncover 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 L denote the listing price set by the seller; $\widehat{P}$ be a measure of the "expected" or "fundamental" property value;[8,9] $\ell = L - \widehat{P}$ be termed the listing premium; let α denote the probability that a willing buyer will be found; and P denote the final sale price resulting from the negotiation between buyer and seller where $P(\ell) = \widehat{P} + \beta(\ell)$.

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

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

The seller decides on the extensive margin of whether $(s = 1)$ or not $(s = 0)$ to list, as well as the listing premium ℓ, to maximize expected utility from final sale of the property. For a listed property, there are two possible outcomes in period 1, which depend on ℓ. With probability $\alpha(\ell)$ the negotiation succeeds, and the seller receives utility from selling the property for an equilibrium price $P(\ell) = \widehat{P} + \beta(\ell)$. 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.

When making these listing decisions, the seller takes $\alpha(\ell)$ and $\beta(\ell)$, i.e., the "demand"

[8]Guren (2018) assumes that the buyer's expected 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/hedonic 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.

[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 percentages (which we later map to logs, relying on the usual approximation), to be consistent with the definitions of gains/losses and home equity employed in our empirical work.

functions, as given; 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. As in Guren (2018), we note that sufficient statistic formulas (Chetty, 2009) for equilibrium outcomes are mappings between sale probabilities $\alpha(\ell)$, final sale prices $P(\ell) = \widehat{P} + \beta(\ell)$, and listing premia ℓ. In particular, the realized premium $\beta(\ell)$ of the final sales price P 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. This assumption simplifies the model, and allows us to more closely focus on our goal, namely, extracting the underlying parameters of seller utility and constraints.[10]

The functions $\alpha(\ell)$ and $\beta(\ell)$ 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, but decreases the probability that a willing buyer will be found within a reasonable time frame.[11] 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 $\beta(\ell)$ collectively as *concave demand*, following Guren (2018), who documents using U.S. data that above average list prices increase 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 patterns in the Danish data.

We next describe the components of $U\left(P(\ell),\cdot\right) = u(P(\ell),\cdot) - \kappa(P(\ell),\cdot)$, 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 Köszegi and Rabin (2006, 2007), as well as down-payment constraints $\kappa(P(\ell),\cdot)$ à la Stein (1995).

[10]As we describe later, we do 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 maximizes subject to their perceived $(\alpha(\ell) + \delta)$ probability.

[11]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, 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), the seller's reference price level is R, which we simply assume is fixed, and in our empirical application, we set R to the original nominal purchase price of the property.[12] Realized gains $G(\ell)$ relative to this reference level are then given by $G(\ell) = P(\ell) - R$.

The parameter η captures the degree of reference dependence. Sellers derive utility both from the terminal value of wealth (i.e. the final price P realized from the sale), as well as from the realized gain G 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 to study and rationalize results found in the lab (e.g., Ericson and Fuster, 2011), as well as in the field (e.g., Anagol et al., 2018).

2.2.1 State Variables

In the model, seller decisions are determined by four state variables, namely, the moving shock θ, the hedonic value of the property $\widehat{P}$, the reference point R, and the outside option level $\underline{u}$. To map model quantities more directly to estimates in the data, and to make our setup more directly comparable to extant empirical and theoretical literature, we calculate the seller's expected or "potential" gains $\widehat{G} = \widehat{P} - R$ as a transformation of

[12] 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.

two of the state variables.[13] Realized gains $G(\ell)$ arise from their "potential" level $\widehat{G}$ plus the markup/premium $\beta(\ell)$, i.e.:

$$G(\ell) = \widehat{G} + \beta(\ell).$$

The remaining two state variables θ and $\underline{u}$ are unobserved, but only the wedge between them $(\underline{u} - \theta)$ is relevant for the seller's decision. Without loss of generality, we therefore set the outside option $\underline{u} = \widehat{P}$, which implies that absent any additional reasons to move $(\theta = 0)$, and with costless and frictionless listings, the seller will be indifferent between staying in their home and receiving the hedonic value in cash. This assumption can equivalently be mapped onto a specification in which the seller does not receive any gains from moving, but experiences a $-\theta$ shock in the event of a failed sale (i.e. the outside option is then rewritten as $\underline{u} = \widehat{P} - \theta$).

We also note that the model implicitly specifies conditions on the relationship between $\underline{u}$ and R. In the online appendix, we discuss this issue in detail. We show there that (i) assuming that R enters (or equals) the outside option (i.e., the consumption utility of households in the event of no sale) generates implausible predictions that we can reject in the data, (ii) if R is used by the seller to "rationally" forecast $\widehat{P}$ (given our normalization of $\underline{u} = \widehat{P}$), the result is innocuous, and doesn't affect any inferences from the model, and (iii) it is potentially possible to reinterpret the model as one of non-rational belief formation (i.e., the seller might view R as the "correct" outside option value), but it is potentially more difficult to rationalize several of the patterns we find in the data (i.e., bunching at just positive gains) with such a model of beliefs.

We next discuss selected predictions of the model to build intuition, and to guide our

[13] We capture listing behavior by studying the listing premium $\ell = L - \widehat{P}$, which is an innocuous normalization of the listing price L. One way to see this is to note that the regression $\underbrace{L - \widehat{P}}_{\ell} = \rho \underbrace{(\widehat{P} - R)}_{\widehat{G}}$ is equivalent to $L = (1 + \rho)\widehat{P} - \rho R$. We estimate a version of this regression in the online appendix and verify the original inferences of Genesove and Mayer (2001) using our sample.

choice of key moments of the data with which to structurally estimate key parameters.

2.2.2 Optimal Listing Premia

To begin with, consider only the intensive margin decision of the optimal choice of listing premium, and assume that $U\left(P(\ell), \cdot\right) = u(P(\ell), \cdot)$:

$$\max_{\ell} \left[\alpha(\ell)\left(u(P(\ell), \cdot) + \theta\right) + (1 - \alpha(\ell))\underline{u}\right] \tag{3}$$

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

To aid interpretation, we analytically solve a version of the simple model in equation (3), under the assumption that demand functions $\alpha(\ell) = \alpha_0 - \alpha_1 \ell$ and $\beta(\ell) = \beta_0 + \beta_1 \ell$ are linear in ℓ (this is an assumption that we later relax to account for concave demand). In this case, the model yields an optimal listing premium schedule which is piecewise linear:

$$\ell^*(\widehat{G}) = \begin{cases} \frac{1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\eta}\right) - \frac{1}{2\beta_1}\frac{\eta}{1+\eta}\widehat{G}, & \text{if } \widehat{G} \geq \widehat{G}_0 \\ -\frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\widehat{G}, & \text{if } \widehat{G} \in (\widehat{G}_1, \widehat{G}_0) \\ \frac{1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\lambda\eta}\right) - \frac{1}{2\beta_1}\frac{\lambda\eta}{1+\lambda\eta}\widehat{G}, & \text{if } \widehat{G} \leq \widehat{G}_1, \end{cases} \tag{4}$$

where $\widehat{G}_0$ and $\widehat{G}_1$ are levels of potential gains determined by underlying model parameters.[14]

Figure 1 illustrates how equation (4) varies with the underlying parameters characterizing preferences.

In the case of no reference dependence ($\eta = 0$), utility derives purely from the terminal house price. In this case, the top left-hand plot shows that ℓ^* is unaffected by the reference

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

price R.

In the case of linear reference dependence ($\eta > 0$, $\lambda = 1$), there is a negatively-sloped linear relationship between ℓ^* and $\widehat{G}$. In this case, R does not affect the marginal benefit of raising ℓ^*, but it does affect the marginal cost, as it affects 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 (i.e., when R is sufficiently low), the utility from a successful sale rises. The resulting ℓ^* will therefore be lower, as the household seeks to increase the probability that the sale goes through. The opposite is true when the household faces a loss in the event of a completed sale (i.e., when R is sufficiently high), which consequently results in a higher ℓ^*.[15]

In the case of (reference dependence plus) loss aversion ($\eta > 0$, $\lambda > 1$), the kink in the piecewise linear utility function leads to a more complex piecewise linear pattern in ℓ^*, which determines the gains that sellers ultimately realize. There is a unique level of potential gains, $\widehat{G}_0$, which maps to a realized gain of *exactly* zero (recall that $G(\ell^*) = \widehat{G} + \beta(\ell^*)$). Sellers with potential gains below $\widehat{G}_0$ want to avoid realizing a loss, meaning that they adjust ℓ^* upwards. However, this upward adjustment increases the probability of a failed sale. Beyond some lower limit $\widehat{G}_1$, the costs in terms of the failure probability become unacceptably high relative to the benefit of avoiding a loss, and it becomes sub-optimal to aim for a realized gain of zero. The seller has no choice but to accept the loss at levels of $\widehat{G} < \widehat{G}_1$, but still sets a marginally higher listing premium for each unit loss beyond this point.

[15] As mentioned earlier, it is important 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.

2.2.3 Bunching Around Realized Gains of Zero

The model reveals that household listing behavior also has material implications for quantities. Loss-averse preferences show up in non-linearities in the schedule of ℓ^* along the $\widehat{G}$ dimension, as well as on the likelihood of transaction completion, and the final price at which these transactions occur. This shows up as shifts in mass in the distribution of completed transactions along the G dimension, additional moments which allow us to pin down underlying utility parameters. In the simple version of the model (assuming linear demand) discussed above, the equation relating potential gains $\widehat{G}$ with final realized gains G is:

$$G(\widehat{G}) = \begin{cases} \beta_0 + \frac{\beta_1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\eta}\right) + \left(1 - \frac{1}{2}\frac{\eta}{1+\eta}\right)\widehat{G} & \text{if } \widehat{G} > \widehat{G}_0, \\ 0 & \text{if } \widehat{G} \in [\widehat{G}_1, \widehat{G}_0], \\ \beta_0 + \frac{\beta_1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\lambda\eta}\right) + \left(1 - \frac{1}{2}\frac{\lambda\eta}{1+\lambda\eta}\right)\widehat{G} & \text{if } \widehat{G} < \widehat{G}_1. \end{cases} \tag{5}$$

The two bottom panels of Figure 1 illustrate how this relationship varies with underlying utility parameters.

When $\eta = 0$, sellers choose a constant listing premium ℓ^*, which results in a constant realized premium $\beta(\ell^*)$ of actual gains G over potential gains $\widehat{G}$ (bottom left plot), meaning that the distribution of G is a simple parallel shift of the distribution of $\widehat{G}$ (bottom right plot, the black dotted line becomes the purple line).

In the linear reference dependence model ($\eta > 0, \lambda = 1$), sellers with $\widehat{G} < 0$ choose relatively higher ℓ^*. This lowers the likelihood that willing buyers will be found, meaning that the likelihood of observing transactions in this domain of $\widehat{G}$ is lower. However, if these transactions *do* go through, the associated G will be higher, shifting mass in the final sales distribution towards $G > 0$ (bottom right plot, the black dotted line becomes the green line).

The effect mentioned above is especially pronounced if sellers are loss averse, i.e., when $\lambda > 1$, in which case the model predicts bunching $(F(\widehat{G_0}) - F(\widehat{G_1}))$ in the final distribution of house sales precisely at $G = 0$ (bottom left plot, black line and bottom right plot black solid line), and greater mass in the distribution when $G > 0$, coming from even less mass when $G < 0$ (bottom right plot, the black dotted line becomes the black solid line).

In the discussion thus far, to build intuition about the effect of the underlying parameters characterizing preferences, we focused on the intensive margin, made several assumptions about the shape of demand, and assumed away other frictions and constraints. We next outline the predictions of the model in the broader case when we consider the extensive margin decision, and then turn to discussing two important potential confounds, namely, concave demand, and the effect of financial constraints.

2.2.4 Extensive Margin

In the discussion thus far, we ignored the seller's decision of whether or not to list. In the model, any force inducing a wedge between the expected utility from a successful listing and the outside option $\underline{u}$ affects decisions along the intensive margin, but can also push the seller towards deciding that listing is sub-optimal. In particular, the model predicts that sellers with lower $\widehat{G}$ are less likely to list. This clear prediction allows us to exploit the relationship of the listing propensity and $\widehat{G}$ as an additional moment to inform structural estimation of underlying preference parameters.[16]

Another important observation here here is that modeling the extensive margin decision is also important to 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

[16] Bunching in the distribution of realized house sales captures *ex post*-negotiation outcomes, and extensive margin decisions capture sellers' *ex ante* listing behaviour, i.e., these two moments are informative about different phases in the listing/selling decision.

list are more conservative (i.e., they set lower listing premia), and those who decide to list are more aggressive (i.e., setting 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.[17]

The moving shock θ (which alters the distance between the outside option and the utility from a successful listing) is a key model component that helps to capture such selection effects. Conditional on the moving shock, the listing decision is a simple binary choice. This means that accounting for the distribution of shocks, as we do in the model, allows us to capture the variation in listing decisions and to calculate average listing premia in the population. These average listing premia incorporate the endogenous first-stage selection effects and can be mapped directly back to the data.

There are more subtle implications of the model linking the extensive and the intensive margins. High realizations of θ affect the listing decision, and push the seller towards setting higher listing premia. However, this force can move ℓ into regions of concave demand (which we discuss in detail in the next subsection) in which the response of buyers is more (or less) pronounced, because of nonlinearities in $\alpha(\ell)$. This in turn means that θ variation can affect the observed magnitude of the seller's responses to $\widehat{G}$, smoothing and blurring the kinks in the model-implied ℓ^* profile. The online appendix illustrates this with a specific example, showing that the characteristic "hockey stick" shape of the average listing premium profile can result from averaging the three-piece-linear form of the listing premium profile in the case of $\lambda > 1$ across the distribution of θ.

2.3 Concave Demand

The demand functions $\alpha(\ell)$ and $\beta(\ell)$ are a critical determinant of listing behavior and the expected shape of ℓ^* in this model. This can be seen even in the simple case of the linear

[17] We thank Jeremy Stein for useful discussions on this issue.

demand functions posited earlier. Equation (4) shows that when the probability of sale is less sensitive to ℓ (i.e., when α_1 is lower), the marginal cost of choosing a larger listing premium is lower, and therefore the optimally chosen ℓ^* is higher. This intuition carries over to a case in which $\alpha(\ell)$ has the concave shape first identified by Guren (2018), and has important implications for the relationship between ℓ^* and $\widehat{G}$. Figure 2 graphically illustrates this mechanism, positing a concave shape for $\alpha(\ell)$ and considering the effect of varying $\alpha(\ell)$ around $\underline{\ell} = 0$, i.e., the point at which $L = \widehat{P}$ (solid and dashed red lines, right-hand plot).

The left-hand plot in Figure 2 documents the relationship between the optimal listing premium ℓ^* and $\widehat{G}$ in the presence of concave demand. When $\widehat{G} > 0$, the seller's incentive is to set ℓ^* low, since they are motivated to successfully complete a sale and capture gains from trade θ. However, in the presence of concave demand (i.e., as illustrated in the right-hand plot, horizontal $\alpha(\ell)$ when $\ell < \underline{\ell}$; combined with $P(\ell) = \beta_0 + \beta_1\ell$), lowering ℓ below $\underline{\ell}$ does not boost the sale probability $\alpha(\ell)$, but doing so does negatively impact the realized sale price $P(\ell)$. It is thus optimal for ℓ^* to "flatten out" at the level $\underline{\ell}$.

The tradeoff faced by sellers facing losses $\widehat{G} < 0$ is different—raising ℓ^* helps to offset expected losses, but lowers the probability of a successful sale. When demand concavity increases, i.e., $\alpha(\ell)$ is more steeply negative, the probability of a successful sale falls at a faster rate with increases in ℓ. Figure 6 illustrates this force—moving from the dashed $\alpha(\ell)$ schedule to the solid $\alpha(\ell)$ schedule in the right-hand plot in turn leads to dampening of the slope of ℓ^* in the left-hand plot. 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: First, the optimal ℓ^* in a linear reference-dependent model ($\eta > 0$, $\lambda = 1$) in the presence of concave demand exhibits a flatter slope in the domain $\widehat{G} > 0$ relative to the case of linear demand. This means that the graph of ℓ^* against $\widehat{G}$ can exhibit a characteristic "hockey stick" shape of the type detected by Genesove and Mayer (2001) even if there is no loss aversion, i.e., $\lambda = 1$. Second, the model predicts a tight link between the shape of $\alpha(\ell)$ and the slope of ℓ^*. We later use this insight to exploit cross-sectional variation in the concavity of demand across different segments of the Danish market to aid structural parameter identification.[18] Third, while we have focused our discussion on how concave demand can generate a non-linear listing premium profile, it will also result in effects on transactions volume. That is, concave demand can result in additional shifts of mass towards positive values of realized gains, depending on the level of $\underline{\ell}$, though it will not be associated with sharp bunching of the type associated with loss aversion.

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, potentially 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 turns out to be a reasonable assumption—we observe significant evidence in the data of bunching using this assumption about R, confirming its relevance to sellers.

2.4 Down-payment Constraints

A well-known confound for the estimation of preference parameters from listing premia (see, e.g., Genesove and Mayer (1997, 2001)) is the effect of down-payment constraints,

[18]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.

which we account for in the model through the function $\kappa(P(\ell), \cdot)$ (recall that $U\left(P(\ell), \cdot\right) = u(P(\ell), \cdot) - \kappa(P(\ell), \cdot)$). 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 down payment 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 is for households to downsize to a less expensive home than $P(\ell)$, or indeed, to move to the rental market—either decision 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); this is typically expensive.[19] A third (usually unobservable) possibility is that households can bring additional funds to the table by liquidating other assets.[20] We therefore assume that violating the down-payment constraint does not lead the seller to withdraw the sale offer, assuming instead that the seller incurs a monetary penalty of μ per unit of realized home equity below the

[19]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.

[20]In Stein (1995), M represents the outstanding mortgage debt net of any liquid assets that the household can put towards the down payment. The granular data that we 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 taking these additional funds into account.

constraint threshold:[21]

$$\kappa(P(\ell)) = \begin{cases} \mu(\gamma - H(\ell)), \text{ if } H(\ell) < \gamma \\ 0, \text{ if } H(\ell) \geq \gamma \end{cases}. \tag{6}$$

We turn next to describing the data and key estimated moments as a precursor to more rigorous structural estimation of the underlying parameters of the model.

3 Data

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 each sold or listed property, rich hedonic characteristics of each 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. Furthermore, we merge the data on the entire housing stock captured 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. This allows us to assess the fraction of the total housing stock that is listed, and to condition observed listing propensities on functions of the predicted sales price, such as the prospective seller's potential gains relative to the original purchase price, or the prospective seller's potential level of home equity in the property.

Our data link administrative datasets 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 data sources, con-

[21] i.e.,

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

struction, filters, and the process of matching involved in assembling the dataset.

3.1 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.[22,23]

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.

3.2 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% of all sale transactions have associated listing data.[24] For each property

[22]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.

[23]Tax-assessed property values are used for determining tax payments on property value. The appendix describes the property taxation regime in Denmark in greater detail including inheritance taxation; we simply note here that there is the usual "principal private residence" exemption on capital gains on real estate, and that property taxation does not have important effects on our inferences.

[24]We more closely investigate the roughly 25% of transactions that do not have an associated electronic listing. 10% of these transactions can be explained by the different (more imprecise) recording of addresses

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.

3.3 Mortgage Data

To establish the predicted/potential 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.[25]

3.4 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

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.

[25]The online appendix provides a detailed description of several features of the Danish mortgage market including the conditions under which mortgages are assumable, as well as the effects of the Danish refinancing system (studied in greater detail in Andersen et al. (2020)) on sale and purchase incentives. These features do not materially impact our inferences.

can aggregate individual data on wealth and income to the household level.[26] 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.

3.5 Final Merged Data

We only keep transactions for which we can measure both nominal losses and home equity, and since the mortgage data run from 2009 to 2016, this imposes the first restriction on the sample. The sample is further 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. Transactions data are available from 1992 to the present, meaning that we can only measure the purchase price of properties that were bought during or after 1992.[27] We exclude foreclosures (both sold and unsold),[28] properties with a registered size of 0, and properties that are sold at prices which are unusually high or low (below 100,000 DKK and above 20MM DKK in 2015, accounting for roughly 0.05% of the total housing stock in Denmark).[29] For listings that end in a final sale, we also drop within-family transactions, transactions that Statistics Denmark flag as anomalous or unusual, and transactions where the buyer is the government, a company, or an organization.[30]

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

[27]In Appendix Table A.2 and Appendix Figure A.39 we further examine properties traded before 1992. Since these properties have no known purchase price, we match them to otherwise similar properties for which we know the purchase price, using two approaches that we describe in the online appendix, with a reasonable success rate. Figure A.39 shows that the main relationships that we find in the main dataset essentially hold in the matched sample using this approach.

[28]The online appendix describes the Danish foreclosure process in detail.

[29]We apply this filter to reduce error in our empirical work, because the market for such unusually priced properties is extremely thin, meaning that predicting the price using a hedonic or other model is particularly difficult.

[30]We apply this filter, as company or government transactions in residential real estate are often conducted at non-market prices—for tax efficiency or evasion purposes in the case of corporations, and

We also restrict our analysis to residential households, in our main analysis dropping summerhouses and listings from households that own more than three properties in total, as they are more likely property investors than owner-occupiers.[31]

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 $214,508$ listings of Danish owner-occupied housing in the period between 2009 and 2016, for both sold (70.4%) and retracted (29.6%) properties, matched to mortgages and other household financial and demographic information.[32] These listings correspond to a total of $191,843$ unique households, and $179,262$ unique properties. Most households that we observe in the data sell one property during the sample period, but roughly 9% of households sell two properties over the sample period, and roughly 1.5% of households sell three or more properties. In addition, we use the entire housing stock, filtered in the same manner as the listing data, comprising $5,540,376$ observations of $807,666$ unique properties to understand sellers' extensive margin decision of whether or not to list the properties for sale.

3.6 Hedonic Pricing Model

To calculate potential gains $\widehat{G}$ (and potential home equity $\widehat{H}$), we require a measure of the expected sale price $\widehat{P}$ for each property-year in the data. To arrive at this measure, we estimate a standard hedonic pricing model on our sample of sold listings and use it to predict prices for the full sample of listed properties, including those that are not sold.[33]

for eminent domain reasons in the case of government purchases, for example.

[31] 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.

[32] The data comprises $173,065$ listings that have a mortgage, and $41,443$ listings with no associated mortgage (i.e., owned entirely by the seller)—we later utilize these subsamples for various important checks.

[33] 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 on a larger data set, including unlisted properties, in order to conduct these additional tests.

The hedonic model predicts the log of the sale price P_{it} of all sold properties i in each year t:

$$\ln(P_{it}) = \xi_{tm} + \beta_{ft}\mathbb{1}_{i=f}\mathbb{1}_{t=\tau} + \boldsymbol{\beta}\mathbf{X_{it}} \\ + \boldsymbol{\beta}_{\mathbf{fx}}\mathbb{1}_{i=f}\mathbf{X_{it}} + \Phi(v_{it}) + \mathbb{1}_{i=f}\Phi(v_{it}) + \varepsilon_{it}, \tag{7}$$

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, ln(the age of the building), dummy variables for whether the property is located in a rural area, or has been marked as historic, and ln(distance of the property to the nearest major city). (Most property characteristics in $\mathbf{X_{it}}$ are time-varying, which contributes to the accuracy of the model). ξ_{tm} are year cross municipality fixed effects (there are 98 municipalities in Denmark), 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.[34] $\Phi(v_{it})$ is a third-order polynomial of the previous-year tax assessor valuation of the property.[35] We interact the apartment dummy with time 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 0.88 in the full sample.[36] The

[34]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.

[35]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-evaluated 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 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.

[36]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 0.89. Moreover, the R^2 when we eliminate the tax assessor valuation from the hedonic characteristics is 0.77. 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.

large sample size allows us to include many fixed effects in the model, helping to deliver a better fit. 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 later describe.

4 First Inferences about Model Parameters

In this section, we document patterns in listing premia and sales transactions volumes in the data in relation to measured G and $\widehat{G}$, and informally discuss how these patterns relate to the predictions of the model, especially regarding the primary parameters of interest η and λ. We also explore how the patterns in the data and possible inferences about underlying parameters vary when we account for three important factors. These are: (i) sellers' down-payment constraints, (ii) concave demand, and (iii) robustness to changes in measurement. Before turning to structural estimation that takes the model's predictions to the data more rigorously in the next section, we discuss the robustness of the patterns seen in the data to various estimation approaches and controls.

4.1 Listing Premia in the Data

Armed with the hedonic pricing model, we estimate listing premia in the data as $\ell = \ln L - \widehat{\ln P}$, where L is the reported initial listing price observed in the data.[37] Mean (median) ℓ is 12.7% (11.3%), and $\ell > 0$ (< 0) for 75% (25%) of the sample. We also estimate potential gains $\widehat{G} = \widehat{\ln P} - \ln R$, where R is set to the nominal purchase price of the property. Mean (median) $\widehat{G}$ estimated in this way is 36% (28%), and 23% (77%) of

[37] 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 robust to using the alternative approach of Genesove and Mayer (2001), and discuss identification and measurement concerns in greater detail below as well.

property-years have $\widehat{G} < 0$ ($\widehat{G} > 0$). The online appendix plots the distributions of these and other variables.

In Figure 3 we plot the average observed listing premium (on the vertical axis) for each percentage bin of potential gains (on the horizontal axis). Sellers who hold properties that have appreciated (declined in value) since the initial purchase choose lower (higher) listing premia. Importantly, this negative relationship is visible not only in the potential loss domain (i.e., $\widehat{G} < 0$), but also across different values in the potential gain domain (i.e., $\widehat{G} > 0$). This is consistent with the predictions of a model with reference dependence $\eta > 0$. Moreover, as we move from the gain to the loss domain, the slope becomes much more pronounced, i.e., sellers react much more aggressively to every unit decrease in potential returns when $\widehat{G} < 0$. For potential gains in the neighbourhood of zero, this "hockey stick" pattern is consistent with the predictions of a model with loss aversion $\lambda > 1$. However, in the piecewise linear formulation that we consider, loss aversion also predicts a flattening out of the listing premium profile deeper into the loss domain, which is not visible in the plot.

While these patterns provide prima facie evidence of the underlying parameters of the seller's utility, we must be wary of such inferences given the influence of three important confounding factors discussed above, namely: (i) concave demand, (ii) the extensive margin, which smooths out the locations of kinks, and can lead to selection effects, and (iii) sellers' financial/down-payment constraints. Keeping these issues in mind, we next discuss additional evidence available from the analysis of transactions volumes.

4.2 Bunching of Realized Sales

Figure 4 plots the distribution of property sales across the dimension of realized gains ($\ln P - \ln R$)—each dot shows the empirical frequency of sales (y-axis) occurring in each 1 percentage point bin of realized gains (x-axis). We overlay on this plot (as a dotted line) the empirical frequency of realized sales (i.e., the same y-axis) occurring in each 1 percentage point of *potential* gains $\widehat{\ln P} - \ln R$ (i.e., a different x-axis). Observing the counterfactual is difficult in most settings which attempt to estimate loss aversion using bunching estimators (e.g., Rees-Jones 2018 cleverly extracts evidence of loss aversion from U.S. tax returns data, where it is difficult to measure "expected tax avoidance costs and benefits"). The distribution of sales with respect to pre-listing potential gains can serve as one possible counterfactual, as we describe in greater detail below.[38]

Figure 4 shows significant bunching of transactions in the positive domain of realized gains G, with a sharp "spike" around $G = 0$, and with significant mass extending further into the domain $G > 0$.[39] While the spike is clearly evident, more information can be extracted about model parameters from the broader distribution of sales across realized gains, especially when we compare it to the distribution of sales across potential gains $\widehat{G}$. This is because in the model when $\eta > 0$, as mentioned earlier, the mapping between $\widehat{G}$ and G occurs through the choice of ℓ^*, and the associated probability of sale. This mapping results in mass in the final sales distribution shifting towards sales with realized $G > 0$. In contrast, when $\eta = 0$, the model predicts that the distribution of G is simply a constant linear transformation of the distribution of $\widehat{G}$. The precise position of the pronounced jump in the distribution at $G = 0\%$, and the distribution of mass to the left and right of this point relative to the counterfactual are also informative about λ. When $\lambda > 1$, the model predicts a jump in the final distribution of house sales precisely at $G = 0$, additional mass in this distribution just to the right of this point, and relatively lower mass in the loss domain, to the left of $G = 0$. The pronounced bunching that we observe precisely at the point $G = 0$ also offers empirical support (which is essentially

[38]We also use alternative approaches to measure this counterfactual density, following Chetty et al. (2011) and Kleven (2016), and fitting a flexible polynomial to the empirical frequency distribution. When doing so, we exclude bins near the threshold, and extrapolate the fitted distribution to the threshold, excluding one bin on each side of the zero gain bin, i.e. $j \in \{-1\%, 1\%\}$, with a polynomial order of 7. The results, reported in the Online Appendix, are robust to other polynomial orders and to variations of the excluded range, and generate similar (but less cleanly estimated) results on the excess bunching mass.

[39]The plot also reveals a small but visible "hole" just to the left of $G = 0$, that may be evidence of a notch in preferences—an important additional feature of the data that we are currently investigating.

non-parametric, since it does not require reliance on a hedonic or other model) for the choice of R as the nominal purchase price (see Kleven, 2016, for a discussion of bunching at reference points).

4.3 Extensive Margin: Probability of Listing

As discussed earlier, understanding the seller's decision of whether or not to list is important for at least two reasons. First, the model makes predictions about this decision, in addition to predicting patterns of listing premia and transactions volumes. Second, accounting for this decision helps to correct for possible selection effects that may drive patterns of observed intensive margin listing premia in the data. This is 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.

To understand the decision to list, we turn to data on the total housing stock in Denmark, corresponding to $12,565,190$ property-years in the data, once merged with the listings data. We compute the unconditional average annual listing propensity, which is 3.75% of the housing stock (corresponding to between 2% and 4% of the housing stock listed across sample years).[40] Figure 8 plots the listing propensity at each level of $\widehat{G}$, which comes from estimating $\widehat{\ln P}$ for all properties in Denmark for which we have data on the nominal purchase price R. The figure shows a mild, but visible increase in the probability of listing as $\widehat{G}$ increases, which is evident when $\widehat{G} > 0$, but more pronounced when $\widehat{G} < 0$. This pattern is once again apparently consistent with levels of $\eta > 0$ and $\lambda > 1$.

[40]We do not attempt to use the model to explain the average propensity to list, as this exercise is beyond the scope of this paper. It would require us to take a strong stance on the factors that drive the moving decision, which we currently summarize using our estimates of θ.

4.4 Confounding Factors

4.4.1 Down-payment Constraints and Home Equity

To account for the role of down-payment constraints, for each observation in the data, we calculate the seller's potential home equity level $\widehat{H} = \widehat{\ln P} - \ln M$, where $\widehat{\ln P}$ is estimated using our hedonic model as before, and M is the outstanding mortgage balance reported by the household's mortgage bank each year.[41] Mean (median) $\widehat{H}$ is 27% (25%), and 77% (23%) of property-years have $\widehat{H} < 0$ ($\widehat{H} \geq 0$). Modal $\widehat{H}$ is around 22%, 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 is restricted to be 80% or lower.[42] Clearly, $\widehat{G}$ and $\widehat{H}$ are jointly dependent on $\widehat{\ln P}$, but there are multiple other factors that influence this correlation, including the LTV ratio at origination (i.e., variation in initial down payments), and households' post-initial-issuance remortgaging decisions. In the online appendix, we plot the joint distribution of $\widehat{G}$ and $\widehat{H}$, and show that there is substantial variation in the four regions defined by $\widehat{G} \lesseqgtr 0$ and $\widehat{H} \lesseqgtr 0$, which permits identification of their independent impacts on listing decisions.[43]

To assess the extent to which any variation in ℓ attributed to $\widehat{G}$ might be confounded

[41] The online appendix plots the distributions of $\widehat{G}$ and $\widehat{H}$ in the data. Both $\widehat{G}$ and $\widehat{H}$ are winsorized at the 1 percentile point; $\widehat{G}$ is also winsorized at the 99 percentile point. We winsorize $\widehat{G}$ because of several large values of given the substantial time elapsed since the purchase of some properties in the data. We set $\widehat{H}$ to 100% in cases in which households have substantial home equity ($\succeq$ 60%), meaning that we consider households to be essentially unconstrained at high levels of home equity. This is necessary to avoid $\widehat{H}$ levels greater than 1, given the log difference approach that we use to compute it. These filters make no material difference to our results—we confirm that our structural estimates are unaffected by these choices.

[42] This constraint does not change over our sample period, though it must be noted that as mentioned earlier, households can engage in non-mortgage borrowing to effectively increase their LTV, but at substantially higher rates. The online appendix documents the changes in the Danish Mortgage Act over the 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.

[43] The online appendix also contains a fuller discussion of additional evidence that we uncover which is consistent with households exhibiting aversion to downsizing. We are able to link sale transactions with future purchase transactions for a subset of households, and show that the future purchase is almost always of higher value than the sale.

by simultaneous variation in $\widehat{H}$, the top left plot in Figure 5 shows a 3-D representation of ℓ against both $\widehat{G}$ and $\widehat{H}$ in the data, averaged in bins of 3 percentage points. The plot reveals that ℓ declines in both $\widehat{G}$ and $\widehat{H}$, consistent with the patterns previously identified in the literature. Unusually, given the large administrative dataset that we have access to, the plot captures the variation ℓ along both dimensions simultaneously, and clearly reveals both *independent* and *interactive* variation along both dimensions. To better see the independent variation, the dotted lines on the 3-D surface indicate two cross-sections in the data ($G = 0\%$ and $H = 20\%$), which we also use later for structural estimation. Clearly, the "hockey stick" profile of ℓ along the $\widehat{G}$ dimension survives, controlling for $\widehat{H}$, and there is also a pronounced downward slope in ℓ along the $\widehat{H}$ dimension, controlling for $\widehat{G}$. In terms of the interactive variation, Panel B of Figure 9 plots how the "marginals" of the listing premium vary as we vary the control variable in each case (i.e., $\widehat{H}$ in the left plot and $\widehat{G}$ in the right plot); we discuss these in more detail towards the end of the paper, where we also evaluate the extent to which we can match these relationships using the model.[44]

4.4.2 Concave Demand

Using the underlying data on the time-on-the-market (TOM) that elapses between sale and listing dates, the left plot in Figure 6 calculates the probability of a house sale within six months (this maps to $\alpha(\ell)$ in the model), which we plot on the y-axis, as a function of ℓ on the x-axis.[45] To smooth the average point estimate at each level of ℓ, we use a simple nonlinear function which is well-suited to capturing the shape of $\alpha(\ell)$, namely, the generalized logistic function or GLF (Richards, 1959, Zwietering et al., 1990, Mead,

[44]The online appendix reports sale transaction frequencies (to show the degree of bunching) in a similar 3-D fashion. We confirm that regardless of the level of $\widehat{H}$, there is a visible shift of mass from the $\widehat{G} < 0$ domain to the $\widehat{G} > 0$ domain.

[45]Mean (median) TOM in the data is 37 weeks (25 weeks). We pick six months in the computation of $\alpha(\ell)$ to match the median TOM observed in the sample. The online appendix shows the distribution of TOM, which is winsorized at 200 weeks, meaning that we view properties that spend roughly 4 years on the market as essentially retracted.

2017).[46] The solid line corresponds to this set of smoothed point estimates.

The right-hand plot in Figure 6 shows how $\ln P(\ell) - \widehat{\ln P}$, i.e., the "realized premium" of the final sales price over the hedonic value (which corresponds to the "markup" $\beta(\ell)$ in the model) varies with ℓ. The plot shows that $\beta(\ell)$ rises virtually one-for-one with ℓ when ℓ is low, but flattens out as ℓ rises. The solid line shows a simple polynomial fit of this relationship that we use in the model.

From the two plots, we can see that in Denmark low list prices appear to reduce seller revenue with little corresponding decline in time-on-the-market. This is virtually identical to the patterns detected by Guren (2018) in three U.S. markets, which he terms "demand concavity".[47]

This evidence of demand concavity serves as a confound for estimating λ, as described earlier. This is because 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 top panel of Figure 6 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 $\alpha(\ell)$ and the slope of ℓ when $\widehat{G} < 0$, which cannot be seen in Figure 6, which is estimated using the entire dataset. To estimate the impact of demand concavity on the shape of the listing premium "hockey stick," we therefore exploit regional variation across sub-markets of the Danish housing market.

To illustrate the predicted correlation between the shape of the listing premium

[46] We describe the GLF in more detail in the online appendix. It is useful for our purposes as it is (i) bounded both from above and below, and it (ii) allows us to easily capture the degree of concavity observed in the data in a convenient way, through a single parameter. In our estimation of the parameters, we restrict the lower bound of the GLF to be equal to zero, to impose that the probability of sale asymptotically converges to 0 for arbitrary high levels of ℓ.

[47] These plots also show that Danish sellers who set high ℓ suffer longer TOM, but ultimately achieve higher prices (i.e., high realized premia) on their house sales, confirming the original finding of Genesove and Mayer (2001), who analyze the Boston housing market between 1990 and 1997.

"hockey stick" and the degree of demand concavity (i.e., the shape of $\alpha(\ell)$) in the data, we separately estimate the slope of ℓ in the domain $\widehat{G} < 0$, as well as separate $\alpha(\ell)$ functions (in particular, the slope of $\alpha(\ell)$ when $\ell \geq 0$) in different local housing markets, namely, different municipalities of Denmark.[48]

The bottom panel of Figure 7 shows results when we sort municipalities by their estimated demand concavity (i.e., the slope of $\alpha(\ell)$ when $\ell \geq 0$). 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 -0.1 and -0.3. The left-hand plot in Figure 7 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 ℓ in the domain $\widehat{G} < 0$ (-0.1 to -0.4) than markets with weak demand concavity (-0.5 to -0.9).[49] Towards the end of the paper, we describe a validation analysis that we undertake to confirm the model-predicted mechanism in the data using instruments for demand concavity.

[48]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.

[49]For the purposes of our current investigation, we focus on the slope differentials, and to show these, Figure 7 normalizes sub-markets to have the same level of the listing premium. We also observe important differences between the *levels* of $\alpha(\ell)$ across these markets i.e., there are both "hot" and "cold" municipalities à la Ngai and Tenreyro (2014). Un-normalized plots in the online appendix reveal that the *level* of ℓ is lower when the level of $\alpha(\hat{\ell})$ is higher and vice versa; and consistent with Ngai and Tenreyro (2014), the levels of $\alpha(\ell)$ and $P(\ell)$ are strongly positively correlated across sub-markets.

4.5 Robustness

4.5.1 Time-series Variation

While it is reassuring that $\widehat{G}$ and $\widehat{H}$ exhibit independent variation in the data, 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 aggregate Danish house prices. To check this, in the online appendix we plot the shares of the data in each of four groups of properties defined by $\widehat{G} \lesseqgtr 0$ and $\widehat{H} \lesseqgtr 0$, in each of the years in our sample. We find that aggregate price variation does shift the relative shares in each group across years, with price rises increasing the fraction of unconstrained winners ($\widehat{G} > 0$ and $\widehat{H} > 0$) relative to losing and constrained groups. However, the relative shares of all four groups are substantial and fairly stable over the sample period, alleviating concerns that different groups simply come from different time periods, i.e., the plots is reassuring that 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.

4.5.2 Bunching: Round Numbers and Holding Periods

In the online appendix, we verify that the bunching patterns documented earlier are robust to commonly expressed concerns in this literature (e.g., Kleven 2016, Rees-Jones 2018). We find that the spike in sales volumes at $G = 0$ and the patterns of excess mass relative to the counterfactual do not appear to be driven by bunching at round numbers, as they remain striking and visible when we exclude sales at prices ending in multiples of $10,000$, $50,000$, $100,000$, and $500,000$ DKK, which (cumulatively) affect roughly 20%, 17%, 5%, and 2% of all observations, respectively. We also show that these bunching patterns are robust when we split the sample into five groups (< 3, $3-6$, $6-9$, $9-12$, > 12 years) based on the time between sale and purchase, i.e., the holding period of the property. Except for the sub-sample with the longest holding period ($>$ 12 years, 20% of the data), we find strong evidence of bunching. Finally, we also find strong evidence of bunching in all cases when we split the sample into quintiles based on the level of R, with quintile cutoffs ranging from around $658,000$ DKK to 1.9MM DKK. Together, these checks assuage concerns that bunching could result from differences in the underlying characteristics of properties—for instance these tests suggest that it is implausible that bunching results from a combination of small properties with shorter holding periods clustering around $G = 0$, and larger properties with longer holding periods showing up at values of $G > 0$.

4.5.3 Unobserved Quality

An important and often-repeated concern in the literature is that the relationships that we observe between ℓ and $\widehat{G}$ (and indeed $\alpha(\ell)$ and ℓ), can be spuriously affected by measurement error in the underlying model for $\widehat{P}$. In particular, if properties with $\widehat{G} < 0$ are deemed to be such as a result of underestimated $\widehat{P}$, we would also see higher listing premia for such properties, resulting in the hockey-stick shape that we observe. Moreover, such an issue could also upwardly bias the true (decreasing) relationship between the probability of a quick sale and ℓ, especially when $\ell > 0$, as houses with mismeasured high listing premia would be expected to transact faster.

We assess the robustness of our results to these concerns in a number of ways, all of which we describe in detail in the online appendix. First, we show that the relationships between ℓ, $\widehat{G}$, and $\alpha(\ell)$ are robust to a battery of changes to the underlying model used to estimate $\widehat{P}$. We do so in several ways. We employ a repeat sales model to difference out time-invariant unobserved property quality; we instrument variation in $\widehat{P}$ using regional house price indices; we control for demographics, financial wealth, and further interactions in the hedonic model using granular data that have previously been unutilized in this manner, and which are potentially informative about the seller's response to earlier under-or over-payment on the property; we use the external tax assessor value of the property instead of our estimated hedonic model; and finally, we verify that our inferences hold even when we use out-of-sample estimated hedonic coefficients.[50]

Second, we implement the bounding approach proposed in Genesove and Mayer (2001) to account for unobserved quality, and confirm that our inferences are robust to doing so.

Third, while the tests just described focus on showing robustness of the magnitudes of the nonlinear relationships observed in the data between ℓ, $\widehat{G}$, and $\alpha(\ell)$, we also document evidence in line with the key predictions from the model. That is, we are able to demonstrate that the observed nonlinearities are in fact discontinuous and sharp around the respective thresholds of $\widehat{G} < 0$ and $\ell > 0$, using a regression kink design (RKD) originally suggested by Card et al. 2015b and implemented e.g., by Landais, 2015, Nielsen et al. 2010, and Card et al. 2015a. In line with the identifying assumptions of this research design, we also show that property-and household-specific observable characteristics are smooth around the respective thresholds.

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 Figure 6 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.

[50]This last variation helps to assuage concerns of overfitting or mechanical correlation arising from our hedonic model being estimated using the sample of sold listings. The model fits relatively precisely out of sample, with R^2's ranging between 0.80 to 0.88 when predicting between 1% to 50% of the data out-of-sample, and the patterns in the relationships between ℓ, $\widehat{G}$, and $\alpha(\ell)$ are robust to using the oos coefficient estimates.

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 $\mathbf{M_m}(\mathbf{x})^{7\times 1}$ corresponding to the moments in the data $\mathbf{M_d}^{7\times 1}$ 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 $\kappa_{\widehat{G}\geq 0}$ the analogous slope for $\widehat{G} \geq 0$ ($\kappa_{\widehat{G}<0}$ and $\kappa_{\widehat{G}\geq 0}$ 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 $\frac{\kappa^+_{\widehat{G}\geq 0} - \kappa^-_{\widehat{G}\geq 0}}{\tilde{\delta}}$.

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 $[-5\%, +5\%]$.[51]

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:

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

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

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 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:

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

5.3 Parameter Estimates

Table 2 shows the estimated parameters and associated standard errors. The data favor a model of reference dependence with $\eta = 0.948$ with a degree of loss aversion $\lambda = 1.576$. 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 $\lambda = 1.59$).[52]

The parameter $\mu = 1.060$ best matches the average $\hat{\ell}$ slope with respect to $\widehat{H}$, i.e., there is an 106 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.[53] The relatively larger number

[52]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 $\lambda = 1.522$ (*s.e.* 0.479), $\mu = 1.158$ (*s.e.* 0.218), $\delta = -0.093$ (*s.e.* 0.0183), $\theta_{\min} = 0.235$ (*s.e.* 0.148), $\theta_{\max} = 1.052$ (*s.e.* 0.131) and $\varphi = 0.039$ (*s.e.* 0.025).

[53]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.

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.

We find that $\delta = -0.097$, which corresponds to a perceived relative reduction of the probability of sale of 9.7%, for a household listing at $\ell = 10\%$, and that the distribution $\theta\ \sim$Uniform$(\theta_{\min}, \theta_{\max})$ has parameters $\theta_{\min} = 0.217$ and $\theta_s = 1.005$. These "moving shocks" correspond to the present discounted value of future benefits from successfully selling and/or moving, and are on the order of 21.7% 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 3.7% of the hedonic value of the house.

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 the online appendix we propose a simple and less formal application of this idea, 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 Validation and Open Questions

6.1 Interactions

The top panel of Figure 9 compares the 3-dimensional patterns of optimal listing premia in the data (left-hand plot) and the model (right-hand plot). The model matches the pronounced increase in $\hat{\ell}$ for $G < 0$, and the similar increase in $\hat{\ell}$ when $\widehat{H}$ declines. A striking feature of this plot is that it seems to indicate that the position of any reference point is not uniquely determined by $\widehat{G}$ or $\widehat{H}$ alone. As we briefly mentioned earlier, there is 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.

The bottom panel of Figure 9 explores these interaction effects in more detail. We plot 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.[54] The left-hand plot in the bottom panel shows that there is a change in the slope of the ℓ - $\widehat{G}$ relationship as $\widehat{H}$ varies, and the right-hand plot, that there seems to be a change in the inflection point in the ℓ - $\widehat{H}$ relationship as $\widehat{G}$ varies. Note that the average level of ℓ 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 ℓ 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 ℓ 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

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

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 ℓ along the $\widehat{G}$ dimension. The right-hand plot in the bottom panel of the figure shows the ℓ - $\widehat{H}$ relationship, where again, the level differences reflect the ℓ - $\widehat{G}$ 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 ℓ *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 ℓ - $\widehat{G}$ 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 ℓ - $\widehat{H}$ 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 down payment 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 down payment data. For this limited subsample, we show a binned scatter plot of the ℓ on the subsequently sold listing against the realized down payment on the subsequent house, controlling for the level of $\widehat{H}$ on the subsequently sold listing. We find evidence that the down payment on the new house is correlated with ℓ, which, given our evidence of $\widehat{G}$ predicting ℓ, is consistent with the idea that households shifting their reference level of housing on the basis of expected gains.

6.2 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 7 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.[55] 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 ℓ with respect to $\widehat{G}$ and $\widehat{H}$, we also include specifications which control for the average age, education length, financial assets, and income of sellers in a given sub-market.

Figure 7 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

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

slope of $\alpha(\ell)$ to the right of $\ell = 0$). Table 3 reports the results of the more formal IV exercise. Column 1 shows the simple OLS relationship between the slope of ℓ for $\widehat{G} < 0$ on demand concavity slope (slope of $\alpha(\ell)$ for $\ell \geq 0$) across municipalities,[56] with a baseline level of -0.407. 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 -0.520. With both instruments (i.e., including the distance to the largest cities as well), the overidentified 2SLS specification gives a result of -0.504 without, and -0.346 with controls for average household characteristics in the municipality. The first-stage F-statistics are between 17 and 25, 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.[57] These results appear to validate the mechanism that we propose in the model.

7 Conclusion

We structurally estimate a new model of house listing decisions on comprehensive Danish housing market data, and acquire new estimates of key behavioral parameters and household constraints from this high-stakes household decision. Underlying preferences seem well characterized by reference dependent around the nominal purchase price plus modest loss aversion, and there is also evidence of the important role of down-payment constraints on household behavior.

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-

[56]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.

[57]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.

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. Home equity constraints also appear to loom larger for households facing nominal losses. However, for households facing nominal gains, there is 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. From 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, Schulhofer-Wohl, 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 help to 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, T. Olsen, and L. Pistaferri (2011): "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records," *The Quarterly Journal of Economics*, 126(2), 749–804.

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.

Engelhardt, G. V. (2003): "Nominal loss aversion, housing equity constraints, and household mobility: evidence from the United States," *Journal of urban Economics*, 53(1), 171–195.

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.

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

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.

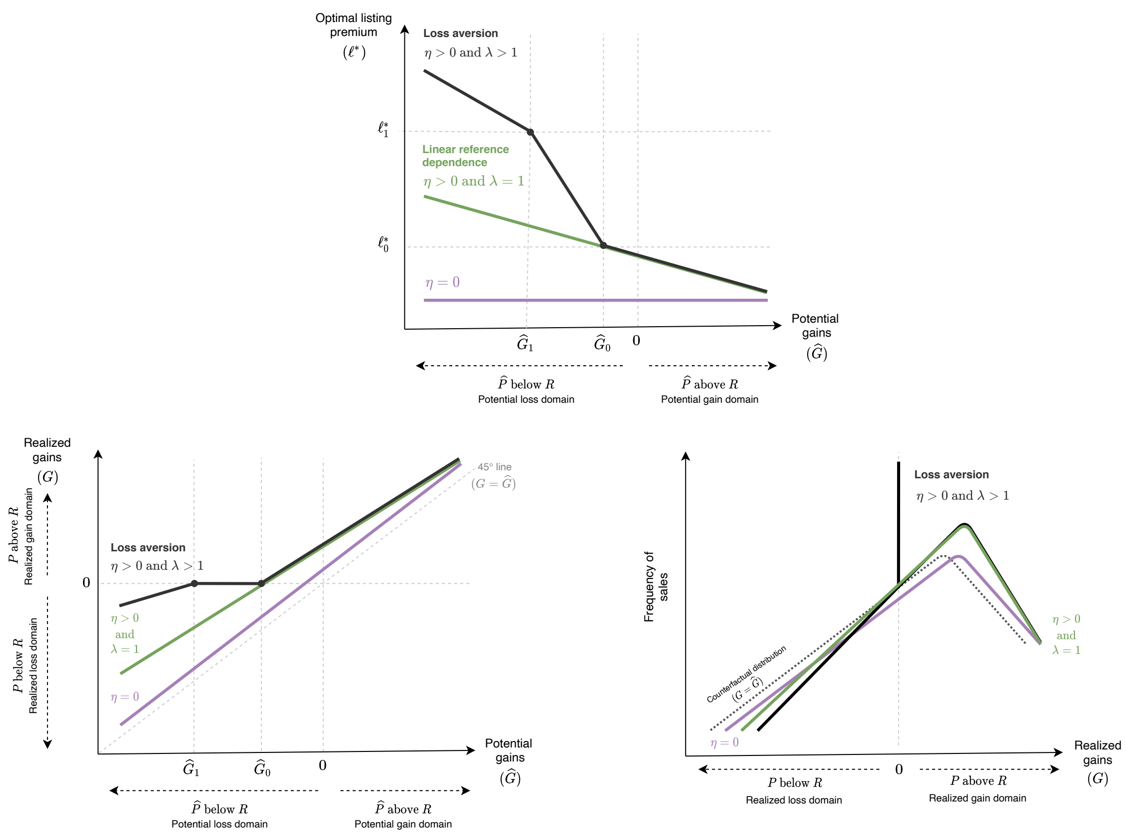


Figure 2

Concave demand

This figure illustrates the link between concave demand and the choice of optimal listing premia. 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. A steeper slope of demand translates into a general flattening out of the listing premium profile.

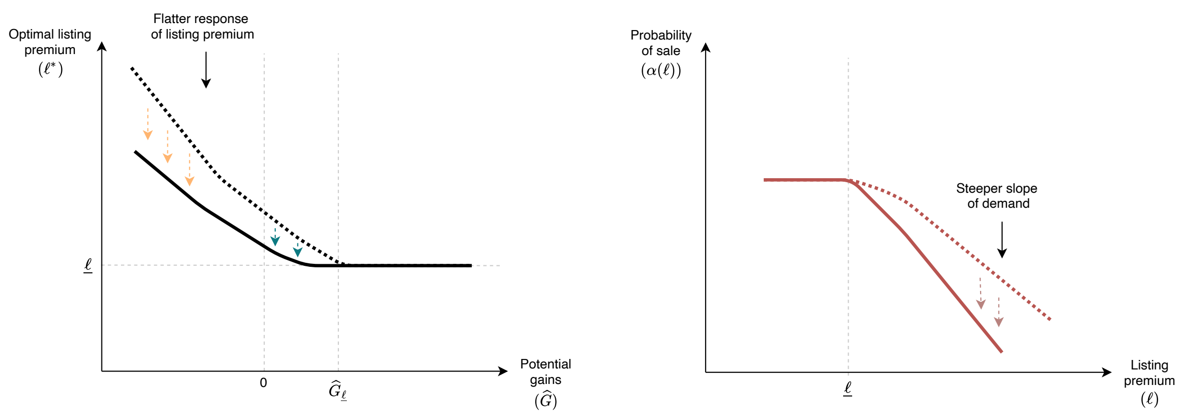


Figure 3

Listing premia and potential gains

The figure reports binned average values (in 1 percentage point steps) for the listing premium (ℓ) for different levels of potential gains ($\widehat{G}$). The green line corresponds to a polynomial of order three, fitted in the positive domain of potential gains. The red line corresponds to an equivalent polynomial fit in the potential loss domain.

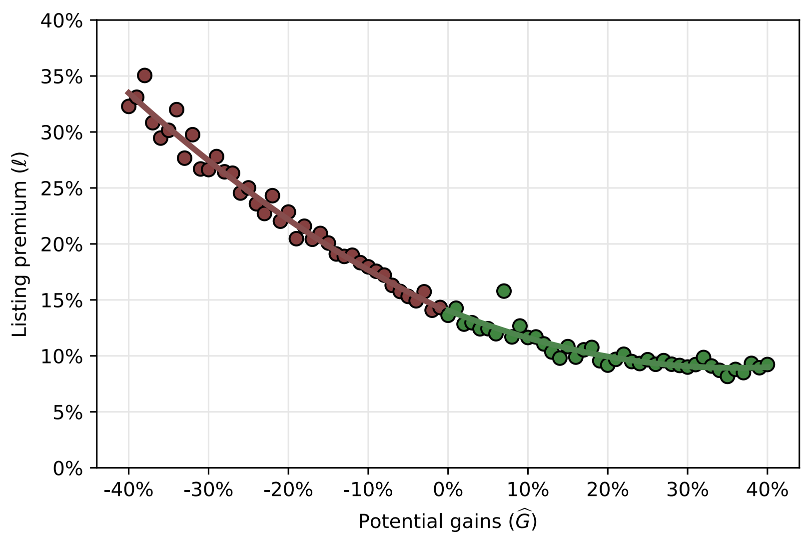


Figure 4

Bunching around realized gains of zero

The figure reports binned frequencies of observations (in 1 percentage point steps) for different levels of realized gains (G). The dotted line shows the counterfactual corresponding to the distribution of potential gains ($\widehat{G}$) in the sample of realized sales.

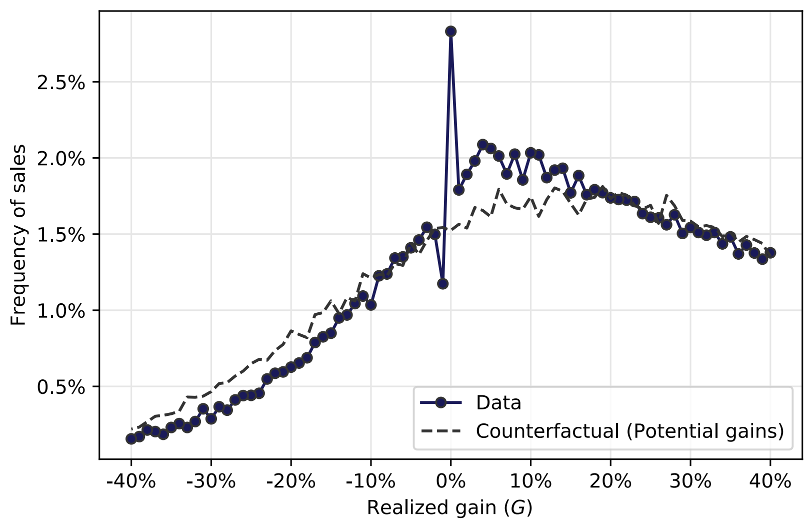


Figure 5

Gains vs. home equity

The figure reports binned average values (in 3% steps) for the listing premium (ℓ) along both levels of potential gains and home equity, and the observed frequency of sales along levels of realized gains and home equity. The dotted lines show the binned values for two cross-sections, where we condition on a home equity level of 20%, and a level of gains of 0%, respectively. We use these two representative cross-sections to generate the empirical moments used in structural estimation.

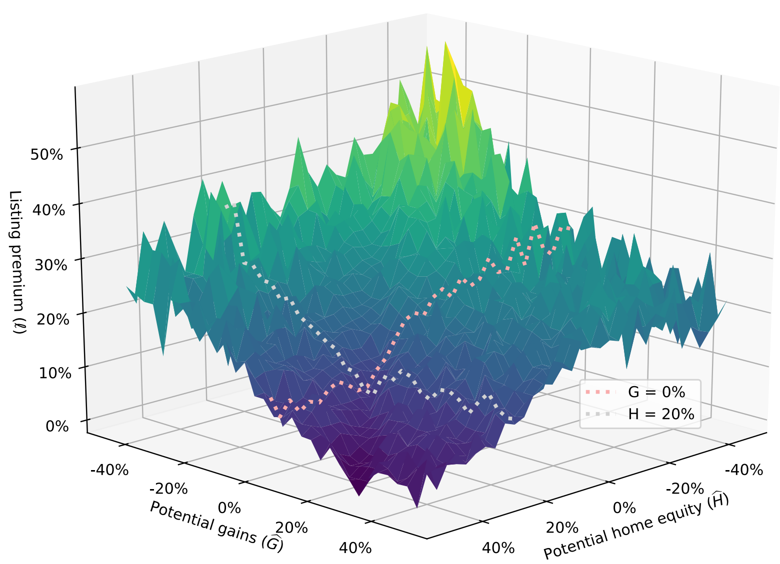


Figure 6

Concave demand in the data

The left-hand side of the figure reports the average probability of sale within six months $\alpha(\ell)$ across 1 percentage point bins of the listing premium in the sample. The solid line indicates fitted valued corresponding to a generalized logistic function (GLF). The right-hand side of the figure shows the average realized premium $\beta(\ell)$ across bins of the listing premium. The solid line indicates fitted values corresponding to a polynomial of order three.

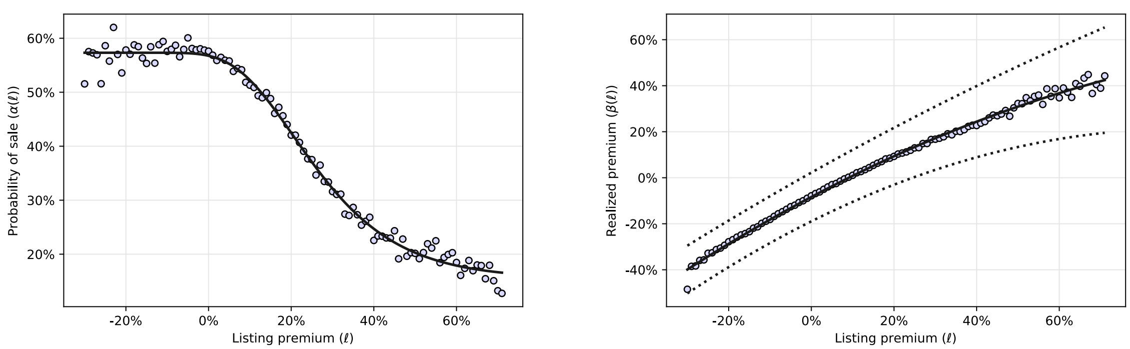
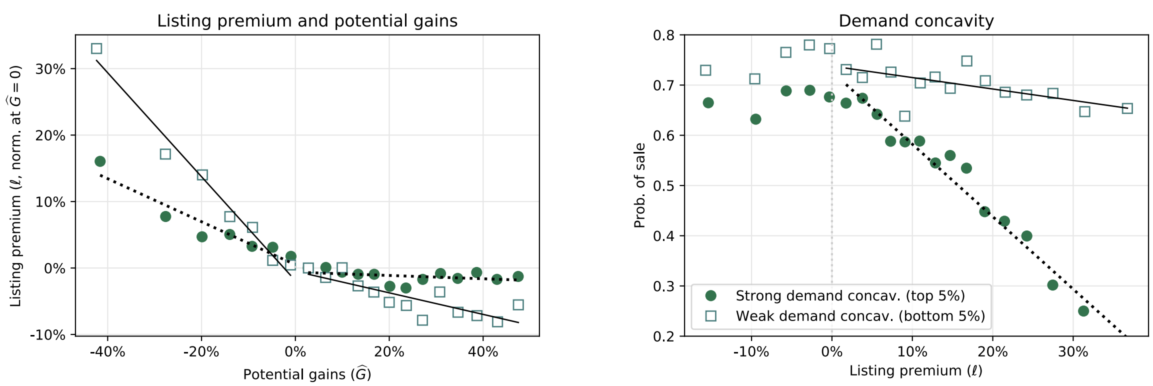


Figure 7

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 $\ell > 0$. 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

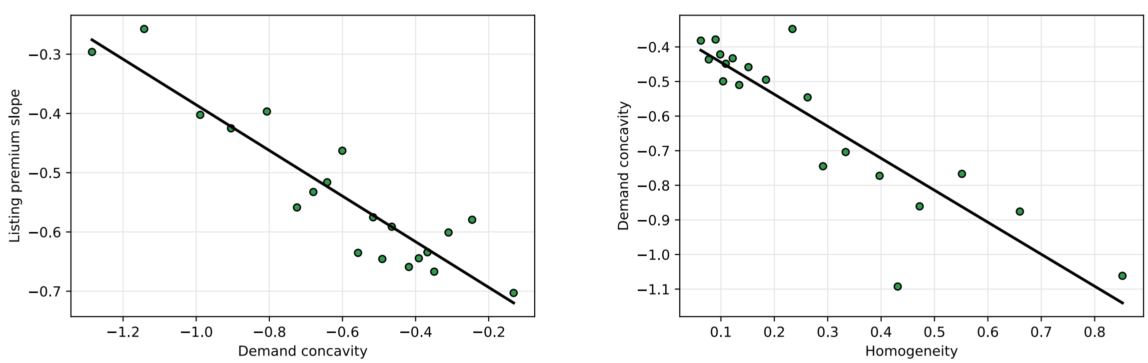


Figure 8

Extensive margin

The figure reports the average yearly probability of listing a property for sale. We first calculate the potential gain level for each unit in the stock of properties in Denmark, for each year covered by our sample of listings. We then divide the number of properties which have been listed for sale by the number of total property × year observations in the stock of properties, for each 1 percentage point bin of potential gains.

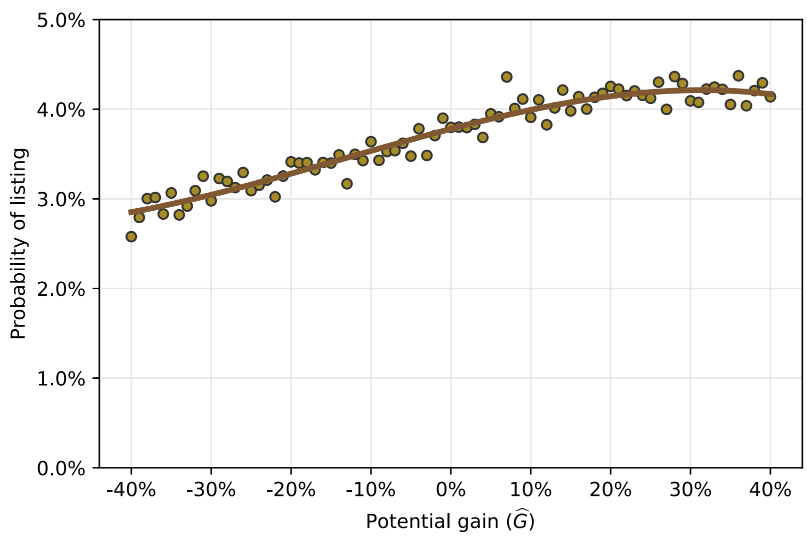
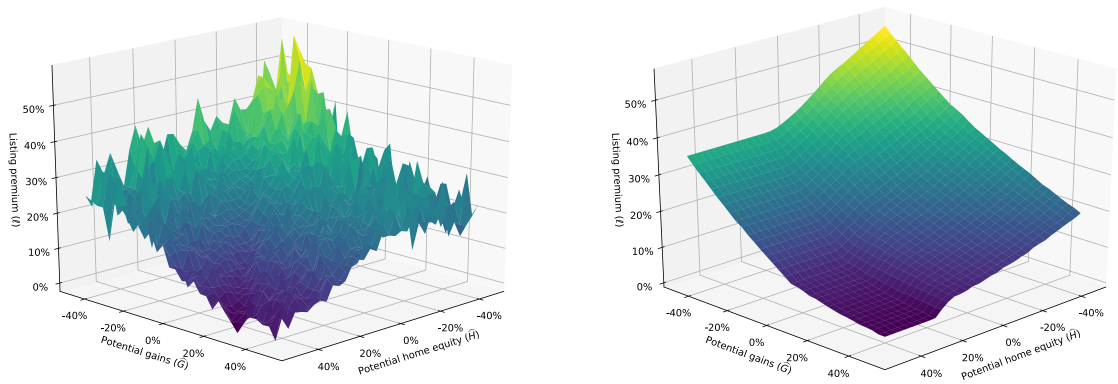


Figure 9

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

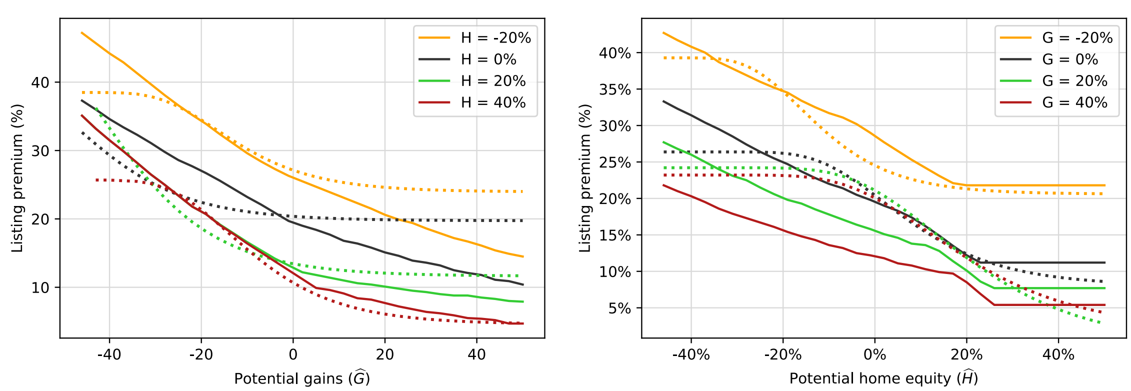


Table 1

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 $\ell - \widehat{G}$ 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 $\widehat{H} = 20\%$. 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 ℓ for $\widehat{G} = 0\%$	0.106***	(0.005)
2.	Slope ℓ–$\widehat{G}$ for $\widehat{G} < 0\%$	-0.490***	(0.047)
3.	Slope ℓ–$\widehat{H}$ for $\widehat{H} < 20\%$	-0.333***	(0.030)
4.	Cross-sectional slope ℓ–$\widehat{G}$–α for $\widehat{G} < 0\%$	-0.407***	(0.065)
5.	Cross-sectional slope ℓ–$\widehat{G}$–α for $\widehat{G} \geq 0\%$	-0.122**	(0.043)
6.	Slope of list. prob. by $\widehat{G}$	0.005**	(0.002)
7.	Bunching above $G = 0\%$	0.302***	(0.050)

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	=	0.948^{***}	(0.344)
λ	=	1.576^{***}	(0.570)
μ	=	1.060^{***}	(0.107)
δ	=	-0.097^{***}	(0.009)
$\theta_{\min}$	=	0.217	(0.165)
$\theta_{\max}$	=	1.005^{***}	(0.197)
φ	=	0.037	(0.011)

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.[58] 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	-0.407***	-0.520***	-0.504***	-0.346
	(0.067)	(0.111)	(0.087)	(0.259)
Household controls				✓
Observations	95	95	95	95
R^2	0.432			
First-stage F-stat		35.96	16.94	25.376
Hansen J-stat (p-val)			0.175	0.199

Reference Dependence in the Housing Market

Online Appendix

(For online publication)

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

July 21, 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:

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

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:

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

For a model with loss aversion, the optimal listing premium is given by:

$$\ell^*(\widehat{G}) = \begin{cases} \frac{1}{2} \left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1} \frac{\theta}{1+\eta} \right) - \frac{1}{2\beta_1} \frac{\eta}{1+\eta} \widehat{G}, & \text{if } \widehat{G} \geq \widehat{G}_0 \\ -\frac{\beta_0}{\beta_1} - \frac{1}{\beta_1} \widehat{G}, & \text{if } \widehat{G} \in (\widehat{G}_1, \widehat{G}_0) \\ \frac{1}{2} \left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1} \frac{\theta}{1+\lambda\eta} \right) - \frac{1}{2\beta_1} \frac{\lambda\eta}{1+\lambda\eta} \widehat{G}, & \text{if } \widehat{G} \leq \widehat{G}_1. \end{cases} \quad (4)$$

1.3 Mapping Between Potential and Realized Gains

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

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

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.

1.4 Extensive Margin Decision

When evaluated at the optimal level of the listing premium ℓ^*, expected utility is given by:

$$U^*(\widehat{G}) = \widehat{P} + \left[\alpha_0 - \alpha_1 \ell^*(\widehat{G})\right]\left[\eta \widehat{G} + (1+\eta)\left(\beta_0 + \beta_1 \ell^*(\widehat{G})\right) + \theta\right] \tag{9}$$

Ignoring search costs, a sufficient statistic to capture the extensive margin decision is a cut-off level of the moving shock $\widetilde{\theta}$ for which:

$$U^*(\widehat{G}) = \underbrace{\widehat{P}}_{\underline{u}}, \text{ i.e.:}$$

$$\widetilde{\theta}(\widehat{G}) = -\eta\widehat{G} - (1+\eta)\left(\beta_0 + \beta_1 \ell^*(\widehat{G})\right) \tag{10}$$

[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 $\ell\lambda^*(\widehat{G}) = \widetilde{\gamma}_{\lambda,0} - \widetilde{\gamma}_{\lambda,1}\widehat{G}$.

Assuming that the moving shock is distributed according to the following cumulative distribution function:

$$\theta \sim F(\theta_{\min}, \theta_{\max}),$$

the listing probability s is given by:

$$s(\widehat{G}) = 1 - F(\widetilde{\theta}(\widehat{G})).$$

Substituting out equation (4) in (10), we get:

$$\begin{aligned}\widetilde{\theta}(\widehat{G}) &= -(\eta + (1+\eta)\beta_1\tilde{\gamma}_1)\widehat{G} - (1+\eta)(\beta_0 + \beta_1\tilde{\gamma}_0) \\ &= -\frac{\eta}{2}\widehat{G} - (1+\eta)(\beta_0 + \beta_1\tilde{\gamma}_0)\end{aligned}$$

We then have:

$$\frac{ds(\widehat{G})}{d\widehat{G}} = \frac{d\left(1 - F(\widetilde{\theta}(\widehat{G}))\right)}{d\widehat{G}} > 0.$$

1.5 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.

1.6 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 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.[2] More realistically, a more smooth distribution of θ will blur the effects of both reference dependence and constraints on the intensive margin.

1.7 Details on the Outside Option

To better understand the role of the outside option $\underline{u}$ in the model, we first look at the case in which it is independent of the reference point R. In this case, the decision of the seller is uniquely determined by the wedge between $\underline{u}$ and the magnitude of the search cost φ (if the listing fails), and the moving shock θ (if the listing succeeds). The choice of $\underline{u}$ is therefore immaterial for seller decisions or outcomes, and only affects the estimated magnitude and the interpretation of the search cost and moving shock φ and θ, respectively.

Choosing the normalization $\underline{u} = \widehat{P}$ seems most reasonable, because it implies that absent any additional reasons to move ($\theta = 0$) and with a zero cost of listing ($\varphi = 0$), the seller will be indifferent between staying in their home and getting the hedonic value in cash.

Alternatively, it may of course be that the reference level R is linked to the outside option. For example, a simple assumption is that $\eta = 0$ (i.e., sellers derive utility exclusively from the value of terminal wealth) while the outside option is $\underline{u} = R$, e.g. because the purchase price R is the seller's current estimate of house value. In this case, the optimal listing premium is a generic function: $\ell^* = f(\widehat{P} - R) = f(\widehat{G})$, which is identical to a model with $u = G$. However, there is little support for this specification in the data: In this case (i) the magnitude of reference dependence and the degree of loss aversion do not affect the slope of the listing premium with respect to $\widehat{G}$; this slope is uniquely pinned

[2]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.

down by the demand "markup" functions, according to a set of implausible restrictions, (ii) loss aversion leads to a discrete jump at $G = 0$ and cannot generate the "hockey stick" pattern observed in the data, (iii) this model cannot explain the patterns of bunching at R that we observe.

Another possibility is that R enters the seller's estimation of value in a more refined form, indexed by a weighting factor κ, in addition to a (potentially mis-specified) hedonic value $\overline{P}$ estimated by the econometrician: $\widehat{P} = (1-\kappa)\overline{P} + \kappa R$. To understand this case, note that the property's estimated value $\widehat{P}$ enters the model in two ways: First, it affects the final price $P(\ell) = \widehat{P} + \beta(\ell)$ realized in the market. Second, it affects the seller's outside option.

If the reference point R enters $\widehat{P}$ in the same way that it enters the outside option, R will drop out in the value comparisons that the seller makes and we infer (see Section 1.5). We can of course strongly reject this case, because of the strong impact of the reference point R on the intensive margin (i.e. the observed "hockey stick" in the data), the excess bunching of realized sales prices exactly at R, and the extensive margin effects, which demonstrate an influence of R on the probability of listing.

However, if R enters the seller's property value estimate (denoted by $\widehat{P}^{Seller}$ below) differently from how it enters $\widehat{P}$ we can distinguish between three cases: First, the seller correctly uses R when valuing the property, but we don't. This is possible, but we believe unlikely, given that our results hold strongly and robustly across a large number of alternative models for $\widehat{P}$, including repeat sales. But even if our hedonic model may miss relevant price variation coming from R, this only affects estimated effects in terms of potential gains $\widehat{G}$, and such a model cannot be reconciled with the evidence of excess bunching in *realized* gains G exactly around observed prices $P = R$. Second, sellers misperceive the importance of R, i.e. they weight it differently: $\widehat{P}^{seller} = (1-\kappa)\widehat{P} + \kappa R$. The optimal listing premium function is then given by $\ell^* = f((\eta+\kappa)(\widehat{P} - R)) = f((\eta+\kappa)\widehat{G})$. In this case, reference dependence and irrational over-weighting of R have observationally equivalent effects on the *average slope* of the listing premium with respect to potential gains, but such a model of misspecified seller beliefs cannot explain the *variation* in slopes ("kinks"), and the bunching of realized prices around the reference point. Third, if both the econometrician and the seller incorrectly use R (and in different ways), we still extract the behaviour of interest, albeit potentially with considerable noise. More importantly, such a version of the model is also unable to explain the observed bunching of prices around the reference point.

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 of the property that is provided by SKAT, which assesses property values every second year.[3] 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

[3]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.

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.2 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.[4] 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. 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

[4]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.

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.[5] 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.

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.

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

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.[6]

Once all filters are applied, the sample comprises 214,508 listings of Danish owner-occupied housing in the period between 2009 and 2016, for both sold (70.4%) and retracted (29.6%) properties, matched to mortgages and other household financial and demographic information.[7] These listings correspond to a total of 191,843 unique households, and 179,262 unique properties. Most households that we observe in the data sell one property during the sample period, but roughly 9% of households sell two properties over the sample period, and roughly 1.5% of households sell three or more properties. In addition, we use the entire housing stock, filtered in the same manner as the listing data, comprising 5,540,376 observations of 807,666 unique properties to understand sellers' extensive margin decision of whether or not 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.

3 Summary Statistics

3.1 Liquid Financial Wealth

Figure A.4 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

[6]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.

[7]The data comprises 173,065 listings that have a mortgage, and 41,443 listings with no associated mortgage (i.e., owned entirely by the seller)—we later utilize these subsamples for various important checks.

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.4 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 (50.2%): $\widehat{H} \geq 20\%$ and $\widehat{G} \geq 0$.

2. Constrained Winners (26.7%): $\widehat{H} < 20\%$ and $\widehat{G} \geq 0$.[8]

3. Unconstrained Losers (6.0%): $\widehat{H} \geq 20\%$ and $\widehat{G} < 0$

4. Constrained Losers (17.2%): $\widehat{H} < 20\%$ and $\widehat{G} < 0$

The density of the data in each of the four groups is shown in Figure A.6. 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 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.7 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.

[8] $M > R$ is frequently observed in the data (47.2% 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.

In addition, we note that the notion of constraints applies only if households are reluctant to downsizing. In Figures A.30 and A.31, 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.

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 ℓ-$\widehat{G}$ and ℓ-$\widehat{H}$ 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):

$$E[\ell(V)] = A + \frac{K - A}{(1 + Qe^{-BV})^{1/\nu}}. \tag{11}$$

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 (11). We set V first as gains ($V = \widehat{G}$), and next, as the level of home equity $V = \widehat{H}$. Panel A of the figure has $\widehat{G}$ along the x-axis, and ℓ along the y-axis. However, we now condition on three levels of $\widehat{H}$: the blue line shows the ℓ-$\widehat{G}$ relationship for households with levels of $\widehat{H}$ 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 $\widehat{H}$ is between -5% and 20%, and the solid line when $\widehat{H}$ is between -15% and -5%).

To better understand these plots, we note that the average level of ℓ declines substantially as households become less constrained, and increases substantially as households become more constrained—this is simply the unconditional relationship between ℓ and $\widehat{H}$, seen in a different way in this plot. (Panel B of the figure shows the level differences that reflect the ℓ-$\widehat{G}$ relationship, i.e., higher levels of ℓ 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 ℓ as a function of $\widehat{G}$ is also 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 ℓ across the $\widehat{G}$ 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 $\widehat{G}$ and $\widehat{H}$ on ℓ. 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:

$$\begin{aligned}
\ell_{it} = \mu_t + \mu_m + \boldsymbol{\xi_0}\mathbf{X_{it}} + \boldsymbol{\xi_1}\mathbf{B_{it}} + \alpha_1 \mathbb{1}_{G_{it}<0} + \alpha_2 \mathbb{1}_{\mathrm{H}_{it}<20\%} \\
+ (\beta_0 + \underbrace{\beta_1 \mathbb{1}_{G_{it}<0}}_{\text{Gains}} + \boldsymbol{\beta_2}\mathbf{B_{it}} + \underbrace{\boldsymbol{\beta_3} \mathbb{1}_{G_{it}<0}\mathbf{B_{it}}}_{\text{Conditional effect}})G_{it} \\
+ (\gamma_0 + \underbrace{\gamma_1 \mathbb{1}_{\mathrm{H}_{it}<20\%}}_{\substack{\text{Down-payment}\\ \text{constraint}}} + \boldsymbol{\gamma_2}\mathbf{B_{it}} + \underbrace{\boldsymbol{\gamma_3} \mathbb{1}_{\mathrm{H}_{it}<20\%}\mathbf{B_{it}}}_{\text{Conditional effect}})\mathrm{H}_{it} \\
+ \varepsilon_{it}.
\end{aligned} \tag{12}$$

Equation (12) allows ℓ to depend (piecewise) linearly on both home equity $\widehat{H}$ and gains $\widehat{G}$ (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.7. To capture the conditional behavior, we bin both home equity and gains (as well as the other conditioning variables) and introduce dummy variables $\mathbf{B}$ into the regression of the respective other dimension to capture the different ℓ-$\widehat{G}$ (and ℓ-$\widehat{H}$) relationships for these groups. We allow for $\mathbf{B}$ to modify both the unconditional relationship with $\widehat{G}$ and $\widehat{H}$ (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 (12), the estimates point to interesting conditional variation in the data. The y-axis of Panel A of Figure A.8 shows the point estimate for the slope of the ℓ-$\widehat{G}$ relationship for different bins of household covariates shown on the x-axis.

Panel B of Figure A.9 investigates the effect of down-payment constraints, conditioning the ℓ-$\widehat{H}$ relationship on the level of household covariates.

4 Measuring Concave Demand

Figure A.3 shows the distribution of time-on-the-market (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 37 weeks (25 weeks). This is higher than the value of roughly 7 weeks reported in Genesove and Han (2012).

We next inspect the inputs to the function $\alpha(\ell)$ in the data. The top plot in Figure A.10 shows how TOM relates to the listing premium ℓ in the data using a simple binned scatter plot. When ℓ is below 0, TOM barely varies with ℓ; however, TOM moves roughly linearly with ℓ when ℓ is positive and moderately high. Interestingly, we also observe that the relationship between ℓ and TOM flattens out as ℓ rises to very high values above 40%. This behavior is mirrored in the bottom panel of Figure A.10, which shows the share of seller *retracted* listings, which also rises with ℓ. 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.

In the paper, we simply convert the two plots into a single number, which is the probability of house sale within six months (i.e., $\alpha(\hat{\ell})$) 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 (Richards, 1959, Zwietering et al., 1990, Mead, 2017) of the form:

$$\alpha(\ell) = A + \frac{K - A}{(C + Qe^{-B\ell})^{1/\nu}}, \tag{13}$$

5 Robustness Against Unobserved Heterogeneity

We find an asymmetric relationship between the listing premium ℓ and potential gains $\widehat{G}$ over the loss domain, as well as an asymmetric decrease in the probability of sale

[9]Since we do not want to model any higher-order effects in this context, we exclude the respective gains bins from $\mathbf{B}$ 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.

when listing premia are greater than zero, as captured in the function $\alpha(\ell)$ ("demand concavity"). In the following, we show that the observed non-linearities are robust to a range of underlying models for $\widehat{P}$, and using additional methods to deal with potentially unobserved heterogeneity from different sources.

5.1 Unobserved Quality and the Listing Premium over Potential Gains

We follow Genesove and Mayer (2001) to illustrate potential confounds. For simplicity, the "hockey stick" relationship between listing premia and potential gains can be expressed in a regression framework as measuring the slope of the effect of gains less than 0 on the listing premium. The regression equation is then:

$$\ell_{ijst} = \alpha + \beta \mathbb{I}[G < 0] \underbrace{(\ln P_{it} - \ln R_{ijs})}_{\text{"True" gain } G} + \epsilon_{ijst}, \tag{14}$$

for a property i sold by household j at time t, with initial purchase taking place at time s, yielding reference price R_{ijs}, and where $\ln P_{it} = \widehat{\ln P_{it}} + v_i$, i.e. the "true" market value of the house known to the seller is the estimated hedonic value plus an unobserved property-i specific component. In addition, we assume that the log reference price $\ln R_{ijs}$ can also be decomposed into three components $\ln \widehat{P_{is}} + \nu_i + \omega_{ijs}$, that is the estimated market value, the unobserved quality of the property, and the degree to which buyers under-or overpaid for the property at time s, ω_{ijs}. Instead of the "true" listing premium ℓ that the seller chooses, we measure $\widehat{\ell}_{ijst} = \ln L_{ijst} - \ln \widehat{P_{it}}$. The log listing price $\ln L_{ijst}$ also comprises three components: the estimated market value $\widehat{P_{it}}$, the unobserved quality of the house ν_i, and the true listing premium ℓ_{ijst}. Substituting yields

$$\widehat{\ell_{ijst}} = \ln L_{ijst} - \ln \widehat{P_{it}} = \nu_i + \ell_{ijst}. \tag{15}$$

Equation 15 reflects the inference problem that a potential buyer faces: the buyer would be willing to pay more for quality ν_i, but not for the seller-specific premium ℓ_{ijst}.[10] The

[10]This is captured in our characterization of the demand side: if ν is relatively easy to discern (e.g. given a very homogeneous housing stock), any increase in ℓ gets penalized more strongly, i.e. the probability of a quick sale decreases more sharply.

ideal regression can hence be written as

$$\ell_{ijst} = \alpha + \beta\mathbb{I}[\widehat{\ln P_{it}} + \nu_i - (\widehat{\ln P_{is}} + \omega_{ijs} + \nu_i) < 0](\widehat{\ln P_{it}} + \nu_i - (\widehat{\ln P_{is}} + \omega_{ijs} + \nu_i)) + \epsilon_{ijst} \tag{16}$$

$$= \alpha + \beta\mathbb{I}[\widehat{\ln P_{it}} - (\widehat{\ln P_{is}} + \omega_{ijs}) < 0](\widehat{\ln P_{it}} - (\widehat{\ln P_{is}} + \omega_{ijs})) + \epsilon_{ijst} \tag{17}$$

In contrast, the feasible regression is

$$\widehat{\ell_{ijst}} = \alpha + \beta\mathbb{I}[\widehat{\ln P_{it}} - (\widehat{\ln P_{is}} + \omega_{ijs} + \nu_i) < 0](\widehat{\ln P_{it}} - (\widehat{\ln P_{is}} + \omega_{ijs} + \nu_i)) + \eta_{ijst}, \tag{18}$$

such that

$$\begin{aligned}\eta_{ijst} = \epsilon_{ijst} + \nu_i + \beta(\mathbb{I}[\widehat{\ln P_{it}} - (\widehat{\ln P_{is}} + \omega_{ijs} + \nu_i) < 0](\widehat{\ln P_{it}} - (\widehat{\ln P_{is}} + \omega_{ijs} + \nu_i)) - \\ \mathbb{I}[\widehat{\ln P_{it}} - (\widehat{\ln P_{is}} + \omega_{ijs}) < 0](\widehat{\ln P_{it}} - (\widehat{\ln P_{is}} + \omega_{ijs}))).\end{aligned} \tag{19}$$

As noted by Genesove and Mayer, the two sources of bias that are caused by this substitution are: first, since ν_i and ω_{ijs} are unobservable, and both appear in the term measuring gains, it should generate classical measurement error, biasing β towards zero.[11] Second, the unobserved quality component ν_i induces a negative correlation with gains (intuitively, gains are overestimated if the true value of the property is higher), likely leading to negative bias in the estimate of β. Since β is expected to have a negative sign, this may cause the steepness of the slope to be overestimated.

We propose the following five robustness checks to validate that the asymmetry in the listing premium over potential gains is not driven by unobserved quality and other confounds.

1) We use alternative pricing models of $\widehat{\ln P}$ and in particular a repeat sales model to absorb time-invariant unobserved property quality ν_i. The key intuition is that we have repeat sales of property i in the sample, such that we can estimate a pricing model ($\widehat{\ln P_{it}^{repeat}}$) with property fixed effects that differences out ν_i. The model estimation is further detailed below. This allows us to measure the true listing premium ℓ set by the seller as $\widehat{\ell_{ijst}^{repeat}} = \widehat{\ln P_{it}^{repeat}} + \ell_{ijst} - \widehat{\ln P_{it}^{repeat}} = \ell_{ijst}$. If we further assume that $\omega_{is} = \widehat{\ln P_{is}^{repeat}} - \widehat{\ln P_{is}}$ and that ω does not vary further across households, we can

[11]Note there is a nonlinearity here which is a departure from the classical measurement error framework, but Genesove and Mayer confirm that this source of bias leads to attenuation.

measure

$$\ell_{ijst} = \alpha + \beta \mathbb{I}[\ln \widehat{P_{it}^{repeat}} - (\ln \widehat{P_{is}^{repeat}} + \omega_{is}) < 0](\ln \widehat{P_{it}^{repeat}} - (\ln \widehat{P_{is}^{repeat}} + \omega_{is})) + u_{ijst}, \tag{20}$$

which comes close to the ideal equation. Figure A.14 shows that the listing premium slope for gains less than zero using repeat sales models for $\ln \widehat{P}$ is comparable to that of the baseline hedonic model without property fixed effects, assuaging concerns that the hockey stick shape is driven by unobserved quality.

2) Next, we substitute shire-level house prices for $\widehat{\ln P}$. This is the reduced form exercise to the IV strategy where we instrument $\ln P_{it}$ (and hence $\widehat{G}$) using $\ln P_t^{shire}$, which is correlated with the property-specific $\ln P_{it}$, but plausibly exogenous to ν_i and ω_{ijs}. The intuition behind this is that households may be able to endogenously affect $\ln P_{it}$ via e.g. unobserved maintenance efforts, but not $\ln P_t^{shire}$.

3) We control for further observables in addition to home equity. As we show, ℓ_{ijst} is set by seller j in line with ijt-specific confounds ξ_{ijt}, most prominently home equity, but also liquidity, demographics, municipality-and time specific-factors, which we observe in our granular data.

4) We replicate the bounding approach as suggested by Genesove and Mayer (2001). They show that controlling for the residual from the previous sales price as a noisy proxy for unobserved quality yields a lower bound for the coefficient on loss (in our case, gain), while not controlling yields an upper bound effect in line with the bias described above. Table A.8 which replicates Table 2 in their paper using our data shows that we get a coefficient of 0.44 to 0.53, meaning that a 10 percent increase in the loss faced is associated with a 4.4 to 5.4% higher list price, similar to and even slightly larger than their bounds of 2.5 to 3.5%.

5) Lastly, while 1) to 4) focus on the magnitude of the listing premium slope, we document evidence in line with the key predictions from the model, that the slope changes discontinuously around the threshold, using a regression kink design (RKD). In particular, we show that the slope to the left of zero increases significantly, and that other property- and household-specific observable characteristics are smooth around the threshold, in line with the identifying assumptions of the design. The RKD estimation is described further below.

5.2 Unobserved Quality and Demand Concavity

We find an asymmetric relationship between listing premia and the probability of sale within six months, which is relatively flat for listing premia smaller equal zero, and decreasing for listing premia greater than zero, i.e. demand is concave, in line with Guren (2018). Unobserved quality as described in equation 15 is again a potential confound to this relationship, as properties with greater observed listing premia purely driven by unobserved quality would be expected to sell faster, leading to an upward bias in the observed demand concavity.

In order to provide robustness for our results, we again 1) use a repeat sales model for $\widehat{P}$ to account for time-invariant unobserved quality. The results are shown in Figure A.14 Panel B, where the slope to the right of a zero listing premium is indeed steeper (more negative) using repeat sales models.

2) We also use shire-level house prices to instrument for individual property prices, which are plausibly exogenous to property-specific unobserved quality. We also control for detailed property-specific observables.

3) Lastly, we also perform a regression kink design analysis and find that there is a significant change in slope to the right of zero listing premia, while other property-specific observable characteristics appear smooth around the threshold.

5.3 Hedonic Pricing Model and Alternatives

The following describes the baseline hedonic pricing model and alternative models in more detail.

5.3.1 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

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

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 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.5.

5.3.2 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.

5.3.3 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.3 and the results are compared in Figures A.12 and A.13.

5.3.4 Out-of-sample Testing

The large number of controls and fixed effects in the hedonic model could give rise to concerns about overfitting. To test for overfitting, we conduct out-of-sample testing. Table A.4 reports mean R^s from 1000 iterations of sampling of 50, 75 and 100 percent, respectively, and Figure A.15 show distributions of the R^s from the 1000 iterations. The model performs well even for models estimated on small samples.[12]

Figure A.16, A.17, and A.18 show that the listing premium to gains and the listing premium to home equity relationships as well as the demand cancavity hold, also when the hedonic price is predicted out-of-sample.

5.3.5 Hedonic Model and the Tax-assessed Value

The accuracy of the hedonic model is improved by including the pre-determined tax-assessed value and in addition adjusting for the current local price development, using municipality-year fixed effects. But even without including the tax-assessed value, our hedonic model performs well. Table A.6 decomposes the hedonic model and shows the R^2 contribution from each component. In itself the tax-assessed value explains 78 percent of the variation of sales prices, and municipality-year fixed effects explain 47 percent. Our baseline hedonic model without tax-assessed value explains 76 percent of the variation in sales prices and adding the third degree polynomial of the tax-assessed value increases explanatory power to 87 percent.

The tax-assessed value in itself stems from a very comprehensive model, developed by the Danish tax authorities (SKAT). Relative to our data, the tax-assessment model utilizes more detailed data parameters such as distance to local facilities like schools and public transport. In addition, in some cases (prior to 2013) the assessment is manually adjusted by the tax authorities if the mechanically set value is opposed by owners or if the property is in the right tail of the distribution. Overall we include the tax-assessed value in our model because it adds information beyond what we observe.

However, the tax-assessed value is in itself inferior to our model, since it - especially in the period of our data - underestimates price levels, see Figure A.22 panel (a). Due to systematic misvaluations, in 2013 the tax assessments were frozen at 2011 levels in order to develop a new model of assessment. As of 2020 the new model is not yet in use. Figure A.22 panel (b) and (c) clearly illustrate the shortcomings of the tax assessment in our

[12]It could potentially be because of the inclusion of the tax-assessed value in the model, since the tax-assessed value in itself is a the outcome of a full sample model. However, although excluding the tax-assessed value from the model reduces the R^2, small samples still provide good fits. See Table A.5.

sample period in particular. The figures show how the tax assessment was backwards-looking and as a result lagged behind realized prices in the housing market boom prior to the financial crisis and in the bust during the crisis. Starting from 2013 the tax assessment decouples from market prices.

Figure A.19, A.20, and A.21 show that the relationships between gains and listing premium, beetween home equity and listing premium, and demand concavity hold when using the tax assessment in the years it is most accurate instead of the hedonic price, although the level of the listing premium is higher (due to the assessed value underestimating the market value, as discussed).

5.4 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. The idea is to include the pricing residual relative to the previous sales price as a noisy proxy for unobserved quality to get a lower bound estimate for the coefficient on loss (in our case, gain), while not controlling for it yields an upper bound effect in line with the bias described above. In particular, we replicate Table 2 in their paper in Table A.8. 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 5.4% increase in list prices, compared to their 2.5 lower bound and 3.5% upper bound estimate.

5.5 Regression Kink Design (RKD)

We implement a regression kink design (RKD) 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 $\ell = 0$. The design (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, the resulting differences in behavior above and below the threshold can be interpreted as causal. The identifying assumption relies on other confounds being smooth around the threshold, e.g. in our case, that unobserved property quality should not have a significant kink precisely at the threshold.

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

$$\tau = \lim_{v \to \overline{v}_+} \frac{dE[\ell_{it}|V_{it} = v]}{dv}\bigg|_{V_{it}=v} - \lim_{v \to \overline{v}_-} \frac{dE[\ell_{it}|V_{it} = v]}{dv}\bigg|_{V_{it}=v}, \tag{21}$$

based on the following RKD specification (Landais 2015):

$$E[\ell_{it}|V_{it} = v] = \kappa_m + \kappa_t + \boldsymbol{\xi}\mathbf{X_{it}} + \left[\sum_{p=1}^{\overline{p}} \gamma_p(\nu - \overline{\nu})^p + \nu_p(v - \overline{v})^p \mathbb{1}_{V \geq \overline{v}}\right]. \tag{22}$$

$$\text{where } |v - \overline{v}| < b. \tag{23}$$

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 $V = \widehat{G}$ 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.10 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.24 to A.26 show further robustness for the RKD using gains.

6 Bunching Estimates Robustness

We conduct several robustness checks for bunching in realized gains around the reference point of the previous sales price. First, we show the prevalence of listings and sales

[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 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.27.

at round numbers in Figure A.37. We then show the distribution of realized gains by excluding sales at rounded prices of 10,000 and 50,000 DKK (Figure A.38) and 100,000 and 500,000 DKK (Figure A.39), respectively. We further show that bunching is robust across all quintiles of the previous sales price (Figure A.40) and when splitting into quintiles by holding period, except for the sub-sample with holding periods of greater than 12 years (top quintile), where there is limited mass of households who are around the zero realized gain threshold.

7 Institutional Background

7.1 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.16.

7.2 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.

7.3 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 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.

7.4 Assumability and Refinancing

Mortgages in Denmark are generally assumable, i.e. sellers can transfer their mortgage to the buyer at sale (Berg et al. 2018). Borrowers also have the option to repurchase their fixed-rate-mortgage from the covered bond pool at market or face value. Both market features alleviate potential seller lock-in, in particular in a rising rate environment (Campbell 2012). In our sample period, over 2009-2016, rates are broadly decreasing, which generates incentives to refinance. Another question is if the assumability of mortgages can relax down-payment constraints, and hence generate additional gains by purchasing a house with a specific mortgage value. In general, any mortgage assumption needs the approval from lenders, who enforce the 20% down-payment constraint for the assumed debt. For instance, if a household sells a house with value $P = 90$ and mortgage balance $M = 80$ to buy a house with value $P = 90$ and mortgage balance $M = 80$, the household can only assume $M = 0.8 \times 90 = 0.72$ and hence requires an additional down payment. It is very rare (but possible) to assume a mortgage with an LTV > 80 after negotiation with the lender, but the mortgage interest rate tends to be expensive (e.g. currently, a 140 bp total fee above the coupon rate). Another benefit of assuming the mortgage is to save the 150bp stamp duty due on new mortgage debt, with a maximum 120 bp benefit at 80% LTV, which households would need to trade-off against the increase in search cost to find a house with high assumable debt, given time, location, and preference constraints.

7.5 Property Taxation in Denmark

SKAT assess the property value to determine the amount of tax to be paid on real estate. The exact rate of property taxation varies across municipalities, but the assessed value is set centrally. In Denmark there is no tax on realized capital gains if the owner "has lived" in the house/apartment, under the condition that the house must not be extremely large (lot size smaller than 1400 sqm). It is not necessary for the owner to live in the property at the time of the sale, but she needs to establish that the property was not used under a different capacity (such as renting to a public authority) before the sale. The "substantial occupation requirement" used to be two years, but now requires only documentation of utilities use, registration etc. Capital gains that do not fall under this exception are taxed like other personal income. Taxation on gifts to family members stands at 15% above 65,700 DKK (as of 2019). However, home owners can also give the property to a child with an interest-free, instalment-free debt note terminated at the time of sale. Heirs can inherit houses and any associated tax exemptions on sale in the event of death of the principal resident.

8 Additional Tables and Figures

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

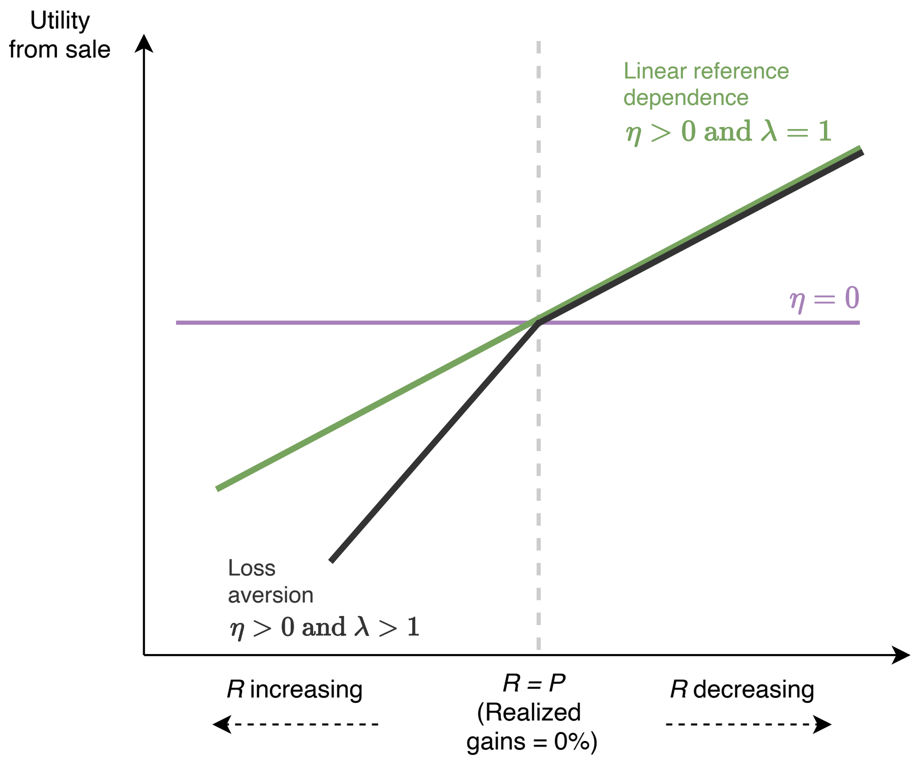


Figure A.2
Price-Volume Correlation

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.

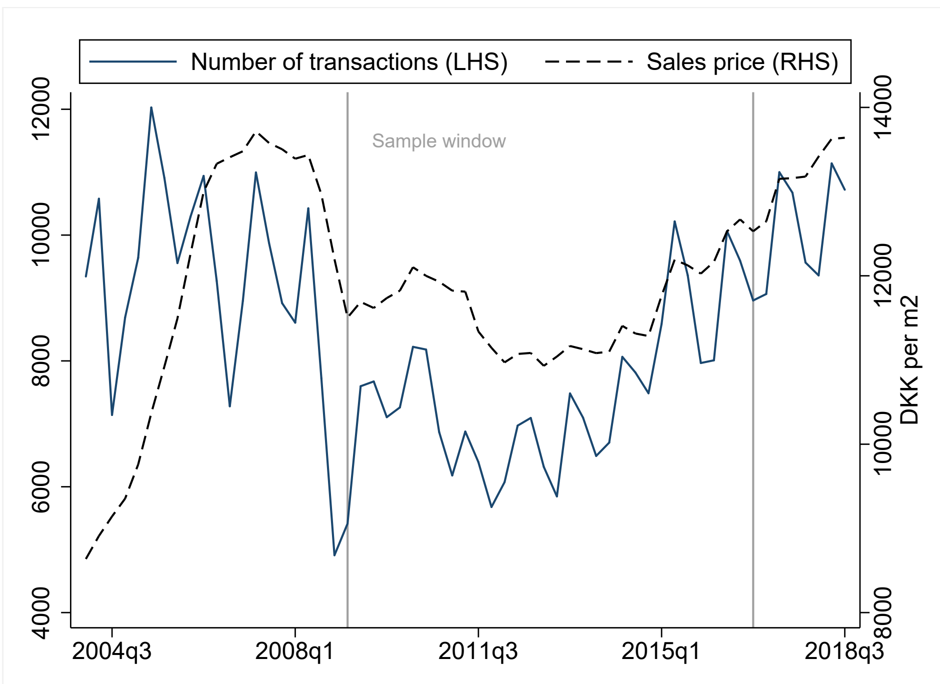
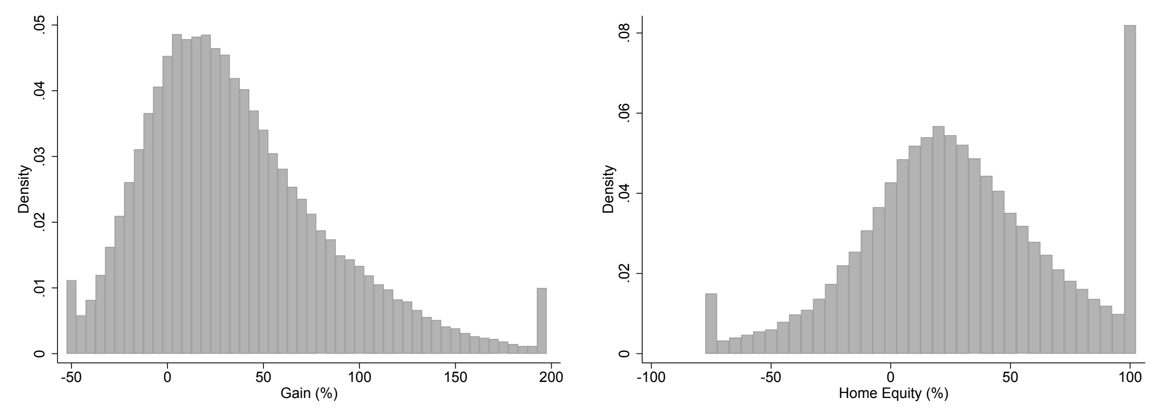


Figure A.3

Summary Statistics: Transaction Characteristics

This figure shows four histograms of main variables of interest. Gain ($\widehat{G}$) is computed as the log difference between the estimated hedonic price ($\widehat{P}$) and the previous purchase price (R), i.e. $\widehat{G} = \ln \widehat{P} - \ln R$, in percent. Home equity ($\widehat{H}$) is computed as the log difference between the estimated hedonic price and the current mortgage value (M), i.e. $\widehat{H} = \ln \widehat{P} - \ln M$, in percent. $\widehat{H}$ is truncated at 100 in order to avoid small mortgage balances leading to log differences greater than 100. The listing premium (ℓ) 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

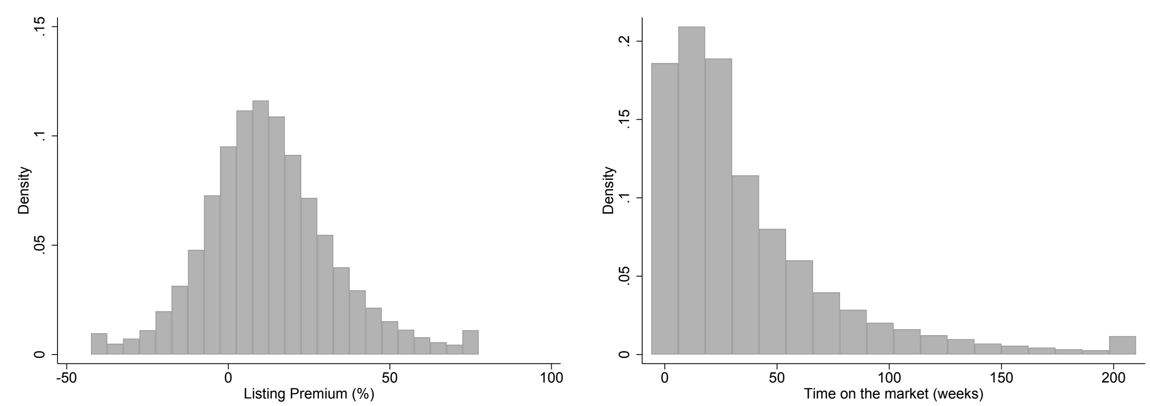
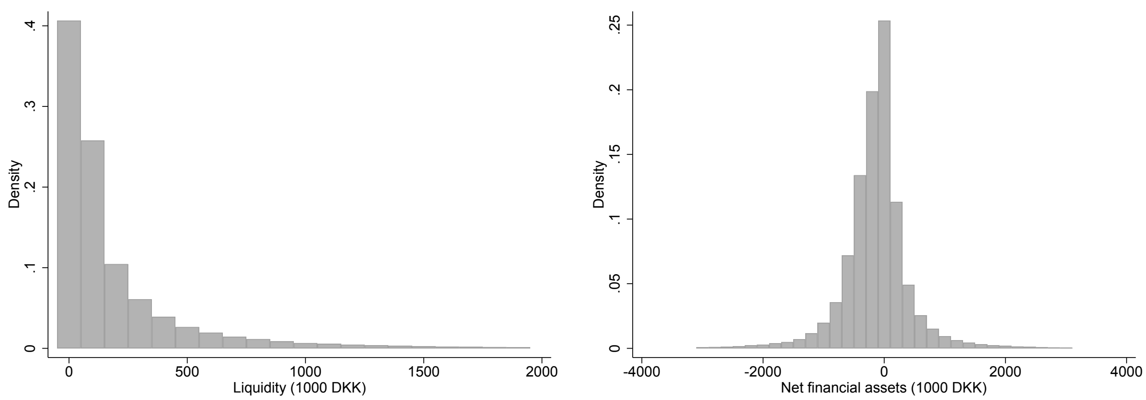


Figure A.4

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

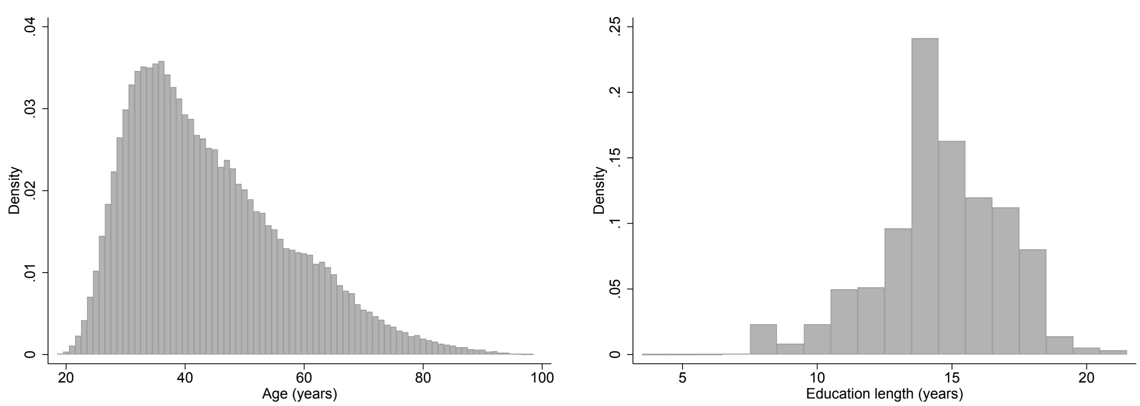


Figure A.5
Actual vs. Predicted Price of Sold Properties

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 ($N = 114,897$). 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 0.88.

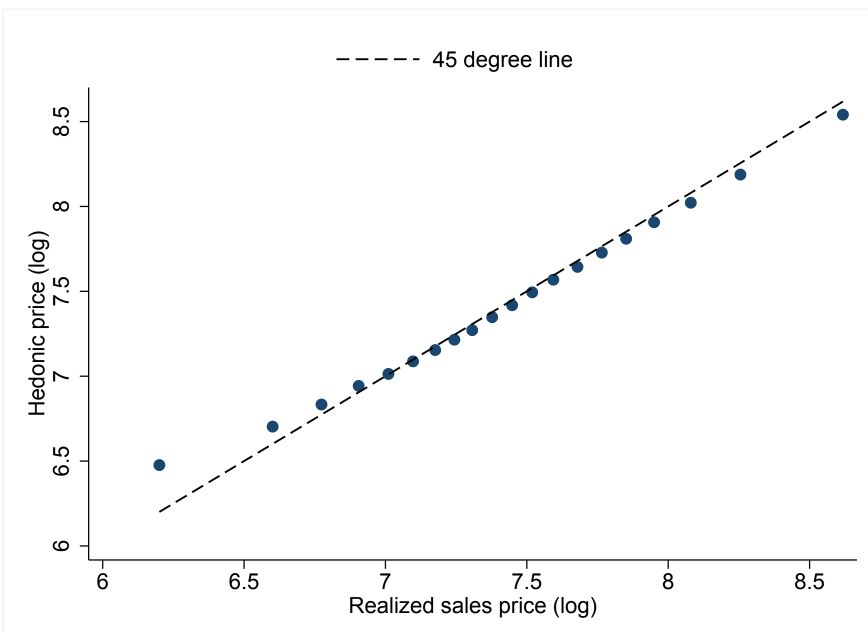


Figure A.6

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 ($\widehat{H}$ ≥20% and $\widehat{G}$ ≥0) covering 48.8% of the sample, (2) Constrained Winners ($\widehat{H}$ <20% and $\widehat{G}$ ≥0) with 26.5%, (3) Unconstrained Losers ($\widehat{H}$ ≥20% and $\widehat{G}$ <0) with 6.2%, and (4) Constrained Losers ($\widehat{H}$ <20% and $\widehat{G}$ <0) accounting for 18.6% of the sample.

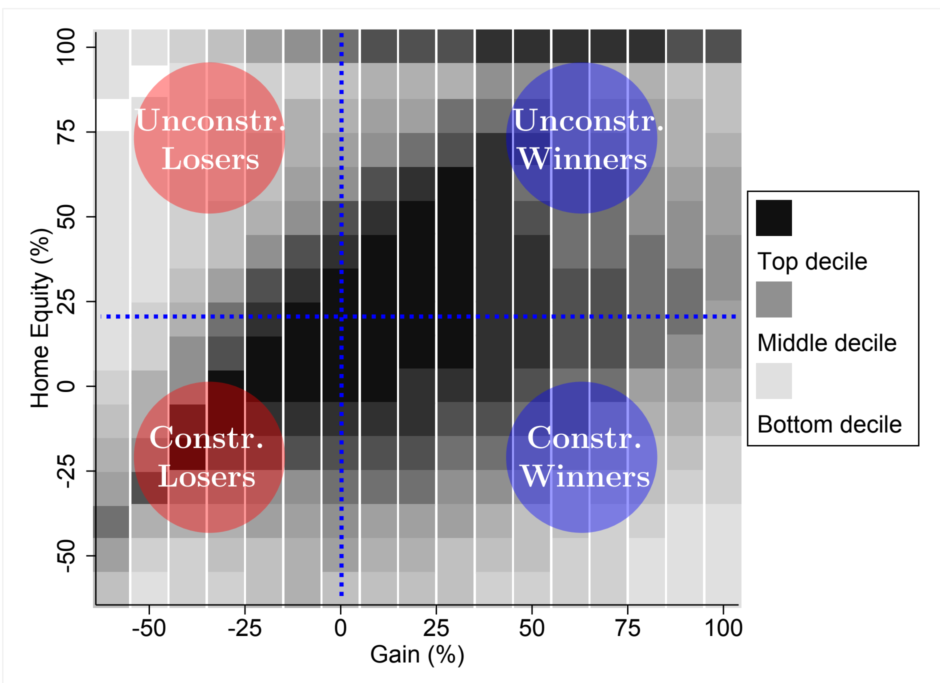


Figure A.7

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 ($\widehat{H}$ $\geq$20% and $\widehat{G}$ $\geq$0), II) Constrained Winners ($\widehat{H}$ <20% and $\widehat{G}$ $\geq$0), III) Unconstrained Losers ($\widehat{H}$ $\geq$20% and $\widehat{G}$ <0), IV) Constrained Losers ($\widehat{H}$ <20% and $\widehat{G}$ <0).

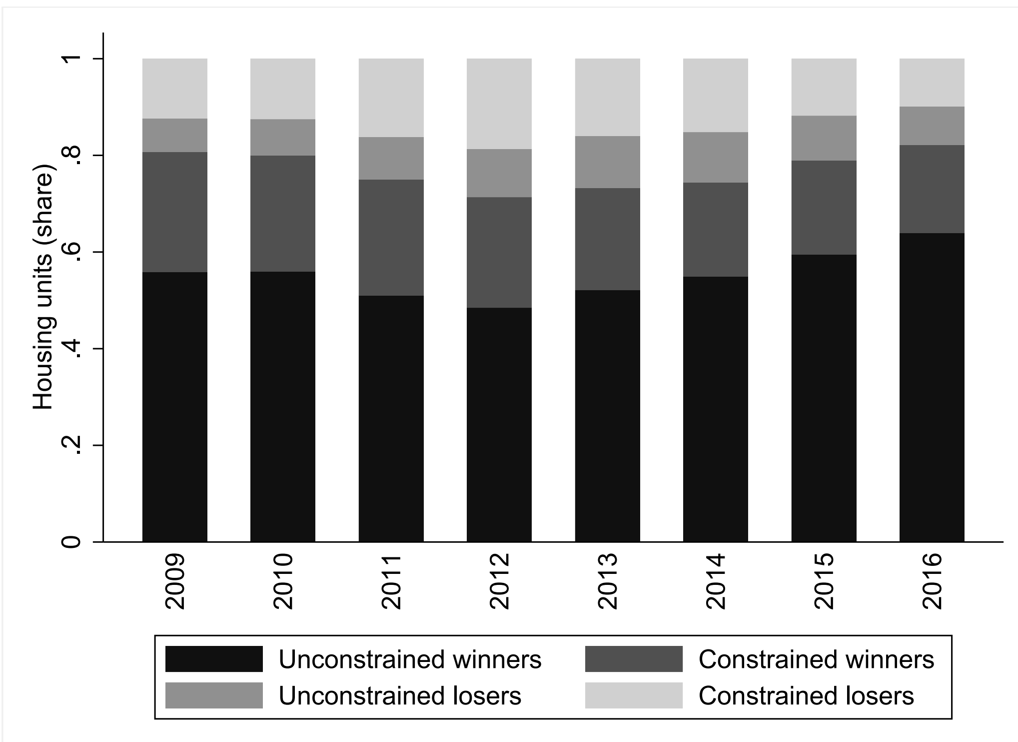


Figure A.8

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 ($\widehat{G} < 0$), 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.

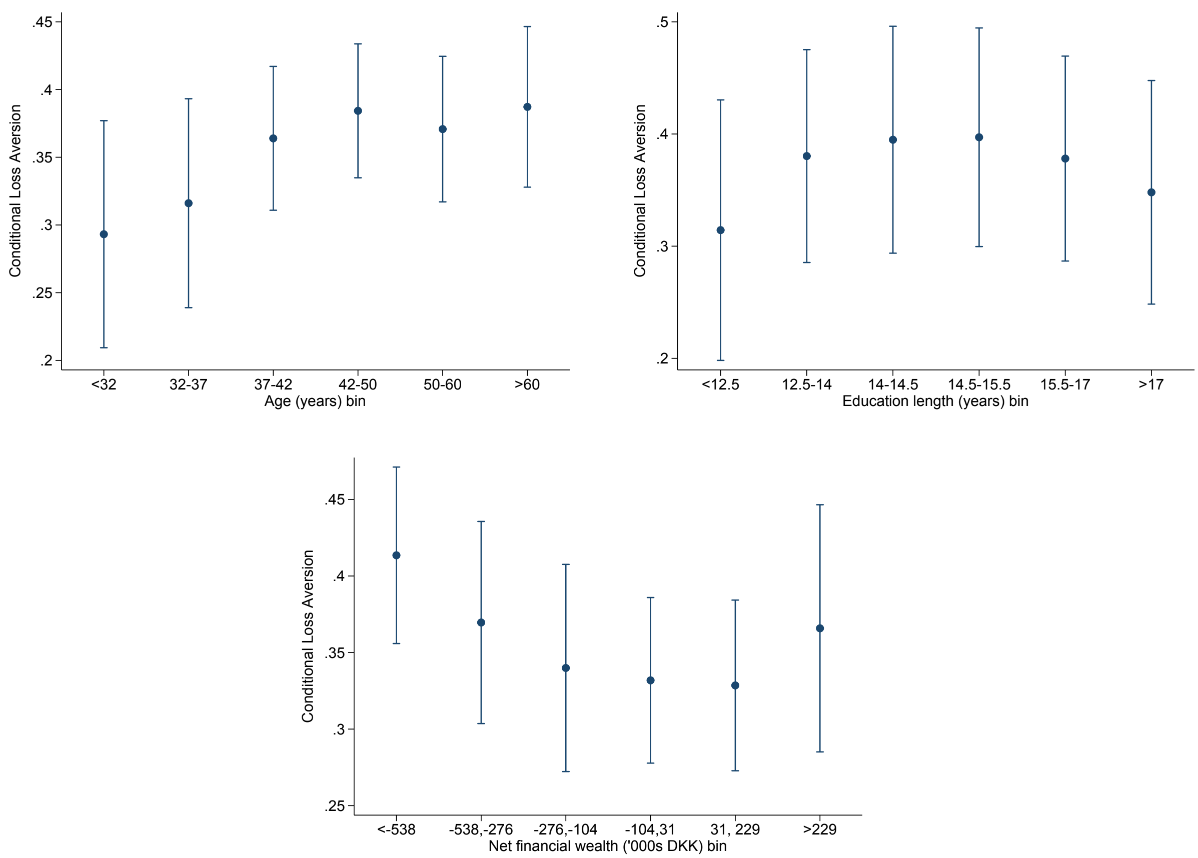
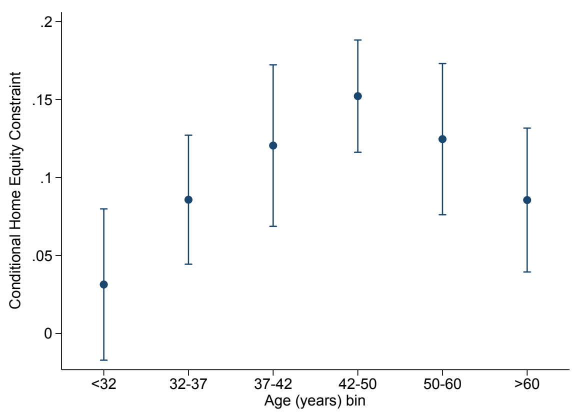
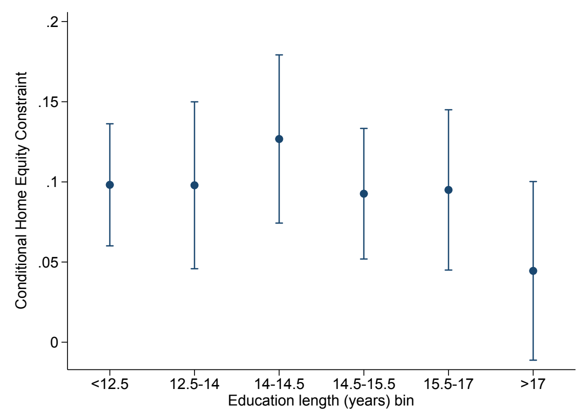


Figure A.9

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 ($\widehat{H} < 20\%$), 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.





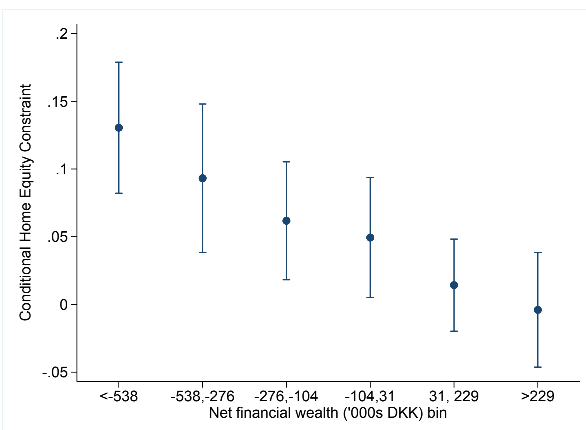
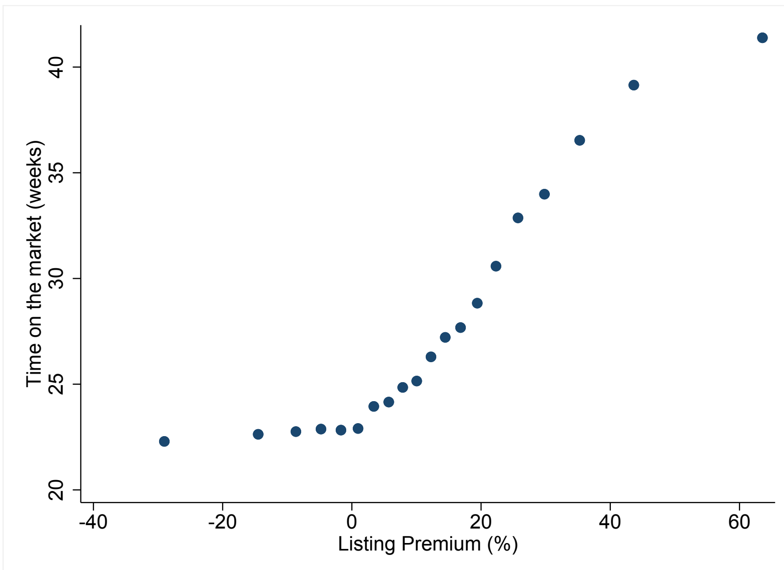


Figure A.10

Time-on-the-market and retraction rate

This figure shows the relationship between (a) time-on-market, and (b) the retraction rate for different levels of the listing premium.



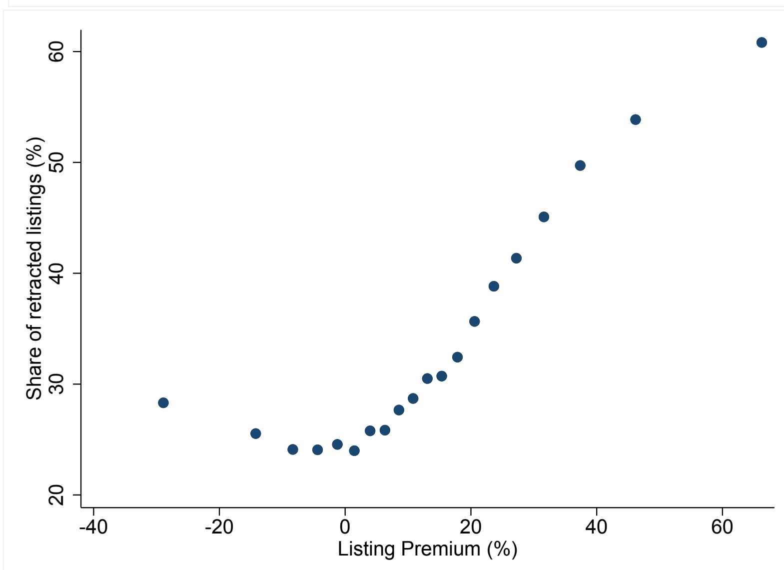


Figure A.11

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 $V = \widehat{G}$ and $V = \widehat{H}$ respectively. The solid dots indicate bin scatter points, for equally spaced bins of experienced gains and home equity.

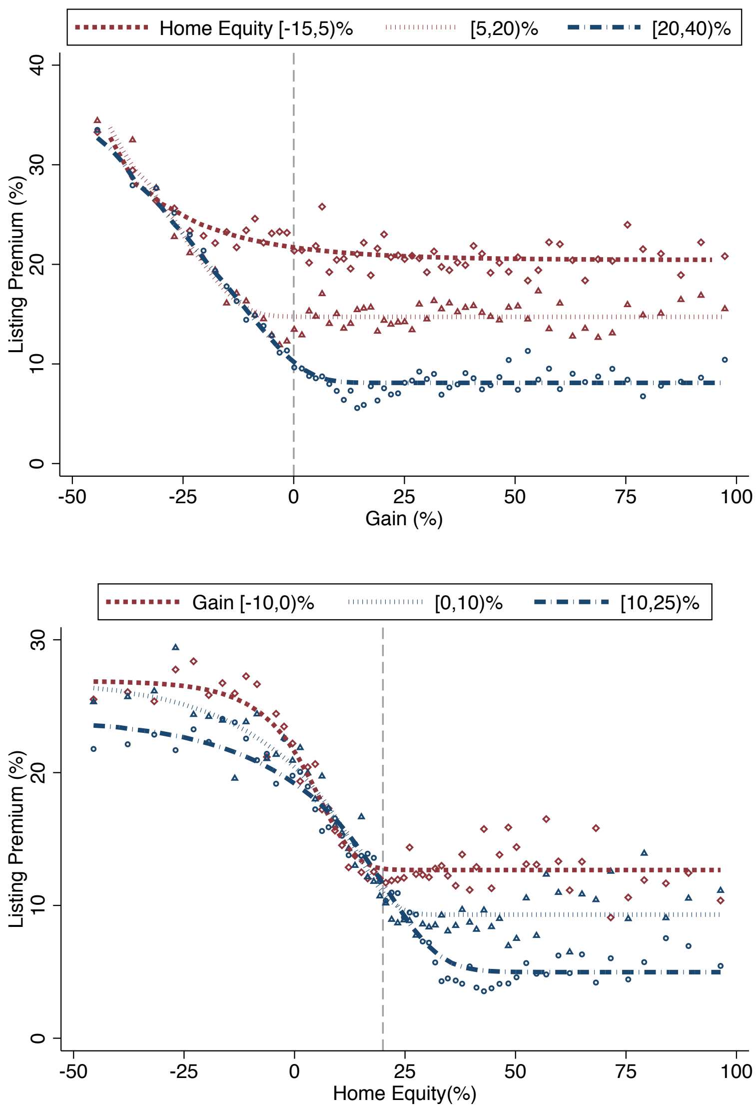


Figure A.12

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.

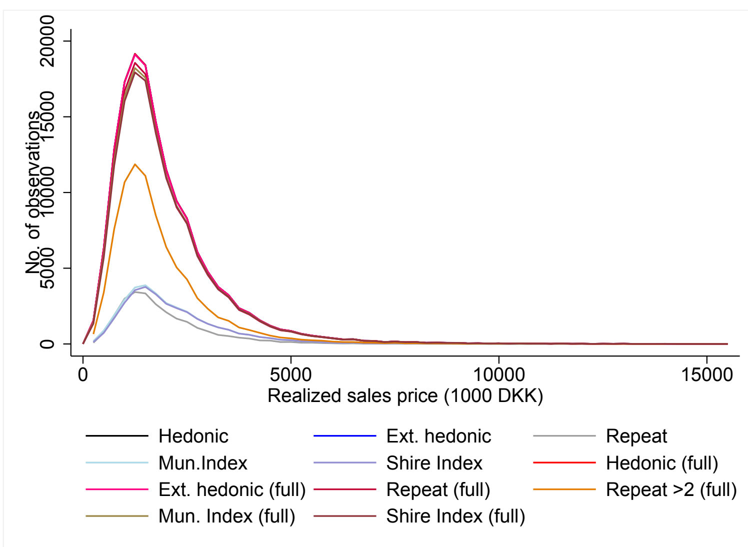
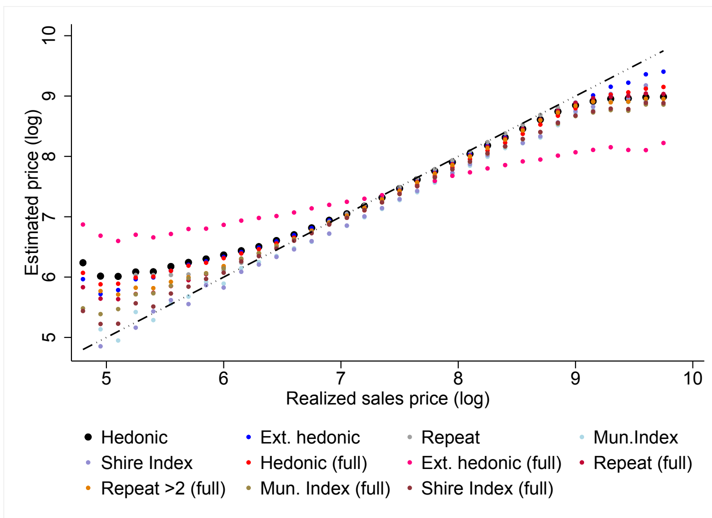


Figure A.13

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

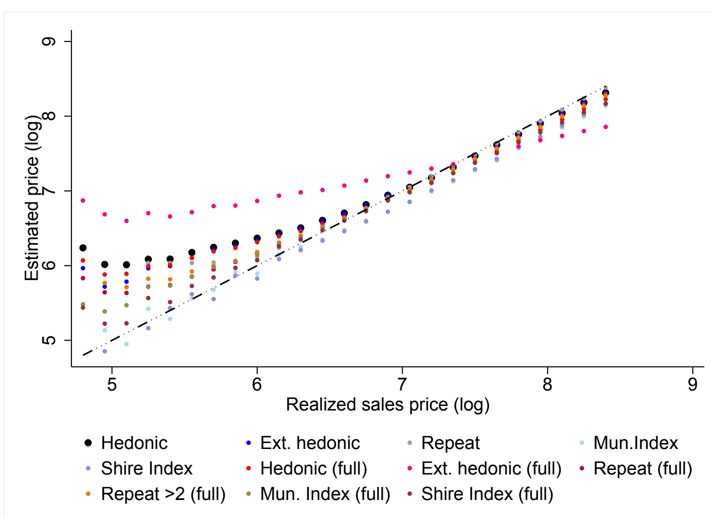
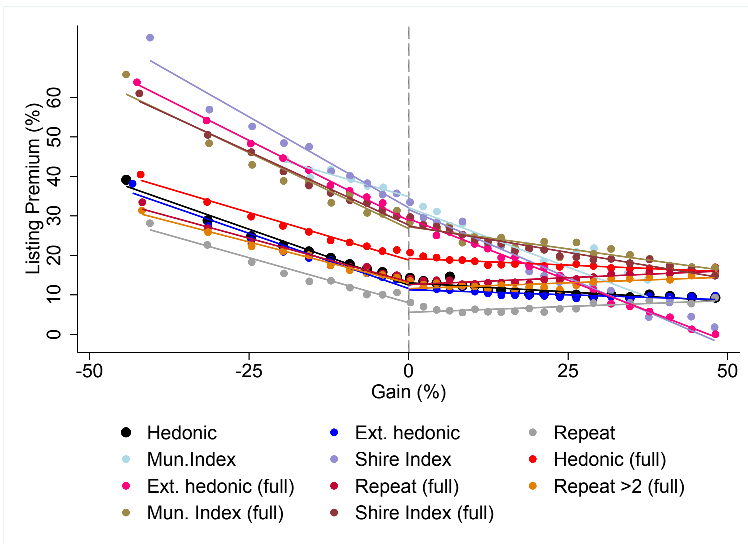


Figure A.14
Robustness to Alternative Models of $\widehat{P}$

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* 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



Panel B

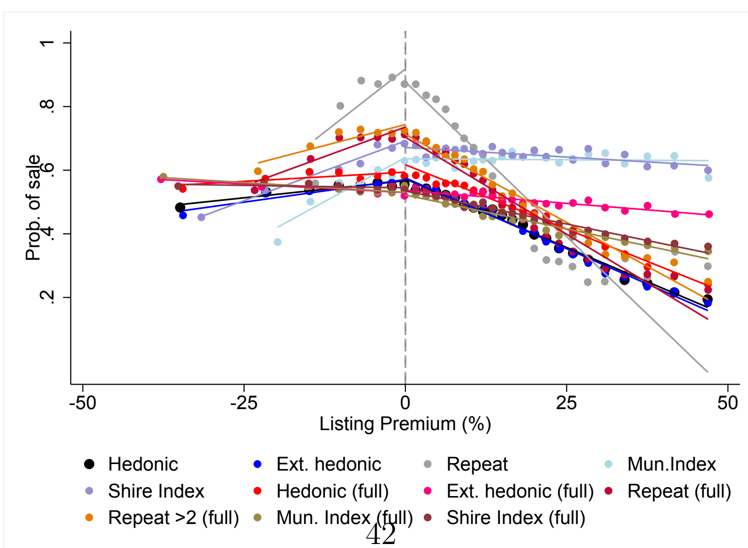
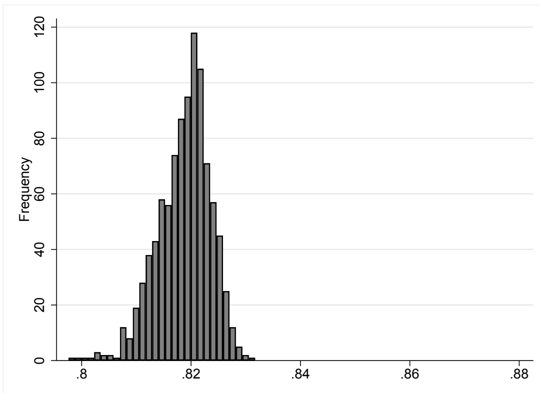


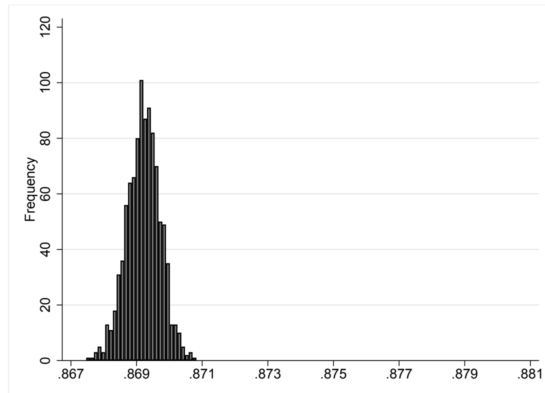
Figure A.15
Distribution of R^2s from out-of-sample estimation of the hedonic model

These figures show the distribution of R^2 from 1000 regressions of realized price on out-of-sample-predicted hedonic prices. Notice the different range of the x-axis in Panel (a) relative to the other panels. In addition, Panel (a) is cropped at 0.79, but in 29 of the regressions, the R^2 was less and in most cases very close to 0, reflecting the vulnerability of a 1 percent sample

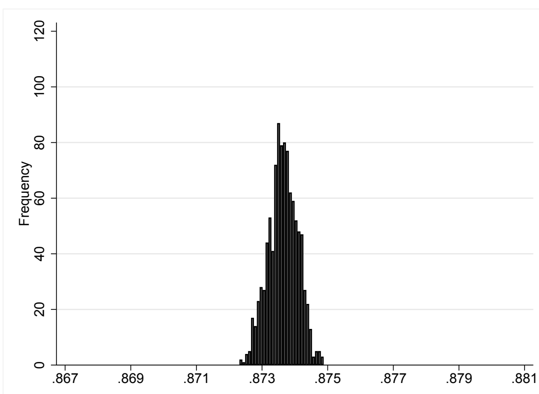
(a) 1 percent sample



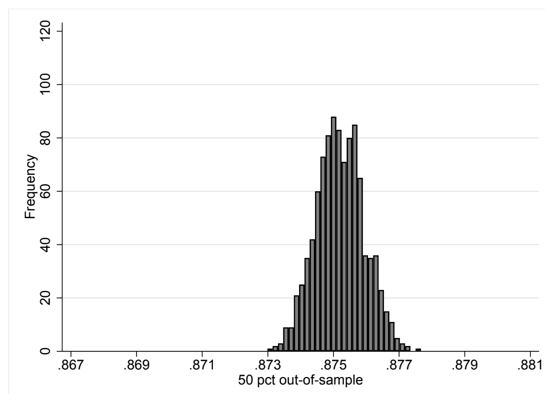
(b) 10 percent sample



(c) 25 percent sample



(d) 50 percent sample



(e) 75 percent sample

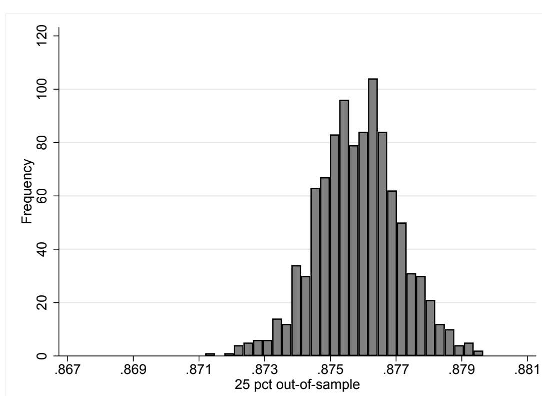
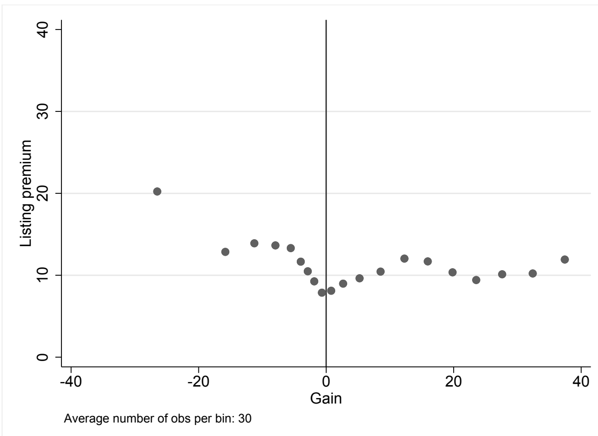


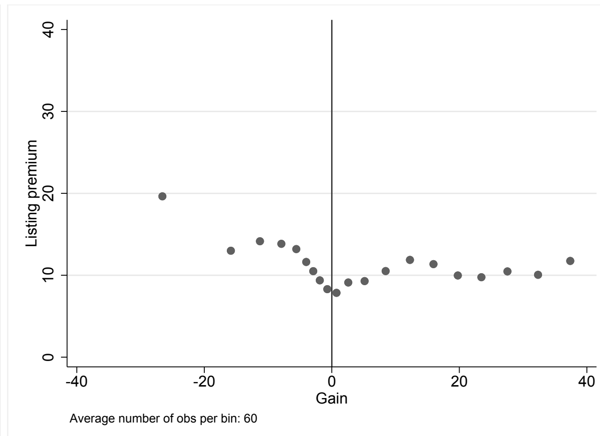
Figure A.16
Listing premium vs. gain at home equity around 20 - out-of-sample predictions

Home equity is between 18 and 22 percent. Notice that the samples are only fractions of sold houses and the sellers have positive mortgage. Bins are averages of 1000 iterations

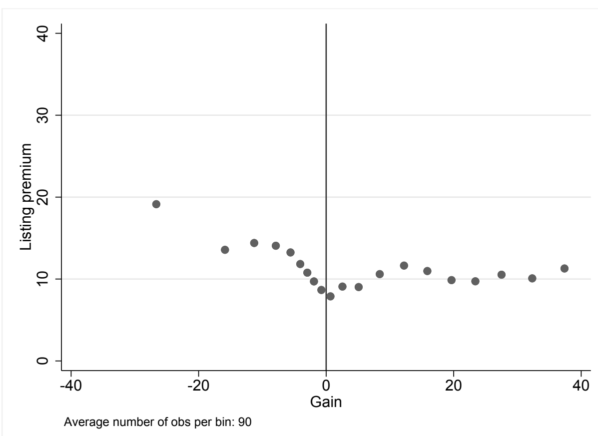
(a) 25 percent out of sample



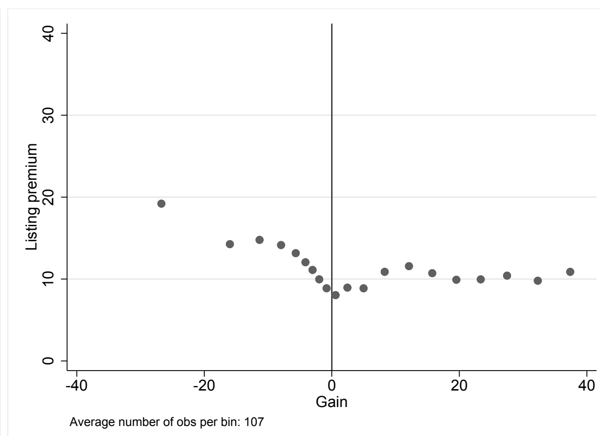
(b) 50 percent out of sample



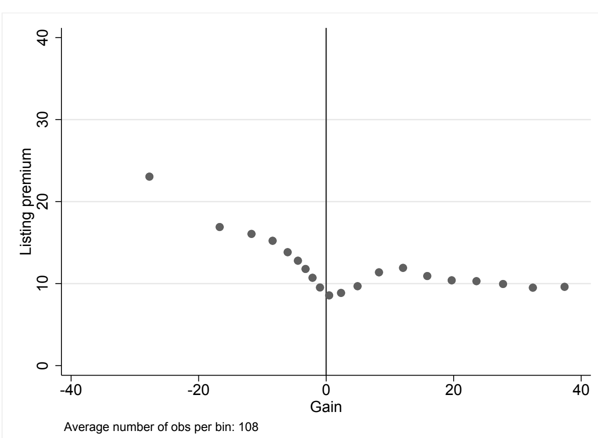
(c) 75 percent out of sample



(d) 90 percent out of sample



(e) 1 percent out of sample



(f) Main data, only sales

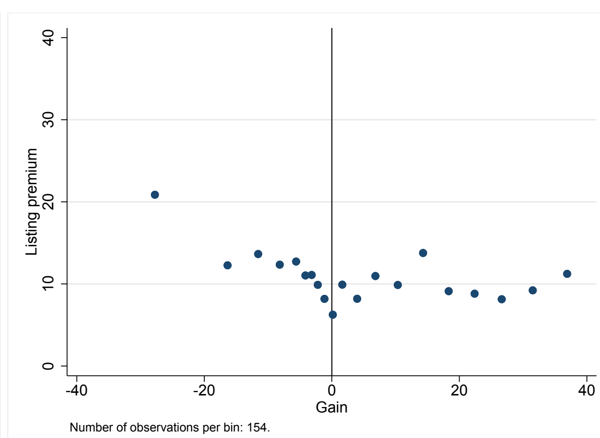
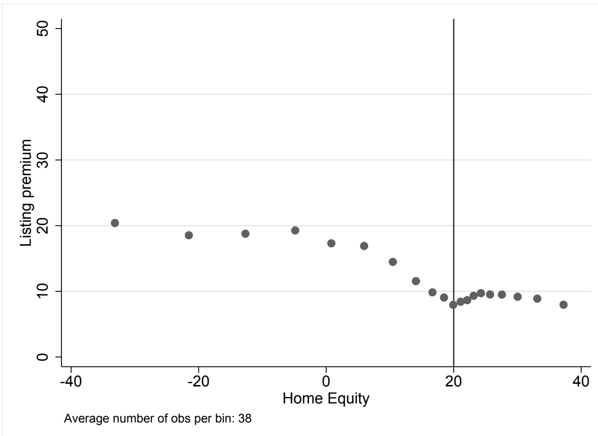


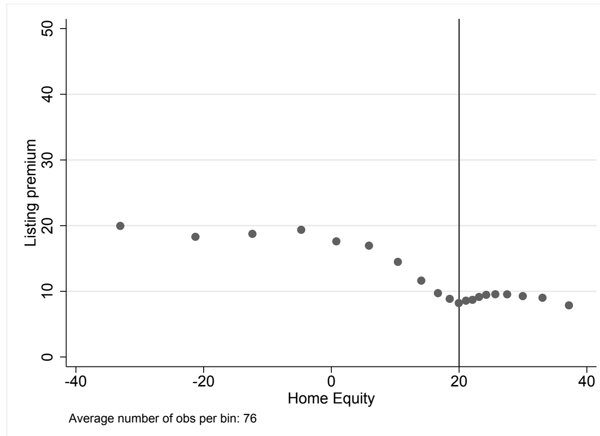
Figure A.17
Listing premium vs. home equity at gain around 0 - out-of-sample predictions

Gain is between -2 and 2 percent. Notice that the samples are only fractions of sold houses and the sellers have positive mortgage. Bins are averages of 1000 iterations

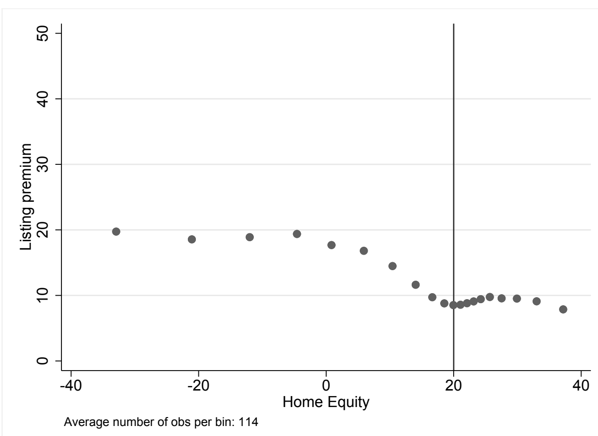
(a) 25 percent out of sample



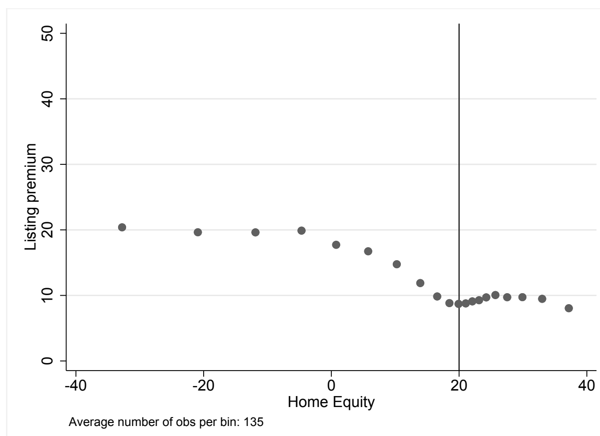
(b) 50 percent out of sample



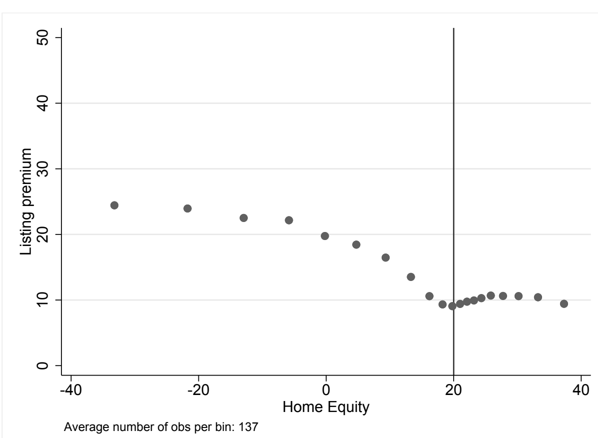
(c) 75 percent out of sample



(d) 90 percent out of sample



(e) 99 percent out of sample



(f) Main data, only sales

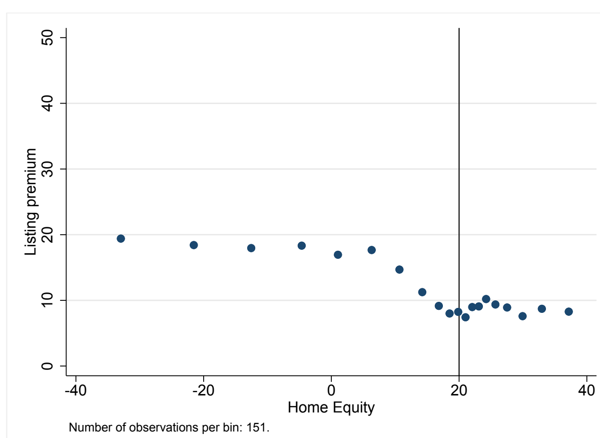
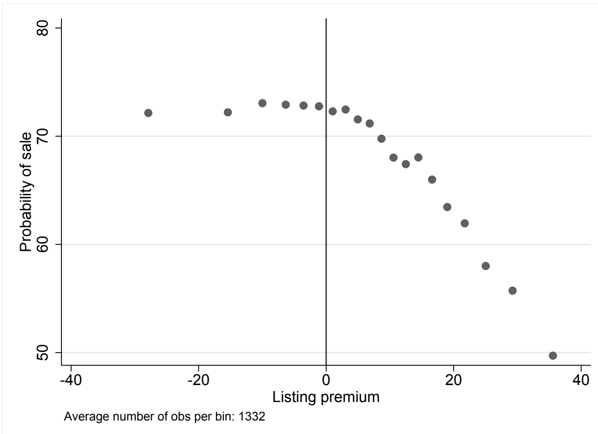


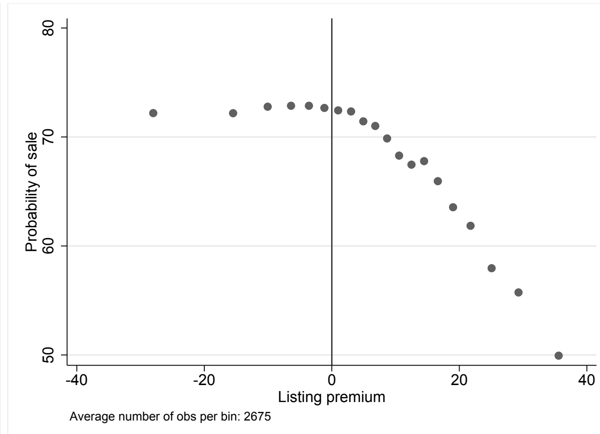
Figure A.18
Prob. of sale vs. listing premium - out-of-sample predictions

Notice that the samples are only fractions of sold houses and the sellers have positive mortgage. Bins are averages of 1000 iterations. Probability of sales refers to the probability of sale within 6 months.

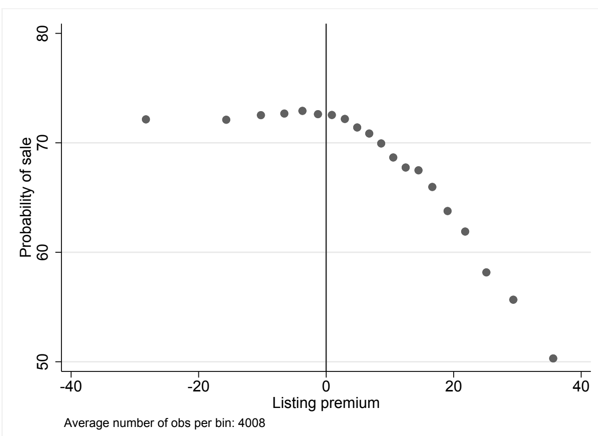
(a) 25 percent out of sample



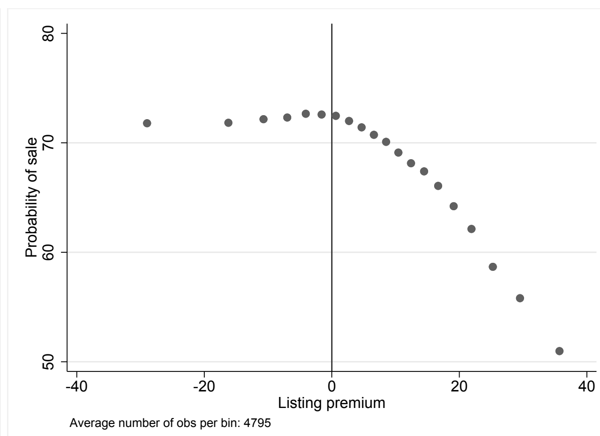
(b) 50 percent out of sample



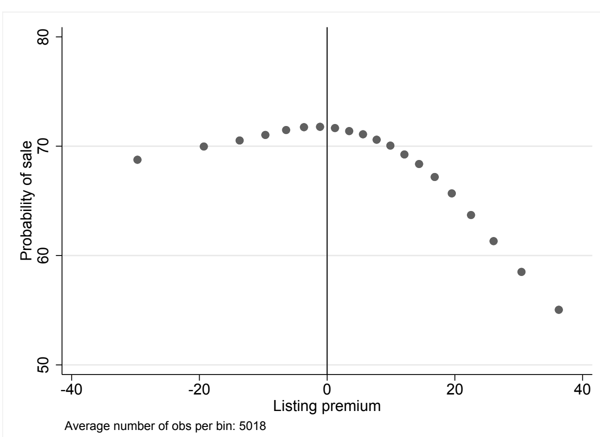
(c) 75 percent out of sample



(d) 90 percent out of sample



(e) 99 percent out of sample



(f) Main data, only sales

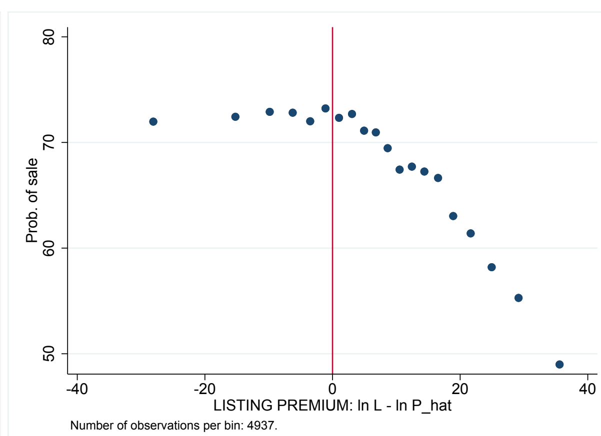
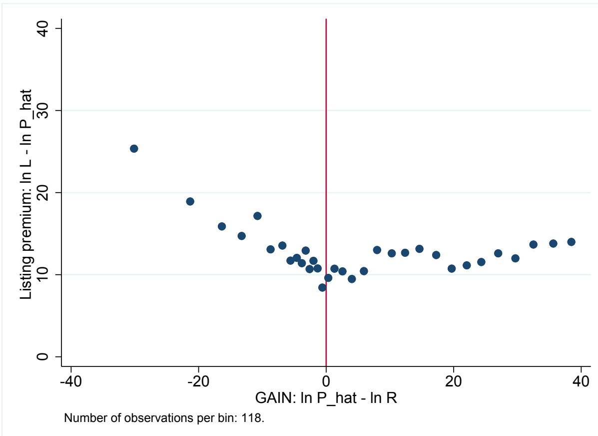


Figure A.19
Listing premium vs. gain at home equity around 20 %

Home equity is between 18 and 22 percent. Panel (b) is restricted to 2010-2012, since this is when tax-assessment is most accurate.

(a) Standard hedonic model



(b) Tax assessed value

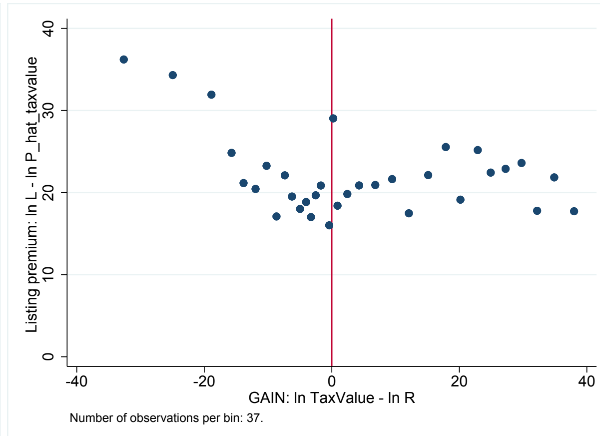
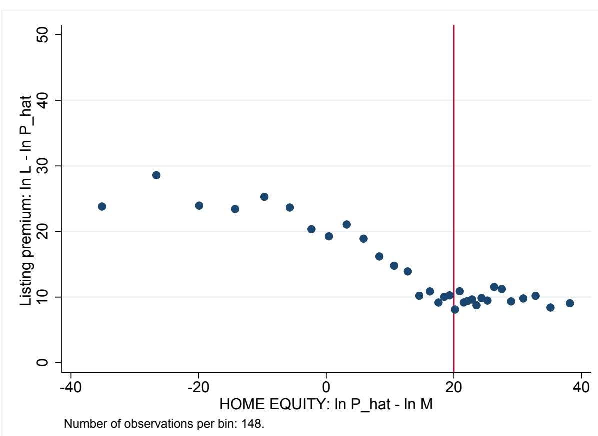


Figure A.20
Listing premium vs. home equity at gain around 0

Gain is between -2 and 2 percent. Panel (b) is restricted to 2010-2012, since this is when tax-assessment is most accurate.

(a) Standard hedonic model



(b) Tax assessed value

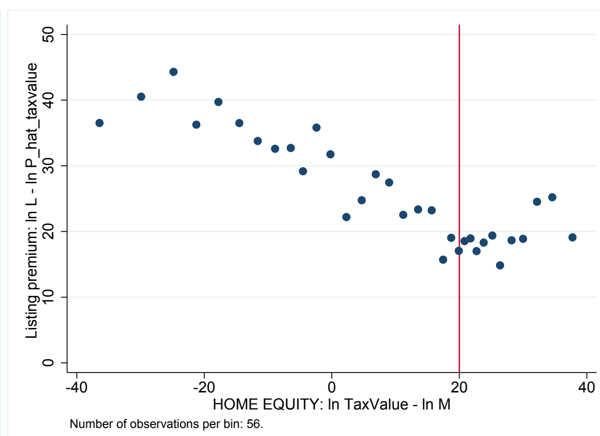
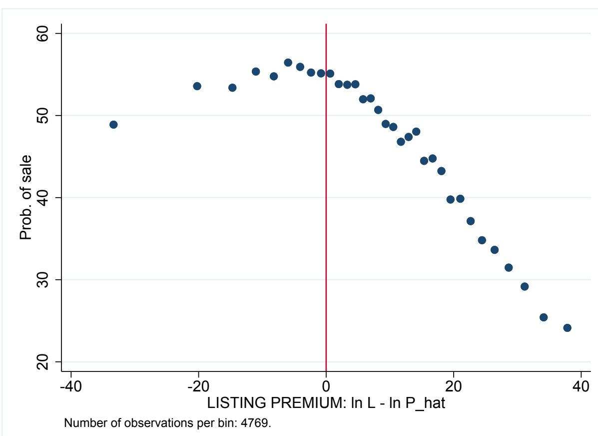


Figure A.21
Probablity of sale vs. listing premium

Panel (b) is restricted to 2010-2012, since this is when tax-assessment is most accurate.

(a) Standard hedonic model



(b) Tax assessed value

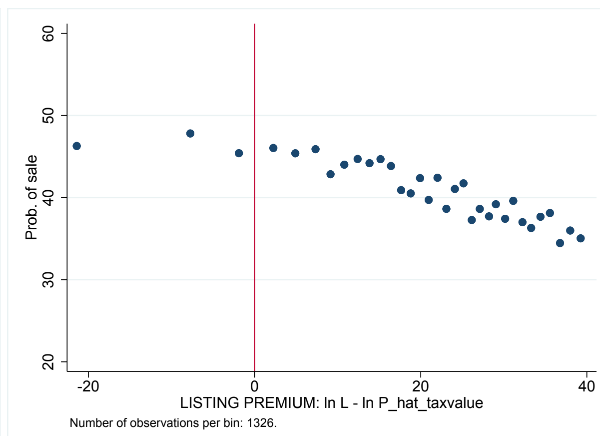
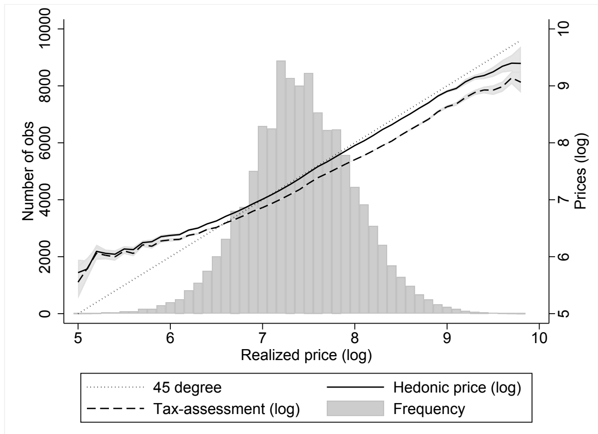


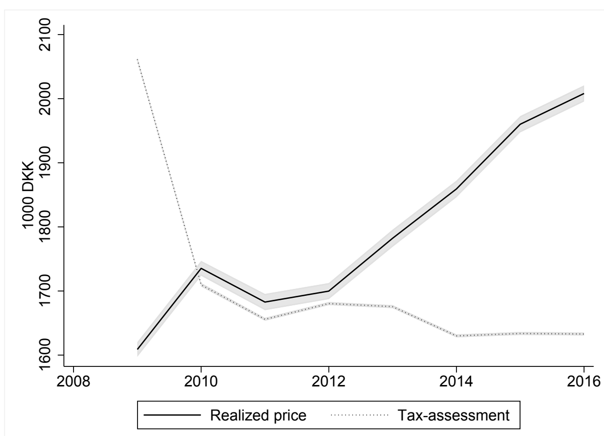
Figure A.22
Quality of the tax-assessed value

Panel (a) shows the tax-assesment relative to the realised sales price as well as the distribution of prices. Panel (b) compares the tax-assessed value to the realised sales prices over our sample period. Panel (c) expands the time period. Data in (a) is the final data of mortgage-holding households from 2009 to 2016. Data in (b) and (c) applies less filters, because they cannot be applied in all years. E.g does it also contains no-mortgage households, since we do not have mortgage data prior to 2009.

(a) Hedonic price vs. tax-assesment



(b) Realized price and tax-assesment by year.



(c) Realized price and tax-assesment by year.

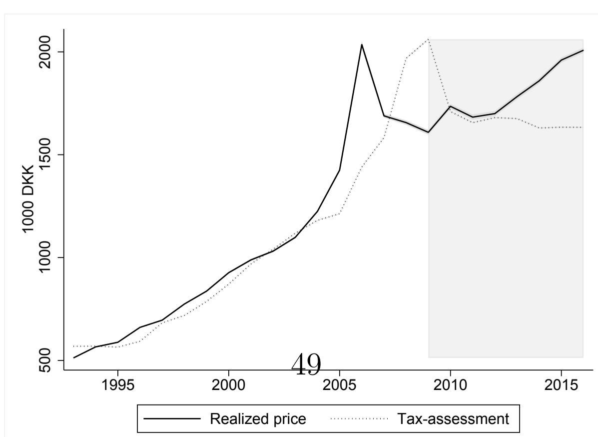
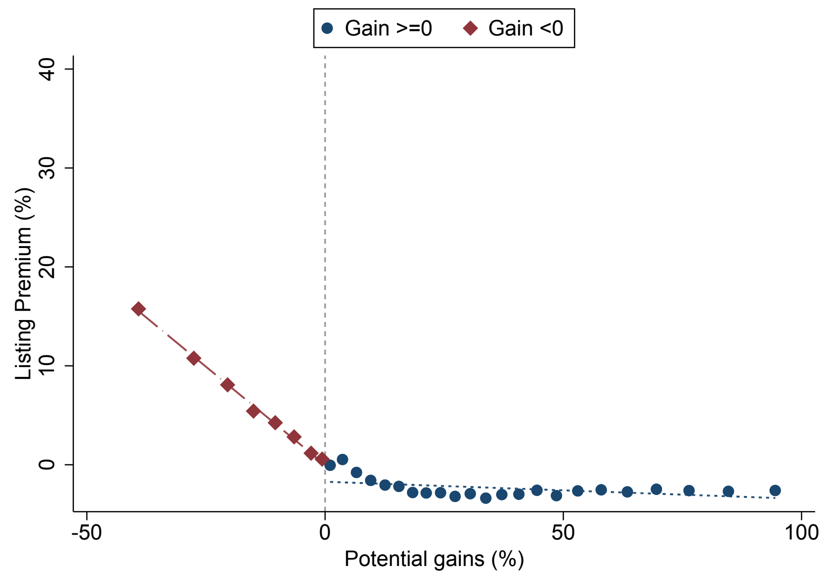


Figure A.23

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.



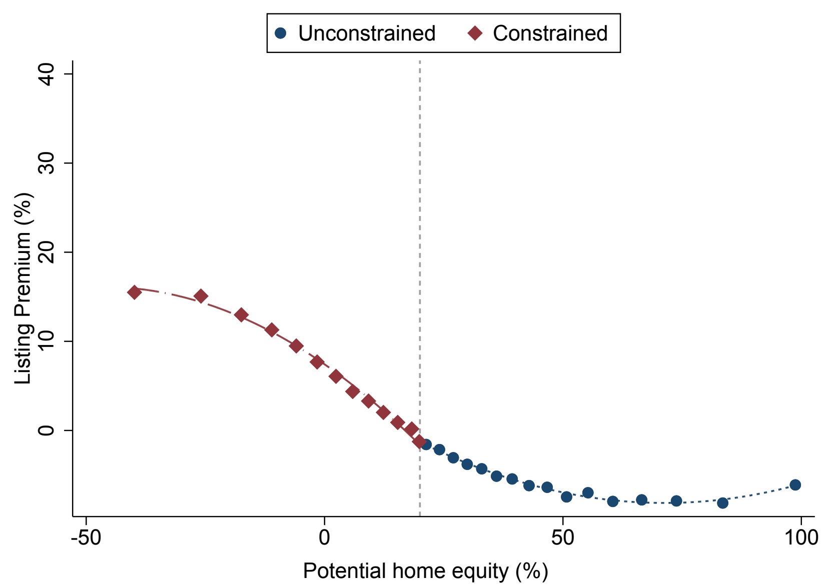


Figure A.24

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]

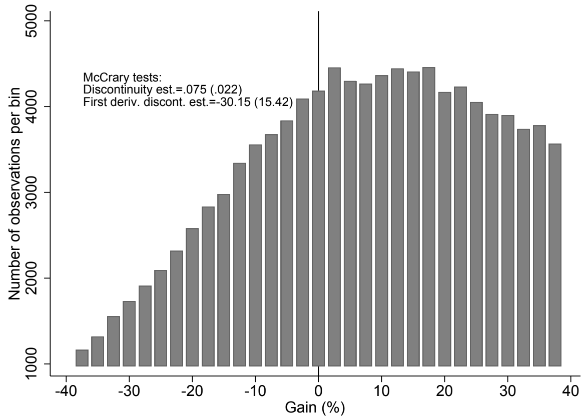


Figure A.25

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.

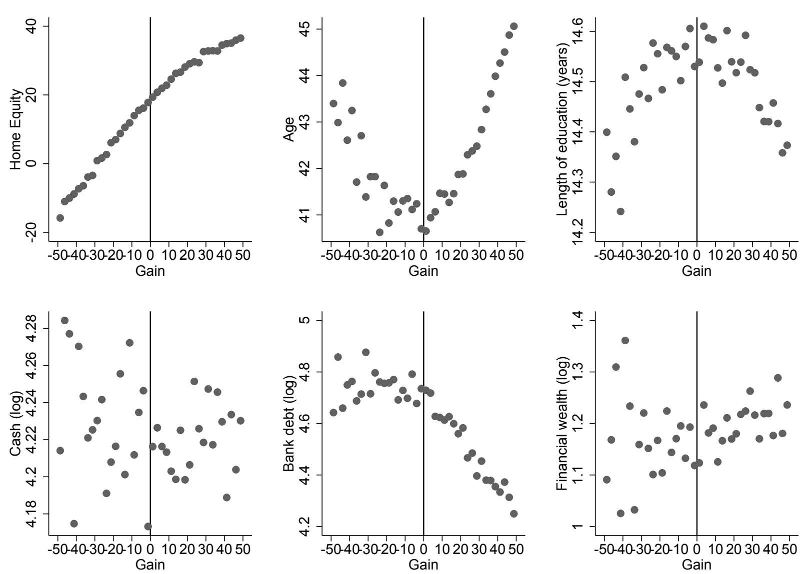


Figure A.26

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).

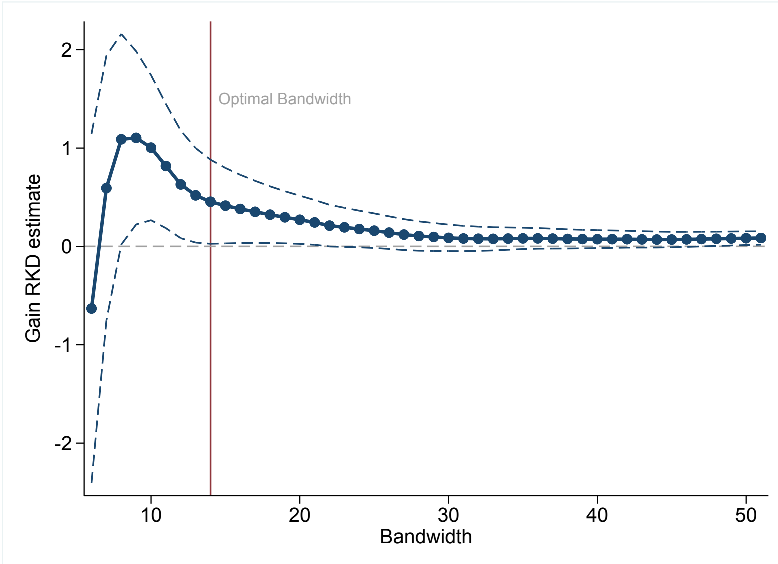


Figure A.27

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).

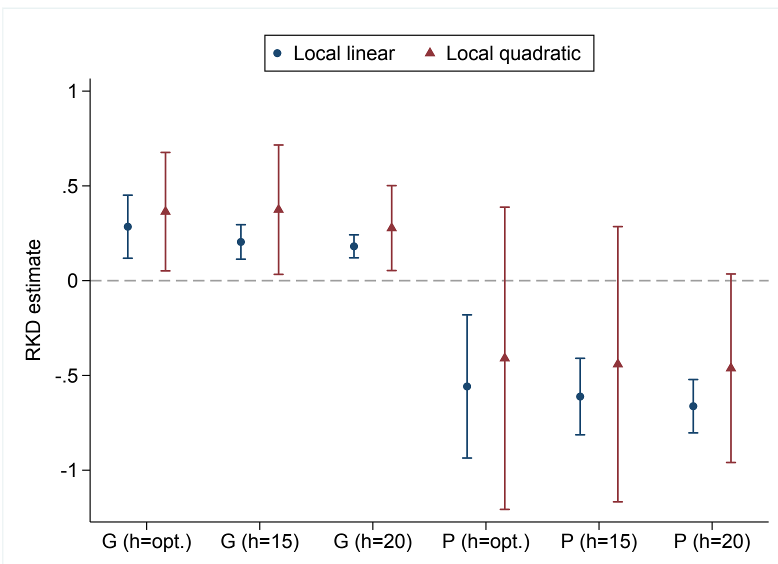


Figure A.28
Non-Mortgage Sample

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 $\widehat{G} \in [-50, 50]$.

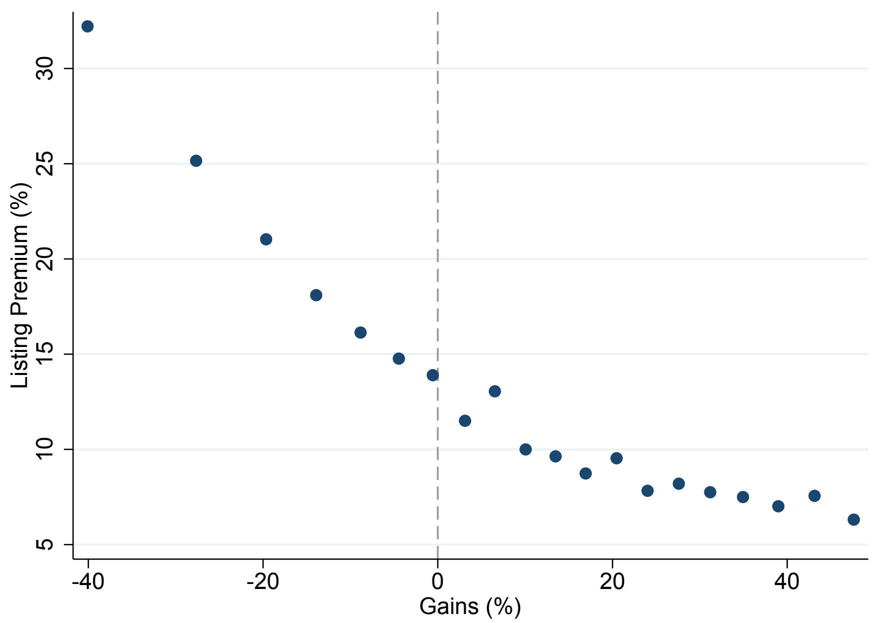


Figure A.29

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.

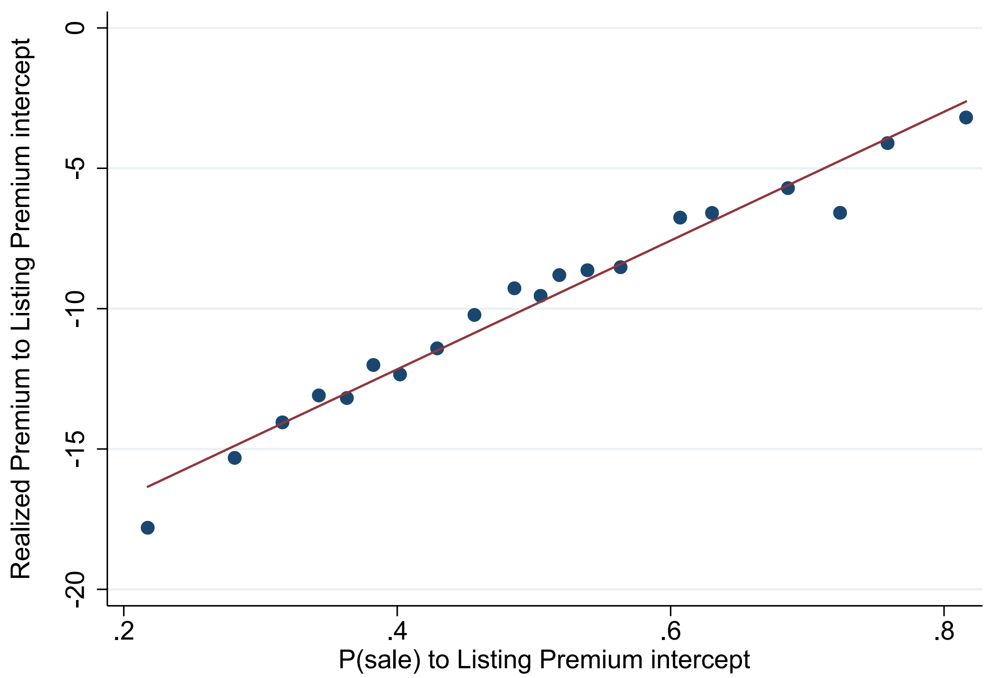


Figure A.30

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 ($\widehat{H}$), based on a sub-sample of the data for which we have information on the next house purchase price and mortgage value ($N = 14,440$).

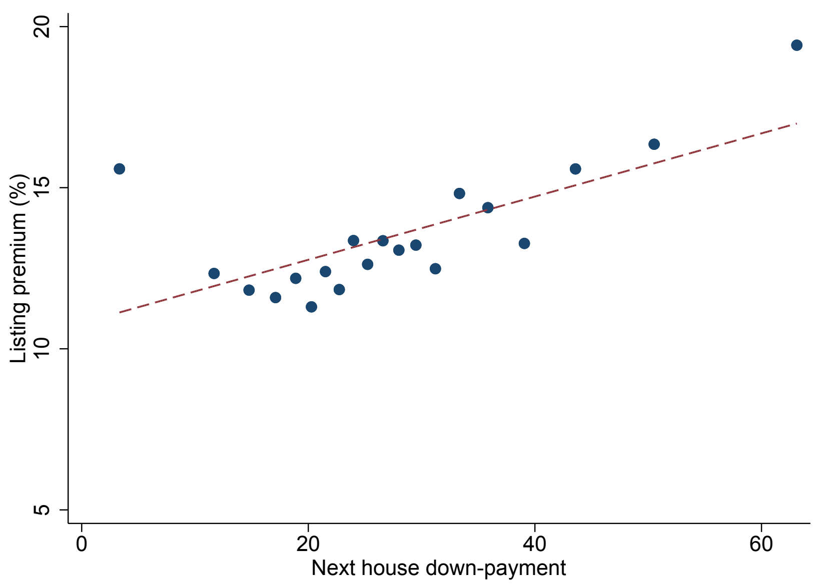


Figure A.31
Current and Next House Price

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 ($N = 14,440$).

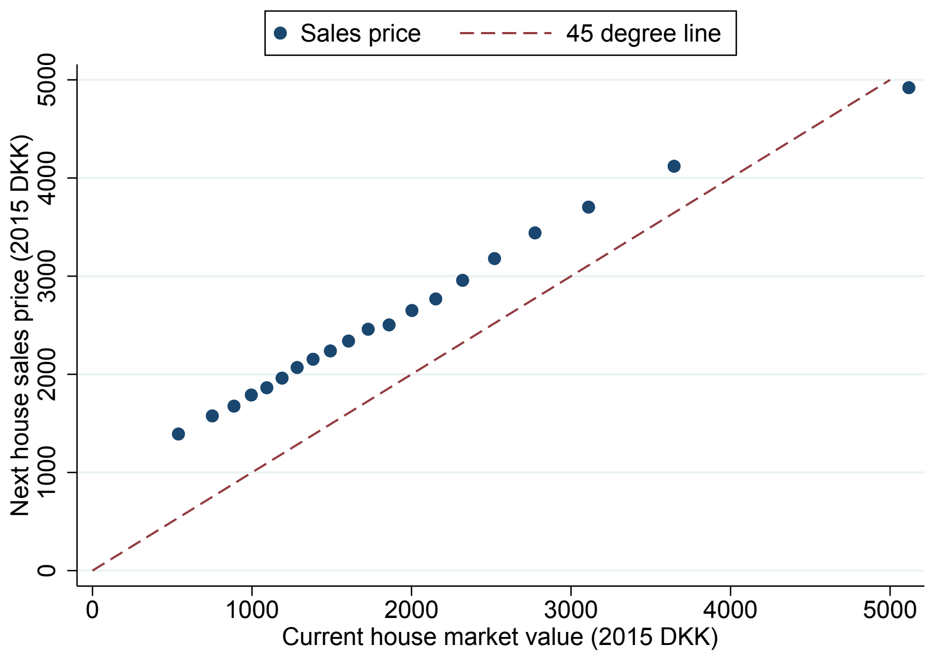


Figure A.32

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.

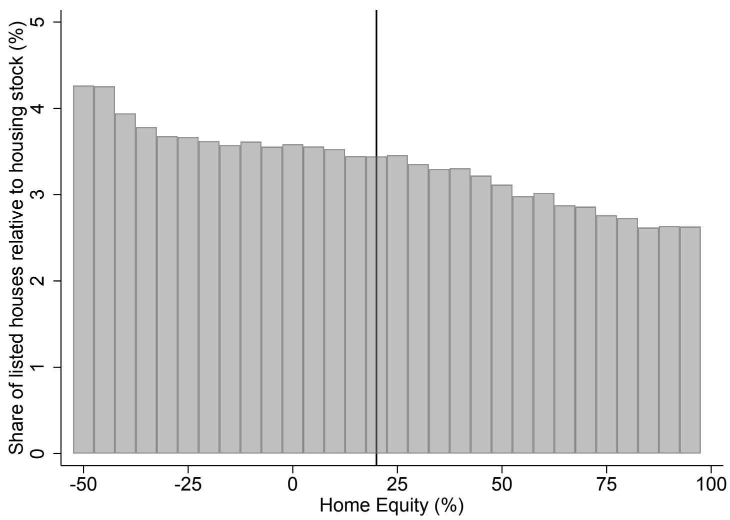


Figure A.33

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.34
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 $\widehat{G} < 0$.

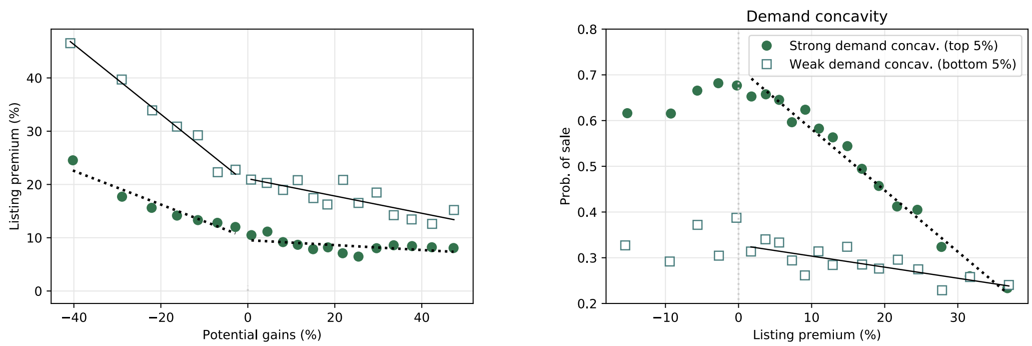


Figure A.35
Model fit

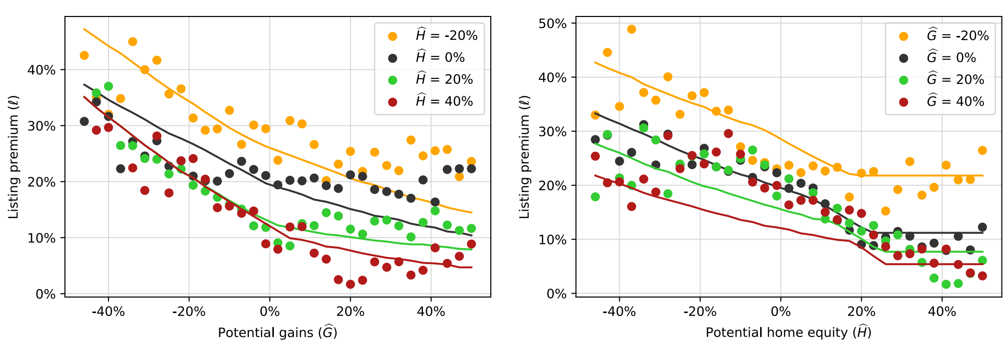


Figure A.36

Bunching around realized gains of zero (polynomial counterfactual)

The figure reports binned frequencies of observations (in 1 percentage point steps) for different levels of realized gains (G). The dotted line shows the counterfactual distribution using a 7th-order polynomial fit, with the excluded range of {-1%,1%}.

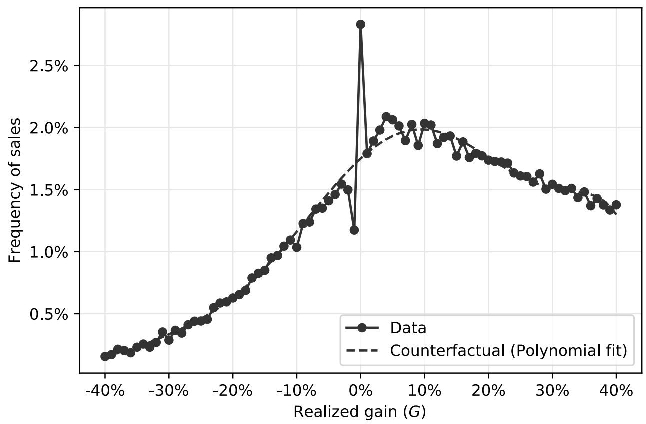


Figure A.37

Incidence of round numbers by rounding multiple

This figure shows the share of listed (sold) houses with a price at a given round number.

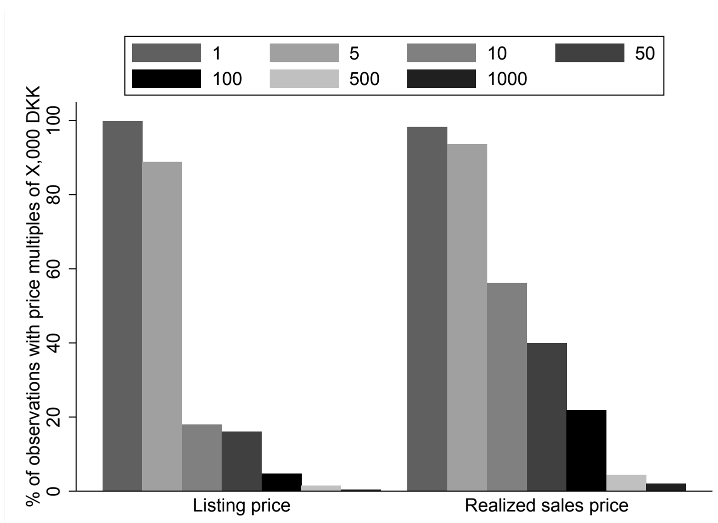


Figure A.38

Bunching robustness: excluding sales at rounded prices (10,000 and 50,000 DKK)

This figure shows robustness for the frequency of sales across realized gains (right-hand panel), against bunching being driven by round sales prices. The frequency is computed without sales that take place at 10,000 and 50,000 DKK, respectively. The blue dots represent the empirical frequency of observations in each 1 percentage point gain bin, and the red line reflects the fitted polynomial counterfactual model.

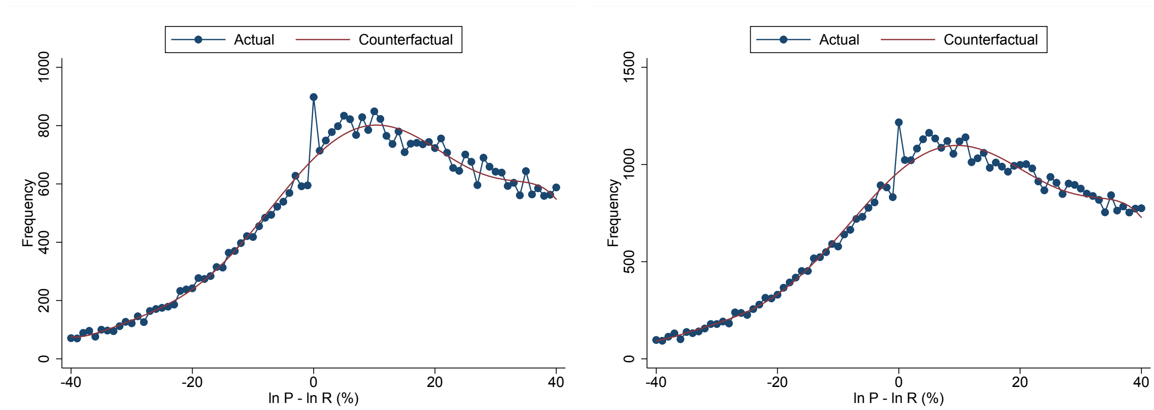


Figure A.39

Bunching robustness: excluding sales at rounded prices (100,000 and 500,000 DKK)

This figure shows robustness for the frequency of sales across realized gains (right-hand panel), against bunching being driven by round sales prices. The frequency is computed without sales that take place at 100,000 and 500,000 DKK, respectively. The blue dots represent the empirical frequency of observations in each 1 percentage point gain bin, and the red line reflects the fitted polynomial counterfactual model.

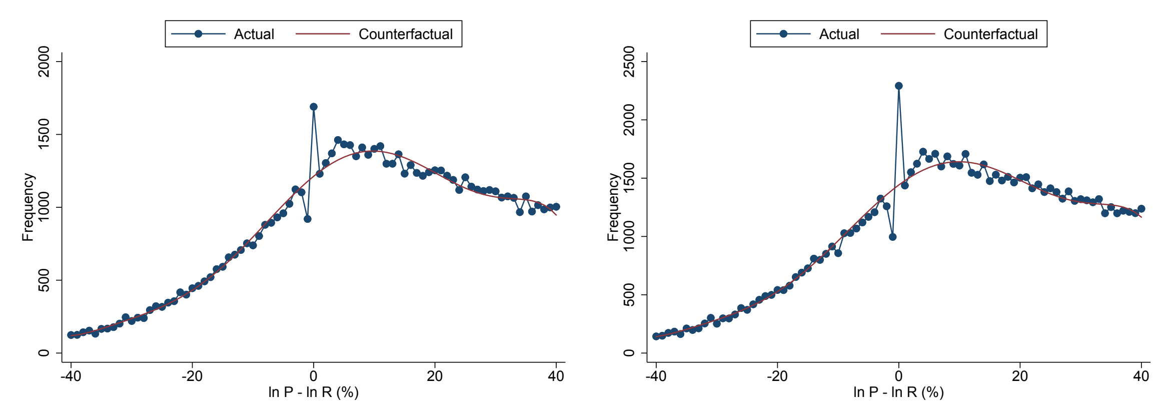


Figure A.40

Bunching robustness: previous sales price - gains at realized price

This figure shows robustness for the frequency of sales across gains at the realized price, by splitting the sample by quintiles of the previous sales price. The blue dots represent the empirical frequency of observations in each 1 percentage point gain bin, and the red line reflects the fitted polynomial counterfactual model.

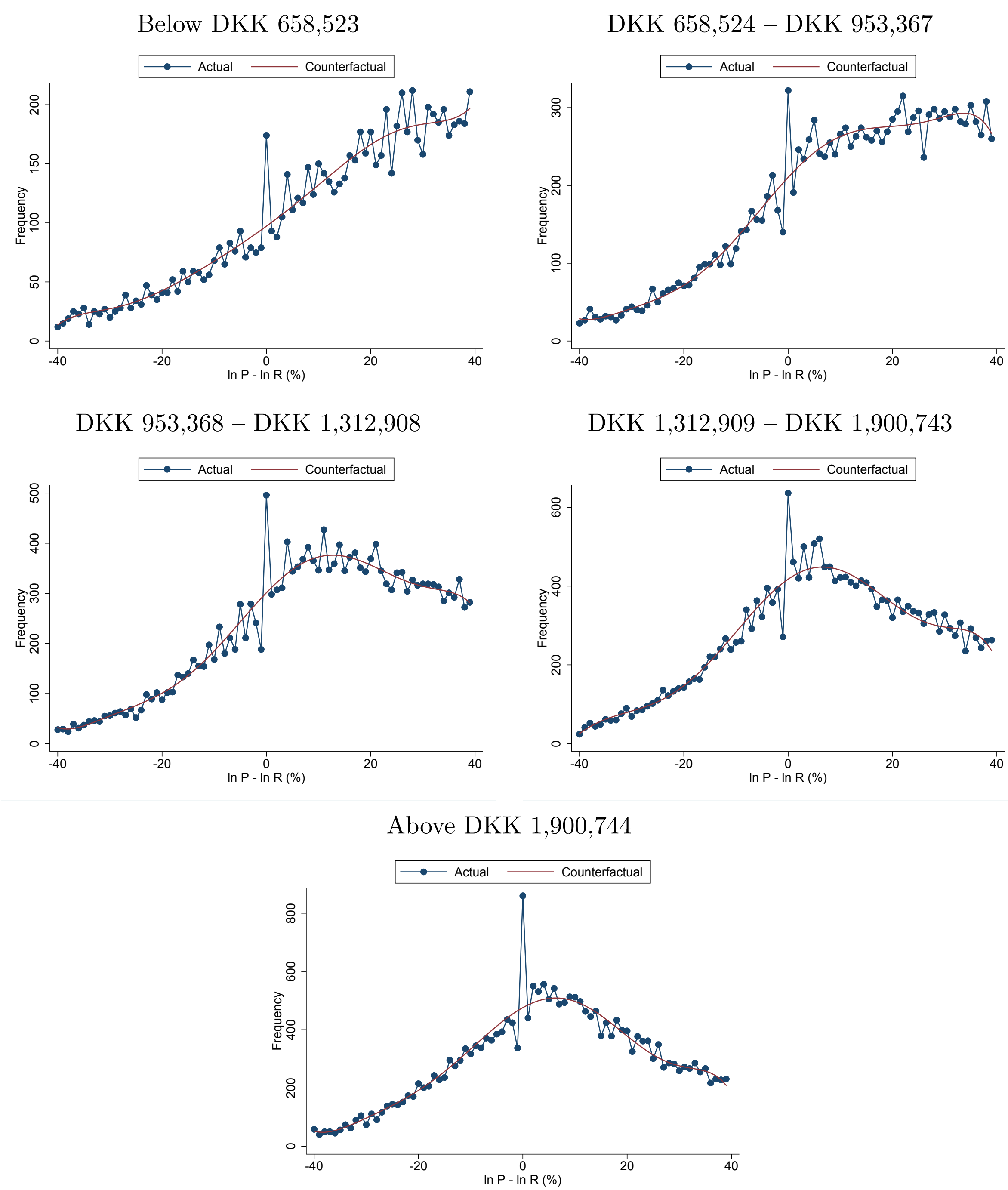
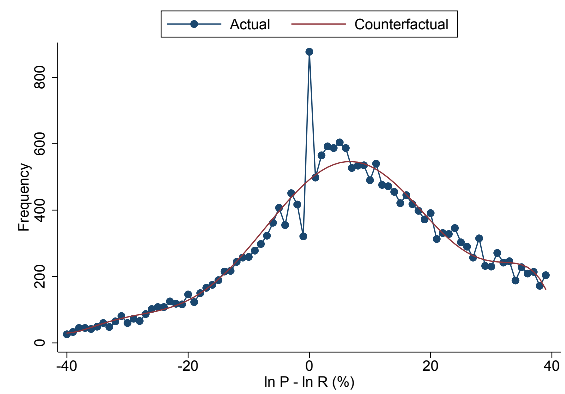


Figure A.41

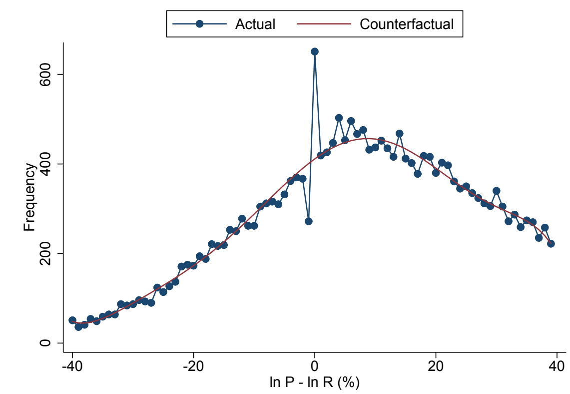
Bunching robustness: holding period - gains at realized price

This figure shows robustness for the frequency of sales across gains at the realized price, by splitting the sample by quintiles of the months since last sale (holding period). The blue dots represent the empirical frequency of observations in each 1 percentage point gain bin, and the red line reflects the fitted polynomial counterfactual model.

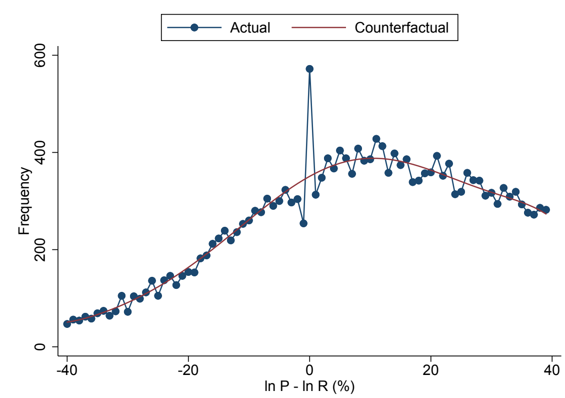
Below 3 years



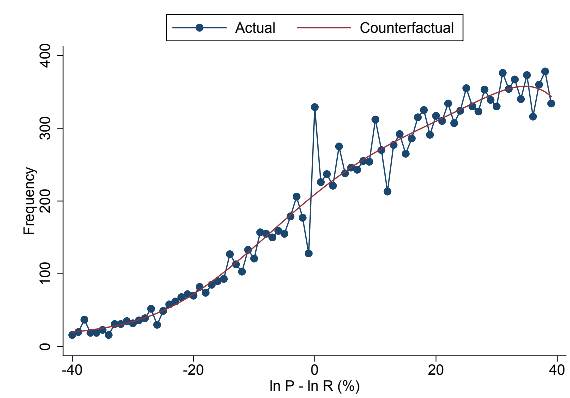
3–6 years



6–9 years



9–12 years



Above 12 years

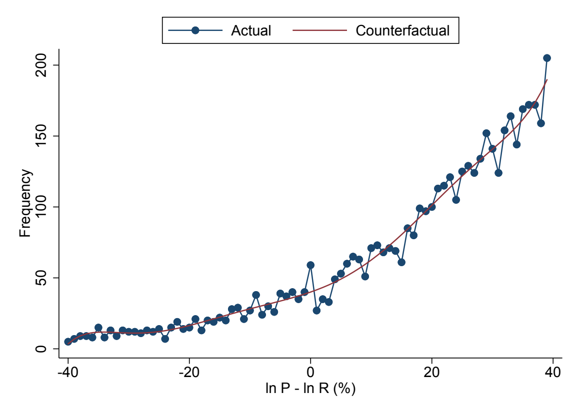
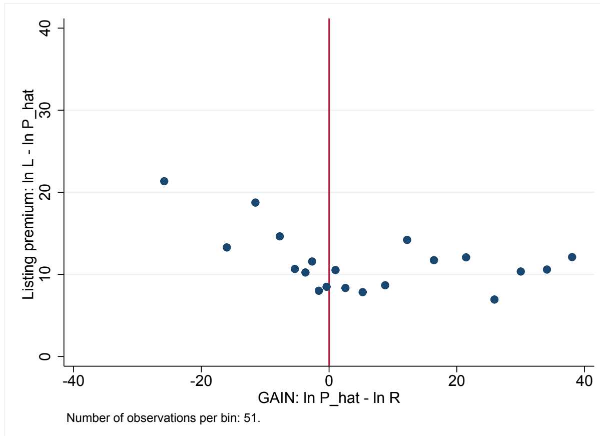


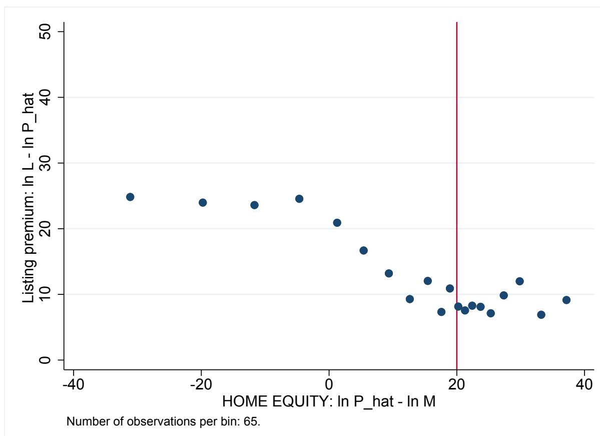
Figure A.42
Main relationships for matched sample

The sample consists of properties from our final sample matched to properties that have been excluded because they where last traded before 1992.

(a) Listing premium - gain



(b) Listing premium - home equity



(c) Demand concavity

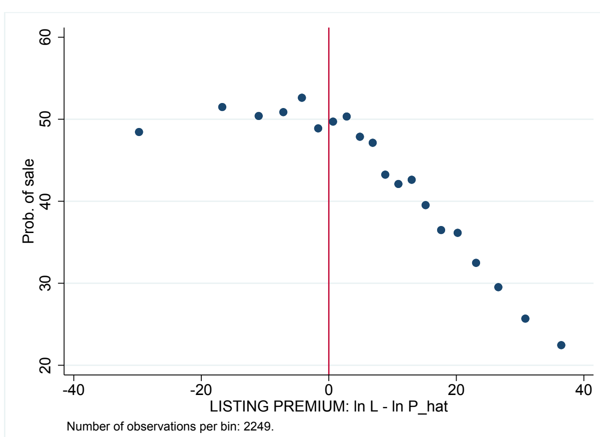


Figure A.43

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 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.

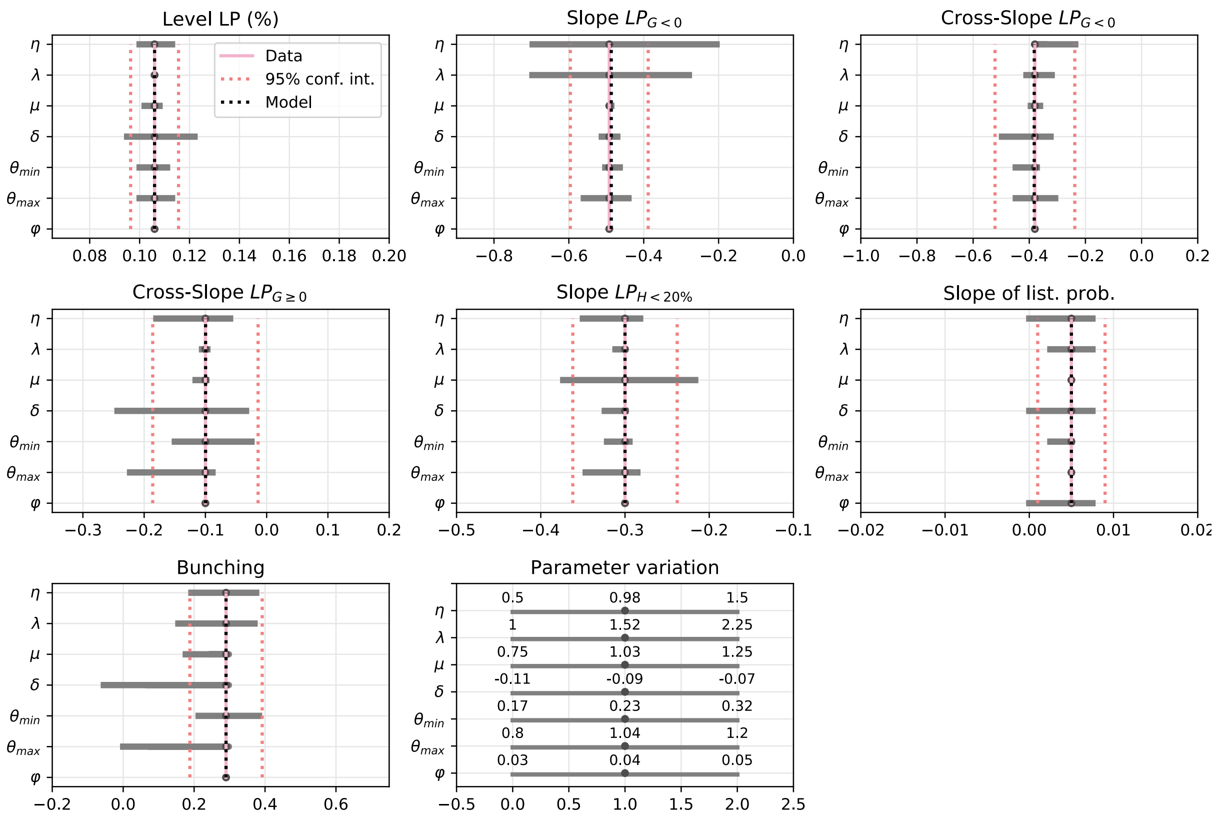


Figure A.44

Gains vs. home equity: Bunching

The figure reports binned average values (in 3% steps) for the observed excess bunching of sales along levels of realized gains and home equity. We calculate the measure of excess bunching as the difference between the frequency of sales in a given bin of *realized* gains and home equity, and the the frequency of sales in the same bin of *potential* gains and home equity. The dotted lines show the binned values for two cross-sections, where we condition on a home equity level of 20%, and a level of gains of 0%, respectively.

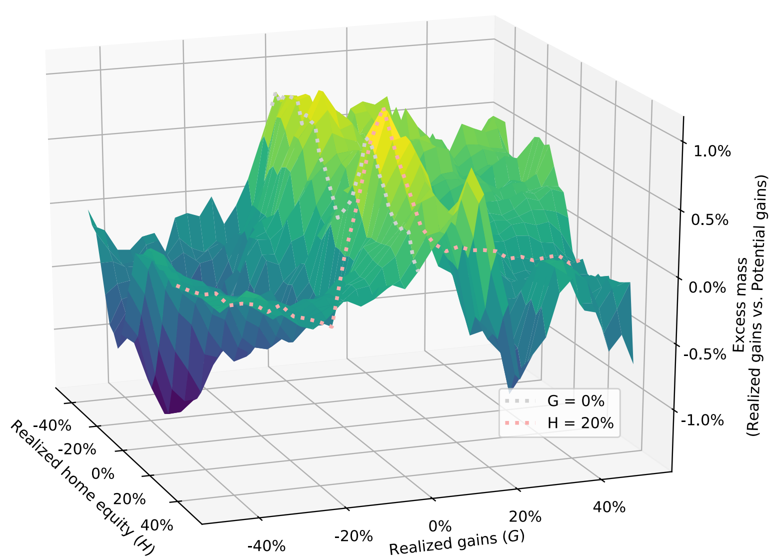


Table A.1

Construction of Main Dataset

This table describes the cleaning and sample selection process from the raw listings data to the final matched data.

All listings of owner-occupied housing[a]	**614,798**
Unmatched in registers[b]	-107,582
	507,216
Cleaning	
No reference price[e]	-144,974
Owner ID not uniquely determined[d]	-71,883
Non-household buyer	-10,175
Foreclosures	-6,310
Extreme price	-5,495
Owner ID not found[c]	-3,982
Missing lot size	-2,814
Error in listing or purchase date[f]	-1,925
Intra-family sale or other special circumstances	-1,733
No listing price	-882
Missing hedonic characteristics	-8
	257,035
Sample selection	
Summer house	-24,138
Professional investor[g]	-18,389
Final data	**214,508**
Of which with a mortgage	173,065
Of which without a mortgage	41,443

[a] *Excluding listings of cooperative housing. We identify cooperative housing using the ownership register, which between version 1 and 2 was updated for 2016 by Statistics Denmark*

[b] *Reasons could be misreported addresses or non-ordinary owner-occupied housing.*

[c] *No owner ID found in registers.*

[d] *E.g. properties with several owners from different households.*

[e] *Purchased before 1992.*

[f] *Listing date is before purchase date.*

[g] *Seller owns more than 3 properties.*

Table A.2
Observables by inclusion in sample

This table show means and standard deviations of properties in our final sample (Column 1), and the corresponding properties that have been sorted out because we do not know the purchase price (Column 2). In addition Column 3 shows characteristics of a matched sample of listed properties. Column 3 properties are sampled from Column 1 and matched to properties in Column 2 using exact matching on listing year, apartment, lot size decile, distance to city decile, and nearest neighbor matching on tax-assessed value. Matching is done with replacement. All properties fulfill all other restrictions, presented in Table A.1

	Sample mean/sd	No found purchase price mean/sd	Matched sample mean/sd
Apartment	0.29	0.11	0.11
	0.45	0.31	0.31
Lot Size	655.47	11984.37	1535.40
	1958.02	53599.92	5368.16
Interior Size	118.81	145.85	144.85
	45.93	50.24	51.80
Rooms	4.08	4.81	4.77
	1.52	1.54	1.54
Bathrooms	1.04	1.06	1.06
	0.22	0.26	0.26
Showers	1.01	1.03	1.02
	0.18	0.21	0.20
Unoccupied	0.06	0.09	0.06
	0.24	0.28	0.24
BuildingAge	63.54	57.55	63.08
	36.98	44.63	42.39
Historic	0.01	0.00	0.01
	0.09	0.07	0.08
Rural	0.11	0.26	0.22
	0.31	0.44	0.42
Distance to city (km)	36.41	43.42	43.32
	34.09	32.75	32.57
Tax assessment	1562.68	1887.73	1826.33
	990.17	1248.42	1057.24
Listing price	2131.28	2456.47	2450.25
	1473.34	1678.92	1599.68
Hedonic price	1896.51		2158.73
	1205.91		1248.11
Holding length	7.67		8.20
	5.07		5.35
Observations	173048	80842	80419

Table A.3

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	214,494	214,405	26,680	-	42,155	39,796
Estimation sample	150,890	150,890	25,386	-	150,889	147,866
R^2	0.8769	0.8828	0.9743	-	0.3544	0.5139
			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	214,494	194,653	185,141	104,053	203,500	199,568
Estimation sample	1,683,001	884,256	1,176,459	661,213	1,797,081	1,795,264
R^2	0.8410	0.6388	0.9192	0.9025	0.4230	0.5341

Table A.4
Out-of-sample test of hedonic model

This table show mean R^2 from 1000 regressions of realized price on out-of-sample-predicted hedonic prices.

	(1)	
	Mean	
50 pct out-of-sample	0.875	(0.0000233)
25 pct out-of-sample	0.876	(0.0000402)
100 pct in-sample	0.878	(.)
Observations	1000	

Standard errors in parentheses

Table A.5
Out-of-sample test of hedonic model w/o tax-assessed value

This table show mean R^2 from 1000 regressions of realized price on out-of-sample-predicted hedonic prices.

	(1)	
	Mean	
50 pct out-of-sample	0.765	(0.0000378)
25 pct out-of-sample	0.766	(0.0000666)
100 pct in-sample	0.769	(.)
Observations	1000	

Standard errors in parentheses

Table A.6

Decomposition of the hedonic model: R^2

This table shows R^2 from different hedonic models: *YearMuni* models realized price using only municipality-year fixed effects. *Location* controls for rural area and distance to city. *Size* controls for interior size, lot size, number of bathrooms and showers. *HistUnoccu* constrols for historic buildings and unoccupied properties. *BuildingAge* for the age of the building. *TaxValue3rd* includes ln(TaxValue) as a third degree polynomial, just like it appears in the baseline hedonic model. *TaxValue* has only ln(TaxValue) in a linear form. Column 1 show a simple model using only the mentioned controls. Column 2 adds controls one by one.

	Simple	Accumulative
YearMuni	.477	.477
Location	.217	.512
Size	.263	.759
HistUnoccu	.041	.764
BuildingAge	.010	.768
TaxValue3rd	.800	.876
TaxValue	.792	.

Table A.7

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.8

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	0.538***	0.442***	0.520***	0.355***	0.557***	0.464***
	(0.016)	(0.016)	(0.026)	(0.026)	(0.016)	(0.016)
LOSS (squared)			0.050	0.238***		
			(0.056)	(0.056)		
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	0.996***	0.993***	0.996***	0.993***	0.996***	0.994***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Market index at listing	0.993***	0.990***	0.993***	0.990***		
	(0.003)	(0.003)	(0.003)	(0.003)		
Residual from last sales price		-0.094***		-0.095***		-0.091***
		(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	0.390***	0.413***	0.391***	0.416***	76.059***	75.883***
	(0.025)	(0.026)	(0.025)	(0.026)	(0.211)	(0.211)
Year-Quarter FE					✓	✓
Observations	173065	157396	173065	157396	173065	157396
R^2	0.883	0.887	0.883	0.887	0.886	0.890

Table A.9

Listing price decomposition

This table shows the simple regression of log listing prices on hedonic price and previous purchase price.

	(1) $\ln L$
$\widehat{\ln P}$	0.897***
	(0.002)
$\ln R$	0.082***
	(0.002)
Observations	214508
R^2	0.87

Table A.10

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.11
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.12

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)
(mean) concav2	-0.223***	-0.333***	-0.338***
	(0.051)	(0.113)	(0.106)
Household controls	✓	✓	✓
Observations	95	95	95
R^2	0.637		
First-stage F-stat	31.268	27.615	27.787
Hansen J-stat (p-val)			0.895

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.13

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	-0.273***	-0.350***	-0.358***
	(0.059)	(0.115)	(0.108)
Household controls	✓	✓	✓
Observations	93	93	93
R^2	0.646		
First-stage F-stat	31.724	28.618	28.832
Hansen J-stat (p-val)			0.830

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.14

Alternative Listing Premia: Premia to Average House Prices

This table reports regression results for the relationship of listing premia (ℓ), the premium to the average house price in a given market (Premium to Avg.), and the ℓ 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***		-0.006***	
	(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				-0.005***				-0.005***
				(0.000)				(0.000)
Observations	173004	173004	173004	173004	171724	171724	171724	171724
R^2	0.048	0.006	0.049	0.037	0.049	0.005	0.049	0.040

Table A.15

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 ℓ for $\widehat{G} = 0\%$, Slope of ℓ-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.16

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.