Universität Zürich[UZH]

CUREM – Center for Urban & Real Estate Management

CUREM Working Paper Series

Globalization of Real Estate Network
Working Paper No. 9

Reference Dependence in the Housing Market

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

January 2021

Key words: Housing, Mortgages, Loss Aversion, Reference Dependence, Down-Payment Constraints.
JEL classification: D03, D12, D14, G02, G50, R21

Steffen Andersen: san.fi@cbs.dk
Cristian Badarinza: cristian.badarinza@nus.edu.sg
Lu Liu: l.liu16@imperial.ac.uk
Julie Marx: jma.fi@cbs.dk
Tarun Ramadorai: t.ramadorai@imperial.ac.uk

Reference Dependence in the Housing Market*

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

January 26, 2021

Abstract

We model listing decisions in the housing market, and structurally estimate household preference and constraint parameters using comprehensive Danish data. In the model, reference-dependent and loss-averse sellers optimize expected utility from property sales, subject to down-payment constraints, and internalize how their choices affect final outcomes. In the data, listing prices vary asymmetrically with gains and losses since purchase; this relationship is modulated by regional demand nonlinearity; and final sales bunch at zero nominal gains. We find that the price-volume correlation in housing markets is primarily driven by households' reference dependence and down-payment constraints, and only modestly by loss aversion.

*We thank 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, four anonymous referees, and seminar and conference 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, Harvard University, Carnegie Mellon, NBER Real Estate Summer Institute, European Finance Association, University of Michigan, MIT, and Columbia Business School 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 market volume is significantly lower in falling than in rising housing markets. A leading behavioral explanation for this fact is that households fixate on achieving specific reference sale prices, and list their houses at levels divorced from market realities. This can make housing markets sluggish when house values decline—making list prices unrealistically high—and more active when markets rise. In support of this explanation, influential work by Genesove and Mayer (2001) shows that list prices rise sharply and asymmetrically when sellers face nominal losses relative to the original purchase prices of their homes.

An alternative rational explanation points to the role of down-payment constraints. Household leverage magnifies the impact of declines in collateral value on home equity, compressing the size of houses into which households can move (Stein, 1995; Genesove and Mayer, 1997). Faced with down-payment constraint-imposed adjustments when housing markets fall, households may set higher listing prices and be prepared to tolerate longer times-on-the-market, or decide not to list at all, generating the price-volume correlation.

To what extent do these competing forces drive housing market liquidity? Assessing their relative strength confronts multiple challenges not fully addressed by subsequent literature (see, e.g., Engelhardt, 2003; Einiö, Kaustia and Puttonen, 2008; Anenberg, 2011; Bokhari and Geltner, 2011; Bracke and Tenreyro, 2020).[1] For one, we need a model to rigorously assign magnitudes to different economic forces at play in the data. A plausible model would incorporate additional realistic constraints, such as the fact that optimizing sellers' listing decisions are disciplined by demand-side responses. And lately, the literature has ingeniously used the behavior of quantities (and not just prices) to better pin down reference dependence in field settings (see e.g., Kleven, 2016; Allen et al., 2017; Rees-Jones, 2018), suggesting a wider set of moments for any such model to match.

In this paper, we develop a new model of house selling decisions, which we structurally estimate using both well-established and newly uncovered facts from a large and granular administrative dataset tracking the entire stock of Danish housing, and the universe of Danish listings and housing transactions between 2009 and 2016. When we feed the model with price movements across Danish regions, it predicts volumes, and delivers a price-volume correlation which realistically approximates the data. The model attributes between 60% and 64% of this correlation to the effect of down-payment constraints, between 23% and 30% to reference dependence, i.e., the extent to which households derive

[1]Online appendix Table A.1 documents a number of outstanding challenges mapped to a comprehensive review of the literature.

utility from realizing gains and losses relative to the original purchase price of their homes, and between 5% and 16% to households' loss aversion, i.e., the asymmetric disutility that households incur from realizing losses.[2]

In the model, sellers decide whether to list their houses, as well as the optimal listing price conditional on listing. Seller utility arises from the final sale price, and additively, from any realized gains or losses measured relative to a reference price, which we simply assume to be the original property purchase price.[3] When optimizing, sellers take into account how their decisions affect the probability that the house is finally sold, as well as the achievable final sale price. Importantly, the model incorporates a parameter which captures seller responses to the effect of home equity down-payment constraints.

To develop an understanding of the model's predictions, first consider sellers who derive utility both from the final sale price of the house, as well as (symmetrically) from any realized gains or losses made relative to the reference price. In this case, we find that sellers optimally set listing price "premia," i.e., markups over a measure of "fair" or hedonic property value, that are linear and downward-sloping in the gains that they expect to make. The intuition for this prediction is that sellers compare the potential gain/loss utility they derive in the event of selling the house to their reservation utility from staying put. When faced with a chance of enjoying gains in the event of a successful sale, they list at lower prices to maximize the probability of the transaction going through. Conversely, when sellers face the prospect of losses, they raise listing prices to try to push their utility in the event of a successful sale to a level higher than their reservation utility.

This prediction changes when sellers are assumed to be loss averse, feeling greater disutility from realized losses than the utility they enjoy from realized gains. This assumed asymmetry in preferences translates into an optimal listing premium profile which is also asymmetric, sloping up more sharply when sellers face losses than when they face gains. In the case with loss-averse sellers, the model also predicts sharp bunching of final sales transactions quantities precisely at realized gains of zero relative to the reference price, and a pronounced shift of transactions away from the realized loss domain.

On the face of it, these insights from the model suggest an easy interpretation of the asymmetric listing price behavior originally detected in Genesove and Mayer (2001)—that this is simply a manifestation of asymmetric preferences. However, several key confounds

[2]The small remainder is accounted for by the effects of price movements on the "extensive margin" decision of households to list at all, which is driven both by preferences and constraints. The ranges on the fractions contributed by the different forces correspond to different underlying assumptions on our structural model.

[3]Throughout the paper, we maintain the assumption that the reference price is the nominal purchase price of the property. We later show evidence to support the validity of this assumption.

can interfere, breaking the direct correspondence between the data and underlying household preferences.

The first problem, mentioned earlier, is that down-payment constraints can manifest similar effects as loss aversion. This creates an empirical confound, because the nominal purchase price of the house (the assumed reference point for loss aversion) can be very close in practice to the outstanding mortgage balance (the point at which down-payment constraints kick in). To cleanly separate these forces, we harness significant independent variation in the Danish data between sellers' home equity position and their gains/losses since purchase.

A second confound comes from the need to assess sellers' gains/losses at the point of their listing decisions. Measuring these "potential gains" requires an estimate of the expected value of houses. As Genesove and Mayer (2001), Anenberg (2011), Clapp, Lu-Andrews and Zhou (2018), and others note, variation in unobservable property quality can obscure empirical inferences about house values. We extend pre-existing empirical approaches to analyze how different sources of measurement error generate specific biases in inferences. Guided by this analysis, we re-estimate all empirical moments using alternatives to the customary hedonic model used to compute value.[4] We then re-estimate all structural parameters using recomputed moments from these alternative empirical models, and confirm robustness of our inferences.[5]

A third confound comes from mapping sale probabilities to listing premia. Guren (2018) shows that U.S. housing markets are characterized by "concave demand," where lowering list prices past a point does not boost sale probabilities, but does negatively impact realized sale prices.[6] We confirm this finding in Danish data, which also exhibit strong demand concavity. In the model, we show that optimizing sellers who face concave demand will set asymmetric listing premium schedules even if their underlying preferences are symmetric across gains and losses. This poses an additional challenge to identifying preference parameters, which we confront by harnessing regional variation across housing markets in Denmark. These regional markets vary in the strength of demand concavity, meaning that comparing the behavior of listing premia in these markets helps us to parse the relative strength of loss aversion and demand concavity.

The main facts in the Danish data are as follows: As in US data, the listing price

[4] To measure potential gains (and, indeed, potential home equity), we employ a standard hedonic model, which fits the data with high explanatory power ($R^2 = 0.88$).

[5] In particular, we estimate repeat sales models to difference out time-invariant unobserved quality, and combine these approaches with time-varying hedonic characteristics and novel data on renovation expenses, to account for potentially time-varying, otherwise unobserved, quality.

[6] Guren attributes this to a signalling issue, in which an unusually low sales price can signal quality issues with the property, leading to a longer time-on-the-market for such properties.

schedule in the Danish data has the characteristic "hockey stick" shape first identified by Genesove and Mayer (2001), rising substantially as expected losses mount, and virtually flat in gains. We find slopes of similar magnitude to those in that paper despite the differences in location, sample period, and sample size.[7] We also see a similar "hockey stick" along the dimension of potential home equity controlling for potential gains, consistent with the distinct effect of down-payment constraints.

The data also show that the shapes of listing premia schedules vary across regional housing markets in Denmark in a manner that mirrors the degree of demand concavity in these markets. In the data, we see listing premia with sharp responses to losses in regional markets with weaker demand concavity, and muted responses to losses in regional markets with strong demand concavity, confirming the model-implied link between demand conditions and sellers' optimal listing decisions.[8]

We find that the distribution of sales exhibits sharp bunching of transactions at realized gains of zero, diffuse bunching mass just to the right of zero, and a significant shift in total mass from realized losses towards realized gains. The spike at precisely zero gains is clear evidence of loss aversion—the only force in the model that can generate this pattern. That said, the magnitude of this spike is relatively small, meaning that fitting just this aspect of the data requires only a modest degree of loss aversion.[9] The inference that loss aversion is modest comes from a very strict interpretation of the model, in which sellers can precisely target and achieve specific realized prices. Therefore, to bound the magnitude of loss aversion and account for the diffuse excess bunching mass visible just to the right of zero, we also allow for the possibility that loss-averse sellers are subject to optimization frictions that can cause targeted and realized prices to differ.

We also estimate listing propensities for the entire Danish housing stock using over 5.5 million property-year observations, and plot them against prospective sellers' potential gains. We see a mild but visible increase in the propensity for homeowners to list their houses as potential gains rise, and the slope appears more pronounced over the potential loss domain than the potential gain domain.

Collectively, these moments in the data allow us to structurally estimate six model

[7]In the original Genesove and Mayer (2001) 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 4.7 and 5.7% for the same 10% nominal loss in the Danish data of apartments, row houses, and detached houses between 2009 and 2016 [N=214,523].

[8]Intuitively, we find that this regional variation in the shape of demand is correlated with the degree of homogeneity of the housing stock. Regions in which the housing stock is relatively homogeneous make listed properties more easily comparable, leading to a steep decline in the probability of sale for houses listed at large positive listing premia, as these properties are more obviously overpriced.

[9]This is consistent with the observation in Kleven (2016) that kinks (and notches) in preferences can lead to clearly visible distortions even when underlying structural elasticities are small.

parameters; we verify identification by sequentially adding model features and moments to pin down these parameters. In our baseline analysis, which ignores the diffuse bunching mass, and only attempts to fit the spike in transactions at zero realized gains, we converge on reference dependence $\eta = 0.871$ (*s.e.* 0.036), meaning that realized gains contribute slightly less than final prices to household utility, and loss aversion $\lambda = 1.220$ (*s.e.* 0.028), meaning that the disutility of losses is only 1.22 times the utility of gains.[10] This is a modest level, lower than the early estimates between 2 and 2.5 (e.g., Kahneman, Knetsch and Thaler, 1990; Tversky and Kahneman, 1992), but closer to those seen in more recent literature (e.g., Imas, Sadoff and Samek, 2017 find $\lambda = 1.59$). When we allow for the possibility of optimization frictions, and account for the diffuse bunching mass as well, we estimate $\eta = 0.545$ (*s.e.* 0.078) and $\lambda = 2.025$ (*s.e.* 0.245). We interpret these pairs of parameter values as useful bounds for the preference parameters that we identify; they generate the range of shares attributed to different forces in our decomposition of the model-implied price-volume correlation.

The remaining estimated parameters imply sizeable down-payment constraints, and suggest that households enjoy significant "gains from trade" from successful sales. Another observation is that in a restricted model where demand is (counterfactually) assumed linear rather than concave, estimated $\eta = 0.325$ (*s.e.* 0.025) and $\lambda = 2.151$ (*s.e.* 0.071). This demonstrates the importance of accounting for real-world frictions when mapping facts to underlying deep parameters (see, e.g., Blundell, 2017), and underlines the merits of a structural behavioral approach (DellaVigna, 2009, 2018) to field evidence.

Overall, the model does a good job of fitting the moments with economically plausible preference parameters. When we feed the model with regionally estimated demand concavity, it nicely predicts the shape of the listing premium schedule in different regional Danish housing markets "out of sample." Most importantly, the model permits a structural decomposition of the price-volume correlation. In the model, changes in fair house values translate into changes in the distribution of potential gains and potential home equity. These affect listing behavior, which in turn translates into realized traded volumes through the estimated concave demand mapping from listings to final sales. Attributing this correlation to the different economic channels, we find that the effect of down-payment constraints is largest, followed by reference dependence, and finally by loss aversion.

In addition to the literature on housing markets, our paper contributes to behavioral economics and household finance. Housing is typically the largest household asset, and

[10]As is standard in the literature, here, the disutility of losses is λ times greater than the utility of gains, so when $\lambda = 1$, there is no loss aversion.

mortgages, typically the largest liability (Campbell, 2006; Badarinza, Campbell and Ramadorai, 2016; Gomes, Haliassos and Ramadorai, forthcoming), making this an important field setting to study households' preferences and constraints. Our strong evidence for reference dependence here complements similar evidence derived from sports settings (Allen et al., 2017), and labor markets (Camerer et al., 1997; Fehr and Goette, 2007; Farber, 2008; Crawford and Meng, 2011; DellaVigna et al., 2017).[11] Our conclusion that loss aversion is relatively modest is also important, given that the original Genesove and Mayer (2001) findings are broadly accepted as large-stakes field evidence for a strong aversion to nominal losses. Moreover, the frequent description of this behavior as a household behavioral "mistake" (see, e.g., Kahneman, 2003; List, 2003, 2004; Fehr and Tyran, 2005; Barberis, 2013) is at variance with our characterization of behavioral but utility-maximizing households.

The paper is organized as follows. Section 2 introduces our model of household listing behavior. Section 3 discusses the construction of our merged dataset, and provides descriptive statistics. Section 4 introduces the moments that we use for structural estimation and uncovers new facts on listing prices and 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 commonly used to explain patterns in listing behavior.[12]

The market consists of a continuum of sellers and buyers of residential property. In period 0 of the two-period model, some fraction of property owners receive a shock $\theta \sim N(\theta_m, \theta_\sigma)$, and decide (i) whether or not to list their property for sale, and if they do list, (ii) the optimal listing price. The "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. It captures a variety of reasons for moving, including moves due to labor market opportunities, or the desire to upsize or downsize.

Our goal is to uncover the structural relationship between listing decisions and seller

[11]Separately identifying and pinning down reference dependence and loss-aversion parameters has posed a challenge even in laboratory settings (see, for example, Ericson and Fuster, 2011).

[12]We describe the intuition of the model here; online appendix B contains a more detailed discussion of model arguments and derivations of equations.

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. We assume that in period 1, buyers visit listed properties, and make offers, which result in sales at final sale prices and properties changing hands with some probability. In the event that a property does not sell, the seller stays in the property, and receives constant utility $\underline{u}$.

More specifically, let L denote the listing price set by the seller; $\widehat{P}$ measure the "expected" or "fundamental" property value; $\ell = L - \widehat{P}$ denote the listing premium; α the probability that a willing buyer is found; and P 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 then be written as:

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

where:

$$U \left(P(\ell), \cdot\right) = u(P(\ell), \cdot) - \kappa(P(\ell), \cdot), \tag{2}$$

which embeds a range of possible preferences $u(P(\ell), \cdot)$, including reference-dependent loss-aversion à la Kahneman and Tversky (1979) and Kőszegi and Rabin (2006, 2007), as well as a penalty for violating down-payment constraints $\kappa(P(\ell), \cdot)$, à la Stein (1995).

The seller decides on the extensive margin, i.e., whether ($s = 1$) or not ($s = 0$) to list, and sets ℓ to maximize the expected utility from final property sale.

Once the property is listed, with probability $\alpha(\ell)$ the sale goes through and the seller receives utility (and pays any penalties arising from down-payment constraints) associated with the final price $P(\ell) = \widehat{P} + \beta(\ell)$, and also receives θ. With probability $1 - \alpha(\ell)$ the listing fails, and the seller falls back to their outside option level of utility $\underline{u}$. Listing incurs a one-time utility cost ϕ, sunk at the point of listing, which captures a range of frictions including psychological "hassle factors,"and search, listing, and transaction fees.

When optimizing, the seller takes $\alpha(\ell)$ and $\beta(\ell)$, i.e., the "demand" functions, as given; we estimate these functions in the data as a reduced-form for equilibrium outcomes in period 1, which the seller internalizes when optimizing utility. This assumption simplifies the model, and allows us to more closely focus on our goal, namely, extracting the underlying parameters of seller utility and constraints. 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.[13] In our estimation, we define a *period* as equal to six months, meaning that $\alpha(\ell)$ captures the probability that the transaction goes through within six months after the initial listing. We later verify robustness to varying the length of this period. In the remainder of the paper, we refer to the two functions $\alpha(\ell)$ and $\beta(\ell)$ collectively as *concave demand*.

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 of the expected value $\widehat{P}$. We later map these percentages to logs, relying on the usual approximation, to be consistent with the definitions of gains/losses and home equity employed in our empirical work.[14]

Next, we describe an important component of $U(P(\ell), \cdot)$, namely, the reference-dependent utility specification $u(P(\ell), \cdot)$, and build intuition using a simple model which focuses on this ingredient.

2.1 Reference-Dependent Loss Aversion: A Basic Model

To build intuition on the role of seller preferences, we first describe a simple version of the model with no extensive margin decision, absent down-payment constraints $\kappa(P(\ell), \cdot)$, and assuming linear demand. We then describe how these model ingredients augment the simple model in subseqent subsections.

2.1.1 Preferences

We adopt a standard formulation of reference-dependent loss-averse preferences, writing $u(P(\ell), \cdot)$ as:

$$u(P(\ell), R) = \begin{cases} P(\ell) + \lambda\eta G(\ell), & \text{if } G(\ell) < 0 \\ P(\ell) + \eta G(\ell), & \text{if } G(\ell) \geq 0 \end{cases}. \tag{3}$$

In equation (3), the seller's reference price level is R, which we simply assume is fixed; in our empirical application, we set it to the original nominal purchase price of the property.[15] Realized gains $G(\ell)$ relative to this reference level are then given by $G(\ell) = P(\ell) - R$.

[13] As we later discuss, $\alpha(\ell)$ is nonlinear in the Danish data; above average list prices reduce final sale probabilities, while below average list prices reduce seller revenue with little effect on sale probabilities. Guren (2018) documents the same patterns in U.S. data.

[14] In our empirical work, we estimate $\widehat{P}$ as the hedonic value of the house. An alternative, followed by Guren (2018), is to assume 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. However, as we show in the online appendix Table A.2, hedonic prices work better (correlate more strongly with changes in the probability of sale) than this alternative, perhaps because Denmark has a relatively homogenous and liquid housing market.

[15] 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

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. η measures the extent to which realized gains affect utility relative to pure terminal wealth.

The parameter $\lambda > 1$ governs the degree of loss aversion. This specification of the problem, a widely-used approach to study and rationalize results found in the lab (e.g., Ericson and Fuster, 2011), as well as in the field (e.g., Anagol, Balasubramaniam and Ramadorai, 2018), assumes that utility is piecewise linear in nominal gains and losses relative to the reference point, with a kink at zero. In online appendix Figure A.1, we graphically describe how both η and λ affect utility, and below, we show how these parameters affect optimal listing decisions in the model.

2.1.2 State Variables

Seller decisions in the simplest version of the model 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 the empirical and theoretical literature, we reduce this set of four state variables to two. First, we calculate the seller's expected or "potential" gains $\widehat{G} = \widehat{P} - R$ as a function of two of the state variables.[16] Realized gains $G(\ell)$ arise from their "potential" level $\widehat{G}$ plus the realized premium $\beta(\ell)$, i.e.:

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

Second, 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 if listing is frictionless, the seller will be indifferent between staying in their home and receiving the hedonic value in cash. This assumption can equivalently be mapped to 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 $\underline{u} = \widehat{P} - \theta$).

---

off a more detailed description of the decision-making problem in the field against stronger assumptions that permit measurement of important underlying parameters.

[16]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 discuss this issue in online appendix section B.6, and estimate a version of this regression in online appendix Table A.3 to verify the original inferences of Genesove and Mayer (2001) using our sample.

The model implicitly specifies conditions on the relationship between $\underline{u}$ and R. In online appendix section B.7, we discuss this issue in detail. A brief summary: (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 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 simplest version of the model to build intuition, and to guide our choice of key moments of the data with which to structurally estimate main parameters.

2.1.3 Optimal Listing Premia, Basic Model

In the simplest version of the model, absent the extensive margin decision, listing costs, or constraints $\kappa(\cdot)$, the maximization problem becomes:

$$\max_{\ell} \left[\alpha(\ell)\left(u(P(\ell), \cdot) + \theta\right) + (1 - \alpha(\ell))\underline{u}\right]. \tag{4}$$

In this case, the first-order condition determining optimal ℓ^* balances the marginal utility benefit arising from a higher realized premium in the event of a successful sale, against the marginal cost associated with an increased chance of the transaction failing—which results in the outside option utility level.

To derive more intuition, we analytically solve the simple model in equation (4), 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 delivers an optimal listing premium schedule which is piecewise linear. We provide a full solution of this model in online appendix sections B.1-B.3, and discuss it graphically below.

Figure 1 illustrates how the optimal listing premium in this simple version of the model varies with the preference parameters. If there is no reference dependence ($\eta = 0$), utility derives purely from the terminal house price. In this case, the left plot (dashed line) shows that ℓ^* is unaffected by the reference price R. In the case of "linear reference dependence" (i.e., $\eta > 0$, $\lambda = 1$), there is a negatively-sloped linear relationship between ℓ^* and $\widehat{G}$. In this case, indicated with a dotted line, 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^* is then 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 results in a higher ℓ^*.[17]

In the case of (reference dependence plus) loss aversion ($\eta > 0$, $\lambda > 1$), indicated with a solid line, 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 this level $\widehat{G}_0$ 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, which we denote as $\widehat{G}_1$, 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 (relative to the $\lambda = 1$ case) for each unit loss beyond this point.

2.1.4 Bunching Around Realized Gains of Zero

Household listing behavior also has implications for transactions quantities. This shows up as shifts in mass in the distribution of completed transactions along the G dimension. We later use this insight to harness an additional set of moments which help to pin down the preference parameters.

The right panel of Figure 1 illustrates how the distribution of realized transactions varies with preference parameters in the simple version of the model (assuming linear demand) discussed above. (Detailed solutions are provided in online appendix section B.2).

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 $\widehat{G}$.[18] 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 are found, meaning that the likelihood of observing transactions in this domain of $\widehat{G}$ is lower. However, if these transactions *do*

[17] 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 section B.4, and note that this prediction is clearly rejected in the patterns seen in the data.

[18] Recall that listing prices are converted to realized prices through the $\beta(\ell)$ function, which is a reduced-form for the matching and negotiations between sellers and prospective buyers.

go through, the associated G will be higher, shifting mass in the final sales distribution towards $G > 0$ (in the right-hand plot, the dashed line becomes the dotted line).

The mass shift is especially pronounced and distinctive if sellers are loss averse, i.e., when $\lambda > 1$, in which case the model predicts bunching in the final distribution of house sales at $G = 0$, coming from even less mass when $G < 0$ (in the right-hand plot, the dotted line becomes the solid line).

We next consider the seller's extensive margin decision of whether to list, and subsequently turn to discussing the effects of adding concave demand and financial constraints into the model.

2.2 Extensive Margin

We have thus far 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 both intensive and extensive margins. In particular, the model predicts that sellers with lower $\widehat{G}$ are less likely to list. This prediction points to the relationship between the propensity to list and $\widehat{G}$ as an additional moment to inform structural estimation of preference parameters.[19]

Modelling the extensive margin decision also helps to account for any selection effects that may drive patterns of observed intensive margin listing premia in the data, an issue that prior literature (e.g., Genesove and Mayer, 1997, 2001; Anenberg, 2011; Guren, 2018) has been unable to control for as a result of data limitations. For example, if sellers that decide not to list are more conservative (i.e., they set lower listing premia), and those who decide to list are more aggressive (i.e., 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.[20]

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 segment of the population that chooses to list. These average listing premia incorporate endogenous first-stage selection effects and can be mapped directly back to

[19] 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.

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

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 variation in θ can affect the observed magnitude of the seller's responses to $\widehat{G}$, smoothing and blurring any kinks in the model-implied ℓ^* profile. The online appendix section B.1 illustrates this analytically with a specific example, showing how a smooth "hockey stick" average listing premium profile can result from averaging the three-piece-linear listing premium profile (for $\lambda > 1$) across the distribution of sellers with different θ.

2.3 Concave Demand

The demand functions $\alpha(\ell)$ and $\beta(\ell)$ are an important determinant of listing behavior, and their functional form can affect the shape of the ℓ^* schedule in this model. In the simple case of the linear demand functions posited earlier, when the probability of sale $\alpha(\ell)$ is less responsive to ℓ, the marginal cost of choosing a larger listing premium is lower, and therefore the level of ℓ^* will be higher. There are even stronger implications for the relationship between ℓ^* and $\widehat{G}$ when $\alpha(\ell)$ has the concave shape first identified by Guren (2018).

The optimal ℓ^* in a linear reference-dependent model ($\eta > 0$, $\lambda = 1$) when demand is concave has a flatter slope in the domain $\widehat{G} > 0$, relative to the case of linear demand. A way to think of "concave" $\alpha(\ell)$ is that it remains constant for ℓ below a level that we denote as $\underline{\ell}$. Lowering listing premia below $\underline{\ell}$ thus results in reductions in final sale prices (since $\beta(\ell)$ increases in ℓ), but does not result in increases in the sale probability. Even if the seller is "linearly reference dependent" with *no* loss aversion ($\eta > 0$ and $\lambda = 1$) the logic of optimization in this case will generate a graph of ℓ^* against $\widehat{G}$ with a "hockey stick" shape. This is an important confound for λ that has not previously been considered in the literature.

More generally, the model predicts a tight link between the shape of $\alpha(\ell)$ and the slope of ℓ^*. A steep negative slope of $\alpha(\ell)$ for ℓ above $\underline{\ell}$ leads to a gradual slope of ℓ^* in the loss domain, since the marginal cost of increasing the listing premium is higher in this case, and vice versa.[21]

[21] 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. In the online appendix section B.5 we illustrate this mechanism, positing a concave shape for $\alpha(\ell)$ and showing the effect of varying $\alpha(\ell)$ around $\underline{\ell} = 0$, i.e., the point at which $L = \widehat{P}$.

Concave demand also has effects on transactions volumes in the model, resulting in shifts of mass in final sales towards positive values of realized gains, depending on the level of $\underline{\ell}$, the point at which concavity "kicks in," though it is not associated with sharp bunching of the type associated with loss aversion, as demand concavity is assumed (and seen in the data to be) smooth.[22]

2.4 Down-Payment Constraints

We can now describe our approach to modelling down-payment constraints $\kappa(P(\ell), \cdot)$, an integral part of our analysis (recall that in our setup, sellers maximize $U(P(\ell), \cdot) = u(P(\ell), \cdot) - \kappa(P(\ell), \cdot))$. Let M denote the household's outstanding mortgage balance, and γ the required down-payment on a *new* mortgage origination. The "realized" home equity position of the household following a sale at price $P(\ell)$ is then $H(\ell) = P(\ell) - M$. We note that the potential home equity position becomes another state variable in our setup, and analogous to gains, realized home equity $H(\ell)$ arises from its "potential" level $\widehat{H}$ plus the realized premium $\beta(\ell)$, i.e.:

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

Assuming that the seller uses H as the down payment on their next home, we can distinguish between potentially constrained (i.e., downsizing-averse) households who expect $H(\ell) < \gamma$, and potentially unconstrained households who expect $H(\ell) \geq \gamma$. If down-payment constraints bind, only unconstrained sellers can move to another property of the same or greater value. However, there are several ways in which Danish households can relax these constraints despite legal restrictions on LTV at mortgage initiation (which, as we discuss later, is strictly set at 20% in Denmark). The first way 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.[23] A third

[22]A more subtle point here is that any change in the precise specification of the reference point R in the presence of loss aversion will also change the location at which bunching is observed, and 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 $\underline{\ell}$. 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.

[23]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

(typically unobservable) possibility is that households can bring additional funds to the table by liquidating other assets, or by borrowing from friends and family.[24] Taking these realistic features into account, we assume that violating the down-payment constraint does not lead the seller to withdraw the sale offer, but instead that the seller incurs a monetary penalty for levels of realized home equity below the constraint threshold:

$$\kappa(P(\ell)) = \begin{cases} \mu(\gamma - H(\ell))^2, \text{ if } H(\ell) < \gamma \\ 0, \text{ if } H(\ell) \geq \gamma \end{cases}. \tag{5}$$

We choose a smooth quadratic penalty function to avoid a discontinuity at the threshold level $H(\ell) = \gamma$. Such a discontinuity would predict bunching in realized prices at a level of home equity of 20%, which we can firmly reject in the data.

2.5 A Note on Structural Parameter Identification

In the table below, we provide an overview of key model variables and parameters. Our goal is to estimate the structural parameters using moments in the data:

State variables:	$\widehat{G}$	Potential gain (measured, $\widehat{P} - R$)
	$\widehat{H}$	Potential home equity (measured, $\widehat{P} - M$)
	θ	Magnitude of the moving shock (unobservable)
Exogenous market conditions:	$\alpha(\ell)$	Concave demand (estimated in the data)†
	$\beta(\ell)$	Realized premium (estimated in the data)†
Endogenous model variables:	ℓ^*	Optimal listing premium
	s^*	Optimal listing decision (extensive margin)
	G	Realized gain (observed, $P - R$)
	H	Realized home equity (observed, $P - M$)
Calibrated model parameters:	γ	Down-payment constraint (set at 20%)
Structural model parameters:	η	Reference dependence
	λ	Loss aversion
	μ	Financial constraint
	θ_m	Average value of the moving shock
	θ_σ	Standard deviation of the moving shock
	ϕ	Magnitude of the listing and search cost

†Exogenous market conditions/demand functions are estimated in the data as described below.

As discussed, the first three structural parameters (η, λ, and μ) have a clear economic interpretation, and the model provides guidance as to how they can be separately identified

mortgage rate. For reference, see categories *DNRNURI* and *DNRNUPI* in the Danmarks Nationalbank's statistical data bank.

[24]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.

in the data: (a) $\eta > 0$ leads to a negative slope of the listing premium profile along the entire range of potential gains, (b) $\lambda > 1$ leads to excess bunching of transactions for realized prices at the nominal reference point, and an additional contribution to the slope of the listing premium profile in the loss domain, (c) $\mu > 0$ leads to a negative slope of the listing premium profile with respect to potential home equity.

The three other parameters (θ_m, θ_σ, and ϕ) can be thought of as "fitting parameters," which allow us to pin down important quantities that are not the main focus of our model. A simple characterization is that (a) θ_m mainly determines the average level of the listing premium, (b) θ_σ drives the degree to which the kinks and non-linearities that our assumptions on preferences and constraints imply will be smoothed out in the data, (c) ϕ is a key determinant of the extensive margin probability that a given property will be listed for sale. Of course, beyond these simplified characterizations, there are complex interactions between different parameters which determine ultimate aggregate outcomes.

We next describe the data and key moments visible in the data as a precursor to a more rigorous structural estimation of the model's parameters using these moments.

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. We merge these data with data on the entire housing stock captured in the Danish housing register to estimate the propensity to list for all properties in Denmark over the sample period, enabling us to empirically capture sellers' extensive margin decisions.

Our high-quality administrative data are linked from different official sources; all data other than the listings data are made available to us by Statistics Denmark. We briefly describe these data below, and online appendix C contains detailed descriptions of data sources, data construction and filtering, and the process of matching involved in assembling the final dataset.

3.1 Property Transactions and other Property Data

We acquire comprehensive administrative data on the ownership and hedonic characteristics of the housing stock of all registered properties in Denmark between 1992 and 2016, as well as associated property transactions from the Danish Tax and Customs Administration (SKAT) register and the Danish housing register (Bygnings-og Boligregisteret, BBR). 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.[25]

3.2 Property Listings Data

Property listings from 2008 to 2016 are provided to us by RealView. We link these transactions to the cleaned/filtered sale transactions in the official property registers; 79.5% of all sale transactions have associated listing data. We describe these data more fully in the online appendix; we note here that unmatched transactions generally occur off-market as direct private transfers.

3.3 Mortgage Data

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 LTV 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.[26]

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), the records also contain a family identification number that links members of the same household. This allows us to aggregate individual data on wealth and income to the household level. We also 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

Our mortgage data run from 2009 to 2016, making this the sample period that we consider. The sample is further restricted, as we only keep transactions for which we

[25] Tax-assessed property values are used for determining tax payments in Denmark. Online appendix D.1, 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. As we later describe, we follow Genesove and Mayer (1997, 2001) in including this variable; removing it does not greatly affect the fit of the hedonic model or our substantive inferences.

[26] Online appendix D.2 provides a 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 measure both nominal losses and home equity. Transactions data are available from 1992 to the present, meaning that we can only measure the purchase price (i.e., reference price) for properties that were bought during or after 1992. We also restrict our analysis to properties for which we know both the ID of the owner, as well as that of the owner's household; this last is to match to wealth and income data. We exclude data from foreclosure transactions,[27] 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, or marked as having an extreme price by Statistics Denmark).[28] 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.[29] 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 to be property investors than owner-occupiers.[30]

In online appendix C.5, we describe the data construction filters and their effects on our final sample in more detail. Once all filters are applied, the sample comprises 214,523 listings of Danish owner-occupied housing between 2009 and 2016, for both sold (70.5%) and retracted (29.5%) properties, matched to mortgages and other household financial information.[31]

These listings correspond to a total of 191,884 unique households, and 179,215 unique properties. Most households sell one property during the sample period, but roughly 9% of households sell two, and roughly 1.5% of households sell three or more properties. As mentioned, we also use the entire housing stock, filtered in the same manner as the listing data, comprising 5,540,349 observations of 807,556 unique properties to capture the determinants of the propensity to sell.

[27]Online appendix D.3, describes the Danish foreclosure process.

[28]We apply this filter as the market for such unusually priced properties is extremely thin, meaning that predicting the price using a hedonic or other model is particularly prone to error.

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

[30]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.

[31]172,399 listings have an attached mortgage, and 42,124 listings have no associated mortgage (i.e., are owned entirely by the seller).

3.6 Hedonic Pricing Model

To calculate listing premia ℓ, potential gains $\widehat{G}$ and potential home equity $\widehat{H}$, we need to measure the expected house value $\widehat{P}$ for each property-year in the data. We do so by estimating a standard hedonic pricing model on our sample of sold listings, and predicting prices for the full sample of listed properties, including those that are not sold.

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

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

where $\mathbf{X_{it}}$ is a vector of property characteristics,[32] 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(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 to the nearest major city). 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. $\Phi(v_{it})$ is a third-order polynomial of the previous-year tax assessor valuation of the property.[33] 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. Online appendix E.1 explains the baseline specification in greater detail, and subsequent appendix sections do so for alternative model specifications.

3.6.1 Hedonic Model Fit

The R^2 of the model equals 0.88 in the full sample, a high degree of accuracy which we verify in various ways, described in more detail in online appendix E.[34]

An important confound when structurally estimating model parameters is noise or

[32] The t subscript indicates that most characteristics are captured afresh each year, which contributes to the accuracy of the model.

[33] 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. We further document the accuracy and role of the tax-assessed value in online appendix E.2.

[34] Briefly, when we estimate the model in levels rather than logs, we obtain an R^2 of 0.85. The R^2 when we eliminate the tax assessor valuation from the hedonic characteristics is 0.77. We achieve a similarly high fit when estimating the model on random 50% samples of the data and fitting the remainder out-of-sample, to address overfitting concerns.

bias in the estimation of $\widehat{P}$, arising from factors such as time-varying unobserved house quality (see e.g., Genesove and Mayer, 1997, 2001; Anenberg, 2011; Clapp, Lu-Andrews and Zhou, 2018). We describe our approach to dealing with these issues later in the paper.

4 First Inferences about Model Parameters

As a precursor to full-blown structural estimation, we document patterns in listing premia and sales transactions volumes in the data in relation to measured G and $\widehat{G}$. We informally discuss how these patterns relate to the predictions of the model, with a specific focus on 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 (i) sellers' down-payment constraints (i.e., $\widehat{H}$), (ii) concave demand, and (iii) robustness to changes in measurement.

4.1 Fact 1: Listing Premia and Potential Gains ("Hockey Stick")

We compute listing premia as $\ell = \ln L - \widehat{\ln P}$, where L is the reported initial listing price observed in the data, and $\widehat{\ln P}$ is estimated using the hedonic model.[35] Mean (median) ℓ is 12.1% (10.7%), and $\ell \geq 0$ (< 0) for 74% (26%) of the sample. Potential gains are computed as $\widehat{G} = \widehat{\ln P} - \ln R$, where R is set to the nominal purchase price of the property. Mean (median) $\widehat{G}$ estimated in this way is 38% (29%), and 23% (77%) of property-years have $\widehat{G} < 0$ ($\widehat{G} \geq 0$). Online appendix Figure A.2 plots the distributions of these variables.

In Figure 2 we plot the average observed listing premium on the y-axis, for each percentage bin of potential gains on the x-axis. The figure shows that prospective sellers of 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., $\widehat{G} > 0$). Overall, the downward slope is consistent with the predictions of a model with reference dependence $\eta > 0$. Moving from the gain to the loss domain, the slope becomes much more pronounced, i.e., listing premia 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 $\lambda > 1$. In the piecewise linear formulation of preferences, however, loss aversion also predicts a flattening out of the listing premium profile deeper into the loss domain, which is not visible in the plot.

[35] In the online appendix Table A.3, we estimate Genesove and Mayer's (2001) specifications on our data and confirm that the coefficient on $\widehat{\ln P}$ in our data using their regression, controlling for a range of other determinants, is close to 1.

While these patterns are apparently consistent with reference-dependent and loss-averse preferences, as discussed earlier, we must keep various confounds in mind before accepting this conclusion. We investigate the influence of these issues below.

4.1.1 Listing Premia, Down-Payment Constraints, and Home Equity

To assess the effects of home equity constraints, for each observation in the data, we calculate $\widehat{H} = \widehat{\ln P} - \ln M$, 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. Mean (median) $\widehat{H}$ is 41% (37%), and 81% (19%) of property-years have $\widehat{H} < 0$ ($\widehat{H} \geq 0$). Modal $\widehat{H}$ conditional on having a mortgage 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.[36] 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 online appendix Figure A.4, 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 $\widehat{H} \lesseqgtr 20$, which permits identification of their independent impacts on listing decisions.[37]

To assess the extent to which any variation in ℓ attributed to $\widehat{G}$ might be confounded by simultaneous variation in $\widehat{H}$, the top left plot in Figure 3 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 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 $H = 20\%$). 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, Figure 10 (left plot) shows how the "marginals" of the listing premium along the $\widehat{G}$ dimension vary

[36] This constraint does not change over our sample period. The online appendix table A.4 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.

[37] Online appendix F also contains a fuller discussion of additional evidence on downsizing/upsizing. We link sale transactions with future purchase transactions for a subset of households, and show that future purchases are almost always of higher value than current sales.

as we vary the level of $\widehat{H}$; we discuss this and the other marginal again at the end of the paper, where we evaluate the extent to which we can match these interactive relationships using the model.[38]

4.2 Fact 2: "Hockey Stick" and Concave Demand

The left-hand plot in Figure 4 shows the probability of a house sale within six months (this maps to $\alpha(\ell)$ in the model) on the y-axis, as a function of ℓ on the x-axis. These probabilities are derived from underlying data on the time-on-the-market (TOM) that elapses between sale and listing dates for the properties in the sample.[39] 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, 2017).[40] The solid line corresponds to this set of smoothed point estimates. The right-hand plot in Figure 4 shows how $\ln P(\ell) - \widehat{\ln P}$, i.e., the "realized premium" of the final sales price over the hedonic value (i.e., $\beta(\ell)$ in the model) varies with ℓ. The plot shows that $\beta(\ell)$ rises virtually one-for-one with ℓ. The solid line shows a simple linear fit of this relationship that we use in the model, with $\beta_0 = -0.068$ and $\beta_1 = 0.835$.

The two plots together reveal 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.[41] We use these estimated $\alpha(\ell)$ and $\beta(\ell)$ functions as the exogenous inputs into sellers' calculations in our structural estimation.

As described earlier, this "demand concavity" is a confound for estimating λ, because the model predicts two distinct drivers for the differential slopes of ℓ^* across gains and losses. When $\lambda > 0$, there are kinks in ℓ^* around $\widehat{G} = 0$, which can be smoothed into

[38]Online appendix Figure A.7 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.

[39]Mean (median) TOM in the data is 35 weeks (24 weeks). We pick six months in the computation of $\alpha(\ell)$ to match the median TOM observed in the sample, but the shape is robust to using other windows such as 3, 9, or 12 months. Online appendix Figure A.2 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.

[40]We describe the GLF in more detail in online appendix G. 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.

[41]These plots also show that Danish sellers who set high ℓ suffer longer time-on-the-market, 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.

the observed pattern by variation in θ. The other possibility is buyer sensitivity to ℓ, captured by the shape of $\alpha(\ell)$.

How plausible is this latter channel? To check, we exploit regional variation across sub-markets of the Danish housing market. 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 municipalities of Denmark.[42] The left-hand plot of Panel B of Figure 2 shows how $\alpha(\ell)$ varies with ℓ for municipalities in the top and bottom 5% ranked on the magnitude of the slope of $\alpha(\ell)$ when $\ell \geq 0$. This slope for municipalities with strong demand concavity (top 5%) lies between -1.4 and -1.2, while the slope for municipalities with weak demand concavity (bottom 5%) lies between -0.1 and -0.3, a substantial difference. The left-hand plot in Figure 2 Panel B shows the corresponding figure for the relationship between $\hat{\ell}$ and $\widehat{G}$ for these municipalities. Consistent with the predictions of the model, in areas with weak (strong) demand concavity, there is a relatively steep (flat) slope in the listing premium schedule.

To deal with this confound, therefore, we use two approaches. One is to include these regional moments to better pin down the "pass through" of demand concavity to listing premia, cleaning up the estimation of the preference parameters in the model. The other is to estimate the model using average demand concavity and then validate the model by checking to see if it can generate pass-through and predict regional variation in listing premia given local demand conditions across regions of Denmark.

One possible response to these patterns is that demand may be responding to anticipated supply in addition to supply responding to demand. To dig deeper, we use the insight that the degree of demand concavity should vary with the ease of value estimation and hence price comparison in a given market. If comparable properties are readily available in a local area, buyers in that area should be quick to penalize unusually high premia with sharp declines in the probability of a quick sale. Conversely, if the market is less homogeneous, buyers face a more difficult inference problem to discern whether high listing premia are indeed warranted for specific properties. In support of this logic, we show in Figure 5 that geographic variation in demand concavity is strongly positively related to the homogeneity of the local housing stock, as measured by the share of apartments and row houses listed in a given sub-market, and an IV strategy confirms that instrumented demand concavity predicts variation in the listing premium slope across different regions of Denmark.[43]

[42]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.

[43]Row houses in Denmark are houses of similar or uniform design joined by common walls, and

4.3 Fact 3: Bunching in Realized Sales

The left-hand panel in Figure 6 plots the frequency 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). Observing a counterfactual for this distribution is difficult, as for other settings which attempt to estimate loss aversion using bunching estimators. Our approach is to 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), as a counterfactual distribution. The counterfactual captures the shape of the realized sales distribution if households sold their properties precisely at the model-implied hedonic value, and we discuss this choice in more detail below.

Using this counterfactual, we see evidence of both reference dependence and loss aversion in Figure 6. First, as in the theory-implied right panel of Figure 1, the broader distribution of sales across realized gains in Figure 6 is shifted to the right when compared to the counterfactual distribution. This is consistent with $\eta > 0$ in the model. Second, 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. Recall the theory-implied right panel of Figure 1—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 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). The plot also clearly reveals two important additional features of the data: sizeable "diffuse bunching" excess mass just to the right of zero, which we discuss in greater detail below, and a small but visible "hole" just to the left of $G = 0$, which we also later attempt to rationalize by adding a "notch" to seller preferences.[44]

apartments have less scope for unobserved characteristics such as garden sheds and annexes than regular detached houses (we show pictures of typical row houses in Denmark in online appendix H). We also discuss the IV strategy in detail in online appendix H.

[44]The online appendix Figure A.3 shows the distribution of *listing prices* around the nominal reference point. The model predicts an agglomeration of mass at listing prices *above* the reference price, coming from loss-averse sellers with potential gains between $\widehat{G}_0$ and $\widehat{G}_1$ who set listing premia to arrive at gains $G = 0$, and take $\beta(\ell)$ into account when optimizing. We do see this in the data, consistent with the model. We also see clear evidence of listing prices bunching *precisely at* the reference price, which is not predicted by the model, but which suggests a separate, additional role for the salience of the reference point in some sellers' listing decisions.

Our counterfactual is based on the hedonic model. One key advantage of our setting is that this model (or a very similar one) is widely used in this market setting for valuation, and can be verified using realized prices. The standard polynomial counterfactual approach would not allow us to gauge the extent of shifts in mass which are predicted in the presence of reference-dependent loss-aversion, as that approach by definition fits the observed empirical distribution up to the range over which the counterfactual is extrapolated.[45] In this sense, our approach is more similar to Rees-Jones (2018), who cleverly extracts evidence of loss aversion from U.S. tax returns data, and employs a model to gauge expected tax avoidance costs and benefits—which would otherwise be difficult to measure.

In the right-hand panel of Figure 6 we calculate a measure of excess mass in each bin, relative to the counterfactual. This allows to be more precise about the exact magnitude of bunching. For example, for realized gains of zero, we find that the observed frequency of realized gains is 76.0% higher than under the counterfactual, with a bootstrap standard error of 7.6%. In the one-percentage point bin immediately to the left of zero, the observed frequency is 27.3% lower than under the counterfactual, with a bootstrap standard error of 3.8%.

4.3.1 Diffuse Bunching

Figure 6 shows a clear spike at $G = 0$, but also a significant build-up of excess mass in the positive domain of realized gains. Such "diffuse bunching" is routinely observed in the literature, and often attributed to optimization frictions affecting agents, such as adjustment or attention costs (see, e.g., Garicano, Lelarge and Van Reenen, 2016; Rees-Jones, 2018; Kleven, 2016). In the presence of such optimization frictions, to obtain plausible inferences on the behavioral elasticity (in our case, the magnitude of loss aversion), a standard approach is to extend the bunching interval beyond the strict point predicted by the underlying model (in our case, $G = 0$). To obtain an upper bound on the magnitude of λ in our structural estimation, we therefore add all excess mass above the counterfactual to the right of $G = 0$ (while this seems like an aggressive assumption, the plot reveals the vast majority of the excess mass is at or below 10% realized gains).

What specific optimization frictions could rationalize diffuse bunching in our setup? We can exclude pure measurement error, because the transaction prices that we observe are from officially registered deeds. Second, we view it as unlikely that a simple version of

[45] We also implement this approach in the online appendix Figure A.8, following Chetty et al. (2011) and Kleven (2016). 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 are robust to other polynomial orders and to variations of the excluded range, and generate similar results on the excess bunching mass at zero.

imperfect targeting of final outcomes (driven, for example, by attention costs) is at work. This is because a) there are high stakes involved in selling a property, b) the seller has ultimate control over the realized price, since he/she always has the option to accept or reject specific offers, and perhaps most importantly, c) pure noise in realized outcomes would imply *symmetric* excess bunching around the reference point, which is not visible in the plot.

Our preferred alternative mechanism is that sellers observe $\beta(\ell)$ imperfectly, and set ℓ^* accordingly. If an offer is received which results in $G > 0$, they are happy to accept it, but their loss aversion results in a reluctance to drop below $G = 0$ (resulting in the asymmetric and visibly higher propensity to accept offers with $P \geq R$).

Figure A.6 in the online appendix provides validation for this assumption. We first compute percent differences between listing prices and realized prices in the full sample of listings ($\ln L - \ln P$). The distribution of these differences is bimodal, with roughly 13% of sellers for which ($\ln L - \ln P = 0$), with the remaining transactions distributed around ($\ln L - \ln P = -8\%$).[46] We assume that this distribution of ($\ln L - \ln P$) is a good mapping between realized prices and listing prices. This mapping leads to a counterfactual distribution of realized gains ($G = \ln P - \ln R$) associated with each "listing gain" ($G_{list} = \ln L - \ln R$). Comparing counterfactual realized gains to actually observed ones, online appendix Figure A.6 shows that regardless of the level of listing gains, there is always substantial excess mass in realized gains to the right of zero, mainly coming from a diminution of mass in transactions with realized gains to the left of this point.

Finally, while optimization frictions might reasonably cause diffuse bunching, another reasonable alternative interpretation is that diffuse bunching reflects heterogeneity in reference points (see, e.g., Rees-Jones (2018)). We limit ourselves in this paper to the assumption that the nominal purchase price is the reference point, and consequently interpret estimated loss aversion obtained using diffuse bunching mass as an upper bound under this assumption.

4.4 Fact 4: Extensive Margin: Probability of Listing

To understand the decision to list, we turn to data on the total housing stock in Denmark, corresponding to $5,540,349$ property-years in the data, once merged with the listings data. The unconditional average annual listing propensity is 3.87% of the housing stock (corresponding to between 3% and 4.5% of the housing stock listed across sample

[46] This observation is consistent with prior real estate literature, see, e.g., Merlo and Ortalo-Magne (2004).

years).[47] Figure 7 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.[48] The figure shows a mild increase in the probability of listing as $\widehat{G}$ increases, which is consistent with levels of $\eta > 0$, and potentially, $\lambda > 1$.

4.5 Robustness

4.5.1 Unobserved Quality and Measurement Error

Relationships between ℓ and $\widehat{G}$ (and indeed $\alpha(\ell)$ and ℓ), can be spuriously affected by measurement error in the underlying model for $\widehat{P}$. To provide a specific example, consider the case of unobserved property quality that causes underestimation of $\widehat{P}$ for a group of properties. In such a case, potential gains for such properties would also be underestimated, possibly leading to $\widehat{G} < 0$, as well as overestimates of listing premia for such properties. This alone could spuriously result in the hockey-stick shape that we observe when plotting ℓ against $\widehat{G}$.

We make a concerted effort to address these measurement error concerns, outlining in detail in online appendix I how different sources of measurement error would affect our inferences about underlying structural parameters, and proposing specific fixes in each case. We summarize the results of this analysis briefly here. Following Genesove and Mayer (2001), we differentiate between two types of measurement error, namely, (potentially time-varying) unobserved quality (ν_{it}), and idiosyncratic over- or under-payment by the seller at the point of purchase (ω_{it}). We outline how these sources of error can bias estimated moments and yield misleading inferences about true underlying relationships. We then outline assumptions under which alternative empirical models of $\widehat{P}$ can unwind these biases in inferences, permitting recovery of clean estimates of true underlying relationships. More specifically, we compare how key moments vary when we a) employ a standard hedonic model; b) employ repeat sales models to difference out time-invariant unobservable components; c) use as regressors time-varying hedonic characteristics and novel tax exemption data on home renovation expenses to narrow in on time-varying changes in unobservable quality. Our preferred alternative model combines all of these features.[49]

We show that the important patterns in the data, including the hockey stick in listing

[47] 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.

[48] In online appendix Figure A.9 we report a residualized version of the listing probability across potential gains, controlling for the level of home equity.

[49] In particular, our pairwise repeat sales model extends the bounding approach that Genesove and Mayer (2001) originally proposed.

premia over potential gains, and the relationship between sale probabilities and listing premia (i.e., concave demand) are robust across these different models of $\widehat{P}$, and quantitatively similar. Perhaps more importantly, we later show how our structural parameter estimates change (preview: they are not substantially affected) when we use these alternative models to estimate $\widehat{P}$.[50]

4.5.2 Sources of Variation in Potential Gains and Home Equity

To identify the independent effects of reference-dependent loss-averse preferences and downsizing aversion induced by home equity constraints, we need to verify that there is independent variation in $\widehat{G}$ and $\widehat{H}$. We do so in online appendix Figures A.4 and A.5, and show that this variation is not confined to one particular part of the sample period. In particular, we show that there is substantial mass in all four quadrants of the joint distribution defined by thresholds in $\widehat{G} \lesseqgtr 0$ and $\widehat{H} \lesseqgtr 0$, and that the relative mass remains fairly stable over the sample period. This alleviates concerns that identification simply comes from different time periods in the data, meaning that identification is likely to arise mainly from the large cross-section rather than the relatively more limited time-series. We also confirm that the inclusion of cohort and cohort-cross-municipality fixed effects in the hedonic model does not materially affect our inferences.

4.5.3 Bunching: Round Numbers, Holding Periods and Reference Price

We show in online appendix J.1 that the bunching patterns documented in Figure 6 are robust to a number of issues previously identified in this literature (e.g., Kleven, 2016; Rees-Jones, 2018). The spike in sales volumes at $G = 0$ and the patterns of excess mass relative to the counterfactual distribution do not appear to be driven by bunching at round numbers, remaining striking and visible when we exclude up to 20% of all observations. We also show that these bunching patterns are robust when we split the sample into five groups 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 at $G = 0$. Finally, we find strong evidence of bunching in all cases when we split the sample into quintiles based on the level of

[50]We also verify that the asymmetric shapes of the listing premium hockey stick and demand concavity are not driven by nonlinearities in observables by checking that property-and household-specific characteristics including renovation expenses are balanced across the domains of potential gains and listing premia, and smooth around $\widehat{G} = 0$ and $\ell = 0$. This is similar to verifying the identifying assumptions behind a regression kink design (RKD) for a discontinuous increase in the slope along a forcing variable, originally suggested by Card et al. (2015*b*) and implemented e.g., by Landais (2015), Nielsen, Sørensen and Taber (2010), and Card et al. (2015*a*). We also implement the RKD for the listing premium hockey stick in online appendix I.6, with the caveat that we do not predict a sharp kink due to the blurring factors described earlier, and that we use zero for the kink threshold, even though the listing premia slope increase begins at $\widehat{G} > 0$.

R, with quintile cutoffs ranging from around 658,000 DKK to 1.9MM DKK. Together, these checks assuage concerns that bunching could result purely from these differences in underlying properties.

We now turn to describing our structural estimation approach.

5 Structural Estimation

5.1 Moments

In the previous section, we highlighted a range of facts in the data. To transparently map these facts back to underlying parameters, the moments we use are the "dots", i.e. binned averages, visible in these plots. This is also to avoid any auxiliary estimation of moments which would then impose parametric assumptions on the data that might not be directly consistent with the model.

First we use the average listing premium $\ell^*(\widehat{G})$ computed in each 1-percentage point bin of potential gains $\widehat{G}$ using all listed properties with potential gains in the interval between -40% and +40% (79 moments, Figure 2 Panel A, labelled as "Hockey stick" in Table 1).

Second, we use the average listing premium $\ell^*(\widehat{H})$ in 1-percentage point bins of potential home equity $\widehat{H}$ for listings attached to a mortgage, covering the same -40% to +40% interval (79 moments, integrating along one dimension of Figure 3, "Home equity").

Third, we use the frequency $f_{sale}(G)$ of realized sales in each 1-percentage point bin of *realized* gains G in the full sample of sales (79 moments, Figure 6, "Bunching").

Fourth, we use the frequency of listings $f_{list}(\widehat{G})$ which fall in any given 1-percentage point bin of potential gains in the full sample of listings, and the frequency $f_{stock}(\widehat{G})$ of property $\times$ year observations in our housing stock dataset for which the potential gain (repeatedly updated each year for each property) falls within each 1-percentage point bin. Dividing these two quantities, we calculate the probability of listing $s^*(\widehat{G}) = f_{list}(\widehat{G})/f_{stock}(\widehat{G})$ for each bin of potential gains in the same interval (79 moments, Figure 7, "Extensive margin").

Finally, for each 1-percentage point bin of potential gains in the full sample of listings, we calculate the average listing premium $\ell^*_{k\in\{High,Low\}}(\widehat{G})$ in two municipality groups, distinguishing between the top 5% locations where the slope of demand is decreasing sharply ("high" demand concavity) and the 5% of locations where it is most flat ("low" demand concavity) (79×2 moments, Figure 2 Panel B, "Cross-sectional variation").[51]

[51] We later exclude these regional moments and attempt to predict them out-of-sample using parameters estimated from the remaining moments.

We collect all $79 \times 6 = 474$ empirical moments obtained in this way in the vector $\mathbf{M_d}$:

$$\mathbf{M_d} = \begin{bmatrix} \ell^*(\widehat{G}) \\ \ell^*(\widehat{H}) \\ f_{sale}(G) \\ s^*(\widehat{G}) \\ \ell^*_{k \in \{High,Low\}}(\widehat{G}) \end{bmatrix}$$

5.2 Moments in the Model

To establish notation, let $\mathbf{M_m}(\mathbf{x})$ with $\mathbf{x} = \left[\mathbf{x}^s, \mathbf{x}^c, \mathbf{x}^d\right]$ denote the vector of model-implied moments. $\mathbf{x}^s$ collects the six structural parameters to be estimated, $\mathbf{x}^c = \gamma = 20\%$ is a calibrated parameter which captures the level of the down-payment constraint in the Danish mortgage market, and $\mathbf{x}^d$ collects quantities (e.g., demand concavity parameters) that are exogenous to the sellers' decisions in the model, and are estimated using the data.

In the baseline version of the model, the joint distribution $f_{stock}(\widehat{G}, \widehat{H})$ of potential gains and home equity in the housing stock is sufficient to generate all endogenous quantities in the model. However, in some estimation variants, as described below, we abstract from the extensive margin decision, and use the distribution $f_{list}(\widehat{G}, \widehat{H})$ of potential gains and home equity observed in our listings dataset.

The model always takes as given the demand conditions that sellers face, either when using the entire set of listing premia in the data (i.e., $\alpha(\ell)$, $\beta(\ell)$), or when "regional" listing premia are included, the appropriate regional "high" or "low" concavity demand conditions as in the left plot in Figure 2 Panel B (i.e., $\alpha_{k \in \{High,Low\}}(\ell)$, $\beta_{k \in \{High,Low\}}(\ell)$). Section 4.2 above describes the construction of these functions in detail.

We note that modeling the unconditional listing probability is outside the scope of this work. We therefore set the listing probability for a potential gains level of $\widehat{G}_+ = 41\%$ to equal its observed value in the data (denoted by $\overline{s}(\widehat{G}_+)$). Put differently, we use the identifying assumption that the distribution of moving shocks θ when gains are at $\widehat{G}_+\%$ is an adequate counterfactual for the distribution of moving shocks for all levels of potential gains below $\widehat{G}_+\%$.

For given values of structural parameters and inputs from the data, we compute the solution to the seller's problem numerically for each level of potential gains ($\widehat{G}$), potential home equity ($\widehat{H}$), and the realization of the moving shock (θ):

$$\left[s^*(\widehat{G},\widehat{H},\theta,\mathbf{x}),\ell^*(\widehat{G},\widehat{H},\theta,\mathbf{x})\right] = \arg\max_{s\in\{0,1\}} \left\{(s)\max_{\ell}\left\{EU(\ell,\widehat{G},\widehat{H},\theta,\mathbf{x})\right\} + (1-s)\underline{u}\right\}, \tag{7}$$

using expected utility as described in equation (1). We obtain the policy functions $s^*(\widehat{G},\widehat{H},\theta,\mathbf{x})$ and $\ell^*(\widehat{G},\widehat{H},\theta,\mathbf{x})$ for points on a three-dimensional grid where potential gains and home equity vary between -100% and 100%, and the moving shock covers 99% of the normal distribution $\theta \sim N(\theta_m, \theta_\sigma)$ (to limit to a finite support), with the weights f_N determined by the corresponding value of the probability density function for each value of θ, and imposing $\sum_\theta f_N(\theta) = 1$. This allows us to compute the complete set of model-implied moments $\mathbf{M_m}(\mathbf{x})$ corresponding to the moments in the data.

5.3 Estimation

We use a variant of classical minimum distance estimation to recover structural parameters from the data. Defining a vector of estimation errors:

$$\mathbf{g}(\mathbf{x}^s) = \mathbf{M_m}(\mathbf{x}^s) - \mathbf{M_d},$$

we seek to estimate the structural parameters $\mathbf{x}^s$, by minimizing the objective function:

$$\widehat{\mathbf{x}}^s = \arg\min_{\mathbf{x}^s} \underbrace{\mathbf{g}(\mathbf{x}^s)'\mathbf{W}\mathbf{g}(\mathbf{x}^s)}_{F(\mathbf{x}^s)},$$

conditional on a choice for the weighting matrix $\mathbf{W}$.

We compute standard errors for the structural parameters using a bootstrap procedure in which we draw random samples (clustered at the shire level) from the underlying data (including the housing stock data), re-estimate the hedonic price $\widehat{P}$, the associated potential gains and home equity, and all values of the moments, and re-estimate the structural parameters for each such vector of moments. In each bootstrap draw, we also re-estimate $\alpha(\ell)$ and $\beta(\ell)$ alongside the structural parameters to account for this source of estimation error.

We start with an equal-weighting scheme in which $\mathbf{W}$ is the identity matrix. In order to avoid a situation in which the scale of the moments influences their weighting in this case, we normalize each binned value for each moment by its average value across all bins. Put differently, we evaluate the model fit in terms of relative prediction errors. We later check the robustness of our inferences to inverse-variance-weighting, i.e., we use the bootstrap draws to construct a diagonal matrix of moment variances $\mathbf{V}$, and set $\mathbf{W} = \mathbf{V}^{-1}$. Online appendix section B.8 describes our numerical optimization procedure in greater detail.

5.4 Identification and Robustness

The model is complex, and there are many different patterns in the data. To understand the forces in the model and the sources of parameter identification, we therefore approach full structural estimation gradually, building from more simple model variants with fewer parameters to more complicated models, and adding subsets of moments as we go along. This culminates in estimating the entire set of model parameters using the full set of 486 empirical moments. Table 1 shows the 10 model variants that we estimate in this fashion, listed in rows. The columns labelled "Parameters" show the specific subsets of parameters estimated in each model along with associated bootstrap standard errors, the column labelled "Model structure" indicates whether or not we allow for concave demand, or assume linear (i.e., linear, not concave α) demand, and the remaining columns indicate the subset of moments used to estimate the parameters in each row.

Model variant 1 only uses the "hockey stick" moments in Figure 2. We match these moments using a stripped-down version of the model involving a representative seller with no financial constraints, in a market with linear demand.[52] In this special case, we solve analytically for the optimal ℓ^* (see online appendix section B.1), and find that η drives the listing premium slope when $\widehat{G} > 0$; λ generates a difference in slope in the domain $\widehat{G} < 0$; and the level of ℓ^* is driven by both η and θ_m, the level of gains from trade (i.e., the moving "shock" in this simple case). The first row of Table 1 shows that this model delivers $\eta = 0.325$, $\lambda = 2.151$ and $\theta_m = 0.489$, all of which are highly statistically significant.

We then check how correctly modelling demand using the concave α function visible in Figure 4 affects parameter estimates. When we incorporate this feature, Model 2 in Table 1 shows that $\eta = 1.030$ is estimated higher, loss aversion $\lambda = 1.297$ is lower, and the fitting parameter θ_m adjusts upwards to a level of 0.806.[53] As surmised, accounting for concave demand substantially decreases measured loss aversion (which nonetheless remains statistically significantly greater than 1), but it also increases measured reference dependence. This is because the relative "flattening out" of ℓ^* when $\widehat{G} > 0$ (and thus, the asymmetry between the slopes in the gain and loss domains) can partly be explained by optimal responses to demand concavity rather than via loss aversion. This in turn

[52] The precise slope of demand in this case is plotted as a straight dotted line in Figure 4.

[53] When η adjusts upwards, this generates a rotation in the listing premium—which now has a steeper slope. Over the relevant domain, this also manifests in a measured higher *level* of the listing premium. To reduce this level to fit the average listing premium, θ_m must rise. The analytical solution in the online appendix section B.1 provides more detail on this mechanism.

increases the degree of reference dependence needed to explain the negative slope of ℓ^* when $\widehat{G} < 0$.

Listing premia are ex-ante choices by sellers. An alternative ex-post measure is offered by bunching in the distribution of realized prices in actual transactions. Model 3 estimates reference dependence and loss aversion parameters only using bunching as a sufficient statistic. Using this moment substantially increases the precision of the λ estimate, but the point estimate $\lambda = 1.079$ is now *lower*—while the excess mass is clearly visible in the "spike" at $G = 0$ in Figure 6, the magnitude of this spike is consistent with a modest degree of loss aversion.

This strict interpretation of the model means that λ is set to match the spike at $G = 0$, but the model cannot explain the diffuse mass to the right of zero. Model 4 includes this diffuse bunching mass into our estimates (under the assumption of optimization frictions as described earlier). We do so by simply adding all excess mass above the counterfactual to the right of $G = 0$ in Figure 6 to $G = 0$. When we re-estimate parameters under this assumption, as expected, the model matched to this synthetically generated moment in addition to the others produces a lower $\eta = 0.525$ with a higher degree of loss aversion $\lambda = 1.750$. In model variants 5–9 below, we do not account for diffuse bunching in this manner, focusing mainly on the sharp spike at $G = 0$. We then return to including diffuse bunching in Model 10.

Model 5 re-estimates the parameters of seller utility using (sharp) bunching and the hockey stick moments together. This model estimates lower $\lambda = 1.091$, tightly controlled by the bunching seen precisely at zero gains. This lower parameter decreases the fit to the slope of ℓ^* when $\widehat{G} < 0$. This means that a higher $\eta = 1.143$ is now required to match the hockey stick (i.e., the curve rotates, giving up fit in the domain $\widehat{G} > 0$ to better match the domain $\widehat{G} < 0$).

Model 6 in Table 1 adds in the cross-sectional (i.e., regional) moments in Figure 2 Panel B. Here, estimated $\lambda = 1.079$ shrinks once again, though the estimates still strongly reject $\lambda = 1$. In the data, listing premia in municipalities in which demand is very close to linear do not show significant variation in slopes across gain and loss domains; and in municipalities in which demand is highly concave, the listing premium is non-linear and precisely consistent with the non-linearity of demand in those locations. This is consistent with $\lambda = 1$ and all nonlinearity in listing premia being driven by demand concavity. However, the strong rejection of $\lambda = 1$ arises from the sharp bunching seen in Figure 6.

An unrealistic feature of all model variants considered thus far is the assumption of a representative seller, with a single value of the moving "shock" $\theta = \theta_m$. This is very restrictive, because it forces loss averse preferences to generate a particular form of non-linearity, with two kinks in the predicted hockey stick precisely located at fixed levels of potential gains. Model 7 allows for heterogeneity in θ, allowing θ to vary according to a normal distribution with mean θ_m and standard deviation θ_σ. This means that the locations of the predicted kinks vary with the level of θ, and the resulting profile of ℓ^* predicted by the model is more smooth, though still non-linear, once averaged across different θ realizations. Estimated $\theta_m = 1.076$ in this model variant, and heterogeneity $\theta_\sigma = 0.591$.[54] This is accompanied by a higher $\lambda = 1.206$. Reference dependence $\eta = 1.193$ once again rises.[55]

Model 8 in Table 1 adds down-payment constraints into the model, and simultaneously adds in the variation of the listing premium along the home equity dimension as additional moments. This change does not greatly affect estimated $\lambda = 1.224$, but $\eta = 0.890$ decreases slightly, since variation in potential home equity and downsizing aversion is able to explain some of the variation in listing premia previously attributed solely to potential gains. Moreover, estimated $\theta_\sigma = 0.767$ rises, as the model attempts to match the observed smoothness of listing premia along the home equity dimension. Model 8 shows that the estimated penalty parameter for the home equity constraint is $\mu = 4.085$ (similar to its level in "final" Models 9 and 10), which is empirically realistic.[56]

In the model variants estimated thus far, θ_σ is only pinned down by matching the degree of "smoothness" of the listing premium profiles seen in the data. A more precise and model-consistent way to estimate θ_σ (and indeed, an important reason for its inclusion in the model) is to use it to help match the extensive margin, i.e., the slope of the listing probability along the $\widehat{G}$ dimension. More specifically, the extensive margin allows us to

[54] To interpret magnitudes, note that this is normalized relative to the hedonic value of the house, meaning that the lifetime contribution to utility discounted back to the present is 1.076 of the hedonic value of the house, or equivalently 6.0% per year, assuming a discount rate equal to 3.80%–the average return to long-term mortgage bonds between 2009 and 2016, and a holding period of 30 years.

[55] The effect of higher dispersion in θ is asymmetric: A higher level of the moving shock is decreasing the listing premium, and bringing it into a flatter region of concave demand, where the response of the seller to reference dependence is more muted. Hence a larger value of η is necessary to match the observed slope in the data.

[56] This coefficient provides an important validation opportunity for the model. Contrast a seller with a potential home equity level of 20% with one with 5%. The former can finance their mortgage at an annual interest rate of x% (x= 0.85%, as of September 2020), while the latter can finance 80% of their loan with the same rate, and the remaining 15% with an annual interest rate of x+e% (e = 4.34% as of September 2020, assuming all financing is done in the Danish unsecured "top-up" mortgage financing market described earlier and in the online appendix, section D.2). If the property value is normalized to 1, the cost arising from violating the constraint equals $4.34\% \times 0.15 = 0.65\%$ per year. Assuming a discount rate of 3.80%, the NPV of the total lifetime cost is equal to 7.3% for a 15-year contract and 10.38% for a 25-year contract. In the model, the equivalent net lifetime penalty from violating the down-payment constraint is estimated to be equal to $4.085 \times (0.15)^2 = 9.19\%$, roughly at the middle of the interval.

jointly identify two parameters, namely, θ_σ and ϕ, through the level and the slope of the listing probability profile. Model 9 adds in the extensive margin moments in Figure 7, and allows for ϕ in the model. In this model, estimated $\theta_m = 1.221$ meaning that the average moving "gain" is roughly 6.8% of the value of the house per year, $\theta_\sigma = 0.682$, and search costs $\phi = 0.035$, i.e., roughly 3.5% of the property price, which is once again empirically realistic.[57] Figure 8 graphically shows how Model 9 fits selected moments.[58]

Finally, the last row of the table, Model 10, reports structural parameter estimates when we allow for diffuse bunching in Model 9. This results in a magnitude of loss aversion $\lambda = 2.025$, which we view as an upper bound, and a magnitude of reference dependence $\eta = 0.545$, which we view as a lower bound. Online appendix Figure A.9 shows how Model 10 fits selected moments—we note that in some dimensions, such as the excess mass on the right tail of the bunching distribution, this model is more visually appealing in terms of fit than Model 9.

Overall, while the model is well able to capture the broad shapes of the empirical patterns in the data, we note that there are areas in which the model falls short. In the next section, we dig deeper into validating the model, identifying where it is unable to match the data, considering the usefulness of model extensions, and discussing what we learn from any gaps that remain between the model and the rich variation seen in the data.

6 Validating the Model

6.1 Variation Across Local Demand Conditions

The top row in Table 2 excludes the municipality-level moments when estimating parameters. Apart from a (statistically insignificant) decrease in search costs to $\phi = 0.018$, the parameters of the model are broadly unchanged from Model 9. We use the parameters from this version of the estimation, which is fitted using "national" demand conditions, to document the out-of-sample prediction capacity of the model, in particular, its ability to predict regional variation in seller behavior given local demand conditions. Figure 9 plots the data in each of the two locations with high vs. low levels of demand concavity

[57] We find that the distribution $\theta \sim N(\theta_m, \theta_\sigma)$ has parameters $\theta_m = 1.221$ and $\theta_\sigma = 0.682$. These "moving shocks" correspond to the present discounted value of lifetime future benefits from successfully selling and/or moving. We also note here that the magnitude of search costs should not strictly be interpreted as the monetary cost of listing, but as a more general wedge between the value of listing and the outside option, potentially also including soft factors like the hassle of allowing visitors in the house, the value of regret, and more generally factors which may affect the seller's utility value of the property in the event that the listing is not successful.

[58] The models are non-nested as they have different sets of parameters and moments, so we cannot compare them directly in terms of goodness-of-fit.

against the model-implied listing premium, when the model is fed the degree of demand concavity in each local area.[59] The figure shows that there is a remarkable ability to fit the degree of "pass through" of demand concavity to the listing premium, which is useful validation of the broad structure of the model.

6.2 Home Equity

Online appendix Figure A.11 shows that with our specification of financial constraints we are well able to capture the distribution of realized levels of home equity despite this not being included as a moment in estimation.[60]

6.3 Measurement Error and Robustness

We earlier discussed the confound posed by measurement error and unobserved quality, as well as several empirical strategies that we use to deal with this confound, which can affect $\widehat{P}$ and hence a large number of the moments used in estimation. The second to fourth rows of Table 2 alter the estimation of $\widehat{P}$ to account for different sources of measurement error, recompute all affected moments, and then re-estimate structural parameters following these changes. Model 'Renovations' augments the baseline hedonic model in equation 6 with lagged information on tax exemptions sought by owners for renovation expenses to capture time-varying unobserved heterogeneity associated with renovations. Model 'Repeat Sales I' is a pairwise repeat sales model which includes time-varying hedonic characteristics and lagged renovation expenses. This is accomplished by including the lagged pricing residual on the same house's previously traded price, a strategy similar to Genesove and Mayer (2001). Model 'Repeat Sales II' generalizes the repeat sales approach, and includes the average of *all* past pricing residuals available since 1992 for each house where available, to use information from all past repeat sales observed, in addition to hedonic characteristics and lagged renovation expenses.[61] Reassuringly, in all cases, relative to Model 9 in Table 1, the point estimates of most parameters remain similar, with no material changes to the qualitative interpretations from that model.[62]

[59]We adjust the level of the model prediction to correspond to the data, since the level of the listing premium depends on other parameters, most notably the average size θ_m of the moving shock, which may be heterogeneous across locations.

[60]Since we assume a continuous and differentiable penalty function, the mass of realized gains around and above a level of home equity of 20% does not take the form of a spike, but is continuous, consistent with the data.

[61]A more comprehensive discussion of the rationalization for and implementation of these models is in the online appendix section I.

[62]One visible change concerns the estimated magnitude of down-payment constraints. Compared to Model 9 in Table 1, the estimated values of the parameter μ in the case of hedonic models "Repeat Sales I" and "Repeat Sales II" appear slightly lower. We can attribute this to the fact that the repeat sales sample is restricted (it contains only properties for which multiple past sales are observed), and happens

Finally, the fifth row of Table 2 replaces the equal-weighting of moments in the data with inverse-variance weights derived from the bootstrap. In this case, while our inferences about λ do not greatly change, η comes in lower. This is partly because the model does not need to closely match the significant listing premium slopes at extreme gains and losses—which are more noisily measured than observations closer to $\widehat{G} = 0$.

6.4 Alternative Specifications

Figure 6 also shows a clear volume of missing mass for realized gains in the one-percentage point bin just below zero. In the bottom two rows of Table 2, we extend the model with an additional feature to capture the sellers' aversion to losses, a discontinuous jump of magnitude ζ for such realized gains just below zero:

$$u(P(\ell), R) = \begin{cases} P(\ell) + \zeta + \lambda\eta G(\ell), & \text{if } G(\ell) < 0 \\ P(\ell) + \eta G(\ell), & \text{if } G(\ell) \geq 0 \end{cases}. \tag{8}$$

The estimated parameters (reported in the bottom two rows of Table 2) are barely affected by this alternative specification of seller preferences, except that some of the excess mass exactly at zero can now be attributed to the effect of the discontinuity, and therefore the value of λ decreases slightly. The estimated notch parameter ζ is equal to -0.129% ($s.e.$ = 0.042) of the property's hedonic value in the baseline case. When we account for diffuse bunching, the magnitude is somewhat larger (-0.216%), but very imprecisely estimated ($s.e.$ = 0.223).[63] The notch leads to accumulation of mass precisely to the left of $G = 0$, meaning that the model is only able to attribute the sources of excess mass correctly in any particular bootstrap draw, if there is a coincidence between missing mass to the left of zero and excess mass exactly at zero, above and beyond what can be rationalized by λ. This leads to high-variance and left-skewness in the distribution of bootstrap point estimates. Overall, we conclude that the estimated magnitude of the notch parameter suggests a modest quantitative impact of this force around the reference point.

6.5 Interactions Between Preferences and Constraints

The left-hand panel of Figure 10 compares the model predicted and observed listing premium hockey sticks, as the potential home equity position of the seller varies. This is an attempt to see if the model is able to capture the 3-d patterns of listing premia

to have a greater number of younger, urban households with high net financial assets—consistent with lower down-payment constraints for this sample.

[63] The patterns in the data in terms of slope variation and missing mass in the distribution of realized outcomes are also broadly consistent with a standard formulation of diminishing sensitivity (see Rees-Jones, 2018; Allen et al., 2017).

observed in Figure 3. In the data, there is considerable variation in the slope of the relationship between $\hat{\ell}$ and $\widehat{G}$ that depends on $\widehat{H}$. Put differently, in the data, 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.

However, the model is unable to capture this variation in slope. The right-hand panel of Figure 10 checks the statistical validity of this observation, albeit in reduced-form, showing estimation results from a regression of listing premia on potential gains, conditioning the coefficient on the level of potential home equity. The plot confirms that less constrained sellers are likely to respond more strongly in their listing decision to their gain/loss position. 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 along the $\widehat{G}$ dimension.[64]

Qualitatively, such interactions between preferences and constraints can emerge in our model through the role of concave demand, as we describe in greater detail below when discussing the model-implied price-volume correlation, and in online appendix section B.8. However, quantitatively this force is far from being able to match the patterns in the data. To test this conclusion more formally, we repeat the exercise in the right-hand panel of Figure 10 after conditioning on the homogeneity of the housing stock in a given location (a primary driver of demand concavity). We confirm that the point estimates remain unaffected.

Overall, this new fact on the interaction of preferences and constraints appears 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 type of model that may rationalize these findings here, with a view towards motivating theoretical work on a broader class of preference and constraint specifications. Our preferred rationalization 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). We discuss this further in the conclusion to the paper.

[64]Section F of the online appendix discusses the ℓ - $\widehat{H}$ relationship as $\widehat{G}$ varies, which suggests that households facing nominal losses may feel more 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).

6.6 Price-Volume Correlation

A positive correlation between prices and volumes arises endogenously in the model. Consider a positive shock to housing "fair value," say as a result of a demand shock for housing. This increases the mean of the distribution of sellers' potential gains relative to their reference points. There are two ways that higher potential gains lead to increases in selling activity in the model. First, along the intensive margin (i.e., conditional on listing), it is rational for reference-dependent (and loss-averse) sellers to react to the increase in potential gains by decreasing listing premia ℓ^*. This in turn leads to higher sale probabilities through $\alpha(\ell^*)$, thus increasing the number of realized transactions in the market. Second, along the extensive margin, more properties will be listed for sale in the housing market. Even homeowners with low values of the moving shock θ who did not find it optimal to list, now seek to list and benefit from the higher price they expect to realize from a successful transaction.

A similar effect operates through down-payment constraints, since the same shock increases the mean of the distribution of sellers' potential home equity relative to their constraint points—once again leading to effects on listing premia and transaction volumes.

Formally, consider one-percentage point bins i of potential gains $\widehat{G}_i = \widehat{P}_i - R_i$ and potential home equity $\widehat{H}_i = \widehat{P}_i - M_i$, for which the number of properties in the housing stock is given by $N_{stock,i}$. Given the model-implied optimal probability $s^*(\widehat{G}_i, \widehat{H}_i)$ that any such property will be listed for sale, and the probability $\alpha(\ell^*(\widehat{G}_i, \widehat{H}_i))$ that conditional on listing, an actual transaction will get realized—implied by the optimal listing premium plugged into the demand function—the number of observed realized transactions N_i is equal to:

$$N_i = s^*(\widehat{G}_i, \widehat{H}_i) \times \alpha(\ell^*(\widehat{G}_i, \widehat{H}_i)) \times N_{stock,i}. \tag{9}$$

For any given distribution of potential gains and potential home equity in the property stock, equation (9) allows us to calculate average transaction volumes $N = \sum_i N_i$, and therefore a model-implied mapping between hedonic valuations and transaction volumes.

To understand whether our model is able to capture the empirically observed price-volume correlation, we first estimate the coefficient ρ from the following regression in the data:

$$\Delta \ln N_{m,t} = \mu + \rho \Delta \ln \widehat{P}_{m,t} + \epsilon_{m,t}, \tag{10}$$

where $N_{m,t}$ are the numbers of transaction in municipality m in year t, and $\widehat{P}_{m,t}$ is the corresponding average hedonic price level in municipality m in year t. The correlation between prices and volumes measured this way is reported in the left-most column of Panel A in Table 3, and equals $\rho = 0.538$.

We then use the observed change in valuations $\Delta\widehat{P}_{m,t}$ to calculate model-implied changes in transaction volumes $\Delta\widetilde{N}$ implied by equation (9), and re-estimate the model-implied correlation coefficient by substituting $\Delta\widetilde{N}_{m,t}$ into equation (10).[65]

We start with the set of structural parameters estimated in Model 9 in Table 1, with $\eta = 0.871$ and $\lambda = 1.220$. This delivers a price-volume correlation equal to $\rho = 0.417$, which is a modest under-prediction of the value of ρ identified in the data. A similar prediction of $\rho = 0.401$ obtains when we evaluate the model at the set of parameters identified in the "Diffuse bunching" case, where $\eta = 0.545$ and $\lambda = 2.025$. Interestingly, with lower η and higher λ, the magnitude of the co-movement is *lower*—this is because loss aversion affects only properties for which $\widehat{G} < 0$, but reference dependence affects the entire set of properties, including those in the gain domain, accounting for a larger share of the data.

Panels B and C of Table 3 attribute the model-implied correlation to the different model ingredients. We begin by calculating the price-volume correlation by ignoring the extensive margin decision, and later, add it back in. This allows us to evaluate the size and importance of the extensive margin decision for the price-volume correlation (i.e., the fact that constrained or loss-averse sellers withhold properties from the market) relative to the intensive margin (the changes in list price-setting behavior, conditional on listing). Since the contribution of each ingredient to the magnitude of the price-volume correlation is not additively separable, we report results for two alternative orderings (shown in Panels B and C) of the sequential inclusion of structural features.[66]

In the first rows of Panels B and C, we confirm that in a frictionless version of the model with no financial constraints ($\mu = 0$) and no reference dependence ($\eta = 0$), transaction volumes are completely independent of price movements. We then add model features successively along the intensive margin, beginning with financial constraints in Panel B, and reference dependence and loss aversion in Panel C. We find that on average across the two cases, down-payment constraints account for 64% of the co-movement, reference dependence constitutes the lion's share of the remaining effect at 30%, and loss aversion accounts for a modest 5%. When we add back in the extensive margin, we find that

[65]We assume that the parameterization of the model is representative for all municipalities and time periods. This is consistent with our estimate of a common coefficient ρ in the data.

[66]Sellers face concave demand, which makes it more costly to react to any additional incentive to list more aggressively. While preferences and constraints contribute additively to seller utility, the presence of the nonlinear "market constraint" imposed by concave demand generates the lack of additive separability in terms of the effects on the optimal listing premium. For example, the response of a constrained seller to a switch from $\eta = 0$ to $\eta > 0$ will be more muted compared to an identical unconstrained seller. A similar intuition applies in the case of a reference-dependent seller who becomes more constrained. We discuss this issue further in online appendix section B.8.

it contributes roughly 1%, which combines all preference and constraint effects on the propensity to list–a reflection of the modest relationship observed in the data between the level of potential gains and the probability of listing. If we use the "diffuse bunching" estimates which allow for a higher level of λ, this increases the marginal contribution of loss aversion to the price-volume correlation to an average level of 16%, it reduces the contribution of reference dependence to 23%, and slightly decreases the magnitude of the down-payment contribution to the price-volume correlation to 60%.

The important role of down-payment constraints for the aggregate dynamics of prices and volumes is not surprising, given that constrained sellers are very prevalent in the data–they account for a large share of 47% of all sellers with a mortgage outstanding, which translates into 35% of the entire sample. For comparison, a much lower share of 23% of sellers face the possibility of realizing a loss. But more importantly, we also find that in the data constrained sellers respond more strongly to an improvement in their home equity position than loss-averse sellers do to the possibility of realizing a loss. On average, the listing premium decreases by 0.37% in response to a 1% increase in potential home equity, and by only 0.26% for a 1% increase in potential gains.

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.

Our parameter estimates are consistent with a high degree of reference dependence in house sellers; they appear to care greatly about nominal gains and losses relative to the original purchase price. However, our parameter estimates of loss aversion in this important field setting, while statistically significant, are somewhat smaller than those discovered in the lab. While preferences do appear to be kinked, our point estimates for the disutility contribution of losses lie between 1.22 and 2.03 times the utility contribution of gains.

Our model cannot completely match novel facts from the data, which we view as a new target for behavioral economics theory. Nominal losses and down-payment constraints interact with one another, in the sense that reference-dependence is less evident when households face more severe constraints, but quite pronounced when households are relatively unconstrained. 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, 2006; Ashraf, Karlan and Yin, 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.

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). From the perspective of these more macro issues, our work offers several contributions. We find that down-payment constraints and reference dependence (and less so, loss aversion) are important for understanding these aggregate housing market dynamics, and offer a quantitative assessment of the relative importance of these underlying drivers. In future work, incorporating interaction effects such as the effect of expected losses on the household response to constraints into such a model could more accurately capture the seemingly extreme reactions of housing markets to apparently small changes in underlying prices, and further inform mortgage market policy (Campbell, 2012; Piskorski and Seru, 2018).

References

Allen, Eric J, Patricia M Dechow, Devin G Pope, and George Wu. 2017. "Reference-dependent preferences: Evidence from marathon runners." *Management Science*, 63(6): 1657–1672.

Anagol, Santosh, Vimal Balasubramaniam, and Tarun Ramadorai. 2018. "Endowment effects in the field: Evidence from India's IPO lotteries." *The Review of Economic Studies*, 85(4): 1971–2004.

Andersen, Steffen, John Y Campbell, Kasper Meisner Nielsen, and Tarun Ramadorai. 2020. "Sources of Inaction in Household Finance: Evidence from the Danish Mortgage Market." *American Economic Review*, 110(10).

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

Ashraf, Nava, Dean Karlan, and Wesley 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, Cristian, John Y Campbell, and Tarun Ramadorai. 2016. "International Comparative Household Finance." *Annual Review of Economics*, 8(1).

Barberis, Nicholas C. 2013. "Thirty years of prospect theory in economics: A review and assessment." *Journal of Economic Perspectives*, 27(1): 173–96.

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

Bokhari, Sheharyar, and David Geltner. 2011. "Loss aversion and anchoring in commercial real estate pricing: Empirical evidence and price index implications." *Real Estate Economics*, 39(4): 635–670.

Bracke, Philippe, and Silvana Tenreyro. 2020. "History dependence in the housing market." *Bank of England Working Paper.*

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor supply of New York City cabdrivers: One day at a time." *The Quarterly Journal of Economics*, 112(2): 407–441.

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

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

Card, David, Andrew Johnston, Pauline Leung, Alexandre Mas, and Zhuan Pei. 2015*a*. "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, David, David S Lee, Zhuan Pei, and Andrea Weber. 2015*b*. "Inference on causal effects in a generalized regression kink design." *Econometrica*, 83(6): 2453–2483.

Chetty, Raj, John N Friedman, Tore Olsen, and Luigi 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, John M, Ran Lu-Andrews, and Tingyu 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, Vincent P, and Juanjuan 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, Stefano. 2009. "Psychology and economics: Evidence from the field." *Journal of Economic Literature*, 47(2): 315–72.

DellaVigna, Stefano. 2018. "Structural behavioral economics." *National Bureau of Economic Research Working Paper.*

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

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F Schmieder. 2017. "Reference-dependent job search: Evidence from Hungary." *The Quarterly Journal of Economics*, 132(4): 1969–2018.

Einiö, Mikko, Markku Kaustia, and Vesa Puttonen. 2008. "Price setting and the reluctance to realize losses in apartment markets." *Journal of Economic Psychology*, 29(1): 19–34.

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

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

Farber, Henry S. 2008. "Reference-dependent preferences and labor supply: The case of New York City taxi drivers." *American Economic Review*, 98(3): 1069–82.

Fehr, Ernst, and Jean-Robert Tyran. 2005. "Individual irrationality and aggregate outcomes." *Journal of Economic Perspectives*, 19(4): 43–66.

Fehr, Ernst, and Lorenz Goette. 2007. "Do workers work more if wages are high? Evidence from a randomized field experiment." *American Economic Review*, 97(1): 298–317.

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

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

Garicano, Luis, Claire Lelarge, and John Van Reenen. 2016. "Firm size distortions and the productivity distribution: Evidence from France." *American Economic Review*, 106(11): 3439–79.

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

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

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai. forthcoming. "Household finance." *Journal of Economic Literature*.

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

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

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

Imas, Alex, Sally Sadoff, and Anya Samek. 2017. "Do people anticipate loss aversion?" *Management Science*, 63(5): 1271–1284.

Kahneman, Daniel. 2003. "Maps of bounded rationality: Psychology for behavioral economics." *American Economic Review*, 93(5): 1449–1475.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica*, 263–291.

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

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics*, 8: 435–464.

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

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-dependent risk attitudes." *American Economic Review*, 97(4): 1047–1073.

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

List, John A. 2003. "Does market experience eliminate market anomalies?" *The Quarterly Journal of Economics*, 118(1): 41–71.

List, John A. 2004. "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica*, 72(2): 615–625.

Mead, Roger. 2017. *Statistical methods in agriculture and experimental biology.* Chapman and Hall.

Merlo, Antonio, and Francois Ortalo-Magne. 2004. "Bargaining over residential real estate: evidence from England." *Journal of urban economics*, 56(2): 192–216.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher 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, Tomasz, and Amit Seru. 2018. "Mortgage market design: Lessons from the Great Recession." *Brookings Papers on Economic Activity*, 2018(1): 429–513.

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

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

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

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

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297–323.

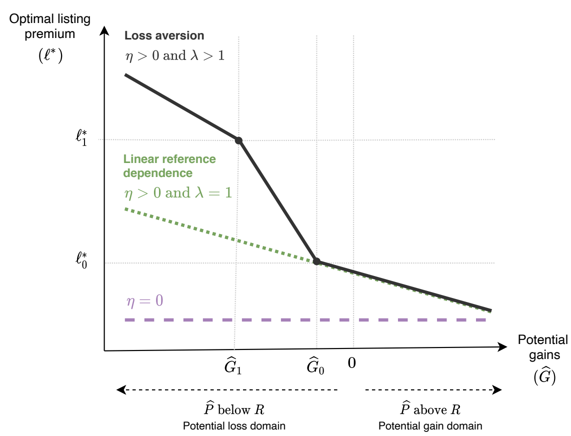
Zwietering, MH, Il Jongenburger, FM Rombouts, and K Van't Riet. 1990. "Modeling of the bacterial growth curve." *Appl. Environ. Microbiol.*, 56(6): 1875–1881.

Figure 1

Reference Dependence and Loss Aversion

The figure illustrates how each specification of the utility function is reflected in the seller's optimal choice of listing premia (left-hand side panel) and distribution of realized gains (right-hand side panel). 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.

Optimal Listing Premium by Potential Gains



Realized Gains Distribution

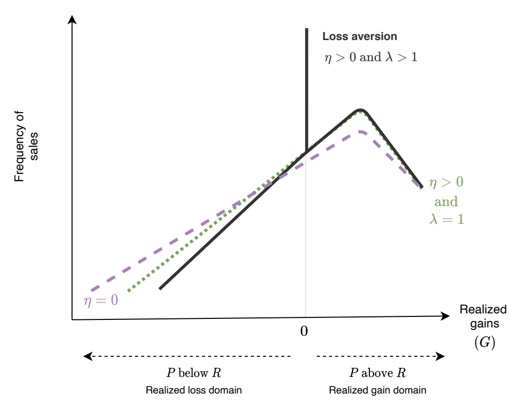
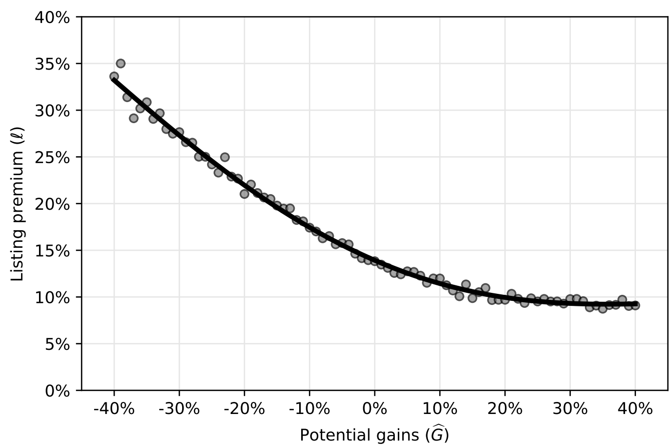


Figure 2

Listing Premia and Potential Gains

Panel A reports binned average values (in 1 percentage point steps) for the listing premium (ℓ) for different levels of potential gains ($\widehat{G}$). The solid line corresponds to a polynomial fit of order three. Panel B shows demand concavity (left-hand side panel), i.e. the probability of sale within six months with respect to the listing premium, and the listing premium over gains (right-hand side panel), when sorting municipalities by the degree of demand concavity, using municipalities in the top and bottom 5% of observations. The degree of demand concavity is estimated as the slope coefficient of the effect of the listing premium on the probability of sale within six months, for positive listing premia ($\ell \in [0, 40]$).

Panel A



Panel B

Cross-Sectional Variation

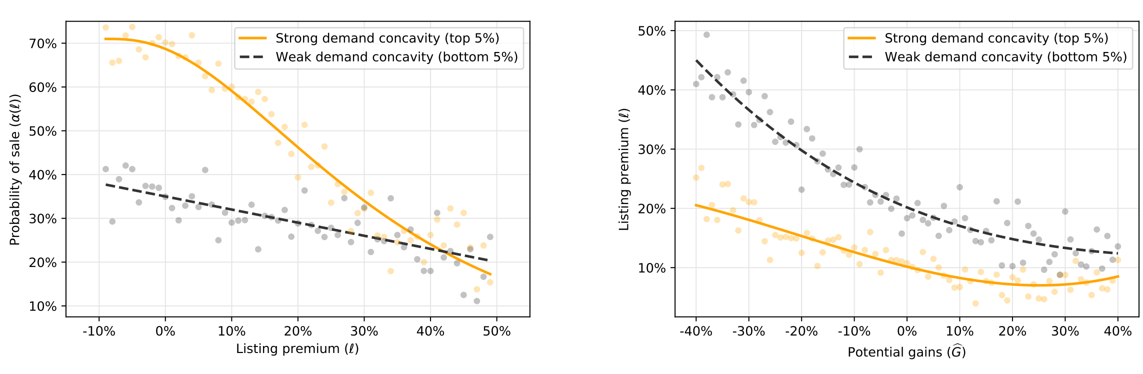


Figure 3

Listing Premia Across Gains and Home Equity

The figure reports binned average values (in steps of 3 percentage points) for the listing premium (ℓ) along both levels of potential gains and home equity.

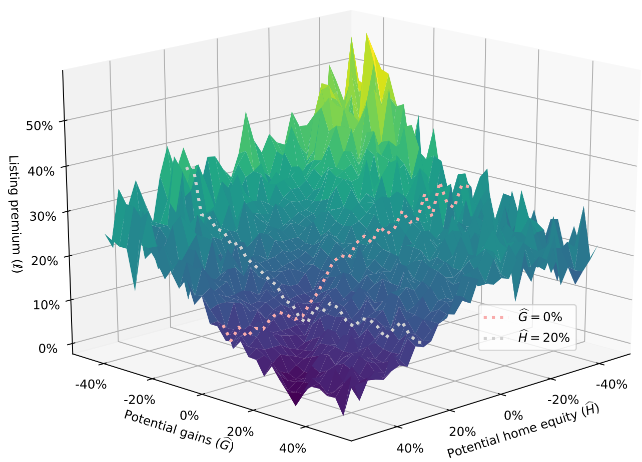


Figure 4

Concave Demand in the Data

The left-hand side 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 right-hand side of the figure shows the average realized premium $\beta(\ell)$ across 1 percentage point bins of the listing premium.

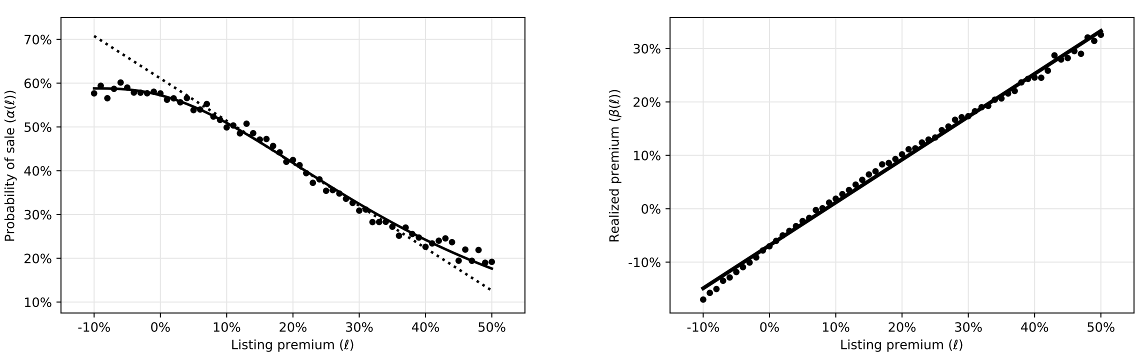


Figure 5

Geographic Variation in Concave Demand and Housing Stock Homogeneity

The figure plots the estimated degree of demand concavity, measured as the slope coefficient of the effect of an increase in the listing premium on the probability of sale within six months, for positive listing premia ($\ell \in [0, 40]$) on the y-axis, against the degree of homogeneity of the housing stock, measured as the share of apartments and row houses, across municipalities in Denmark.

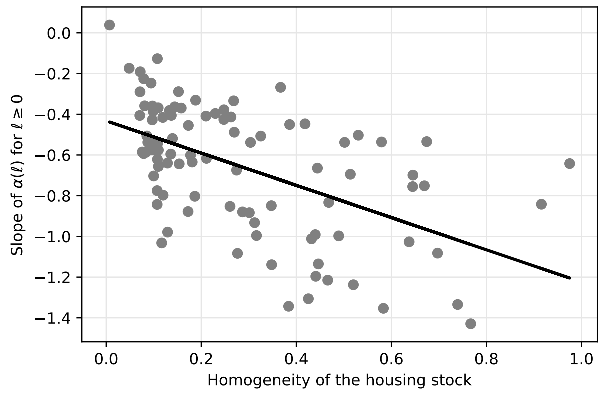


Figure 6

Bunching Around Realized Gains of Zero

The left-hand panel 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. The right-hand panel reports frequencies of observations of realized gains relative to the level of the counterfactual. Dotted lines indicate 95% confidence intervals based on bootstrap standard errors.

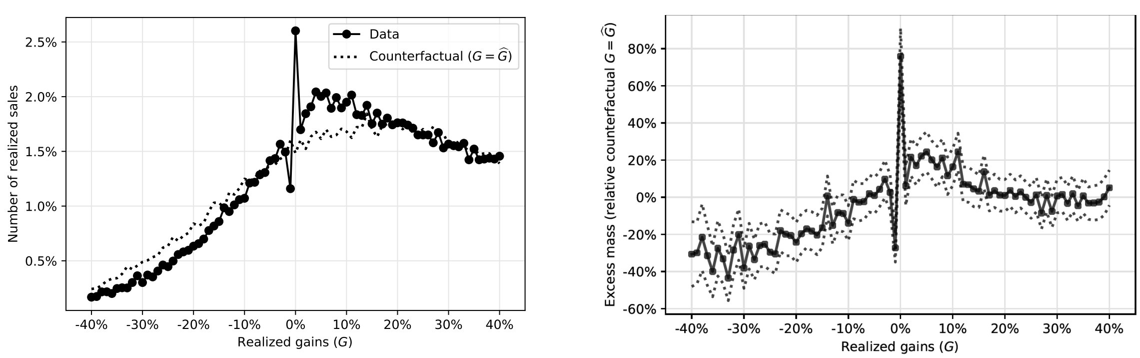


Figure 7
Extensive Margin

The figure reports the average annual 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, using the same hedonic model used to calculate potential gains in the sample of listings. We then divide the number of properties which have been listed for sale by the number of total property $\times$ year observations in the stock of properties, for each 1 percentage point bin of potential gains. Dotted lines indicate 95% confidence intervals based on bootstrap standard errors.

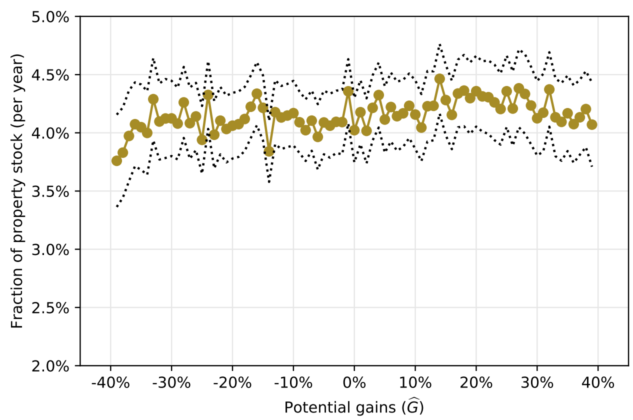


Figure 8

Overview of Model Fit

The figure reports our set of moments in the data and in the model, evaluated at the set of parameters which correspond to the complete version of the model and the complete set of empirical moments, as indicated in line 9 of Table 1. Dotted lines show 95% confidence intervals based on bootstrap standard errors.

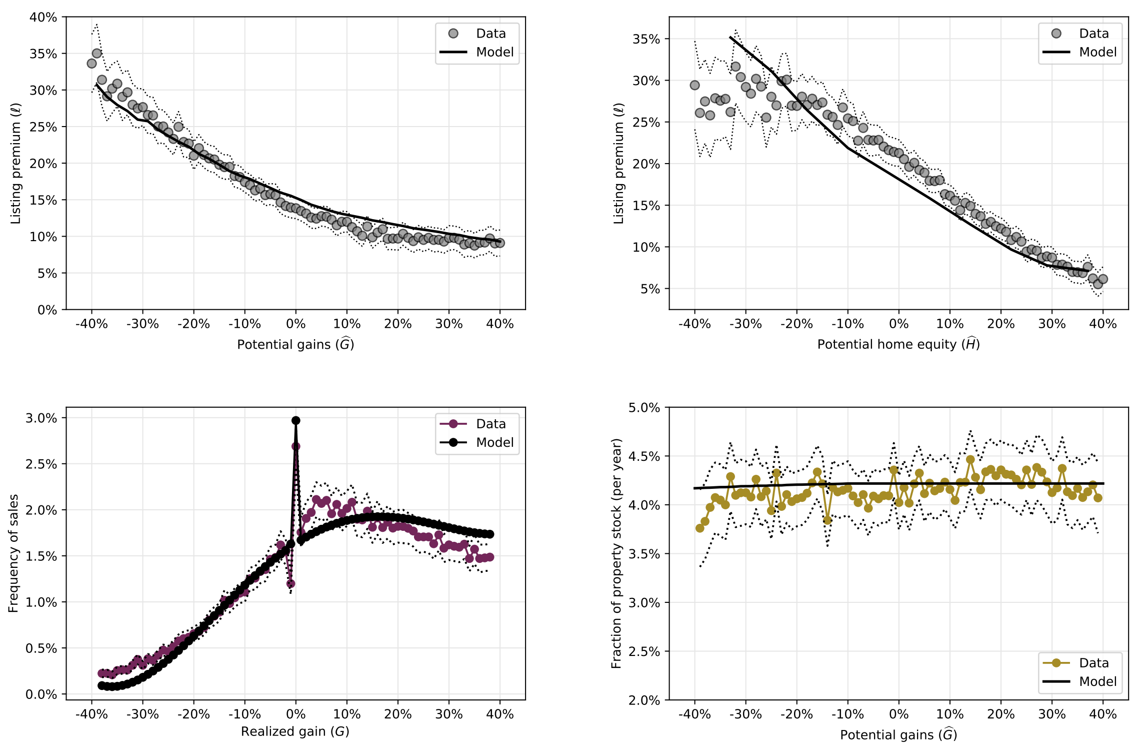


Figure 9

Out-of-Sample Fit: "Hockey Stick" across Municipalities

The figure reports binned average listing premia by potential gains, for two distinct groups of municipalities, sorted by the estimated degree of demand concavity, using municipalities in the top and bottom 5% of observations in the data. The degree of demand concavity is measured as the slope coefficient of the effect of an increase in the listing premium on the probability of sale within six months, for positive listing premia ($\ell \in [0, 40]$), for each municipality. We evaluate the model fit out-of-sample at the set of parameters which correspond to the complete version of the model, and the set of empirical moments indicated in the first row of Table 2. Dotted lines show 95% confidence intervals based on bootstrap standard errors.

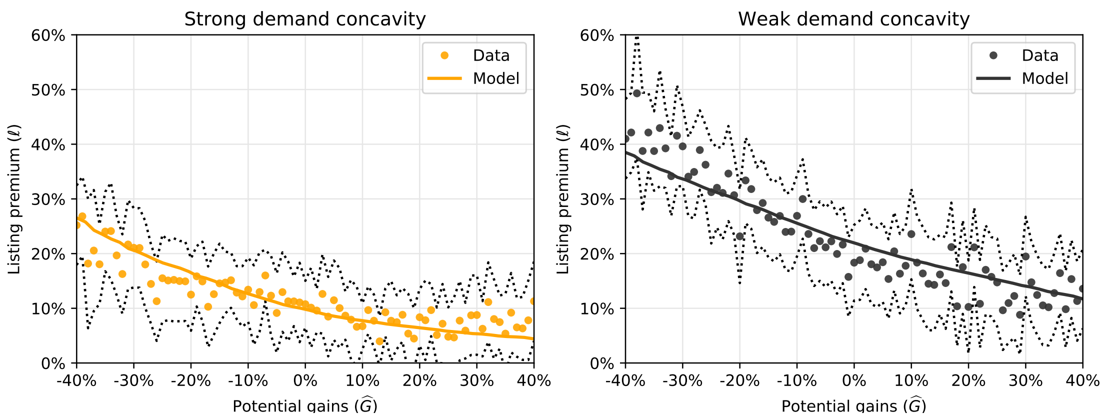


Figure 10

Out-of-Sample Fit: The Home Equity Dimension

The left-hand plot reports the model fit for conditional listing premia profiles, conditioning on different levels of home equity, evaluated at the set of parameters which correspond to the complete version of the model and the complete set of empirical moments, as indicated in line 10 of Table 1. Individual dots indicate observations in the data, and solid lines their model-implied counterparts. The right-hand plot shows a non-parametric estimate of the slope of the listing premium vs. potential gains over the entire range between -40% and 40%, for different levels of home equity.

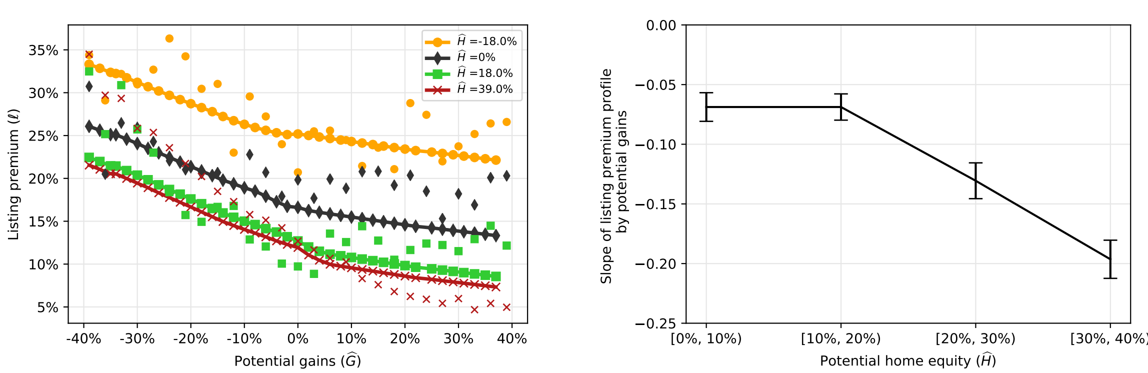


Table 1
Estimated Parameters

The table reports structural parameter estimates obtained through classical minimum distance estimation. We recover concave demand $\alpha(\ell)$ and $\beta(\ell)$ from the data and set the down-payment constraint $\gamma = 20\%$. Each row corresponds to a different model variant (1-10). We report the parameters estimated, for different model structures and moments used in the estimation. In parentheses, we report standard errors based on re-estimating the model across bootstrap draws, clustered at the shire level.

Model variant	**Parameters**						**Model structure**	**Moments used in estimation**					
	Reference dependence	Loss aversion	Distribution of moving shocks		Financial constraints	Search costs	Concave demand	Hockey stick	Bunching		Cross-sectional variation	Home equity	Extensive margin
	η	λ	θ_m	θ_σ	μ	ϕ			Sharp	Diffuse			
1	0.325	2.151	0.489				×	✓	×	×	×	×	×
	(0.025)	(0.071)	(0.012)										
2	1.030	1.297	0.806				✓	✓	×	×	×	×	×
	(0.028)	(0.050)	(0.027)										
3	1.093	1.079	1.036				✓	×	✓	×	×	×	×
	(0.038)	(0.011)	(0.046)										
4	0.525	1.750	0.896				✓	×	×	✓	×	×	×
	(0.136)	(0.229)	(0.069)										
5	1.143	1.091	0.833				✓	✓	✓	×	×	×	×
	(0.027)	(0.017)	(0.027)										
6	1.168	1.079	0.849				✓	✓	✓	×	✓	×	×
	(0.044)	(0.018)	(0.035)										
7	1.193	1.206	1.076	0.591			✓	✓	✓	×	✓	×	×
	(0.064)	(0.037)	(0.046)	(0.059)									
8	0.890	1.224	1.305	0.767	4.085		✓	✓	✓	×	✓	✓	×
	(0.045)	(0.029)	(0.021)	(0.053)	(0.195)								
9	0.871	1.220	1.221	0.682	3.909	0.035	✓	✓	✓	×	✓	✓	✓
	(0.036)	(0.028)	(0.021)	(0.038)	(0.104)	(0.011)							
10	0.545	2.025	1.180	0.689	3.156	0.024	✓	✓	×	✓	✓	✓	✓
	(0.078)	(0.245)	(0.036)	(0.067)	(0.091)	(0.022)							

Table 2

Alternative Model Specifications

The table reports structural parameter estimates obtained through classical minimum distance estimation. We recover concave demand $\alpha(\ell)$ and $\beta(\ell)$ from the data and set the down-payment constraint $\gamma = 20\%$. In parentheses, we report standard errors based on re-estimating the model across bootstrap draws, clustered at the shire level. In the row labelled "Out-of-sample cross-sectional variation" we re-estimate the model version in line 9 of Table 1, excluding the cross-sectional moments from the estimation. We then use three alternative specifications for the hedonic model of $\widehat{P}$. First, we include information on renovation expenses to proxy for potentially time-varying unobserved heterogeneity. Second, we estimate a pairwise repeat sales model with time-varying hedonic characteristics and lagged renovation expenses, by including the lagged past pricing residual, similar to Genesove and Mayer (2001). Third, we generalize the repeat sales approach and include the average of all past pricing residuals to use information from all past repeat sales observed, plus lagged renovation expenses. In the rows labelled "Notch" we allow for a discontinuous jump in preferences at a level of $G = 0\%$ realized gains.

Model variant	**Parameters**						
	Reference dependence	Loss aversion	Distribution of moving shocks		Financial constraints	Search costs	"Notch" (%)
	η	λ	θ_m	θ_σ	μ	ϕ	ζ
Out-of-sample cross-sectional variation	0.871	1.182	1.223	0.681	3.900	0.018	
	(0.031)	(0.023)	(0.012)	(0.035)	(0.097)	(0.007)	
Renovations	0.814	1.220	1.221	0.705	3.844	0.038	
	(0.033)	(0.030)	(0.018)	(0.033)	(0.087)	(0.011)	
Repeat sales I	0.827	1.134	1.336	0.653	3.521	0.058	
	(0.061)	(0.044)	(0.051)	(0.054)	(0.041)	(0.049)	
Repeat sales II	0.838	1.134	1.322	0.661	3.534	0.055	
	(0.064)	(0.038)	(0.052)	(0.048)	(0.058)	(0.055)	
Inverse variance	0.839	1.109	1.223	0.707	3.897	0.017	
	(0.065)	(0.028)	(0.011)	(0.028)	(0.175)	(0.007)	
"Notch" in preferences	0.866	1.182	1.223	0.715	3.834	0.036	-0.129
	(0.040)	(0.030)	(0.024)	(0.016)	(0.088)	(0.016)	(0.042)
"Notch" + Diffuse bunching	0.509	2.007	1.179	0.694	3.200	0.017	-0.216
	(0.077)	(0.225)	(0.041)	(0.056)	(0.091)	(0.030)	(0.223)

Table 3

Price-Volume Correlation in the Data and the Model

Panel A reports the estimated coefficient ρ from the following regression specification:

$$\Delta \ln N_{m,t} = \mu + \rho \Delta \ln \widehat{P}_{m,t} + \epsilon_{m,t},$$

where $N_{m,t}$ are transaction volumes and $\widehat{P}_{m,t}$ is the average hedonic price level in municipality m in year t. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively, based on standard errors clustered by year. We distinguish between the "Baseline" case (Model 9 in Table 1), and the "Diffuse bunching" case (Model 10 in Table 1). We repeat the calculation of the model-implied value of ρ, starting with a frictionless version of the model ($\eta = \mu = 0$), and sequentially adding structural ingredients. The column labeled "Cumulative" reports the magnitude of the price-volume correlation for the respective model version. We consider two alternative ways of ordering the sequential inclusion of model features: in Panel B, we start with the role of financial constraints, and in Panel C with the role of reference dependence and loss aversion. In the column labeled "Marginal", we report the share of the magnitude of the price-volume correlation ρ which can be attributed to the respective model ingredient.

Panel A

	Data	Model	
		Baseline	Diffuse bunching
Price-volume correlation (ρ)	0.538***	0.417	0.401
	(0.196)		

Panel B

Decomposition of ρ in the model	Baseline		Diffuse bunching	
	Cumulative	Marginal (%)	Cumulative	Marginal (%)
Frictionless model	0.000	-	0.000	-
+ Financial constraints ($\mu > 0$)	0.292	70.0%	0.263	65.5%
+ Reference dependence ($\eta > 0$)	0.394	24.5%	0.341	19.5%
+ Loss aversion ($\lambda > 1$)	0.412	4.2%	0.395	13.6%
+ Extensive margin decision	0.417	1.3%	0.401	1.4%

Panel C

Decomposition of ρ in the model	Baseline		Diffuse bunching	
	Cumulative	Marginal (%)	Cumulative	Marginal (%)
Frictionless model	0.000	-	0.000	-
+ Reference dependence ($\eta > 0$)	0.143	34.4%	0.103	25.7%
+ Loss aversion ($\lambda > 1$)	0.169	6.1%	0.177	18.5%
+ Financial constraints ($\mu > 0$)	0.412	58.2%	0.395	54.4%
+ Extensive margin decision	0.417	1.3%	0.401	1.4%

Reference Dependence in the Housing Market

Online Appendix

(For Online Publication)

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

January 26, 2021

*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.

Contents

List of Figures

List of Tables

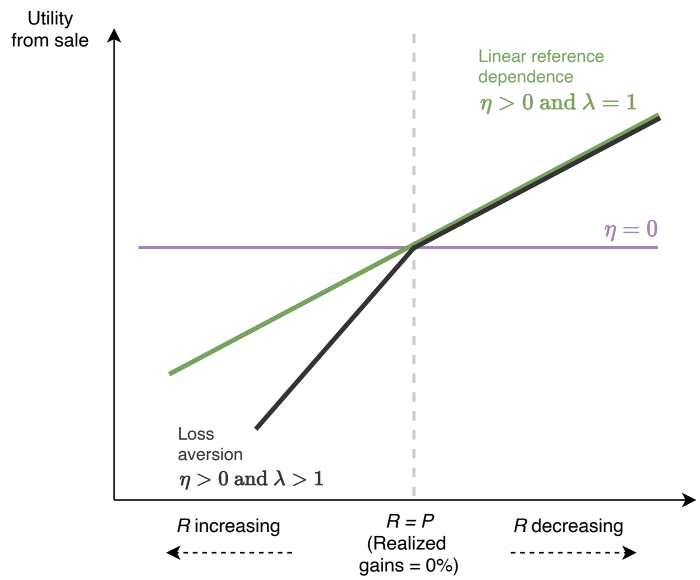
A Additional Tables and Figures

Figure A.1

Reference Dependence and Loss Aversion

The top plot 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. The bottom plot illustrates the mapping between potential gains on the horizontal axis, and realized gains on the vertical axis that result from the optimal choice of listing premia.

Seller Utility over Realized Gains



Realized Gains and Potential Gains given Optimal Choice of Listing Premia

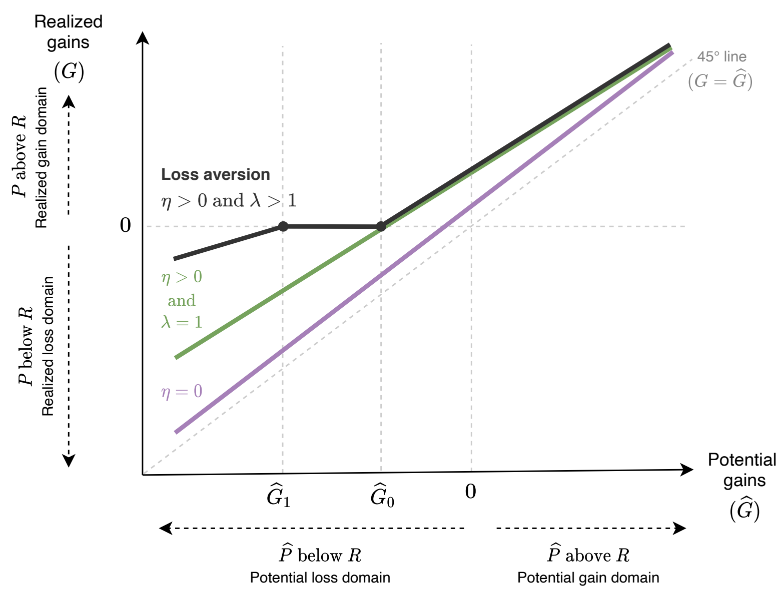


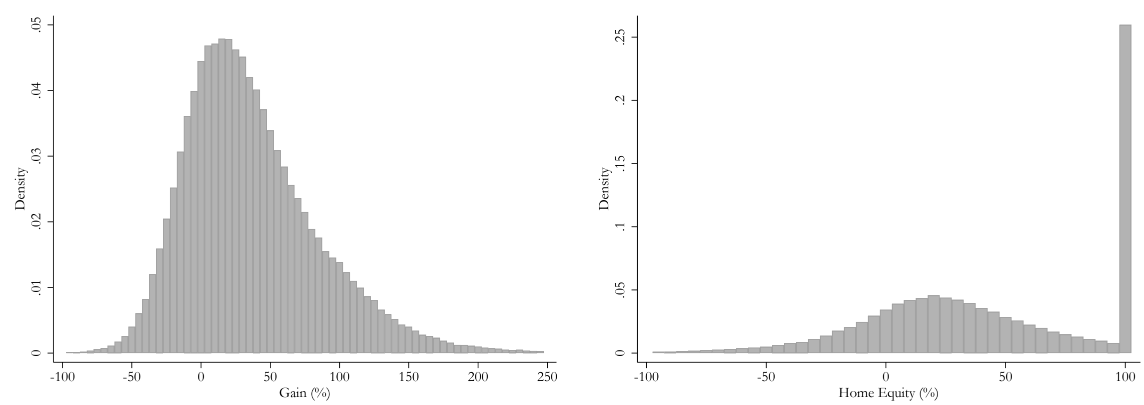
Figure A.2

Graphical Summary Statistics:

Potential Gains, Potential Home Equity, Listing Premium and Time-on-the-Market

This figure shows four histograms of main variables of interest. The potential 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. Potential 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. Time on the market (TOM) measures the time in weeks between when a house is listed and recorded as sold. Each listing spell is truncated at 200 weeks.

Panel A



Panel B

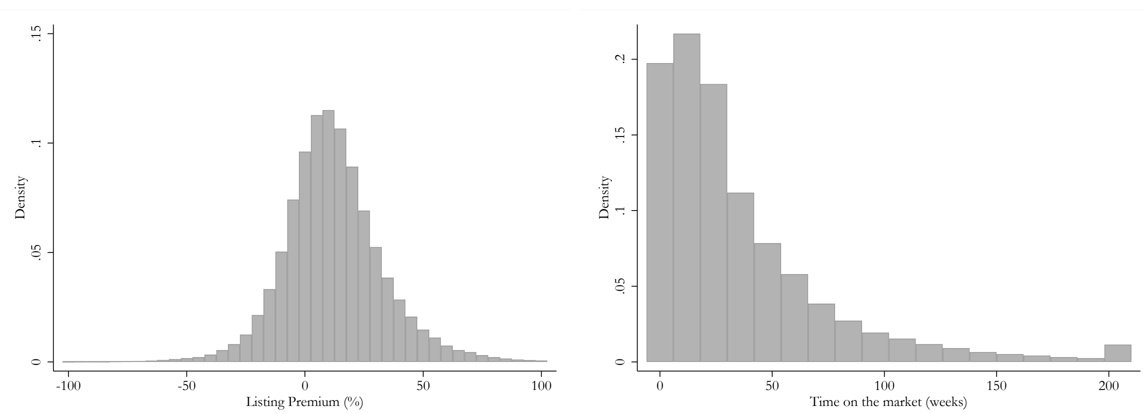


Figure A.3

Bunching of Listing Prices around Reference Point

This figure reports the distribution of listing prices relative to the reference point ($G_{\text{list}} = L-R$) in bins of 1 percentage points. The dotted line shows the counterfactual corresponding to the distribution of potential gains ($\widehat{G}$) across listings.

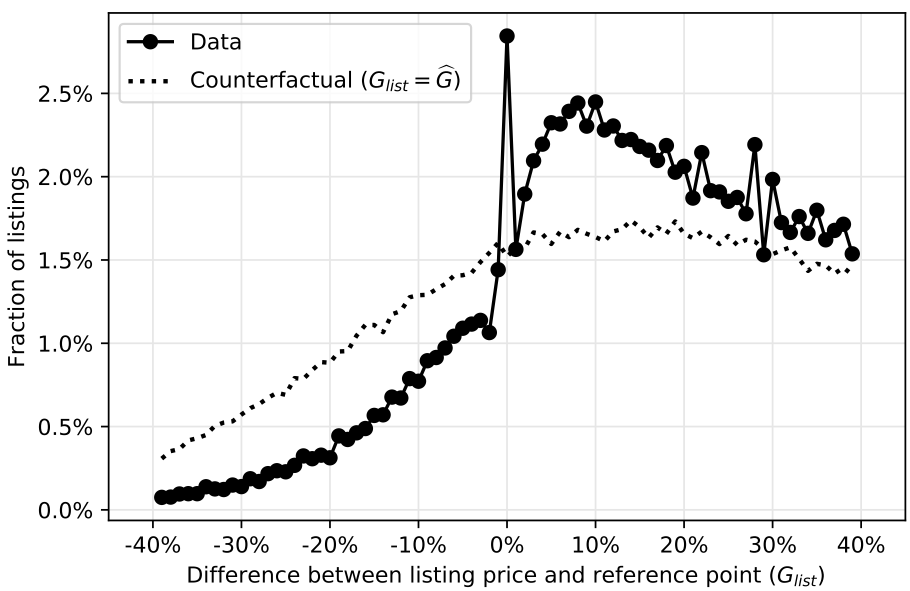


Figure A.4

Joint Distribution of Gains and Home Equity and Regions with $\widehat{G} \lesseqgtr 0$ and $\widehat{H} \lesseqgtr 20$

This figure plots the joint distribution of the potential 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 55.7% of the sample, (2) constrained winners ($\widehat{H}$ <20% and $\widehat{G}$ ≥0) with 21.4%, (3) unconstrained losers ($\widehat{H}$ ≥20% and $\widehat{G}$ <0) with 9.0%, and (4) constrained losers ($\widehat{H}$ <20% and $\widehat{G}$ <0) accounting for 13.9% of the sample.

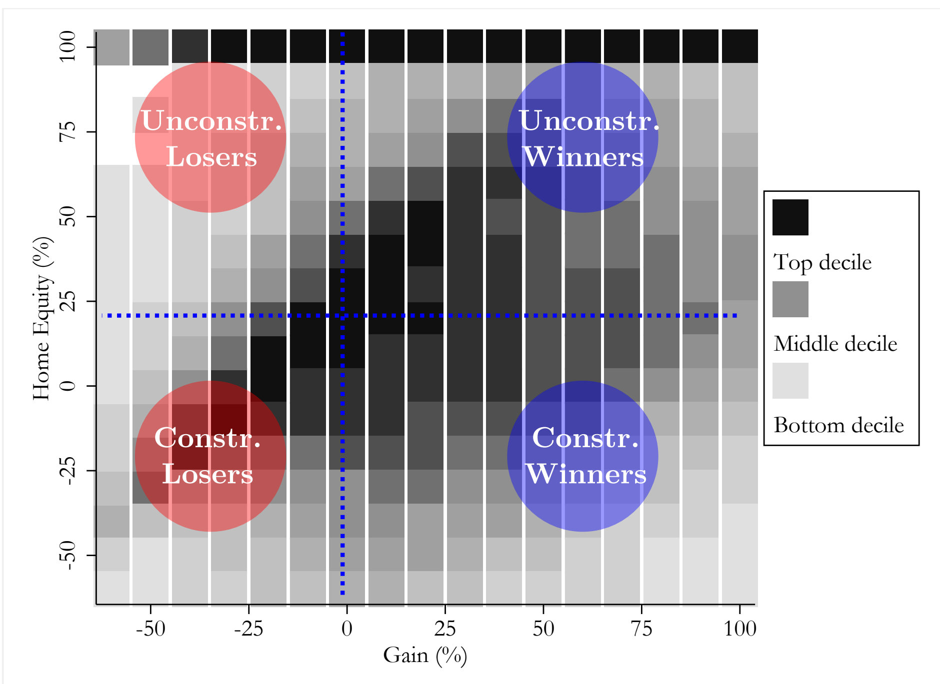


Figure A.5
Seller Groups - Listed (Relative Shares)

This figure shows the relative share of each seller group over time. The four groups are defined as follows: (1) unconstrained winners ($\widehat{H}$ ≥20% and $\widehat{G}$ ≥0), (2) constrained winners ($\widehat{H}$ <20% and $\widehat{G}$ ≥0), (3) unconstrained losers ($\widehat{H}$ ≥20% and $\widehat{G}$ <0), (4) constrained losers ($\widehat{H}$ <20% and $\widehat{G}$ <0).

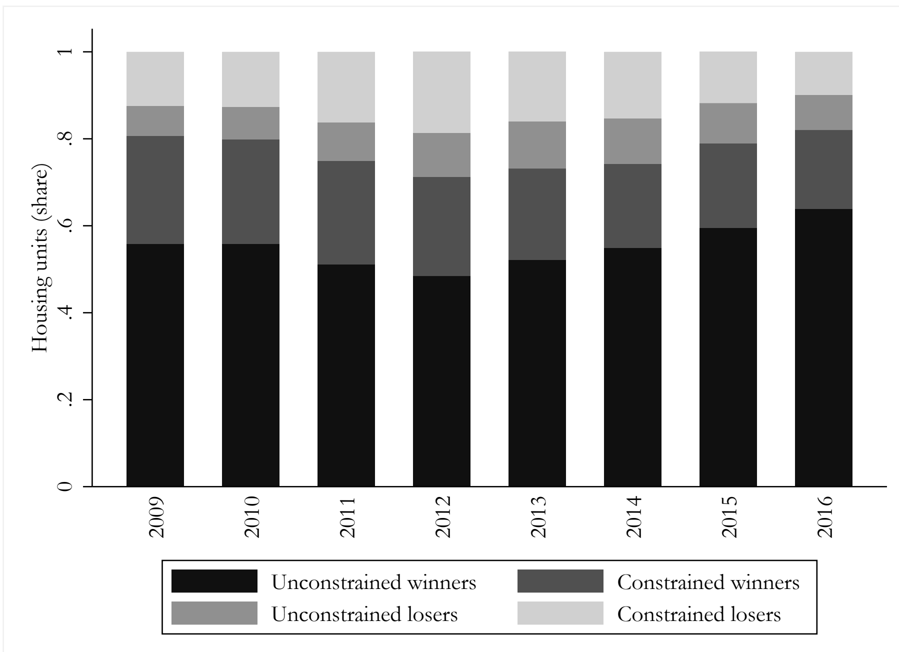


Figure A.6

Realized Gains vs. Listing Gains

This figure reports the distribution of *realized* prices relative to the reference point (i.e. realized gains $G = \ln P - \ln R$), for different levels of *listing* prices relative to the reference point (i.e. listing gains $G_{\text{list}} = \ln L - \ln R$). The dotted line shows a counterfactual corresponding to the aggregate distribution of differences between realized gains and listing gains $G - G_{list}$ in the full sample of listings, rescaled by the respective value of the listing gain G_{list} in each panel.

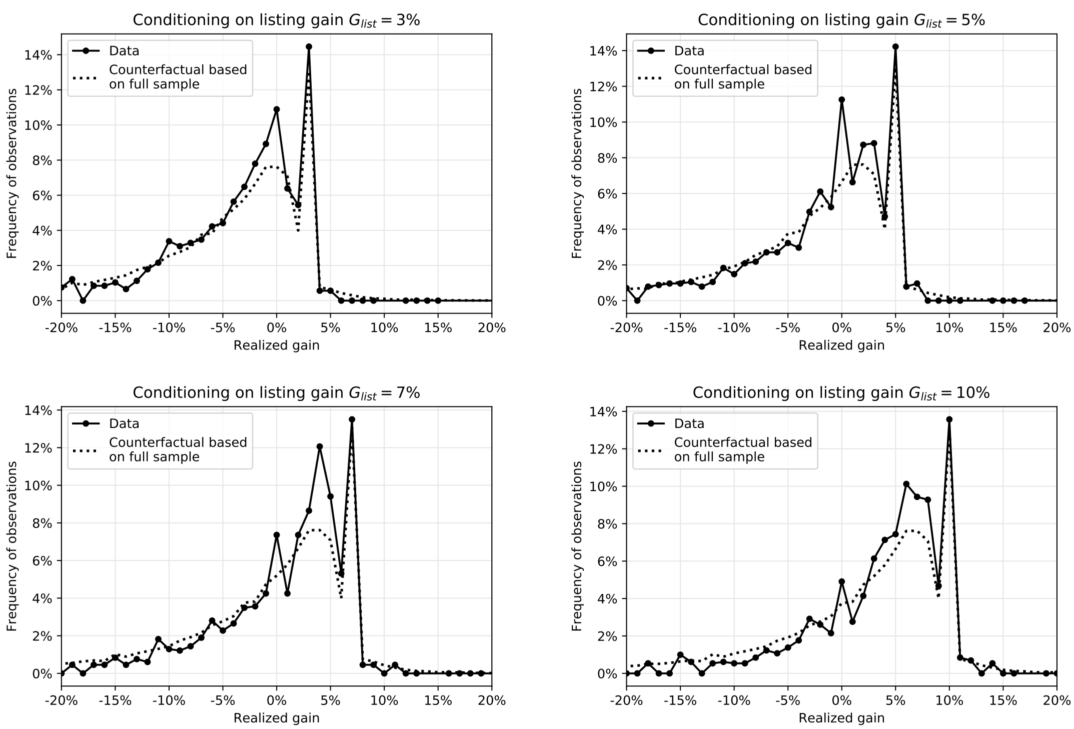


Figure A.7

Realized Gains vs. Realized 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 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.

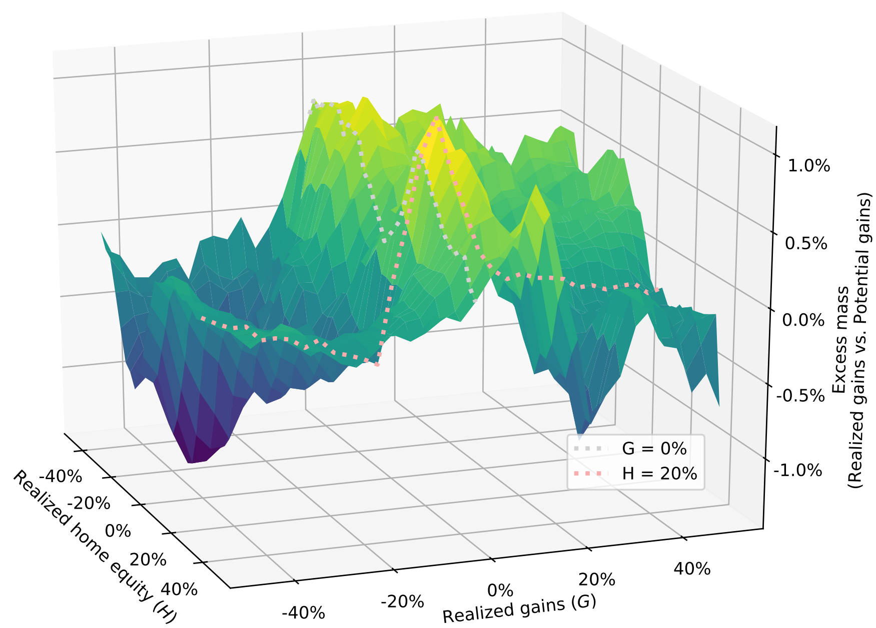


Figure A.8

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%].

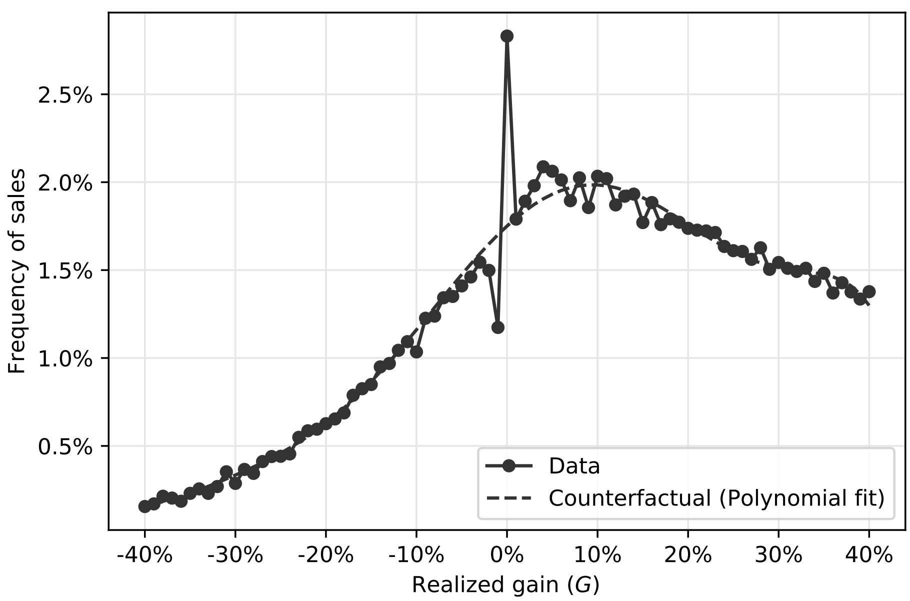


Figure A.9
Overview of Model Fit: Diffuse Bunching

This figure reports our set of moments in the data and in the model, evaluated at the set of parameters which correspond to the complete version of the model and the complete set of empirical moments, as indicated in line 10 of Table 1 in the paper. Dotted lines show 95% confidence intervals based on bootstrap standard errors.

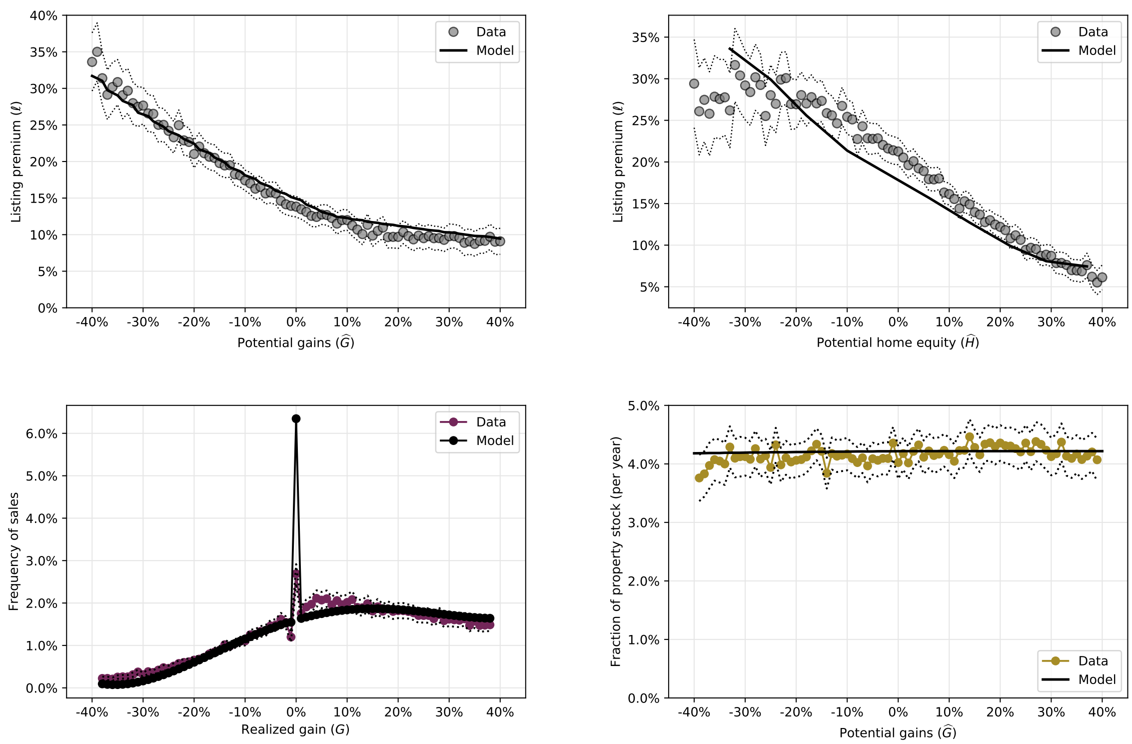


Figure A.10

Extensive Margin - Residualized

This figure reports the average annual probability of listing a property for sale across bins of potential gains, partialling out the effect of home equity. We calculate the potential gain and home equity level for each unit in the stock of properties in Denmark, for each year covered by our sample of listings, using the same hedonic model used to calculate potential gains in the 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 and home equity, yielding the probability of listing across bins, and run a regression of the probability of listing on each bin of potential gains and home equity. The dots shown reflect the bin fixed effect for each gain bin, while controlling for home equity bin fixed effects.

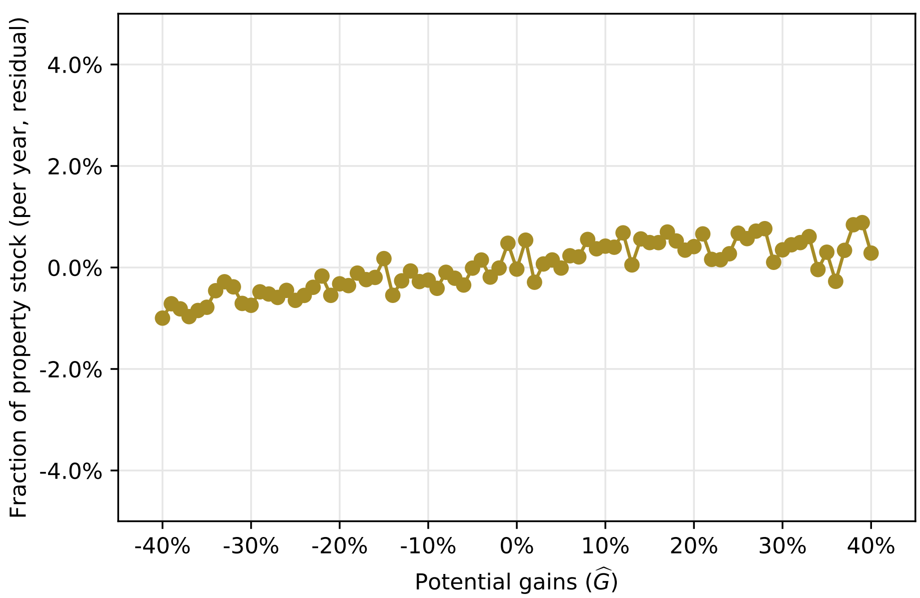


Figure A.11

Out-of-Sample Fit: Volumes Along the Home Equity Dimension

The plot reports the frequency of observations by realized home equity levels in the data and the model, evaluated at the set of parameters which correspond to the complete version of the model and the complete set of empirical moments, as indicated in line 7 of Table 1 in the paper.

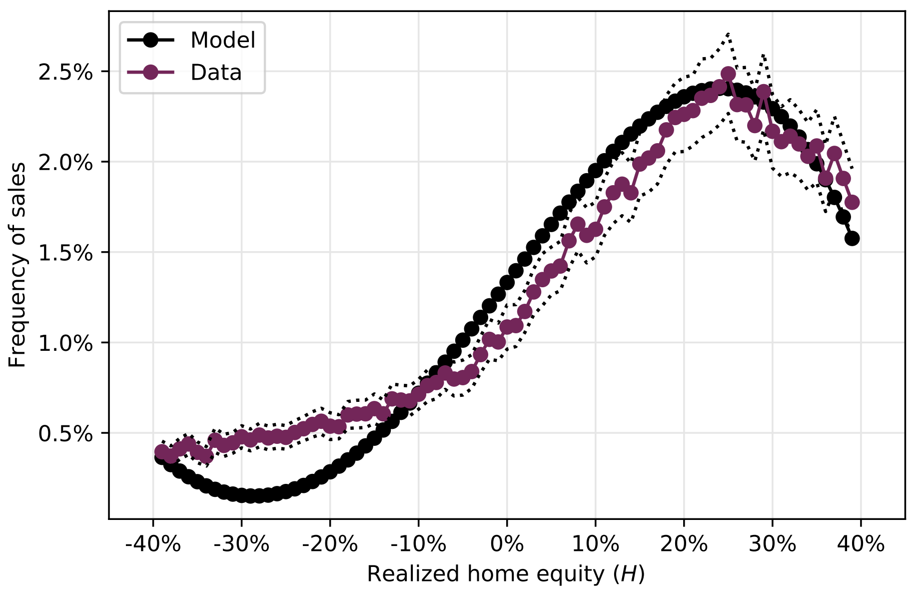


Table A.1

Literature Overview

Paper	Summary	Data	Model with ref. dependence	Bunching evidence	Estimate of λ
Genesove and Mayer (2001)	Sellers facing nominal losses set higher listing prices, attain higher realized sales prices and exhibit a lower sell hazard.	Boston condominium market (1990-1997), N=5,785	×	×	×
Engelhardt (2003)	Nominal loss aversion, rather than down-payment constraints, decreases household mobility.	National Longitudinal Survey of Youth and matched metro-level house price data (1985-1996), N=6,461	×	×	×
Einiö et al (2008)	Find evidence for aversion to realize losses in real estate transactions.	Helsinki (1987-2003), N=79,483	×	✓**	×
Anenberg (2011)	Both down-payment constraints and loss aversion affect final sales prices, using a repeat-sales estimator for prices.	San Francisco Bay Area (1988-2005), N=27,467	×	×	×
Bokhari and Geltner (2011)	Study the role of loss aversion, anchoring, and seller experience in the commercial real estate market.	RCA data on large US commercial property sales in the US (2001-2009), N=6,767	✓	×	×
Bucchianeri and Minson (2013)	Find evidence for anchoring effects in residential home sales, and that higher listing prices lead to higher realized sales prices.	DE, NJ and PA (2005-2009), N=14,616	×	×	×
Hayunga and Pace (2017)	Study the determinants of listing price and the trade-off with time on the market, and find that expected losses matter.	NAR Survey (2010-2012), N=3,302	×*	×	×
Liu and van der Vlist (2019)	Sellers set higher initial list prices and revise their list price downward when facing an expected loss.	MLS data, Randstad area of the Netherlands (2008-2013), N=319,609	×	×	×
Hong et al. (2019)	Properties with a capital gain have higher selling propensities and lower final sales prices.	Singaporean condominium market (1998-2012), N=1,964,907	×	×	×
Bracke and Tenreyro (2020)	Sales prices and selling propensities are affected by past house prices, in line with loss aversion and home equity constraints.	UK price (1995-2014) and listing data (2008-2014), matched dataset N=2,610,073	×	×	×

*Model to set optimal list price, but no reference dependence
**Show frequency distribution of realized gains

Table A.2

Alternative Listing Premia

This table reports regression results for the relationship between probability of sale (within six months) and the listing premium (ℓ), and the log relative mark-up, measured as the premium relative to the average house price in a given market, defined at the municipality-year and shire-year level, respectively. The regression estimates the effect of a positive listing premium (log relative mark-up) up to 40%, on the probability of sale indicator variable. Columns (1) to (3) provide univariate results for each measure, while columns (4) and (5) include the listing premium and the municipality or shire-level log relative mark-up, respectively. 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)
Listing Premium	-0.913***			-0.864***	-0.809***
	(0.014)			(0.024)	(0.023)
Log Relative Mark-up (Muni)		-0.103***		-0.012	
		(0.018)		(0.021)	
Log Relative Mark-up (Shire)			-0.180***		-0.084***
			(0.018)		(0.021)
Observations	119978	58493	60597	42399	44280
R^2	0.034	0.001	0.002	0.030	0.028

Table A.3

Genesove and Mayer (2001) Replication

This table replicates Table 2 from Genesove and Mayer (2001) using our main dataset (with a small reduction in the total number of observations because we cannot measure the pricing residual from the last sales price for all observations). 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 ≥ 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	0.566***	0.471***	0.513***	0.342***	0.588***	0.494***
	(0.016)	(0.015)	(0.026)	(0.025)	(0.016)	(0.015)
LOSS (squared)			0.001***	0.003***		
			(0.001)	(0.000)		
LTV if >=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	0.994***	0.991***	0.994***	0.990***	0.995***	0.991***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Market index at listing	0.991***	0.988***	0.991***	0.987***		
	(0.003)	(0.003)	(0.003)	(0.003)		
Residual from last sales price		-0.097***		-0.099***		-0.094***
		(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	0.409***	0.444***	0.410***	0.448***	76.252***	76.015***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.180)	(0.181)
Year-Quarter FE					✓	✓
Observations	193893	193893	193893	193893	193893	193893
R^2	0.888	0.889	0.888	0.889	0.890	0.892

Table A.4
Amendments to the Danish Mortgage-Credit Loans and Mortgage-Credit Bonds Act Between 2009 and 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.

B Details on Model Framework

B.1 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 utility function is given by:

$$\begin{aligned} U^*(\widehat{G}) = \max_{\ell} (\alpha_0 - \alpha_1 \ell) & \left[P(\ell) + \eta \underbrace{(\widehat{G} + \beta_0 + \beta_1 \ell)}_{G(\ell)} \left(\lambda 1_{\widehat{G}+\beta_0+\beta_1\ell<0} + 1_{\widehat{G}+\beta_0+\beta_1\ell\geq 0} \right) + \theta \right] \\ & + (1 - \alpha_0 + \alpha_1 \ell) \widehat{P}. \end{aligned} \quad (4)$$

To understand the solution to this optimization problem, we distinguish between three types of sellers: "Winners" ($G(\ell^*(\widehat{G})) > 0$) choose an optimal listing premium equal to the one given in equation (3), "Bunchers" ($G(\ell^*(\widehat{G})) = 0$) choose a listing premium exactly as large as necessary to realize a gain of zero:

$$\ell_B^*(\widehat{G}) = -\frac{\beta_0}{\beta_1} - \frac{1}{\beta_1} \widehat{G},$$

and "Losers" choose a listing premium corresponding to equation (3), but for a higher level of reference dependence ($\lambda \eta$):

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

The expression of the optimal listing premium is then given by:

$$\ell^*(\widehat{G}) = \begin{cases} \frac{1}{2} \left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1} \frac{\theta}{1+\eta} \right) - \frac{1}{2\beta_1} \frac{\eta}{1+\eta} \widehat{G}, & \text{if } \widehat{G} \geq \widehat{G}_0 \\ -\frac{\beta_0}{\beta_1} - \frac{1}{\beta_1} \widehat{G}, & \text{if } \widehat{G} \in (\widehat{G}_1, \widehat{G}_0) \\ \frac{1}{2} \left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1} \frac{\theta}{1+\lambda\eta} \right) - \frac{1}{2\beta_1} \frac{\lambda\eta}{1+\lambda\eta} \widehat{G}, & \text{if } \widehat{G} \leq \widehat{G}_1. \end{cases} \quad (6)$$

where $\widehat{G}_0$ and $\widehat{G}_1$ are the threshold levels of potential gains which determine the two limits of the bunching interval, with $\widehat{G}_0 + \beta_0 + \beta_1 \ell^*(\widehat{G}_0) = 0$ and $\widehat{G}_1 + \beta_0 + \beta_1 \ell^*_\lambda(\widehat{G}_1) = 0$. Equation (6) shows that if demand is linear, the solution to the seller's optimal listing premium profile is piecewise linear. If demand is concave, this will be reflected accordingly in the shape of the listing premium. In addition, note that the magnitude of the moving shock θ implicitly determines the values of $\widehat{G}_0$ and $\widehat{G}_1$, i.e., the location of the kink(s) in the listing premium along the potential gains dimension. This implies that the characteristic smooth "hockey stick" shape of the average listing premium profile can result from averaging the three-piece-linear form of the listing premium profile across the distribution of θ.

B.2 Mapping Between Potential and Realized Gains

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

$$G(\widehat{G}) = \widehat{G} + \beta_0 + \beta_1 \ell^*(\widehat{G}). \tag{7}$$

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:

$$G(\widehat{G}) = \gamma_0 + \gamma_1 \widehat{G} \tag{8}$$

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

$$G(\widehat{G}) = \begin{cases} \gamma_0 + \gamma_1 \widehat{G} & \text{if } \widehat{G} > \widehat{G}_0, \\ 0 & \text{if } \widehat{G} \in [\widehat{G}_1, \widehat{G}_0], \\ \gamma_{\lambda,0} + \gamma_{\lambda,1} \widehat{G} & \text{if } \widehat{G} < \widehat{G}_1. \end{cases} \tag{9}$$

Here, we have:

$$\widehat{G}_0 = -\frac{\gamma_0}{\gamma_1} \text{ and } \widehat{G}_1 = -\frac{\gamma_{\lambda,0}}{\gamma_{\lambda,1}}, \tag{10}$$

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

B.3 Extensive Margin Decision

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

$$U^*(\widehat{G}) = \widehat{P} + \left[\alpha_0 - \alpha_1 \ell^*(\widehat{G})\right] \left[\eta \widehat{G} + (1+\eta)\left(\beta_0 + \beta_1 \ell^*(\widehat{G})\right) + \theta\right] \tag{11}$$

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

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

$$\widetilde{\theta}(\widehat{G}) = -\eta \widehat{G} - (1+\eta)\left(\beta_0 + \beta_1 \ell^*(\widehat{G})\right) \tag{12}$$

[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 normally distributed:

$$\theta \sim N(\theta_m, \theta_\sigma),$$

the listing probability s is given by:

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

Substituting out equation (3), expressed in simplified form: $\ell^*(\widehat{G}) = \widetilde{\gamma}_0 - \widetilde{\gamma}_1 \widehat{G}$, in equation (12), 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:

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

B.4 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.

B.5 The Role of Concave Demand

Figure L.1 graphically illustrates the role of concave demand, 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 $L = \widehat{P}$.

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 L.1 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 ($\eta = 0$).

B.6 State Variables: Listing Premia and Potential Gains

In this section, we explain why the listing premium ℓ and the potential gain $\widehat{G}$ are sufficient to characterize the control variable $\ln L$ and the state space spanned by the exogenous variables $\ln \widehat{P}$ and $\ln R$. Consider first a simple version of the model in which the seller chooses the listing price L directly, and there is no reference dependence ($\eta = 0$):

The optimization problem is:

$$\max_L \alpha(L - \widehat{P}) \left(\widehat{P} + \beta(L - \widehat{P}) + \theta\right) + \left(1 - \alpha(L - \widehat{P})\right) \widehat{P}, \tag{13}$$

where concave demand $\alpha(L - \widehat{P}) = \alpha_0 - \alpha_1(L - \widehat{P})$ and $\beta(L - \widehat{P}) = \beta_0 + \beta_1(L - \widehat{P})$ and $P(L) = \widehat{P} + \beta(L - \widehat{P})$ are defined over the listing premium, as in Genesove and Mayer (2001) and Guren (2018).[2]

The first-order condition is:

$$\alpha_0\beta_1 - \alpha_1\beta_0 + 2\alpha_1\beta_1\widehat{P} - \alpha_1\theta = 2\alpha_1\beta_1 L^*, \tag{14}$$

which implies that:

$$L^* = \widehat{P} + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\theta}{2\beta_1} \tag{15}$$

The main message of equation (15) is that the listing price is chosen as a markup over $\widehat{P}$. I.e. in the model the coefficient of L on $\widehat{P}$ is equal to one. Put differently: the listing premium ℓ is uncorrelated with $\widehat{P}$. So we can work with the listing premium $\ell \equiv L - \widehat{P}$ and the potential gains $\widehat{G} \equiv \widehat{P} - R$ both in the data and in the model.

For completeness, note that the optimization problem with $\eta > 0$ becomes:

$$\max_L \alpha(L - \widehat{P}) \left(\widehat{P} + \beta(L - \widehat{P}) + \eta(\widehat{P} + \beta(L - \widehat{P}) - R) + \theta\right) + \left(1 - \alpha(L - \widehat{P})\right) \widehat{P}, \tag{16}$$

and the optimal solution is:

$$L^* = \widehat{P} + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\widehat{P} - R) + \theta}{2(1 + \eta)\beta_1}, \tag{17}$$

so these conclusions carry through to a setup with reference dependence.

B.7 The Role of 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),

[2] This assumption on β implies that the resulting realized price from the negotiation is a weighted average of the hedonic value $\widehat{P}$ and the listing price L. To see this, note that: $P = \widehat{P} + \beta_0 + \beta_1(L - \widehat{P}) = \beta_0 + \beta_1 L + (1 - \beta_1)\widehat{P}$.

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.

We do not need to impose any further restriction on the level of the outside option, but we note that for a listing to be optimal, we have: $\underline{u} < u(P(\ell^*)) + \theta - \varphi$.

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 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.

More generally, the case where R enters the outside option because it is rationally used to determine $\widehat{P}$ is in fact one of the valuation models that we consider – a standard repeat-sales approach. Following on from the analytical results described in the previous subsection, (i.e. a simple model with linear demand and linear reference dependence), we have:

$$\widehat{P} = R + \delta_t - \delta_s, \tag{18}$$

where δ_t is the aggregate price index at the time of the listing, and δ_s is the price index at the time of initial purchase, which implies that:

$$L^* = R + \delta_t - \delta_s + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\cancel{R} + \delta_t - \delta_s - \cancel{R}) + \theta}{2(1+\eta)\beta_1}, \tag{19}$$

and therefore:

$$L^* - \widehat{P} = \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\delta_t - \delta_s) + \theta}{2(1+\eta)\beta_1}. \tag{20}$$

The parameter η can therefore be identified empirically by variation in $\delta_t - \delta_s$, and the precise way in which R enters the valuation model is irrelevant.

For completeness, we note that in the case of reference dependence and loss aversion, the optimal solution is:

$$L^* = \begin{cases} \widehat{P} + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\widehat{P}-R)+\theta}{2(1+\eta)\beta_1}, & \text{if } \widehat{P} - R > \widehat{G}_0 \\ \widehat{P} - \frac{\beta_0}{\beta_1} + \frac{\widehat{P}-R}{\beta_1}, & \text{if } \widehat{P} - R \in [\widehat{G}_1, \widehat{G}_0] \\ \widehat{P} + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta\lambda(\widehat{P}-R)+\theta}{2(1+\lambda\eta)\beta_1}, & \text{if } \widehat{P} - R < \widehat{G}_0 \end{cases} \tag{21}$$

and with repeat sales:

$$L^* - \widehat{P} = \begin{cases} \frac{\alpha_0\beta_1-\alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\delta_t-\delta_s)+\theta}{2(1+\eta)\beta_1}, & \text{if } \widehat{P}-R > \widehat{G}_0 \\ -\frac{\beta_0}{\beta_1} + \frac{\delta_t-\delta_s}{\beta_1}, & \text{if } \widehat{P}-R \in [\widehat{G}_1, \widehat{G}_0] \\ \frac{\alpha_0\beta_1-\alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta\lambda(\delta_t-\delta_s)+\theta}{2(1+\lambda\eta)\beta_1}, & \text{if } \widehat{P}-R < \widehat{G}_0 \end{cases} \tag{22}$$

All conclusions from above also carry through to this case.

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. 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.

B.8 Structural Estimation

B.8.1 Overview of Parameters and Moments

Structural parameters ($\mathbf{x}^s$)	
Reference dependence	η
Loss aversion	λ
Distribution of moving shocks	θ_m, θ_σ
Financial constraints	μ
Listing/search cost	ϕ
Calibrated parameters ($\mathbf{x}^c$)	
Down-payment constraint	$\gamma = 20\%$
Exogenous inputs from the data ($\mathbf{x}^d$)	
Density of observations	$f_{stock}(\widehat{G}, \widehat{H}), f_{list}(\widehat{G}, \widehat{H})$
Concave demand	$\alpha(\ell), \beta(\ell)$
Cross-sectional variation of concave demand	$\alpha_{k\in\{High,Low\}}(\ell), \beta_{k\in\{High,Low\}}(\ell)$
Normalization of listing probability	$\overline{s}(\widehat{G}_+)$
Endogenous moments in the model $\mathbf{M_m}(\mathbf{x})$	
"Hockey stick"	$\ell^*(\widehat{G}, \mathbf{x})$
Variation of listing premia by potential home equity	$\ell^*(\widehat{H}, \mathbf{x})$
Bunching of realized sales	$f_{sale}(G, \mathbf{x})$
Extensive margin decision	$s^*(\widehat{G}, \mathbf{x})$
Cross-sectional variation of listing premium	$\ell^*_{k\in\{High,Low\}}(\widehat{G}, \mathbf{x})$

B.8.2 Numerical Optimization

The algorithm that allows us to find an estimate for the set of structural parameters $\widehat{\mathbf{x}}^s$ can be expressed as: $\widehat{\mathbf{x}}^s = \kappa\left(F(\mathbf{x}^s), \mathbf{x}_0^s\right)$. It takes the form of a gradient search method, which starts from an initial guess $\mathbf{x}_0^s$, calculates the gradient vector for each parameter and adjusts the step size according to the direction of the gradient. To avoid a situation in which ad hoc initial starting values $\mathbf{x}_0^s$ influence the local convergence point, we approximate an annealing procedure by first running a Monte Carlo technique, with a set of $N = 50,000$ draws of parameters $\mathbf{x}^s_{i=1,\ldots,N}$ and evaluate the function $F(\mathbf{x}^s_{i=1,\ldots,N})$ at each draw, choosing as a starting point $\mathbf{x}_0^s$ for the optimization the parameter combination which delivers the best overall model fit across all draws.

Since we consider different classes of models in our estimation (including or excluding particular structural features such as concave demand, financial constraints, and the extensive margin decision), we choose as the starting point for a particular class of models the parameters which minimize the prediction error in the version with the most complicated parameterization. In doing so, we also check that the prediction error of the final reported estimate is lower than the minimum error detected across all Monte Carlo draws for that particular model specification.

B.8.3 Attributing the Price-Volume Correlation to Model Ingredients

In our model, preferences (i.e., reference dependence and loss aversion) and constraints (i.e., down-payment requirements) enter the seller's utility function additively. In a model with linear demand, the intuition that both frictions should therefore contribute additively to the optimal listing premium profile holds, and it is straightforward to separate the contribution of each friction to the implied magnitude of the price-volume correlation. However, when the demand function is non-linear, the marginal contribution of one friction is moderated by the effect of the other.

To develop intuition for this, consider a reference-dependent seller who moves from being unconstrained to constrained. If this seller has accumulated large positive potential gains, the listing premium that he/she initially chooses will be low. This means that when this seller becomes constrained, he/she has considerable room to adjust the listing premium upwards in an attempt to surmount the down-payment threshold, without necessarily suffering a large drop in the probability of sale. Now consider a comparable seller in an initial position with negative potential gains. This seller will optimally choose a high listing premium, potentially high enough to be in a region of concave demand in which the decline in probability of sale for unit increases in the listing premium is very steep. This means that such a seller does not have much room to increase the listing premium in the event of becoming constrained, without having to accept a significant drop in the probability of sale. The same intuition applies when considering the effects on a constrained seller when varying the degree of reference dependence or loss aversion.

Despite the fact that reference dependence and down-payment constraints enter utility additively, market/demand constraints bind nonlinearly. The resulting interactive effects of preferences and financial constraints on listing premia make it necessary to consider two alternative sequential ordering schemes to identify the marginal contributions of the underlying structural ingredients to the price-volume correlation.

C 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.

C.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 Boligregisteret, 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 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 L.3 we show the price-volume correlation in Denmark over a broader period containing our sample period. The plot looks very similar to the broad patterns observed in the US.

C.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

[3] As we describe later, this is the same practice followed by Genesove and Mayer (1997, 2001); it helps improve the fit of the hedonic model, but barely affects our substantive inferences when we remove this variable.

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.

C.3 Mortgage Data

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

C.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.

[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.

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

C.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, and to listings for which the listing price and date is registered correctly.[6]

To restrict data to prices to regular market transactions, we exclude within-household transactions and transactions that Statistics Denmark flag as anomalous or unusual. We also drop foreclosures (both sold and unsold) and transactions where the buyer is the government, a company, or an organization.[7,8]

To ensure broad validity of the hedonic model, we exclude houses with a registered size of 0 or other missing hedonic characteristics. We also drop properties that are sold or listed at prices which are unusually high or low (below 100,000 DKK and above 20MM DKK in 2015-prices, or for other reasons marked by Statistics Denmark as having an extreme price).[9]

In addition, we 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.

Once all filters are applied, the sample comprises 214, 523 listings of Danish owner-occupied housing in the period between 2009 and 2016, for both sold (70.5%) and retracted (29.5%) properties, matched to mortgages and other household financial and demographic information.[10] These listings correspond to a total of 191, 884 unique households, and 179, 215 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, 349 observations of 807, 556 unique properties to understand sellers' extensive margin decision of whether or not to list the properties for sale.

Table L.1 documents the cleaning and sample selection process from the raw listings data to the final matched data.

[6]This implies that we drop listings for which the price is missing, as well as listings that are dated before the previous purchase date.

[7]The Section D.3 describes the Danish foreclosure process in detail.

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

[9]We apply this filter to reduce noise for our predicted hedonic prices, 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. In practice we drop 3,363 properties that Statistics Denmark mark as extremely priced, 1,501 properties with a listing or selling price below 100K DKK, and 639 properties with a listing or selling price above 20M DKK.

[10]The data comprises 172, 399 listings that have a mortgage, and 42, 124 listings with no associated mortgage (i.e., owned entirely by the seller).

D Institutional Background

D.1 Property Taxation in Denmark

SKAT assess the property value to determine the amount of property tax due. The exact rate of property taxation varies across municipalities, but the assessed value is set centrally. In addition, 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 sizes 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, prior to 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 for the sale in the event of death of the principal resident.

D.2 Assumability, Refinancing and Unsecured Mortgage

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 benefits by purchasing a house with a specific mortgage value. In general, any mortgage assumption needs the approval from mortgage 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. Another benefit of assuming the mortgage is to save the 150bp stamp duty due on new mortgage debt, with a maximum 120 basis point benefit at 80% LTV, which households would need to trade off against the potential increase in search cost to find a house with high assumable debt, given time, location, and preference constraints.

To some extent any down-payment gap (to bridge funding gaps between 80% and 95% loan-to-value) can be financed using normal bank/consumption debt lent to the buyers by their financial institution or occasionally from the seller of the property, but this additional mortgage tends to be expensive. 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.

D.3 The Foreclosure Process in Denmark

Home owners who cannot pay their mortgage or property tax may benefit from selling their home -– even if they have negative home equity — to pre-empt being 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 sales prices significantly below market prices. Selling in the market thus potentially allows home owners to repay a bigger fraction of their debt. This provides a rational for "fishing" behavior as mentioned in the main text, as home owners even with negative home equity may find it optional to pick a point on the right of the demand concavity trade-off, i.e. choose a high listing premium at the expense of decreasing the probability of sale prior to the foreclosure process.

A foreclosure takes place if a home owner 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 still fails to pay the creditor after two to three months, 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 foreclosure, 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 the highest bid determines the buyer and the price. If the buyer meets 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 may be preferred to foreclosure auctions by buyers as well, as they have fewer opportunities to assess the house and have to buy the house "as seen", without the opportunity to make any future claims on the seller. In addition, buyers have to pay additional fees of more than 0.5 percent of the price.

E Hedonic Pricing Model and Alternative Models of $\widehat{P}$

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

E.1 Baseline Hedonic Model

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

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

ξ is a constant, ξ_t are year fixed effects, ξ_m are municipality fixed effects (98 municipalities in total), and ξ_{tm} are municipality-year fixed effects. $\mathbb{1}_{i=f}$ is an indicator variable for whether the property is an apartment (denoted by f for flat), rather than a detached 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 0.876. The model fit is shown in Figure L.4.

E.2 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. However, the hedonic model excluding the tax-assessed value performs well in its own right. Table L.2 decomposes the hedonic model and shows the R^2 contribution from each component. By itself, the tax-assessed value explains around 80 percent of the variation in sales prices, and municipality-year fixed effects explain around 48 percent. Our baseline hedonic model without the tax-assessed value explains 77 percent of the variation in sales prices, and adding up to the third degree polynomial of the tax-assessed value raises explanatory power to 88 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 model for tax assessment utilizes some further information such as the distance to local amenities such as schools and public transport. In addition, in some cases (prior to 2013), the assessment is manually adjusted and verified by the tax authorities if the mechanically predicted value from the model is challenged by owners or if the property is in the right tail of the price distribution.

However, the tax-assessed value by itself is inferior to our model, as it underestimates price levels, especially in the period relevant for our data, see Figure L.5 Panel (a). This is because 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 L.5 panel (b) and (c) illustrate the shortcomings of the tax assessment in our sample period in particular. The figures show how the tax assessment is slow to incorporate more recent price developments, and as a result lags behind realized prices in the housing market boom prior to the financial crisis and in the subsequent bust, and is inaccurate from 2013 for the reasons just stated.

Figure L.6 and L.7 show that the relationships between listing premia over potential gains and home equity, and demand concavity, are preserved when using just the tax assessment prior to 2013, with a higher level of the listing premium, reflecting the inaccuracy introduced through the lag between assessed and realized prices.

E.3 Repeat Sales Models

We estimate a simple repeat sales model which does not rely on hedonic estimation, by adjusting the previous purchase price based on changes in the shire-level annual price index ("Simple Repeat"). The price index is the shire-year specific mean square meter price,[11] based on traded properties filtered to match the filtering of the municipality indices provided by Finance Denmark.[12] That is, $\ln \hat{P}_{SimpleRepeat} = \ln(R \cdot index_t / index_s)$, where R is the previous purchase price, t refers to the listing year, and s to the previous purchase year.

Next, we estimate a combined repeat sales models which uses information from time-varying hedonic characteristics, as well as information from repeat sales by adding the (average) pricing residual from previous sales to the baseline hedonic model (as described above). The lagged residuals are $ln(P_l) - \hat{ln(P_l)}$ for lags l up to thirteen past sales. We estimate four variants of the combined repeat sales model: the residual from the last sale, utilizing all pairs of repeat sales ('Repeat Sales ($T = 2$)"); average residuals from all existing previous sales, but only for properties with at least two repeat sales (three sales in total) ("Repeat Sales ($T \geq 3$)"); average residuals from all existing previous sales, but only for properties with at least three repeat sales (four sales in total) ("Repeat Sales ($T \geq 4$)"), and average residuals from all existing previous sales ("Repeat Sales ($T \geq 2$)").

We provide further motivation for the use of these repeat sales models in section I.

E.4 Repeat Sales Model with Renovations Data

We extend the baseline hedonic model to also include recent renovations of the property. Since our repeat sales models are able to account for the time-invariant component of unobserved quality ν_{it}, the renovation expense data are a way to proxy for the potentially time-varying unobserved component. We take advantage of the tax-deductability of renovations from 2011 and include controls for deducted amounts by the seller, and the data is further described in section E.4.1 below. We add $\bar{r}_{it} \equiv \mathbb{1}_{r_{it}} + \mathbb{1}_{r_{it}>0} + d + \mathbb{1}_{i=f}\mathbb{1}_{r_{it}} + \mathbb{1}_{i=f}\mathbb{1}_{r_{it}>0} + \mathbb{1}_{i=f} \cdot r_{it}$ to the baseline hedonic model, where $\mathbb{1}_{r_{it}}$ is an indicator for renovations data being available, $\mathbb{1}_{r_{it}>0}$ indicates that the seller has deducted a positive amount, and r_{it} is the logarithm of the deducted amount. Everything is also interacted with the apartment dummy, $\mathbb{1}_{i=f}$, letting the effect of renovations differ across different property types, i.e. detached houses or apartments.

We estimate model variants that aggregate the renovations data differentially, to reflect that property maintenance and renovation expenses accrue and add to unobserved quality over time. We use one-year lagged renovation deductions, available for the years 2012-2016. We also use

[11] This price index is not available for all observations, which reduces the number of observations slightly.

[12] In calculating their indices, Finance Denmark first exclude all transactions with a square meter price below 1,000 or above 20,000 1992-level DKK, a transactions price below 100,000 or above 25 million 1992-level DKK, and transactions of properties smaller than 25 square meters or bigger than 750 square meters.

three-year lagged cumulative deductions, leaving us with data for 2014-2016, and five-year lagged cumulative deductions, which we can only estimate for observations in 2016.

Lastly, we estimate composite models that add both past past residuals and one-year lagged renovation deductions to the baseline hedonic model, combining the advantages of all three sources of information ("Repeat Sales 1" for pair-wise repeat sales, and "Repeat Sales 2" for all repeat sales). For an overview of all models, see Table L.3.

E.4.1 Renovations Data Description

As a proxy for property maintenance and renovation expenses, we merge administrative data on tax exemptions on services done as part of the property. From 2011, Danish households have been able to deduct expenses for these service works done from the tax bill ("Boligjobordningen"). The initiative was introduced as a measure to reduce tax avoidance and to incentivize private consumption following the 2008/2009 recession, but has later been made permanent. Exemptions apply to incurred labor cost, conducted by external service providers in the home or summer house of the household, but not material cost. Services include property maintenance and renovations, but also other services such as cleaning. From 2011 to 2015, the maximum tax-deductable amount was 15,000 DKK per adult household member. Between 2016 to 2018 the maximum amount was split into 12,000 DKK for maintenance and renovations and 6,000 DKK for other services.

Data on claimed deductions by individuals is obtained from the Danish Tax Authorities and is made available to us by Statistics Denmark. We aggregate the deductions data by households and link them to the seller of a property. In most cases the services will have been conducted in the property for sale, but in some cases it may relate to a summer house or another property by the seller, which we cannot distinguish. From 2011 to 2016, about a quarter of listings are associated with owners claiming some tax deduction for renovation expenses. 27 percent of claims were at the maximum amount and the average claimed exemption per listing was 14,550 DKK, conditional on claiming a positive amount. To get a sense of magnitudes, 14,550 DKK is about one percent of the average list price of around 1,572,000 DKK. It is difficult to get a sense of how much the all-in renovation cost would be as these vary substantially by type of renovation. But to give an example, estimates of the labor cost of a kitchen renovation are between 10,000 to 15,000 DKK, with estimates for the full cost including material at 40,000 to 150,000 DKK[13] (around 6400 to 24,000 USD), which implies a multiple of between 3 to 10 to get an estimate of the all-in renovation cost, translating to about 3 to 10% of the average list price. We caveat that these are very rough estimates, but they illustrate that we should be able to proxy for a significant source of time-varying unobserved property quality, by simply assuming that the value of the renovation capitalizes into the new market value of the property.

We also show binned averages of the renovation expense variable across potential gains and listing premia, cumulated in different ways as described above, in Figure L.8. Renovation expenses are broadly flat across potential gains and listing premia by looking at current and lagged 1-year expenses, assuaging concerns that the hockey stick shape in the listing premium when sellers face negative potential gains, or the shape of demand concavity, is driven primarily by time-varying maintenance expenses. At longer horizons, cumulative renovation expenses appear slightly lower for negative listing premia, which may suggest that listing premia that we estimate as very negative may in fact be less so, i.e. they sell at less of a true discount because

[13] As for instance obtained from https://www.designa.dk/inspiration/koekkenguiden/hvad-koster-et-nyt-koekken.

it reflects the lower degree of maintenance over a longer period, and we directly account for this in our robustness checks by including these variables in the pricing model.

E.5 Out-of-Sample Testing

The large number of controls and fixed effects in the hedonic model could give rise to concerns about overfitting. To assess this, we conduct out-of-sample testing of the model. Table L.4 reports mean R^s from 1000 iterations of sampling 50, 75 and 100 percent of the data, respectively, estimating the model on that sample, and fitting the model to the remaining sample, and Figure L.9 show distributions of the R^s from these 1000 iterations. The model performs well out of sample even for models estimated on small samples.[14] Figure L.10 and L.11 show that the listing premium over gains and home equity relationships, as well as the pattern for demand concavity are preserved when the hedonic price is predicted out-of-sample.

F Downsizing Aversion and Interaction Effects

Figure L.12 documents the change in the position of the kink in the ℓ - $\widehat{H}$ relationship as $\widehat{G}$ varies in the data. It appears as if a household's propensity to engage in "fishing" behavior, i.e. raising listing premia and risking a lower probability of sale, kicks in at a level of $\widehat{H}$ that varies with the level of $\widehat{G}$.

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 a new, larger house.

To provide suggestive evidence on this story, Figure L.13a uses a subsample of the data for which we have information on the households' subsequent down payment ($N = 16,115$). For this limited subsample, we show a binned scatter plot of the listing premium 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 shift their reference level of housing on the basis of expected gains.

In addition, for a subsample of the data for which we have information on households' subsequent house purchase price ($N = 36,952$), we show in Figure L.13b that this price (in 2015 DKK) lies almost always above the previous purchase price, suggesting that households "trade up" their real house value on average, and that downsizing aversion may hence factor into their decision making.

[14] We note that we would expect the out-of-sample fit of our model to be quite high, given that one of the observable variables is the tax-assessed value of the house, which included by itself has an R^2 of 0.8. Excluding the tax-assessed value from the model reduces the R^2, but the fit is not greatly impaired, see Table L.5. The exercise further suggests that the remaining hedonic model coefficients appear relatively stable. As an alternative, we also conduct an out-of-sample test by estimating the model only on one year of the data (e.g. 2009), and fitting it to the remaining observations (e.g. 2010-2016). The difference between in-sample and out-of-sample R^2 for any estimation year and out-of-sample window combination lies between 6 to 9 percentage points, e.g. the in-sample R^2 for 2009 is 0.85, and the OOS R^2 for 2010-2016 is 0.76, a more noticeable drop, but which given the small and disparate in-sample window, likely represents a lower bound on the out-of-sample predictive ability of the model.

G Functional Form of Measured Concave Demand

Figure A.2 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 L.14 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 L.14, 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:

$$\alpha(\ell) = A + \frac{K - A}{\left(C + Qe^{-B\ell}\right)^{1/\nu}}. \tag{23}$$

H Listing Premia, Housing Stock Homogeneity, and Demand Concavity

In the main text, we document 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 right-hand plot of Panel B of Figure 2 (in the main text) 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.

For instance, for a block of identical apartment buildings, we would expect buyers to penalize sellers much more strongly for a given increase in the listing premium, as there is limited uncertainty surrounding the fair valuation of the property. On the other hand, if the housing stock is much less homogeneous, buyers may be more willing to tolerate listing premia as they would be willing to pay for variation in quality and less standard property characteristics. This can be micro-founded in a search and matching framework as done in Guren (2018), in which buyers do not know the true quality of a property ex ante, and decide to view and verify at a search cost, guided by initial listing prices, resulting in high listing premia being more viable in markets in equilibrium where the source of the listing premium is more likely to stem from non-standard property characteristics. Hence we use different measures of the homogeneity of the housing stock in a given geographic market to instrument for the degree of demand concavity. The degree of homogeneity of the housing stock may affect the level of the listing premium, but there is no obvious mechanism to link it to the degree of loss aversion, i.e. no obvious reason to believe that it should affect the slope of the listing premium schedule over potential gains other than through demand concavity. In other words, our identifying assumption is that we believe that variation in the homogeneity of the housing stock relates to differences in the slope of demand concavity, rather than innate differences in loss aversion across sub-markets.

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.[15] 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.

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

We find 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 in Figure L.16 Panel A, with each dot representing a municipality, with more homogeneous municipalities exhibiting stronger demand concavity, i.e. a more sharply decreasing probability of sale for any given increase in the listing premium. And similarly in Panel B, we find that stronger, i.e. more negative values of, demand concavity are correlated with a flatter, i.e. less negative, slope of the hockey stick. Table L.6 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,[16] with a baseline level of -0.323. 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 -0.531. With both instruments (i.e., including the distance to the largest cities as well), the overidentified 2SLS specification gives a result of -0.480 without, and -0.483 with controls for average household characteristics in the municipality.

[15]In Figure L.15, we show pictures of typical row houses in Denmark.

[16]Municipalities are required to have at least 20 observations where $\widehat{G} < 0$, leaving 96 out of 98 municipalities, but results are robust to keeping all municipalities.

I Unobserved Quality

An important concern in the literature is that the "true" $\widetilde{P}$ is imperfectly observed. Following Genesove and Mayer (2001), we differentiate between two types of measurement error, namely, (potentially time-varying) unobserved quality, and an idiosyncratic over- or under-payment by the seller at the point of purchase.

In this section, we show that (i) the simple repeat sales model eliminates the bias coming from time-invariant unobserved quality, (iii) time-varying observables that capture information from the tax-assessment value, time-varying hedonic characteristics and time-varying valuation of hedonics, together with data on property renovations, attenuate the bias coming from time-varying unobserved quality, (iii) a novel generalized repeat sales approach, where we average valuation residuals from all available past sales, attenuates the residual bias (not already captured by the hedonic estimation across all observations) coming from the past history of over- or under-payment.

I.1 A General Formulation of the Problem

Our structural model (assuming a seller with linear reference dependence) implies the following "true" relationship between listing premia and potential gains:

$$\underbrace{L_{ist} - \widetilde{P}_{it}}_{\widetilde{\ell}_{ist}} = \mu_0 + m(\underbrace{\widetilde{P}_{it} - R_{is}}_{\widetilde{G}_{ist}}) + \varepsilon_{ist}, \tag{24}$$

which we aim to estimate in the data. Here, L_{ist} is the listing price chosen by the seller of property i listed for sale in period t, $\widetilde{P}_{it}$ is the "true" hedonic value of the property at the time of listing and $\widetilde{\ell}_{ist}$ is the "true" listing *premium*. R_{is} is the price of the property when initially purchased in period s, and $\widetilde{G}_{ist}$ is the "true" potential gain.[17]

As noted, to begin with, we restrict our focus to the case of linear reference dependence, captured by the parameter m. We will later relax this assumption.

When bringing this model to the data, the problem is that the "true" $\widetilde{P}$ is imperfectly observed. Let $\widehat{P}$ be the "feasible" valuation model, and ξ_{it} the potentially time-varying estimation error:

$$\widetilde{P}_{it} = \widehat{P}_{it} + \xi_{it}. \tag{25}$$

The observed listing premia and potential gains are affected by estimation error in opposite directions:

$$\begin{aligned} \widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\ &= \widetilde{G}_{ist} - \xi_{it}, \end{aligned} \tag{26}$$

$$\begin{aligned} \widehat{\ell}_{ist} &= L_{ist} - \widehat{P}_{it} \\ &= \widetilde{\ell}_{ist} + \xi_{it}. \end{aligned} \tag{27}$$

Assuming that the shocks ε_{it} and ξ_{it} are uncorrelated with "true" potential gains $\widetilde{G}_{ist}$, the

[17]In the main part of the paper, we use P and R without time subscripts to differentiate between prices related to current time t, and previous purchase time s, while here we maintain R to denote the reference price for consistency, but note that $R_{is} \equiv P_{is}$.

estimated coefficient $\widehat{m}$ is then given by:

$$\widehat{m} = \frac{Cov(\widehat{G}_{ist}, \widehat{\ell}_{ist})}{Var(\widehat{G}_{ist})} = \frac{Cov(\widetilde{G}_{it} - \xi_{it}, \widetilde{\ell}_{ist} + \xi_{it})}{Var(\widetilde{G}_{ist} - \xi_{it})} = \frac{Cov(\widetilde{G}_{it} - \xi_{ist}, \mu_0 + m\widetilde{G}_{it} + \varepsilon_{it} + \xi_{it})}{Var(\widetilde{G}_{ist} - \xi_{it})}$$
$$= m \underbrace{\frac{Var(\widetilde{G}_{ist})}{Var(\widetilde{G}_{ist}) + Var(\xi_{it})}}_{\substack{\text{Classical measurement} \\ \text{error}}} - \underbrace{\frac{Var(\xi_{it})}{Var(\widetilde{G}_{ist}) + Var(\xi_{it})}}_{\substack{\text{Over-estimation bias} \\ \text{(because } m<0)}} \quad . \tag{28}$$

Equation (28) shows that in the vein of Genesove and Mayer (2001), unobserved heterogeneity such as unobserved property quality can cause measurement error, and a hockey stick slope estimate that is potentially over-estimated, i.e. too steeply negative.

I.2 Sources of Estimation Error

Following and expanding on Genesove and Mayer, 2001, we specify two sources of estimation error which affect the "feasible" valuation model: (i) Time-varying property unobserved quality[18] (which could have an average component as well as a time-varying component arising, for example, from home improvements), and (ii) Over- and under-payment by buyers in the market at different points in time.

Formally, we start by assuming that realized prices in the market have the following components:

$$P_{it} = \underbrace{X_i\beta + \delta_t + \nu_{it}}_{\substack{\text{"True" hedonic value} \\ \text{of property } (=\widetilde{P}_{it})}} + \underbrace{\omega_{it}}_{\substack{\text{Idiosyncratic over-} \\ \text{or under-payment}}} , \tag{29}$$

where X_i are property characteristics, δ_t is the aggregate price index, ν_{it} is the (potentially time-varying) unobserved quality of the property, and ω_{it} is an idiosyncratic over- or under-payment component relative to the "true" hedonic value of the property. We can write this true hedonic value using the expression:

$$\widetilde{P}_{it} = X_i\beta + \delta_t + \nu_{it}. \tag{30}$$

We assume that both sources of error ν_{it} and ω_{it} are uncorrelated with the observable property characteristics X_i and with the predictable time-varying component of prices δ_t. Moreover, we assume that both the unobserved quality and the over- or under-pricing components of realized prices are distributed randomly across properties, such that, when estimated in sufficiently large samples, they have an expected value of zero:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \nu_{it} = 0, \text{ and } \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \omega_{it} = 0. \tag{31}$$

In addition, the over- and under-pricing error is assumed to be distributed randomly through time, i.e. it has an expected value of zero if a sufficiently large number of periods is observed:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \omega_{it} = 0. \tag{32}$$

[18] In Genesove and Mayer (2001), this component is assumed to be time-invariant, and we relax this assumption and discuss the implications further below.

Note that this assumption may not be plausible for ν_{it}, for instance because permanent property improvements cause trends in ν_{it} over time, and assumptions on the distribution of ν_{it} and ω_{it} over time directly inform which model of prices should be preferred, as discussed further below.

I.3 Feasible Models

We can use several different approaches to estimate hedonic values in the data. In this section, we explore the implications of each of these approaches for the accurate estimation of m.

I.3.1 Model Descriptions

1. Standard hedonic regression:

   a. Time-invariant observables

   $$\widehat{P}_{it} = X_i\beta + \delta_t. \tag{33}$$

   The hedonic value in the above equation is obtained from a regression of the actually realized transaction prices P_{it} on a set of property characteristics X_i and time fixed effects.

   b. Time-varying observables

   $$\widehat{P}_{it} = X_{it}\beta + \delta_t. \tag{34}$$

   Equation 33 easily generalizes to a model with time-varying observables, examples for this type of information that we capture are e.g. the tax assessment value of the property, any time-varying valuation of hedonic characteristics, and changes in hedonic characteristics of the property over time.

2. Repeat sales models: Repeat sales models contain information from past transactions, including on unobserved quality, and are widely used to generate aggregate price indices, as proposed by e.g. Case and Shiller (1987). We apply this intuition to estimate prices for individual properties, using the formulation:

   $$\widehat{P}_{it} = X_i\beta + \delta_t + \bar{\nu}_{it_{\tau<t}} + \bar{\omega}_{it_{\tau<t}}, \tag{35}$$

   where $\bar{\omega}_{it}$ is the average value of past idiosyncratic over- or under-payments, and $\bar{\nu}_{it_{\tau<t}}$ is the average value of past unobserved quality components, i.e. averaged over periods for which $\tau < t$, prior to current period t.

   Denote with T the total number of repeat transactions observed for a given property. For $T = 2$ (one repeat sale), the model simplifies to

   $$\widehat{P}_{it} = X_i\beta + \delta_t + \nu_{is} + \omega_{is}, \tag{36}$$

   which is equivalent to estimating current price levels as the previous purchase price scaled by changes in the aggregate house price index since purchase, d_t/d_s (and analogous to how the aggregate Case-Shiller house price index is implemented):

   $$\widehat{P}_{it}^{level} = R_{is}^{level} \cdot \frac{d_t}{d_s}. \tag{37}$$

To see this, we can use previous notation and express the model in logs:

$$\begin{aligned}
\widehat{P}_{it} &= R_{is} + \delta_t - \delta_s. \\
&= X_i\beta + \delta_s + \nu_{is} + \omega_{is} + \delta_t - \delta_s \\
&= X_i\beta + \delta_t + \nu_{is} + \omega_{is},
\end{aligned} \tag{38}$$

which is equivalent to equation (36).

3. Combined repeat-sales model with time-varying observables: Suppose realized prices are characterized by a time-varying observable component $X_{it}\beta$

$$P_{it} = X_{it}\beta + \delta_t + \nu_{it} + \omega_{it}. \tag{39}$$

As noted above, this term could capture changes in hedonic characteristics of the property, or changes in the valuation of these characteristics over time. A more general way to write the repeat sales model and augment it with time-varying observables is to note that for $T = 2$:

$$\nu_{is} + \omega_{is} = \underbrace{R_{is} - \widehat{P}_{is}}_{\text{Hedonic model pricing residual at time } s} \tag{40}$$

And more generally:

$$\bar{\nu}_{it_{\tau<T}} + \bar{\omega}_{it_{\tau<T}} = \underbrace{\frac{1}{T-1}\sum_{\tau<T} R_{i\tau} - \widehat{P}_{i\tau}}_{\substack{\text{Average of the hedonic model pricing} \\ \text{residuals across repeat sales for which } \tau < T}} \tag{41}$$

So the $T = 3$ repeat sales model with time-varying observables requires to estimate

$$\widehat{P}_{it} = X_{it}\beta + \delta_t + \frac{\nu_{is} + \nu_{is'}}{2} + \frac{\omega_{is} + \omega_{is'}}{2}, \tag{42}$$

where s' refers to the purchase time prior to s, which can be implemented as

$$\begin{aligned}
\widehat{P}_{it} = & \underbrace{X_{it}\beta + \delta_t}_{\text{Hedonic model with time-varying observables}} + \frac{1}{T-1}\sum_{\tau<T} R_{i\tau} - \widehat{P}_{i\tau} \\
&= X_{it}\beta + \delta_t + \frac{R_{is} - \widehat{P}_{is} + R'_{is'} - \widehat{P}_{is'}}{2} \\
&= X_{it}\beta + \delta_t + \frac{X_{is}\beta + \delta_s + \nu_{is} + \omega_{is} - (X_{is}\beta + \delta_s) + X_{is'}\beta + \delta_{s'} + \nu_{is'} + \omega_{is'} - (X_{is'}\beta + \delta_{s'})}{2} \\
&= X_{it}\beta + \delta_t + \frac{\nu_{is} + \nu_{is'}}{2} + \frac{\omega_{is} + \omega_{is'}}{2}.
\end{aligned} \tag{43}$$

This flexible formulation can accommodate other variations of the hedonic model such as including location-time fixed effects and information on renovations, which we implement in our robustness checks. Note that this model only requires us to estimate the baseline model, and collect the residuals, i.e. is estimated in a single step.

4. Renovations as a proxy for unobserved quality: we can include additional information on

renovations (in the form $\overline{r}_{it}$ as described in section E.4),

$$\widehat{P}_{it} = X_i\beta + \delta_t + \overline{r}_{it}, \tag{44}$$

assuming that $Cov(\overline{r}_{it}, \nu_{it'}) \neq 0, \forall t'$, i.e., most importantly, $Cov(\overline{r}_{it}, \nu_{it}) \neq 0$, i.e. these additional time-varying covariates are informative of potentially time-varying unobserved quality.

I.3.2 Model Estimation

We compare coefficient estimates from these four types of feasible models.

1. The standard hedonic regression:

$$\widehat{P}_{it} = X_i\beta + \delta_t, \tag{45}$$

implies that the potential gain estimated on the set of observables is:

$$\begin{aligned}
\widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\
&= X_i\beta + \delta_t - (X_i\beta + \delta_s + \nu_{is} + \omega_{is}) \\
&= \delta_t - \delta_s - \nu_{is} - \omega_{is} \\
&= \widetilde{G}_{ist} - \nu_{it}.
\end{aligned} \tag{46}$$

When using the observable potential gain $\widehat{G}_{ist}$ to estimate equation (24), two biases arise in m. As above, we can replace $\xi_{it} = \nu_{it}$, and unobservable quality nu_{it} causes noise (biassing m towards zero), and a downward bias in m (over-estimation of the hockey stick slope). For completeness, the estimated coefficient is:

$$\begin{aligned}
\widehat{m} &= \frac{Cov(\widehat{G}_{ist}, \widehat{\ell})}{Var(\widehat{G}_{ist})} \\
&= \frac{Cov(\tilde{G}_{ist} - \nu_{it}, \nu_{it} + \tilde{\ell}_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\
&= \frac{Cov(\tilde{G}_{ist} - \nu_{it}, \nu_{it} + \mu_0 + m\widetilde{G}_{ist} + \epsilon_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\
&= m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} - \underbrace{\frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\substack{\text{Over-estimation bias} \\ \text{(because } m < 0\text{)}}}
\end{aligned} \tag{47}$$

2. The simple repeat sales approach with $T = 2$ (one repeat sale) is:

$$\widehat{P}_{it} = X_i\beta + \delta_t + \nu_{is} + \omega_{is}. \tag{48}$$

Recall, the true gain is:

$$\begin{aligned}\tilde{G}_{ist} &= \tilde{P}_{it} - R_{is} \\ &= X_i\beta + \delta_t + \nu_{it} - (X_i\beta + \delta_s + \nu_{is} + \omega_{is}) \\ &= \delta_t - \delta_s + \nu_{it} - \nu_{is} - \omega_{is}\end{aligned} \tag{49}$$

The potential gain is:

$$\begin{aligned}\widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\ &= \delta_t - \delta_s \\ &= \tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}\end{aligned} \tag{50}$$

Hence the coefficient estimate is:

$$\begin{aligned}\widehat{m} &= \frac{Cov(\widehat{G}_{ist}, \widehat{\ell})}{Var(\widehat{G}_{ist})} \\ &= \frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \nu_{it} + \tilde{\ell}_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\ &= \frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \nu_{it} + \mu_0 + m\widetilde{G}_{ist} + \epsilon_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\ &= m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} \underbrace{- \frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\nu_{is}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0\text{, bias}}\end{aligned} \tag{51}$$

If $\nu_{it} = \nu_{is}$, the last two terms cancel out, and assuming $Cov(\omega_{is}, \nu_{it}) = 0$, this type of repeat sales model would be unbiased. Note that the simple repeat sales model hence deals well with time-invariant unobserved property quality ($\nu_{it} = \nu_i$), but otherwise relies on the assumption that unobserved quality does not change much between the previous and the current purchase. In order to relax this assumption, we use the two following models which capture time-varying information on property value.

3. Combined repeat-sales model with time-varying observables:

   Similar to the above, we get

$$\widehat{m} = m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \bar{\nu}_{it_{\tau<T}} + \bar{\omega}_{it_{\tau<T}}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} \underbrace{- \frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\bar{\nu}_{it_{\tau<T}}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0\text{, bias}} \tag{52}$$

First, if time-varying observables are informative for prices, this information will be partialled out from ν_{it} and the bias from the unobserved heterogeneity component is reduced compared to the simple repeat sales model. Second, equation (52) shows that the bias in m is decreasing with the magnitude of the explanatory power of $\bar{\nu}_{it_{\tau<T}}$ for ν_{it}. The choice of repeat sales model hence also depends on what we believe most accurately captures the

data-generating process behind ν_{it} - we should use $T = 2$ if we believe $\nu_{it} \approx \nu_{is}$, but we should use $T = 3$ if we believe that $\nu_{it} \approx (\nu_{is} + \nu_{is'})/2$ etc. If time-varying observables or lagged average residuals *perfectly* capture time-varying unobserved quality, then this model generates a bias-free approach. Third, the term on average idiosyncratic over-or under-payments ($\bar{\omega}_{it_{\tau<T}}$), and hence this component of the measurement error, decreases with T. In our implementation of the models, we hence include different models for $T = 2$, $T \geq 2$, $T \geq 3$, and $T \geq 4$.

4. Time-varying information on renovations:

$$\widehat{P}_{it} = X_i\beta + \delta_t + \overline{r}_{it}, \tag{53}$$

implies that the potential gain estimated on the set of observables is:

$$\widehat{G}_{ist} = \widetilde{G} - (\nu_{it} - \overline{r}_{it}), \tag{54}$$

and hence the estimated coefficient is

$$\widehat{m} = m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \bar{r}_{it}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} \underbrace{- \frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\bar{r}_{it}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0 \text{, bias}} \tag{55}$$

We can think of the inclusion of $\overline{r}_{it}$ as including a direct proxy variable of time-varying unobserved quality into the set of observables X_{it}. Hence the intuition is similar to 3., the bias in m is decreasing with the magnitude of the explanatory power of $\overline{r}_{it}$ for ν_{it}. If the renovations variable perfectly captures time-varying unobserved quality, then this model also generates a bias-free approach.

I.3.3 Discussion

We implement these different models and compare them below.[19] Table L.3 provides an overview of all the models that we implement. Figures L.17 and L.18 provide a graphical overview of the predictive ability of the main models, and a comparison using binned scatter plots for a) the listing premium over potential gains, and b) the probability of sale within six months and the listing premium, respectively. Table L.7 provides a quantitative comparison of the main models, by estimating summary statistics of the moment relationships that we use to estimate our structural model. In particular, we estimate piecewise linear slopes for the listing premium over negative (row 2) and positive (row 3) potential gains, and for the probability of sale within six months for positive listing premia (row 6), using the same support as we use for the individual moments. Our main models are: the baseline model (Ia), the baseline model augmented with lagged 1-year renovation expenses (Ib), a simple repeat sales model based on shire-level house price changes (II), the combined baseline hedonic and repeat sales model with renovation expenses and the lagged pricing residual for one repeat sale (IIIa), and

[19] Note that in the case of any repeat sales model, the estimation requires at least one repeat transaction of the property ($T = 2$), and more for $T > 2$. Our sample contains at least one repeat sale in the transaction register data between 1992 and 2016 by definition, as we need to observe the previous purchase price. For more than one repeat sale within this time window, however, the properties that get traded more often may become less representative of the overall sample. Hence the optimal number of repeat transactions is ambiguous, and we estimate models for properties where $T = 2$, $T \geq 3$ and $T \geq 4$, as well as using the maximum number of repeat sales available for each property ("all", i.e. $T \geq 2$).

the equivalent for any number of repeat sales and average lagged residuals (IIIb). Table L.8 provides the concomitant comparison for variants of the repeat sales models, and sub-samples for which we observe renovation tax expenses.

Focusing on the main models in Table L.7, we find that that the level of the listing premium around zero potential gains is between 12 to 15 percent across our preferred models, but quite high (27 percent) for the simple shire-level repeat-sales model (II) together with a low R^2 of 0.57 (while all other models have R^2s between 0.87 and 0.88), suggesting that it is indeed the least precise model as it does not include time-varying information from observables. The hockey stick of listing premia over negative potential gains is between -0.45 to -0.53 across main models, and between between -0.90 to -0.93 for demand concavity (row 6, slope of probability of sale with respect to positive listing premia), but is almost half as steep for the simple repeat sales model (-0.49) which we discuss further below. Column Ic provides an out-of-sample estimation of the baseline model, by estimating the model on a random 50% sample of the data, and fitting prices on the other half, and using these to generate listing premia, potential gains and home equity. The results show averages and standard errors from 100 bootstrap draws (starting from the extensive margin). It demonstrates that the moment statistics are very robust to the data used when fitting the model, with an average 0.873 out-of-sample R^2.

In Table L.8, columns IVa to IVc show the combined hedonic and repeat sales model split by the number of repeat sales observed (2, 3 or more, and 4 or more, respectively), with broadly similar results, a slightly decreasing hockey stick and increasing demand concavity with the number of repeat sales observed, as well as a higher propensity to list for a given potential gain (row 7) - illustrating that conditioning on higher number of repeat sales also conditions on a subset of possibly more selected and more liquid properties. Column IVd shows the model with any number of repeat sales. Moving from 2 to 4 or more repeat sales increases the model R^2 from 0.881 to 0.895, as the sample is increasingly selected on more liquid and frequently traded properties, improving valuation accuracy using our model, but most moments, in particular the demand concavity slope in row (6), only varies between -0.91 to 1.01, assuaging concerns that unobserved quality causes a substantial bias in demand concavity.

Our preferred models for comparison are IIIa and IIIb, which combine the repeat sales approach with time-varying observable hedonics and information on renovation expenses. IIIa is based on pairwise repeat sales ($T = 2$) and includes the lagged pricing residual to the hedonic model with time-varying observables and renovation expenses, which is similar in spirit to the approach used in Genesove and Mayer (2001). We further generalize the repeat sales approach to also include average lagged pricing residuals for any number of repeat sales observed in IIIb ($T \geq 2$).

In sum, we implement each one of the feasible approaches that we discuss above. None of these approaches invalidates the basic moments that we detect in the data, despite being subject to potentially different sources of underlying error. We also implement different versions of the repeat sales model by varying T, and results remain broadly robust. That should provide some reassurance that our main estimates are not being generated solely by the sources of error, but rather, by the deeper structural forces that make sellers set listing premia in response to underlying potential gains and losses.

I.3.4 Comparison to Genesove and Mayer (2001) Bounding Approach

The pairwise repeat sales model, i.e. including the last pricing residual to the baseline model, in model IVa (Repeat Sales ($T = 2$)) is most comparable in spirit to what Genesove and Mayer

(2001) propose. With only time-invariant unobserved heterogeneity, they show that including the pricing residual from the previous sale (as a noisy proxy for unobserved quality) likely provides a lower bound estimate of the relationship between ask prices and losses and ask prices. We replicate Table 2 in their paper in Table A.3 to compare our results and data directly to theirs. Comparing column (2) and (1), the effect from a 10% increase in potential losses is between a 4.7 to 5.7% increase in list prices, compared to their 2.5 lower bound and 3.5% upper bound estimate. In addition to this approach, our combined repeat sales models (IIIa and IIIb) use time-varying hedonic characteristics and renovation expense data to capture the remaining, possibly time-varying, unobserved heterogeneity, and our results are broadly robust. We hence propose additional model components to narrow in on the remaining variation that could be explained by unobserved quality.

I.4 Reference Dependence with Loss Aversion

In the more general case, in which the seller also exhibits loss aversion, our structural model implies the following data-generating process for listing premia:

$$L_{ist} - \widetilde{P}_{it} = \mu_0 + f(\eta, \lambda, \underbrace{(\widetilde{P}_{it} - R_{is})}_{(\widetilde{G})}) + \varepsilon_{it}, \tag{56}$$

where f is either a piecewise linear function with two kinks somewhere in the neighbourhood of zero, or a convex function which is steeper in the loss domain and flatter in the gain domain.

If we can approximate this smooth function by a series of piecewise linear segments, we can assess the impact of unobserved quality locally analogously to our discussion above. For example, consider an (erroneous, but utilized in the literature earlier) two-piece piecewise linear specification with a kink at a potential gains level of zero:

$$\underbrace{L_{ist} - \widetilde{P}}_{\tilde{\ell}_{ist}} = \mu_0 + m_0 \underbrace{(\widetilde{P}_{it} - R_{is})^-}_{\widetilde{G}^-_{ist}} + m_1 \underbrace{(\widetilde{P}_{it} - R_{is})^+}_{\widetilde{G}^+_{ist}} + \varepsilon_{it}, \tag{57}$$

where a $-/+$ superscript indicates that the value of the respective quantity is negative or positive, respectively.

Focusing on listing premia over negative potential gains, and assuming that ε_{it} and ξ_{it} are uncorrelated with "true" potential gains $\widetilde{G}_{ist}$, our estimated coefficient of interest is:

$$\widehat{m}_0 = \frac{Cov(\widehat{G}^-_{ist}, \widehat{\ell}_{ist})}{Var(\widehat{G}^-_{ist})} = \frac{Cov((\widetilde{G}_{it} - \xi_{it})^-, \widetilde{\ell}_{ist} + \xi_{it})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)} \tag{58}$$

$$= \frac{Cov((\widetilde{G}_{it} - \xi_{ist})^-, \mu_0 + m_0 \widetilde{G}^-_{it} + m_1 \widetilde{G}^+_{it} + \varepsilon_{it} + \xi_{it})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)}$$

$$= m_0 \underbrace{\frac{Cov((\widetilde{G}_{ist} - \xi_{ist})^-, \widetilde{G}^-_{it})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)}}_{\text{Measurement error/Smoothing}} - \underbrace{\frac{Cov((\widetilde{G}_{ist} - \xi_{ist})^-, \xi_{ist})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)}}_{\text{Over-estimation bias (because } m<0)} \tag{59}$$

Equation (59) shows that in the vein of Genesove and Mayer (2001), unobserved quality can cause measurement error, and a hockey stick slope estimate that is potentially over-estimated, i.e. too steeply negative.

I.5 Regression Kink Design (RKD)

In order to verify that our approach is robust to measurement error not only in η (as shown above), but also that there is a significant slope change as predicted by $\lambda > 1$, we employ a regression kink design (RKD), first suggested by Card et al. (2015b) and implemented e.g., by Landais (2015), Nielsen et al. (2010), Card et al. (2015a). We employ this method with the caveat that the model does not predict a sharp kink exactly at $\widehat{G}$, due to the smoothing factors described in the main text, and that we use zero for the kink threshold, even though the listing premia slope increase starts at $\widehat{G} > 0$. We complement this robustness check with our non-parametric evidence on bunching around zero realized gains.

Note that while the realized gain is an outcome of household decision making, households only have imperfect control over potential gains $\widehat{G}$ which we use as the running variable V, with a kink point at zero ($\overline{v} = 0$): as long as households can only imperfectly manipulate on which side of the threshold they are, the resulting differences in behavior above and below the threshold can be interpreted as causal.[20]

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. We show indirect evidence for this by plotting binned averages of observable property characteristics and household characteristics (Figure L.20). We also show that the distribution of the running variable is smooth around the threshold (no bunching in *potential* gains) (Figure L.20).

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

$$\tau = \lim_{v \to \overline{v}_+} \frac{dE[\ell_{it}|V_{it} = v]}{dv}\bigg|_{V_{it}=v} - \lim_{v \to \overline{v}_-} \frac{dE[\ell_{it}|V_{it} = v]}{dv}\bigg|_{V_{it}=v}, \tag{60}$$

based on the following RKD specification (Landais 2015):

$$E[\ell_{it}|V_{it} = v] = \kappa_m + \kappa_t + \boldsymbol{\xi}\mathbf{X_{it}} + \left[\sum_{p=1}^{\overline{p}} \gamma_p(\nu - \overline{\nu})^p + \nu_p(v - \overline{v})^p \mathbb{1}_{V \geq \overline{v}}\right]. \tag{61}$$

$$\text{where } |v - \overline{v}| < b. \tag{62}$$

We estimate versions with and without controls (time (κ_t) and municipality (κ_m) fixed effects, home equity, 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.

Table L.9 reports results across different bandwidths within which we fit a local linear function on each side of the threshold. Figure L.22 provides further robustness checks on using local quadratic estimation and bandwidth choice.[21]

[20]For instance, while households can spend money to renovate their house to achieve a higher market price, they cannot control aggregate house price movements that will also affect the house value.

[21]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

The estimate of the increase in (absolute) slope at zero is about 0.2, which is broadly consistent with our baseline moment summary statistics in which the listing premium slope over positive potential gains is around -0.1, and around -0.5 around negative gains, despite using additional controls and restricting to a narrower estimation range around the threshold.

I.6 Demand Concavity

I.6.1 Estimation

Building on previous notation, let $\alpha(\widetilde{\ell})$ denote the probability of a quick sale, which is decreasing in the *true* listing premium $\widetilde{\ell}$. Again using a piecewise-linear formulation, we have

$$\alpha_{ist}(\widetilde{\ell}) = \mu_1 + n_0(\underbrace{L_{ist} - \widetilde{P}_{it}}_{\widetilde{\ell}^-_{ist}})^- + n_1(\underbrace{L_{ist} - \widetilde{P}_{it}}_{\widetilde{\ell}^+_{ist}})^+ + \epsilon_{ist}, \tag{63}$$

where the coefficient n_1 measures the decrease in the probability of sale α_{ist} for a given increase in the listing premium, and α_{ist} is an indicator variable taking the value 1 if a sale was completed in six months, and 0 otherwise.

With observed listing premia

$$\widehat{\ell}_{ist} = L_{ist} - (\widehat{P}_{it} + \xi_{it}) = \widetilde{\ell}_{ist} + \xi_{it}, \tag{64}$$

the feasible regression is

$$\alpha_{ist}(\widehat{\ell}) = \mu_1 + n_0(\widetilde{\ell}_{ist} + \xi_{it})^- + n_1(\widetilde{\ell}_{ist} + \xi_{it})^+ + \epsilon_{ist}, \tag{65}$$

with estimated main coefficient of interest

$$\begin{aligned} \widehat{n}_1 &= \frac{\alpha_{ist}(\widehat{\ell}), (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+} = \frac{Cov(\mu_1 + n_0\widetilde{\ell}^-_{ist} + n_1\widetilde{\ell}^+_{ist} + \epsilon_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+} \\ &= n_1 \underbrace{\frac{Cov(\widetilde{\ell}^+_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+}}_{\text{Measurement error}} + \underbrace{\frac{Cov(\epsilon_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+}}_{0 \text{ if } \xi_{it} \perp \epsilon_{it}} \end{aligned} \tag{66}$$

Equation 66 shows that the presence of ξ_{it} may cause measurement error. Depending on assumptions about the error term ϵ_{it}, there is not neccessarily a bias when estimating the slope of the probability of sale for positive listing premia. Sources of such correlation could be local housing market conditions that affect local probabilities of sale and could be correlated with e.g. renovation expenses. We deal with this concern by estimating demand concavity *across* different geographic markets, to isolate the relationship between demand concavity and the hockey stick in listing premia within a given sub-market.

higher for the second-order polynomial, reported in Figure L.23.

J Bunching Estimation

J.1 Robustness

We conduct several robustness checks for bunching in realized gains at 0, where the realized sales prices is equal to the reference point of the previous sales price. First, we show the prevalence of sales at round numbers in Figure L.24. We then show the distribution of realized gains by excluding sales at rounded prices of 10,000; 50,000; 100,000 and 500,000 DKK (Figure L.25), respectively. We further show that bunching is present across all quintiles of the previous sales price (Figure L.26) and when splitting into quintiles by holding period (Figure L.27), except for the sub-sample with holding periods of greater than 12 years (top quintile). The degree of bunching seems to be declining with the holding period, consistent with the previous purchase price becoming less salient and memory effects over time, as well as a decreasing absolute mass of households who are around the zero potential gains threshold, given house price appreciation over time.

Lastly, in Figure L.28, we show that the estimate of excess mass is robust when using a hedonic pricing model with cohort (i.e., previous purchase year) fixed effects.

J.2 Bunching of Listing Prices around Nominal Purchase Price

Figure A.3 reports the distribution of listing prices around the nominal reference point. In our model, listing prices will also be more likely to be located *above* the reference point, as a result of loss aversion. Loss averse sellers will aim to realize gains in the positive domain. To do so, they must set listing prices above their reference point, because they take into account market conditions that translate listing premia into realized premia (the $\beta(\ell)$ function). But this does *not* imply bunching of listing prices exactly at the reference point. Indeed, when we solve for the distribution of the differences between listing prices and reference prices predicted by the model, we find that there is an interval to the right of the reference price in which sellers set listing prices which are quite close to one another. However, when we inspect Figure A.3, while we do see that there is such behaviour in a region between roughly 7% and 10% above the reference price, there is also another region visible in the plot. Contrary to the model, there is some bunching of listing prices precisely at the reference price. This suggests a separate, additional role for the salience of the reference point in sellers' listing decisions.

K Household Demographics

K.1 Liquid Financial Wealth

Figure L.29 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 73,000 DKK and the mean in the sample is 222,000 DKK. When we divide *gross* financial assets by mortgage size, we find that households, at the median, could relax their constraints by around 6.15 percent if they were to liquidate all financial asset holdings. However, the right-hand side of the top panel of the figure shows that this would be misleading. Looking at *net* financial assets, once short-term non-mortgage liabilities (mainly unsecured debt) are accounted for, substantially changes this picture. The median level of net financial assets in the sample is -99,000 DKK and the mean is -121,000 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 7.5% 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 7.5%, 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 robustness checks 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.

K.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 L.29 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.

L Additional Appendix Figures and Tables

Figure L.1
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.

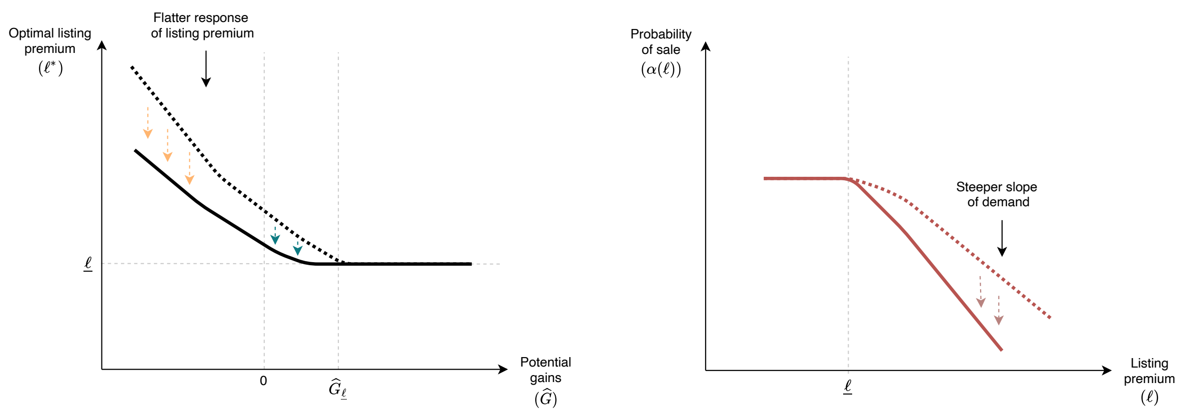


Figure L.2

Reference Dependence and Loss Aversion: Simple Version of the Model

The figure illustrates how each specification of the utility function is reflected in the seller's optimal choice of listing premia (left-hand side panel) and distribution of realized gains (right-hand side panel), implied by a simple version of the model with linear demand estimated in the data ($a_0 = 0.61, a_1 = -0.97, \beta_0 = -0.068, \beta_1 = 0.835$), no financial frictions ($\mu = 0$), and for a stylized set of structural parameters $\eta = 0.5, \lambda = 2$, and $\theta_\mu = 0.5$.

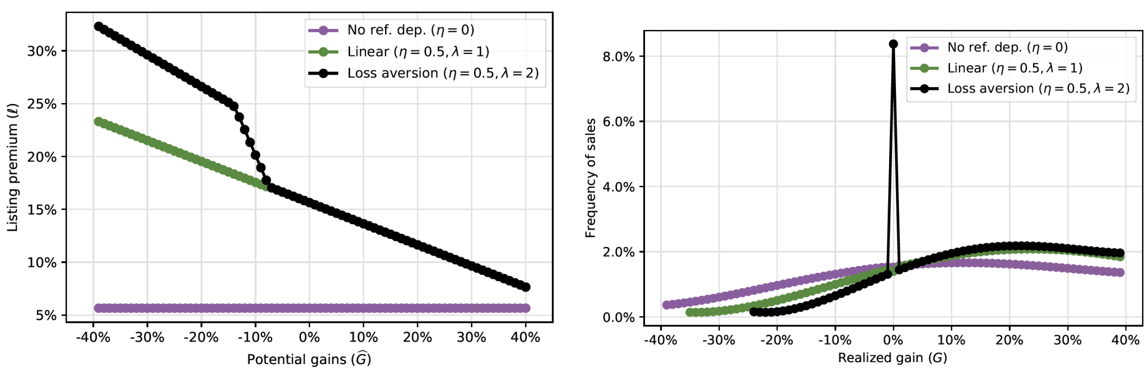


Figure L.3

Price-Volume Correlation

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

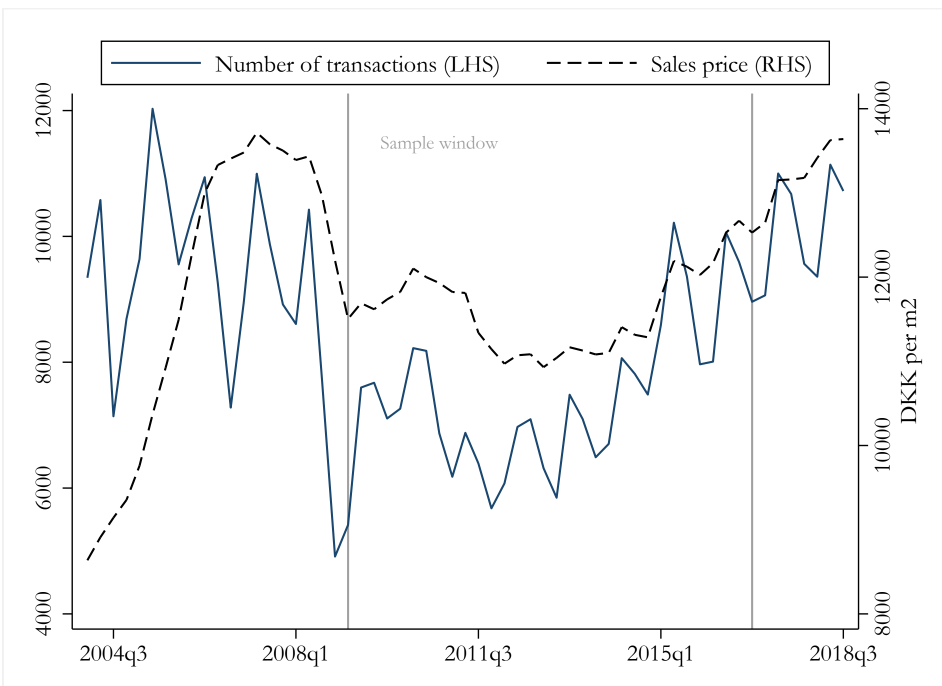


Figure L.4

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 ($N = 114,303$). 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.

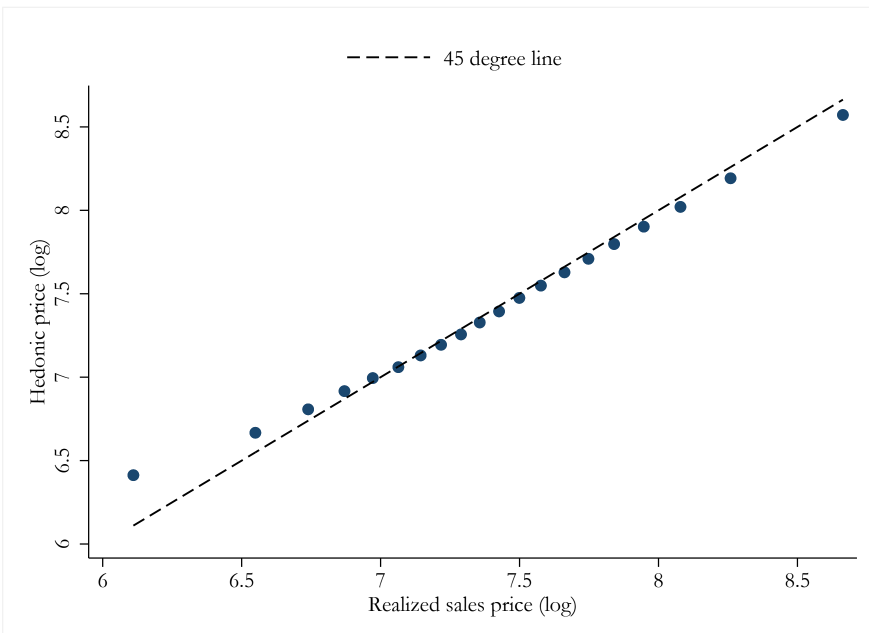
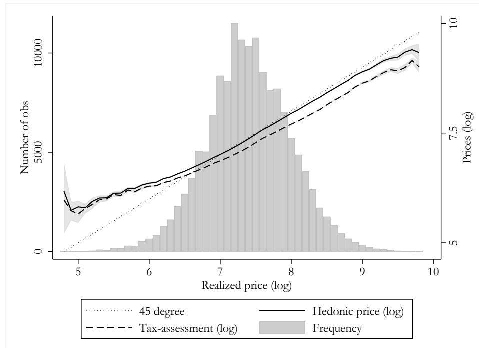


Figure L.5

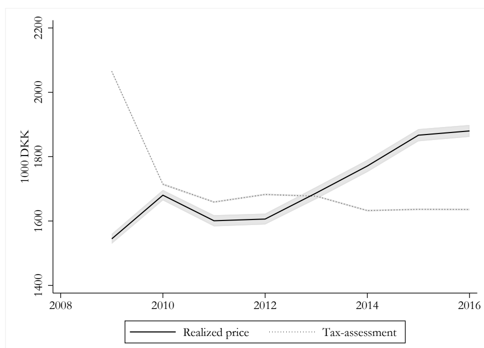
Accuracy of Tax-Assessed Value

Panel (a) shows the tax-assessment relative to the realized sales price as well as the distribution of prices. Panel (b) compares the tax-assessed value to realized sales prices over the full period of time for which we have data. Panel (c) zooms in on our sample period. Data in (a) is the final data set of listings from 2009 to 2016.

(a) Hedonic Price vs. Tax-Assessed Value



(b) Realized Price and Tax-Assessed Value
1992-2016



(c) Realized Price and Tax-Assessed Value
2009-2016

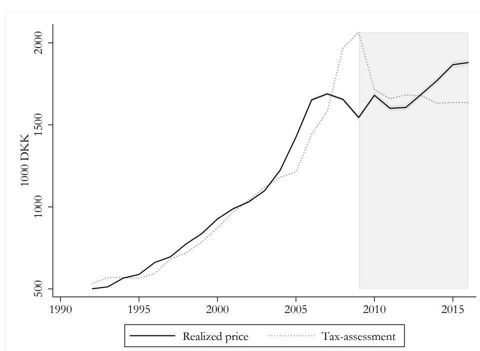


Figure L.6

Listing Premia across Potential Gains and Tax-Assessed Value

This figure compares the listing premium to potential gains relationship for the baseline hedonic model and the tax-assessed value, with data restricted to 2010-2012, when the standalone tax-assessment is most up to date.

(a) Standard hedonic model **(b)** Tax-assessed value

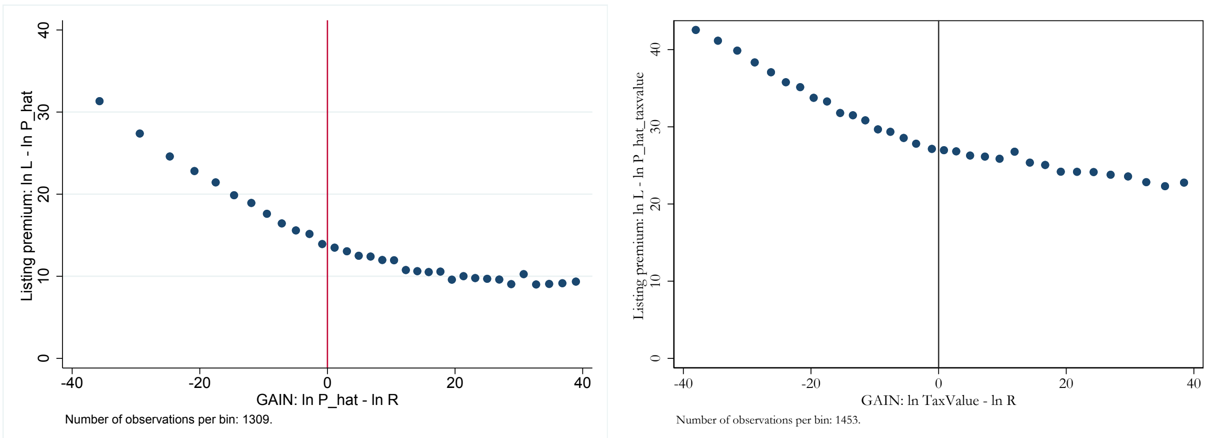


Figure L.7

Probability of Sale by Listing Premia (Concave Demand) and Tax-Assessed Value

This figure compares demand concavity for the baseline hedonic model and the tax-assessed value, with data restricted to 2010-2012, when the standalone tax-assessment is most up to date.

(a) Standard hedonic model **(b)** Tax assessed value

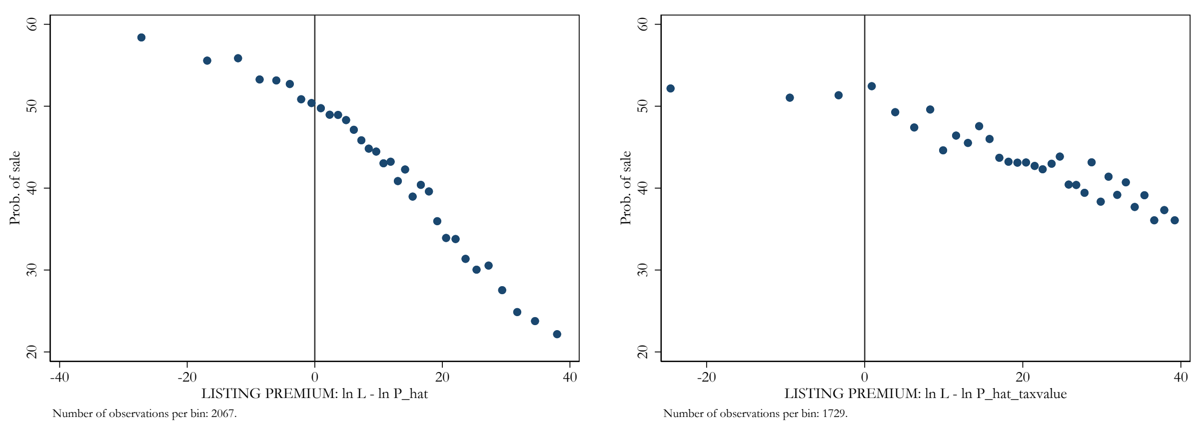
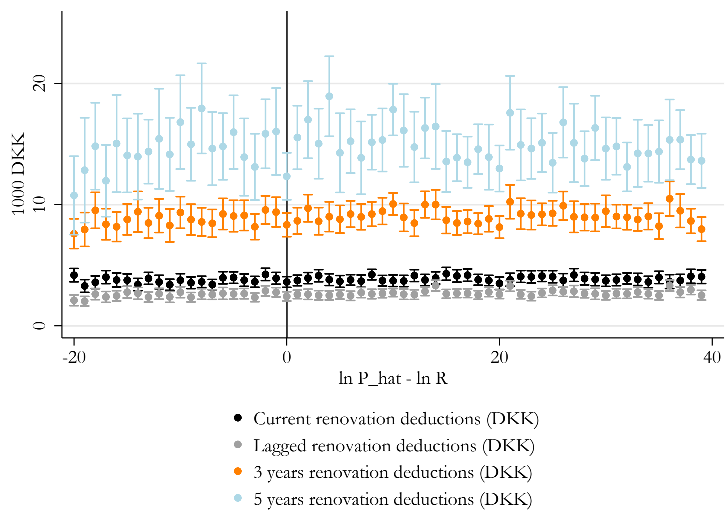


Figure L.8

Renovation Expenses across Potential Gains and Listing Premia

$\widehat{G}$ and $\widehat{\ell}$ These figures show binned averages of different variants of the renovation expense variable, across potential gains $\widehat{G}$, and listing premia $\widehat{\ell}$. Bands reflect 95% confidence intervals.

Panel A: Renovation Expenses by $\widehat{G}$



Panel B: Renovation Expenses by $\widehat{\ell}$

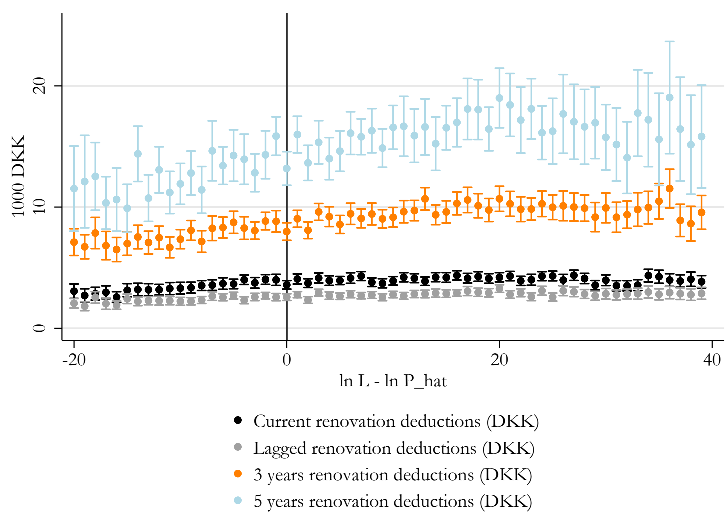
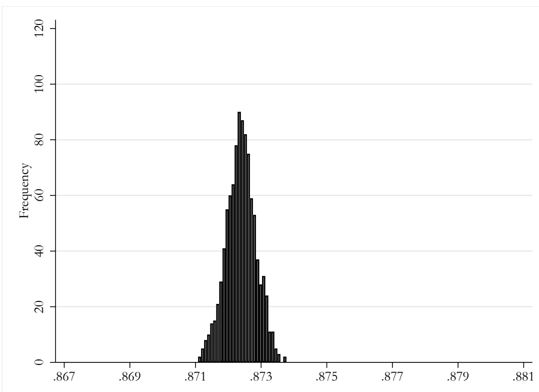


Figure L.9

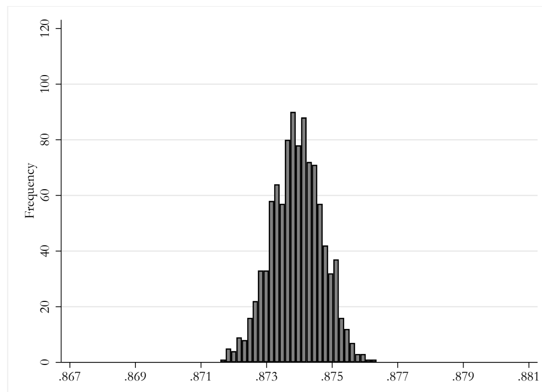
Distribution of R^2s from Out-of-Sample Estimation of the Hedonic Model

These figures show the distribution of R^2 from 1000 regressions of realized price on out-of-sample-predicted hedonic prices.

(a) 25 percent sample



(b) 50 percent sample



(c) 75 percent sample

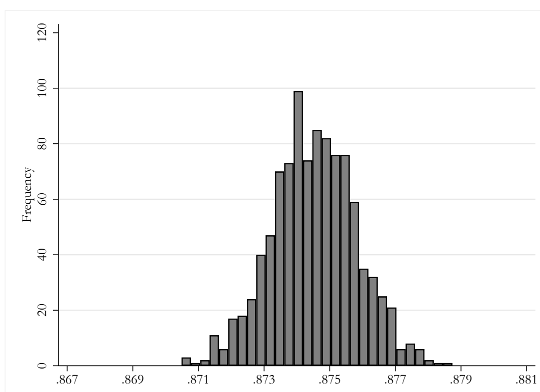
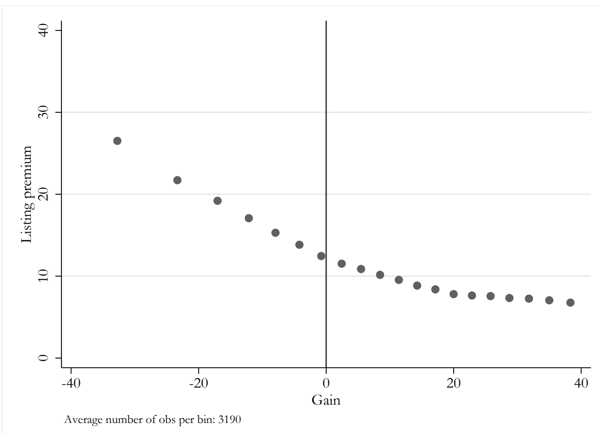


Figure L.10

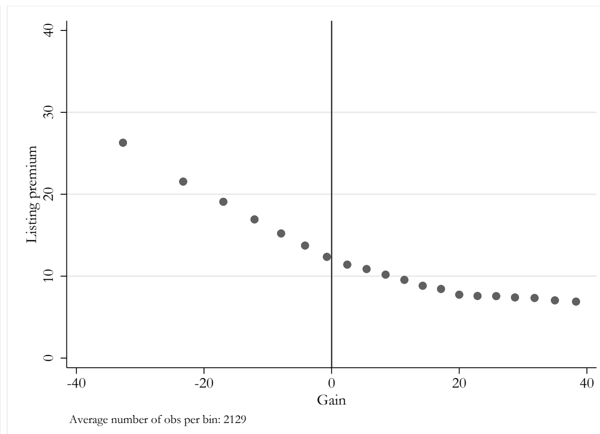
Listing Premia across Potential Gains - Out-of-Sample Predictions

This figure compares the listing premium - potential gains relationship for out-of-sample predictions using different out-of-sample size cut-offs. Dots are averages of 1000 iterations

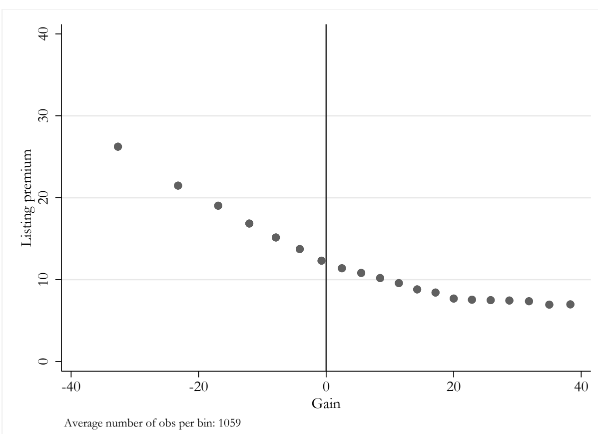
(a) 25 percent out of sample



(b) 50 percent out of sample



(c) 75 percent out of sample



(d) Main data, only sold properties

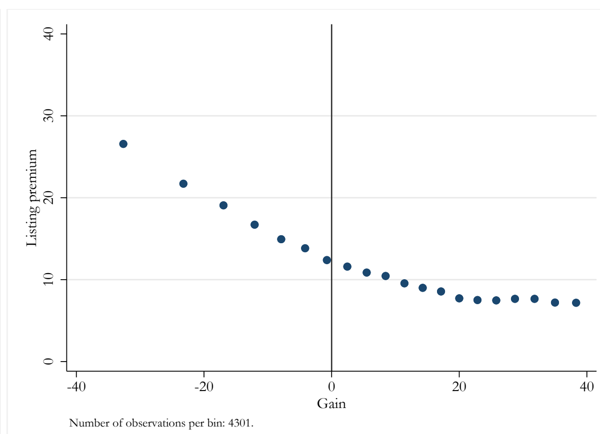
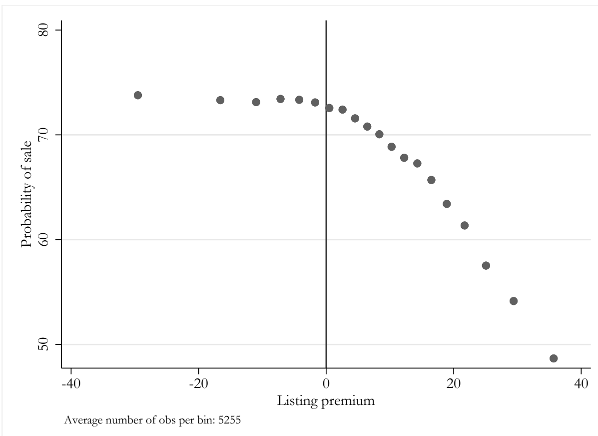


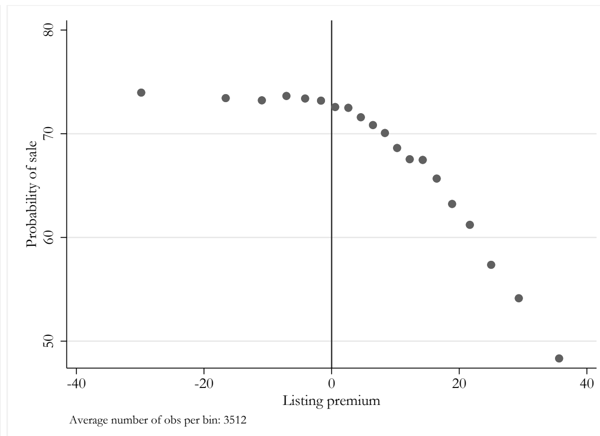
Figure L.11
Probability of Sale by Listing Premium (Concave Demand) -
Out-of-Sample Predictions

This figure compares the demand concavity for out-of-sample predictions using different cut-offs. Dots are averages of 1000 iterations. Probability of sale refers to the probability of sale within 6 months.

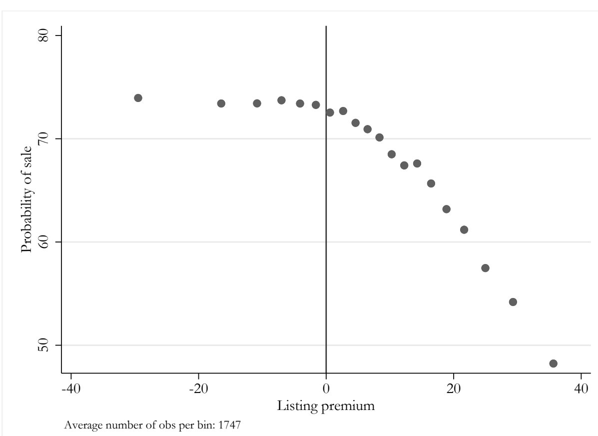
(a) 25 percent out of sample



(b) 50 percent out of sample



(c) 75 percent out of sample



(d) Main data, only sold properties

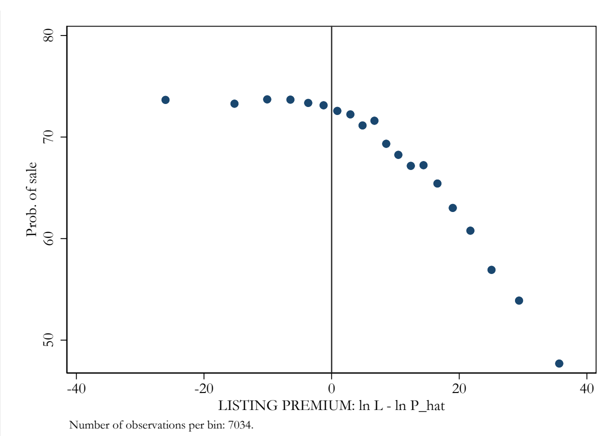


Figure L.12

Out-of-Sample Fit: The Home Equity Dimension

The plot reports the model fit for conditional listing premia profiles along the home equity dimension, conditioning on different levels of potential gains, evaluated at the set of parameters which correspond to the complete version of the model and the complete set of empirical moments, as indicated in line 9 of Table 1 in the paper. Individual dots indicate observations in the data, and solid lines their model-implied counterparts.

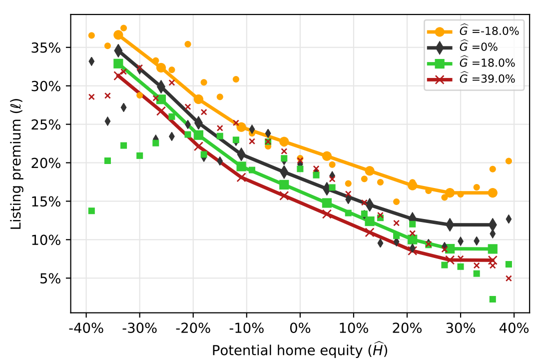
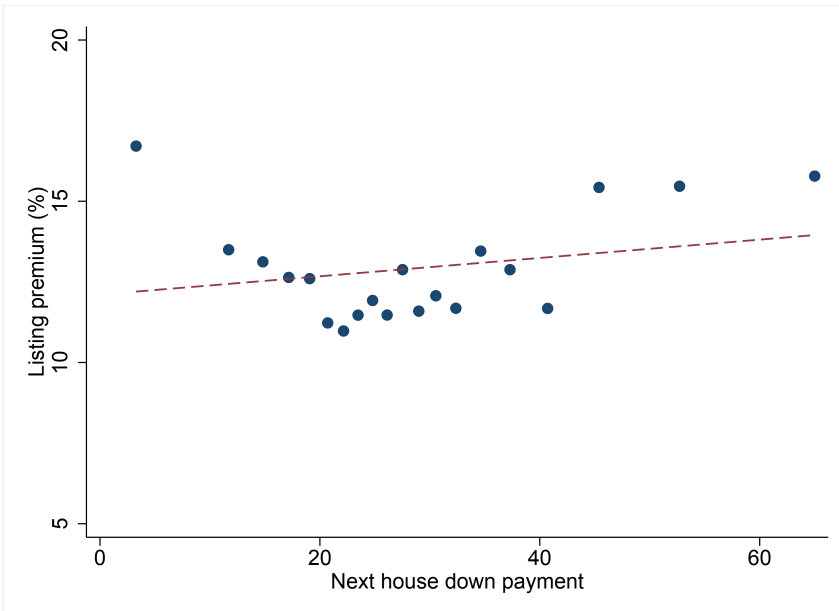


Figure L.13

Listing Premia and Down Payment, and Current and Next House Price

Figure (a) shows a binned scatter plot of the listing 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 ($N = 16,115$). Figure (b) 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 ($N = 36,952$).

(a) Listing Premium Predicts Down Payment



(b) Current and Next House Price

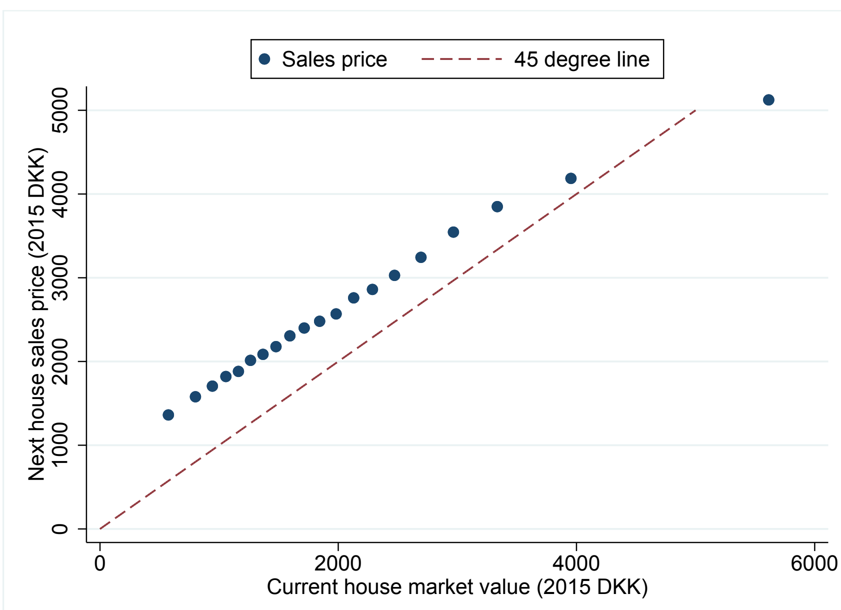


Figure L.14
Time-On-the-Market and Retraction Rate

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

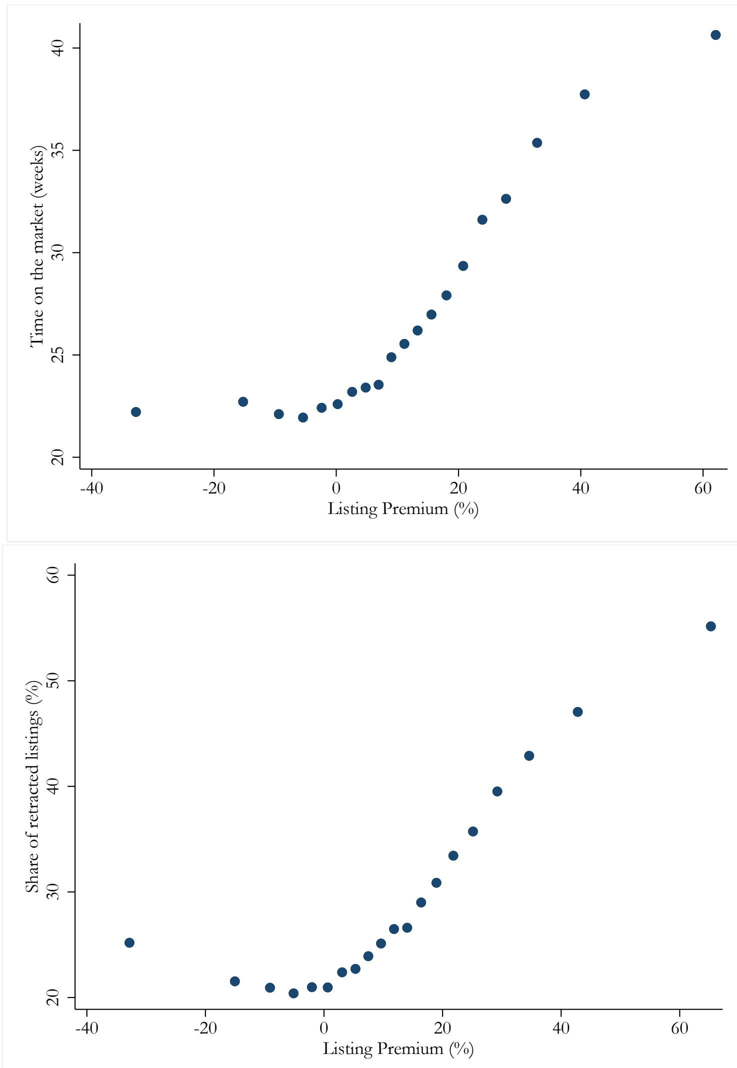


Figure L.15

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

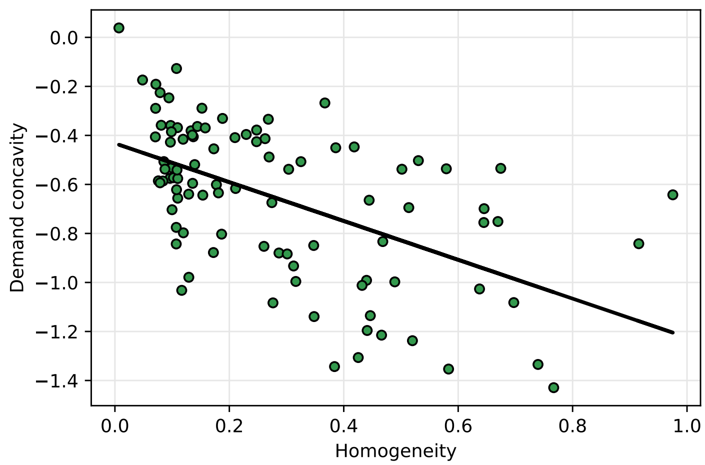
Figure L.16

Regional Variation in Demand Concavity, Listing Premium-Gain Slope and Housing Stock Homogeneity

Panel A shows a scatter plot of the correlation between the main instrument, the share of listed apartments and row houses in a given municipality, and the degree of demand concavity. The degree of demand concavity is measured as the slope coefficient of the effect of an increase in the listing premium on the probability of sale within six months, for positive listing premia ($\ell \in [0, 40]$). Panel B shows the correlation between the estimated listing premium slope over negative potential gains ($\widehat{G} \in [-40, 0)$) and demand concavity across municipalities.

Panel A

Homogeneity of Housing Stock and Demand Concavity across Regions



Panel B

Demand Concavity and "Hockey Stick" Slope across Regions

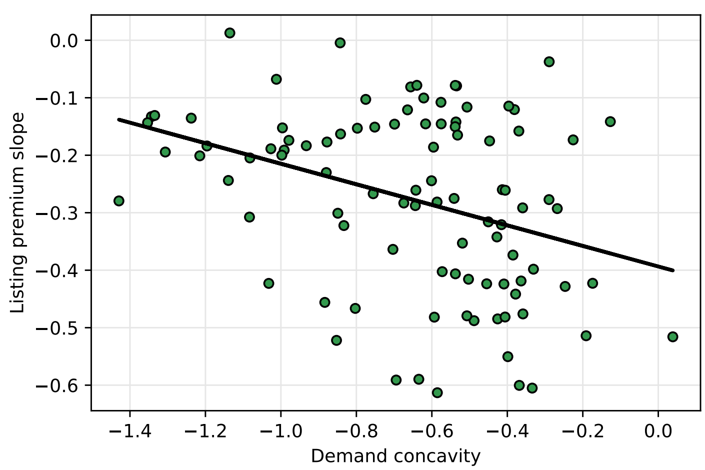
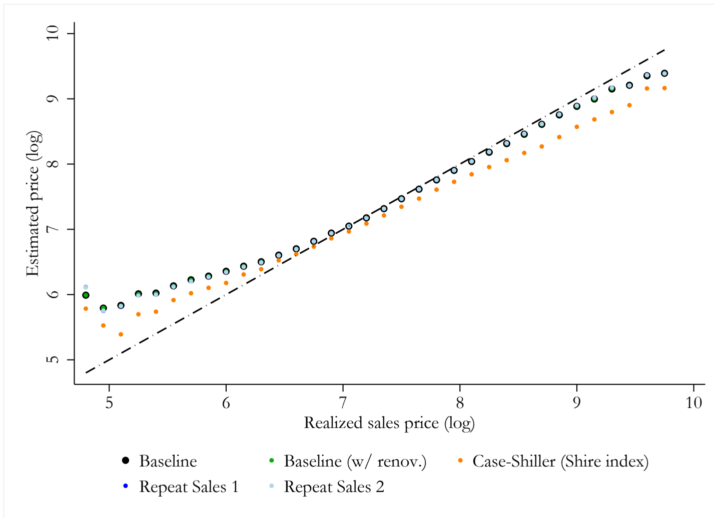


Figure L.17

Estimated vs. Realized ln(price) Across Main Models

This graph compares the main model estimated prices to the realized sales price in logs, across binned averages of the realized sales price. Panel A does this across all properties, while Panel B restricts to properties below 5 million DKK.

Panel A: All



Panel B: Below 5 mil. DKK

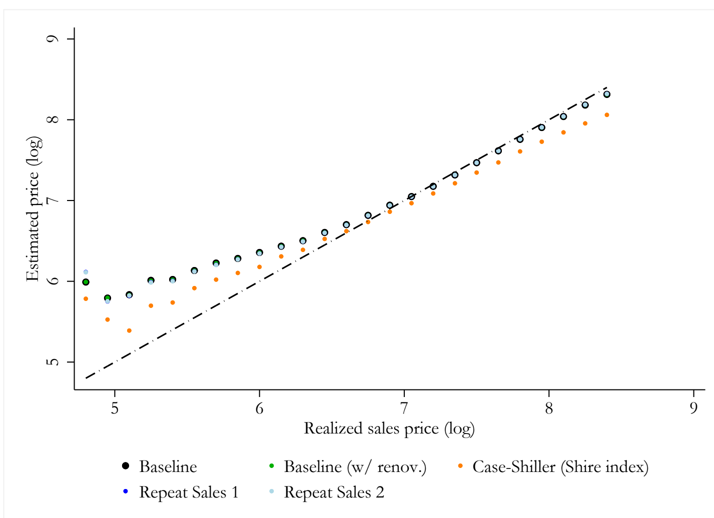
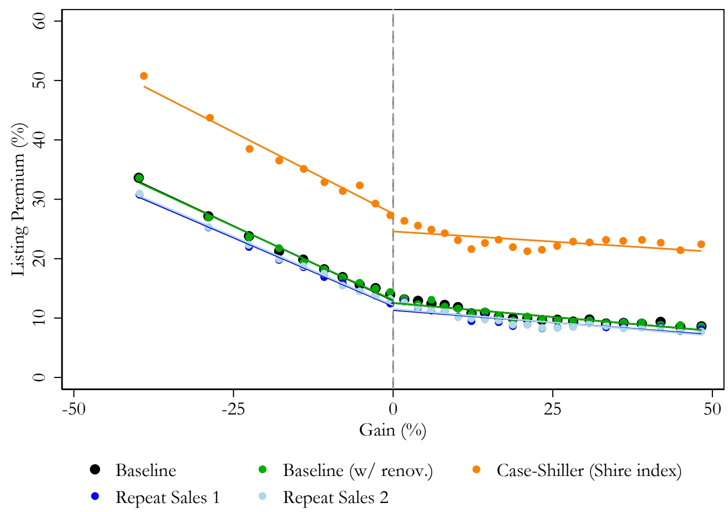


Figure L.18

Hockey Stick and Demand Concavity Across Main Models

These figures compare our two key empirical shapes across our main models of $\widehat{P}$. Panel A shows the hockey stick relationship for listing premia over potential gains, and Panel B shows demand concavity (probability of sale with respect to listing premia).

Panel A: $\widehat{\ell} - \widehat{G}$ Hockey Stick



Panel B: Demand Concavity

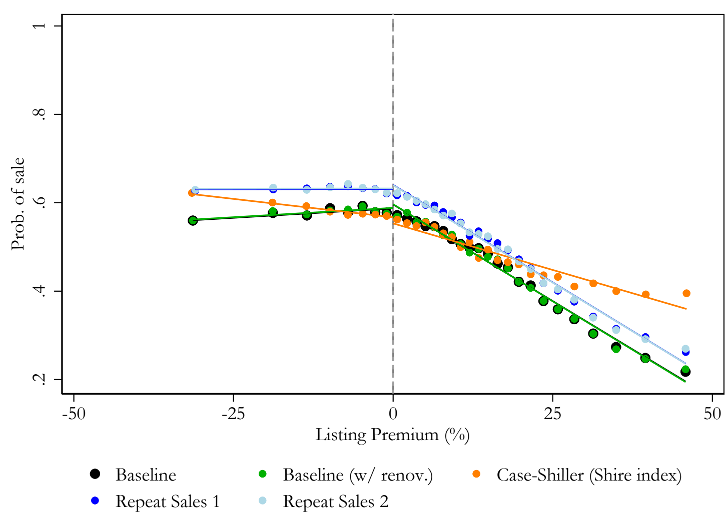
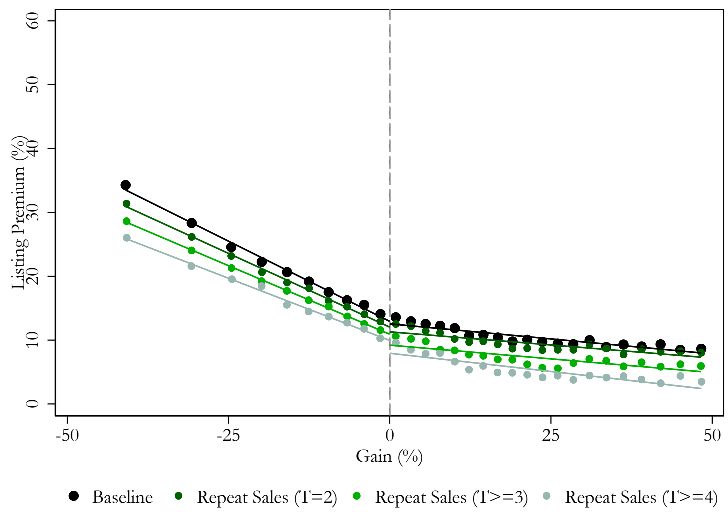


Figure L.19

Hockey Stick and Demand Concavity Across Repeat Sales Models

These figures compare our two key empirical shapes across our repeat sales models of $\widehat{P}$, for differing numbers of repeat sales observations. Panel A shows the hockey stick relationship for listing premia over potential gains, and Panel B shows demand concavity (probability of sale with respect to listing premia).

Panel A: $\widehat{\ell} - \widehat{G}$ Hockey Stick



Panel B: Demand Concavity

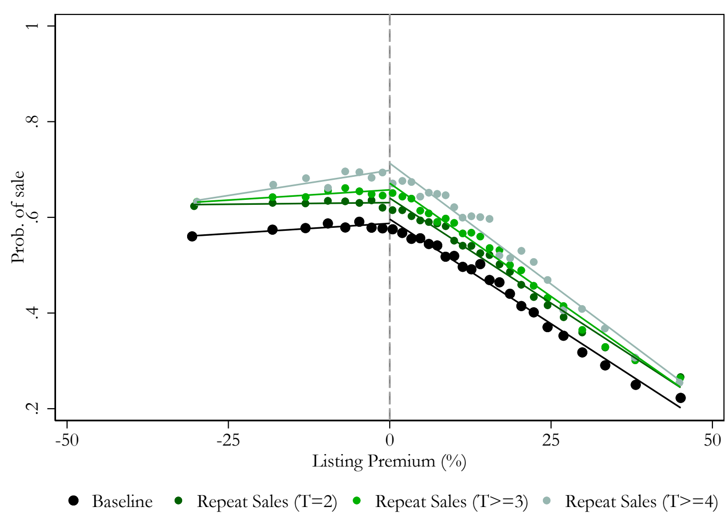


Figure L.20

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.[22]

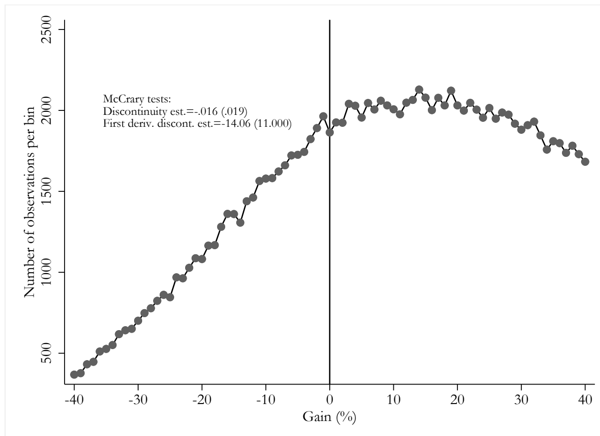


Figure L.21

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.

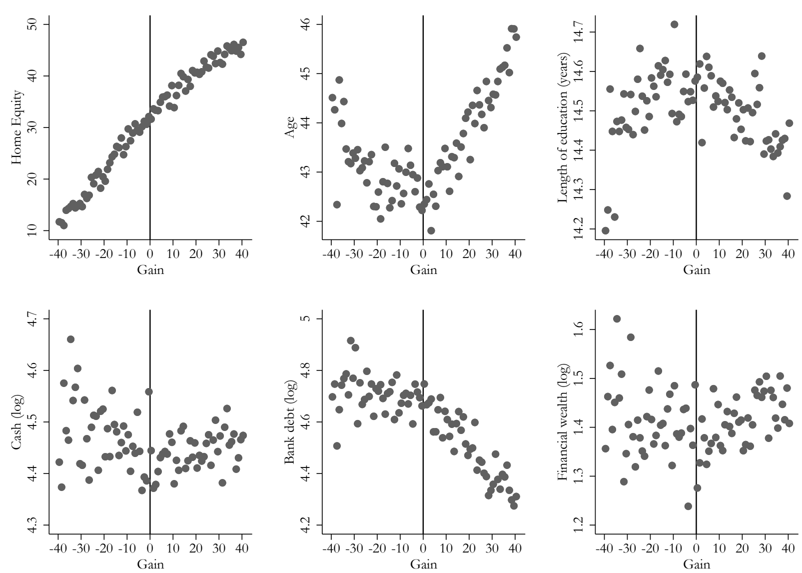


Figure L.22

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).

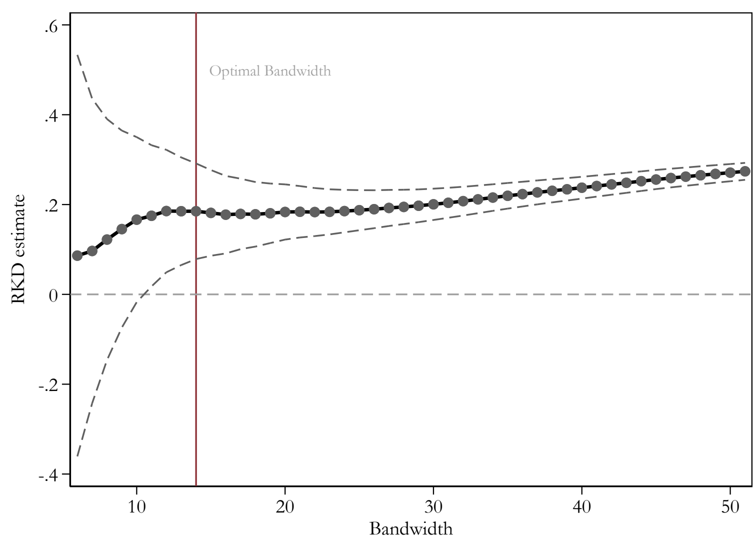


Figure L.23

RKD Estimation: Local Linear vs. Local Quadratic Estimation Results

This figure compares regression kink estimates of listing premia across potential gains, with a cutoff point at 0 potential gains, using a local linear regression with estimates using a local quadratic regression, across different bandwidths $b \in \{b^*, 15, 20, 30\}$. b^* refers to the MSE-optimal bandwidth selector from Calonico et al. (2014).

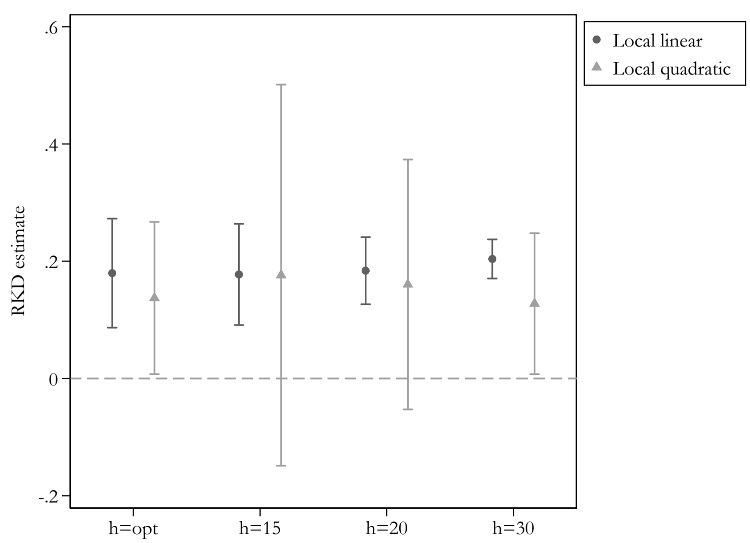


Figure L.24

Incidence of Round Numbers by Rounding Multiple

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

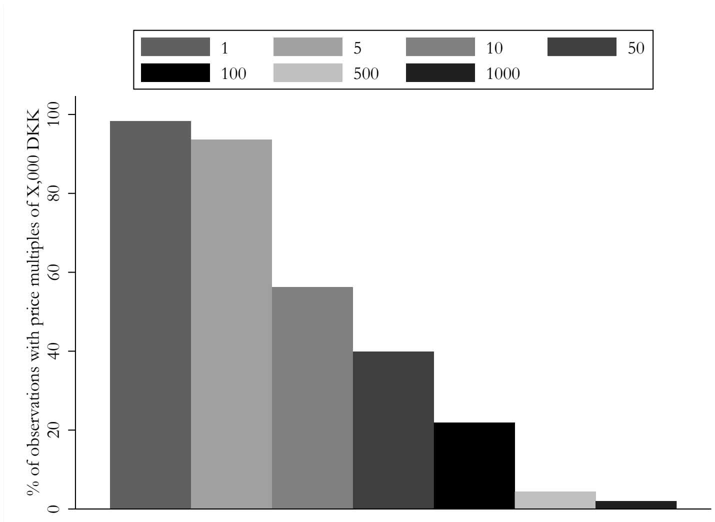
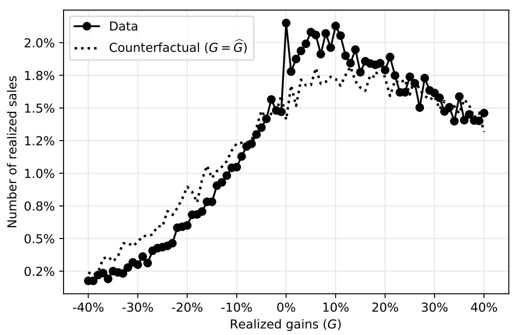


Figure L.25

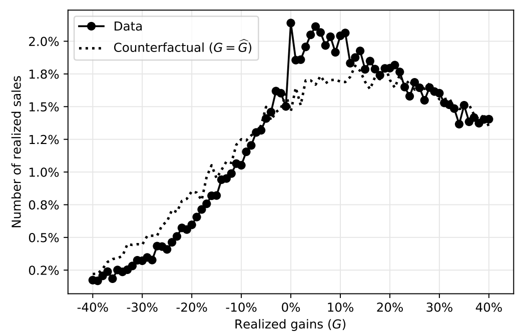
Bunching Robustness: Excluding Sales at Rounded Prices

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; 50,000; 100,000; and 500,000 DKK, respectively. The dots represent the empirical frequency of observations in each 1 percentage point bin of realized gains, and the dotted line reflects the counterfactual frequency based on 1 percentage point bins of potential gains.

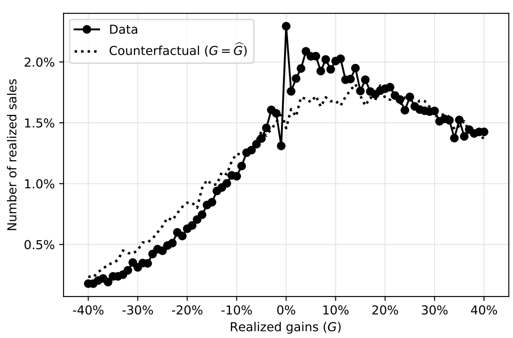
10,000 DKK



50,000 DKK



100,000 DKK



500,000 DKK

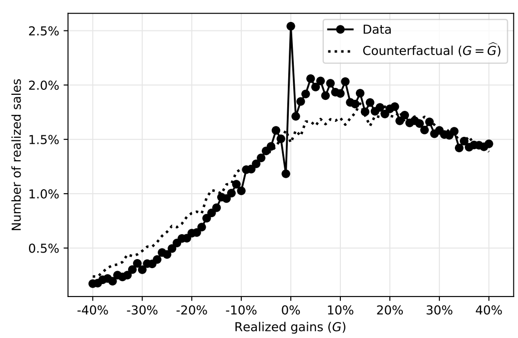
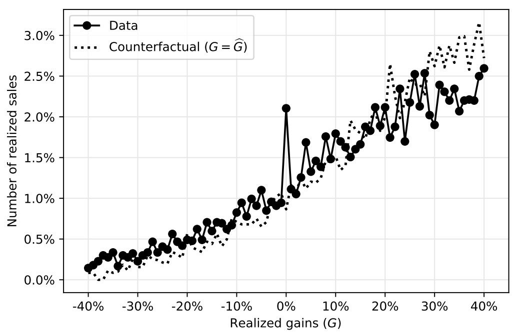


Figure L.26

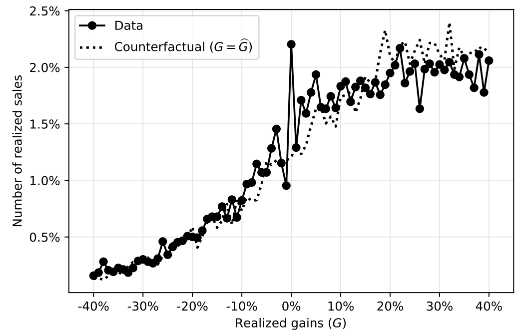
Bunching Robustness: Across Previous Sales Price

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 dots represent the empirical frequency of observations in each 1 percentage point bin of realized gains, and the dotted line reflects the counterfactual frequency based on 1 percentage point bins of potential gains.

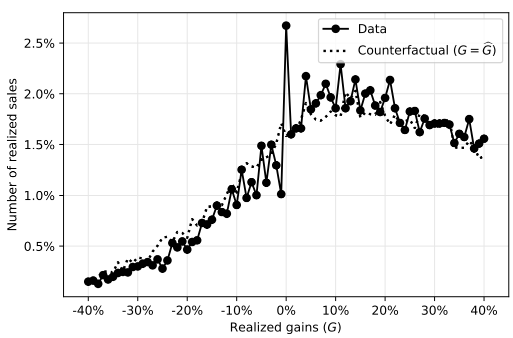
Below DKK 659,000



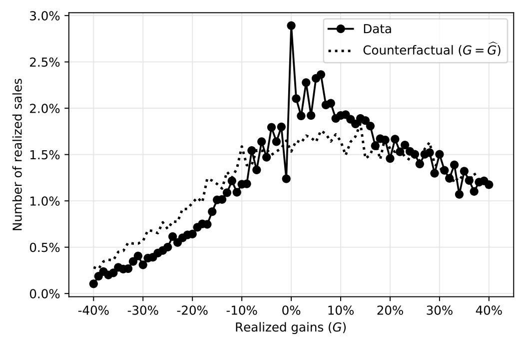
DKK 659,000 – DKK 953,000



DKK 953,000 – DKK 1,313,000



DKK 1,313,000 – DKK 1,901,000



Above DKK 1,901,000

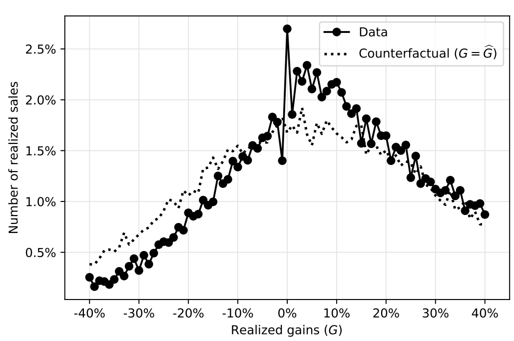
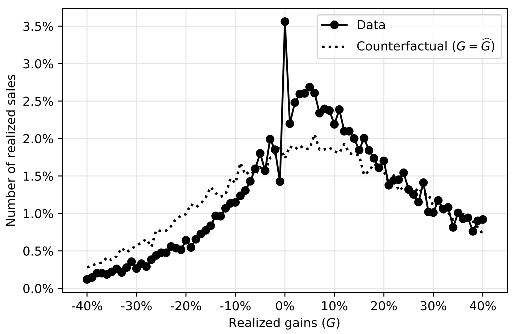


Figure L.27

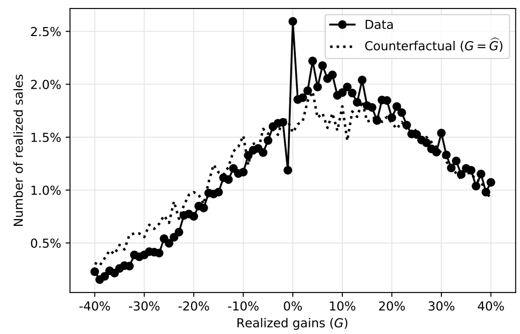
Bunching Robustness: Across Holding Periods

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 dots represent the empirical frequency of observations in each 1 percentage point bin of realized gains, and the dotted line reflects the counterfactual frequency based on 1 percentage point bins of potential gains.

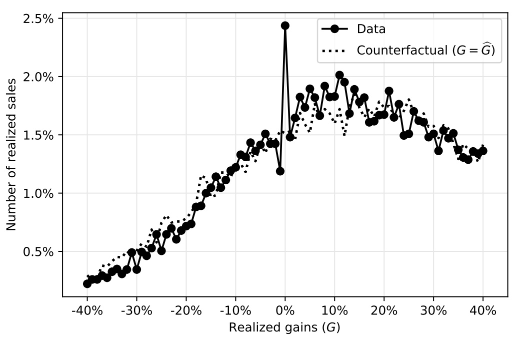
Below 3 years



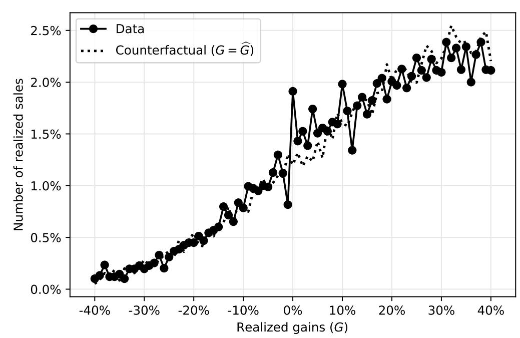
3–6 years



6–9 years



9–12 years



Above 12 years

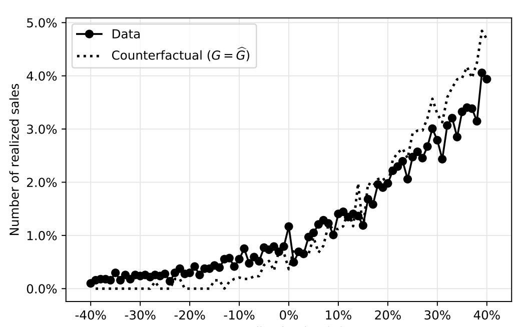


Figure L.28

Bunching Robustness: Model with Cohort Fixed Effects

This figure shows robustness for the excess mass of the frequency of sales across realized gains relative to the potential gains counterfactual, using the baseline hedonic model augmented with cohort fixed effects.

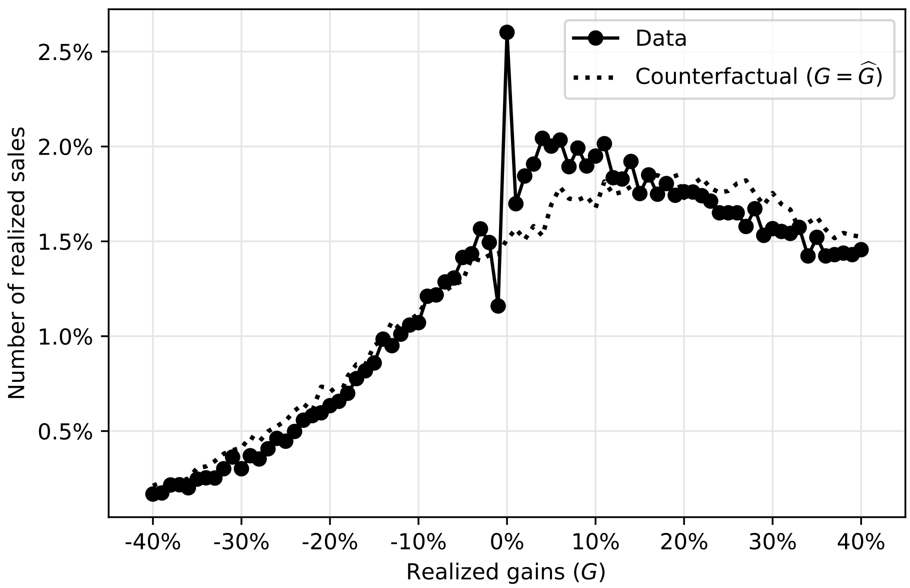
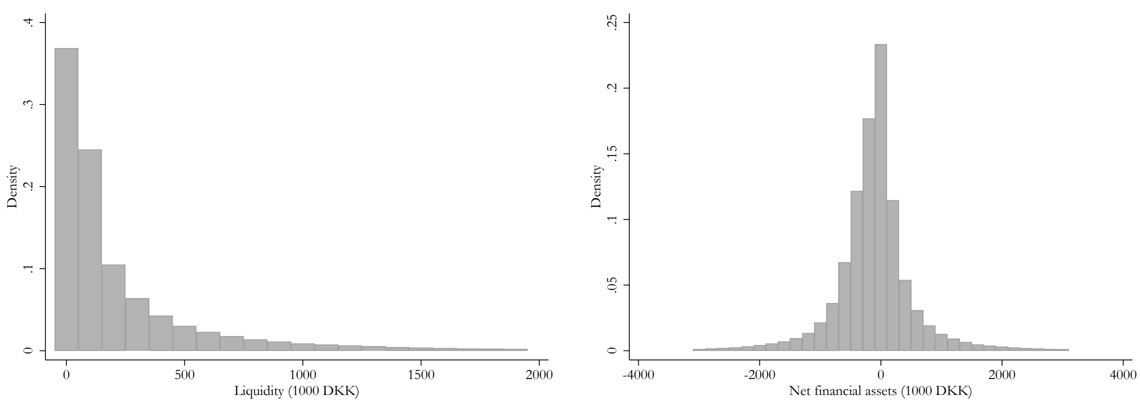


Figure L.29

Summary Statistics: Household Demographics

This figure shows four histograms of household characteristics. Panel A depicts the distribution of available liquid assets (left) and net financial wealth (right). 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. 1.6 percent of households have liquid asset above 2 million DKK, and 1.2 percent have net financial assets below -3 million or above 3 million DKK, but the figures are truncated at these values for better visual representation of the main mass. 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

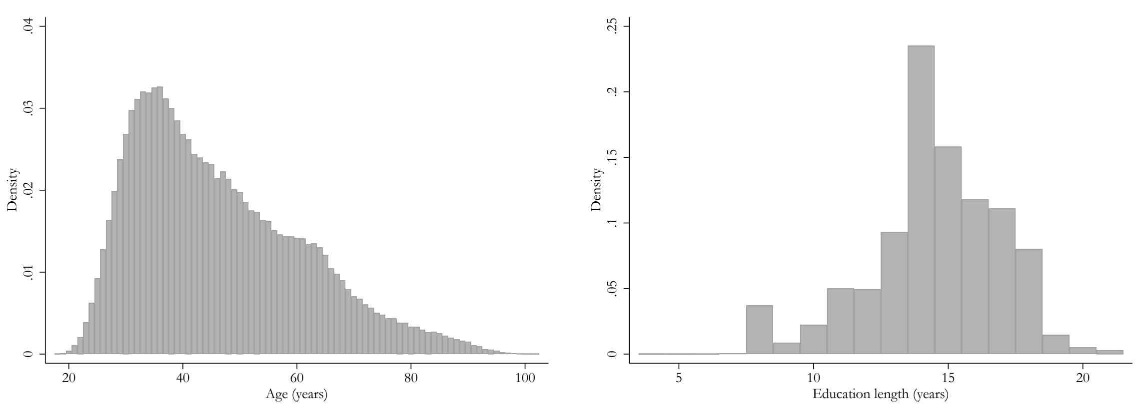


Figure L.30

Residualized 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.

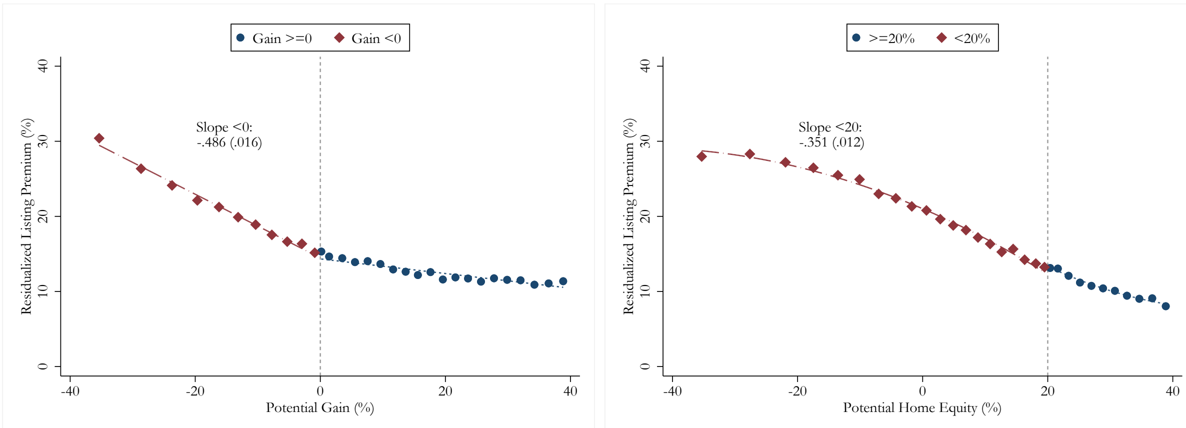


Figure L.31

Listing Premium " Hockey Stick" for Sellers Without Mortgage

This figure shows the relationship between listing premium and potential gains for the sample of households with no mortgage ($N = 42,124$), using a binned scatter plot of equal-sized bins for $\widehat{G} \in [-50, 50]$.

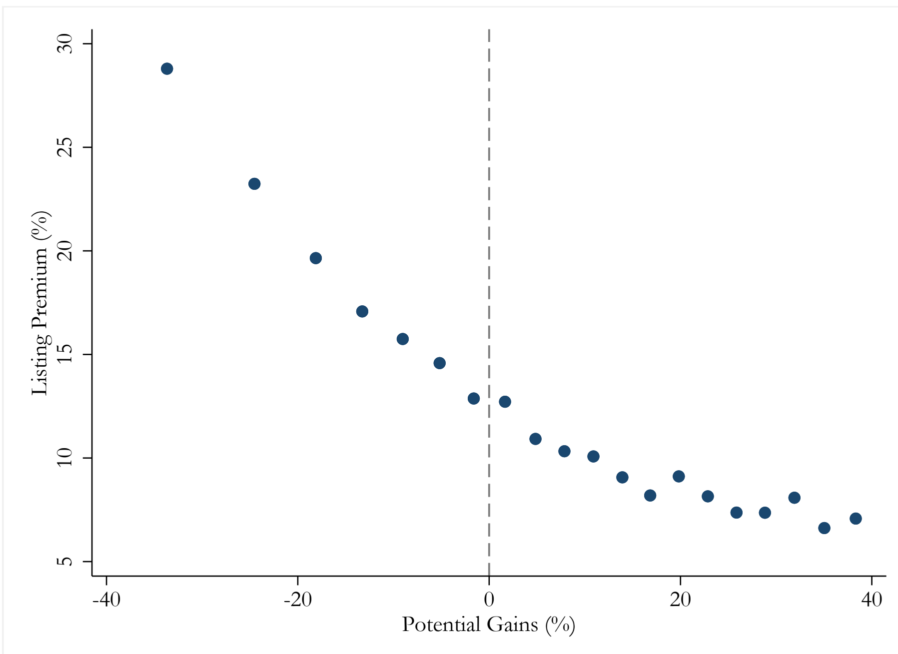


Table L.1

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 real estate[a]	**615,053**
Unmatched in registers[b]	-107,361
	507,361
Cleaning	
No reference price[c]	-145,006
Owner ID not uniquely determined[d]	-71,870
Non-household buyer	-10,206
Foreclosures	-6,340
Extreme price[e]	-5,503
Owner ID not found[f]	-3,988
Missing lot size	-2,825
Error in listing or previous purchase date[g]	-1,941
Intra-family sale and other special circumstances	-1,746
No listing price	-881
Missing hedonic characteristics	-8
	257,047
Sample selection	
Summer house	-24,130
Professional investor[h]	-18,394
Final data	**214,523**
Of which with a mortgage	172,399
Of which without a mortgage	42,124

[a] *Excluding listings of cooperative housing.*
[b] *Reasons could be misreported addresses or non-ordinary owner-occupied housing.*
[c] *Purchased before 1992.*
[d] *E.g. properties with several owners from different households.*
[e] *Listed or sold at prices below 100,000 DKK or above 20,000,000 DKK (2015-prices) or marked as extreme price by Statistics Denmark.*
[f] *No owner ID found in registers.*
[g] *Listing date is before previous purchase date.*
[h] *Seller owns more than 3 properties.*

Table L.2

R^2 of Hedonic Model - Contributions

This table shows the R^2 from different components of the hedonic model for the sample period 2009-2016, as well as the period pre-2013 and post-2013. Row 1 presents the R^2 from using the hedonic characteristics only. Row 2 shows the R^2 from municipality-year fixed effects, and row 3 the R^2 from up to the third-degree polynomial of the tax-assessed property value. Row 4 shows the contribution of lagged renovation tax exemptions for the years 2012 to 2016. Column 1, 3, and 5 show separate contributions of each component, and column 2, 4, and 6 show composite contributions to R^2.

	(1)	(2)	(3)	(4)	(5)	(6)
	Simple	Cumulative	Simple	Cumulative	Simple	Cumulative
	2009-2016		2009-2012		2013-2016	
1) Hedonics only	0.536	0.536	0.551	0.551	0.544	0.544
2) Municipality-year FEs	0.477	0.768	0.477	0.758	0.467	0.775
3) Tax-assessment	0.800	0.876	0.805	0.865	0.834	0.881
4) Renovation exemptions	0.026	0.876	0.009	0.865	0.027	0.882

Table L.3

Overview and Description of Models of $\widehat{P}$

This table provides an overview of the different models of $\widehat{P}$ that we implement and model features. A more detailed description of the estimation methods is provided in the online appendix.

Model	Name	Description	Time-varying observables	Tax-assessed value	Repeat sales	Renovation expenses
Main						
Ia	Baseline	Baseline hedonic model	✓	✓	×	×
Ib	Baseline (with renovation expenses)	Baseline with 1-year lagged renovation expenses	✓	✓	×	✓
Ic	Baseline (OOS)*	Baseline, estimated on 50% of the data and fitted on the remaining 50%	✓	✓	×	×
II	Simple Repeat (Shire index)	Simple repeat sales model using previous purchase price and shire-level house price changes	×	×	✓	×
IIIa	Repeat Sales I	Baseline with 1-year lagged renovation expenses and last pricing residuals $(\nu_{is}+\omega_{is})$ (for $T=2$, one repeat sale)	✓	✓	✓	✓
IIIb	Repeat Sales II	Baseline with 1-year lagged renovation expenses and average past pricing residuals $(\bar{\nu}_{it_{\tau<t}}+\bar{\omega}_{it_{\tau<t}})$	✓	✓	✓	✓
Additional						
IVa	Repeat Sales ($T=2$)	Baseline with last pricing residual, for $T=2$ (one repeat sale)	✓	✓	✓	×
IVb	Repeat Sales ($T\geq 3$)	Baseline with average past pricing residuals, for $T\geq 3$ ($\geq$ two repeat sales)	✓	✓	✓	×
IVc	Repeat Sales ($T\geq 4$)	Baseline with average past pricing residuals, for $T\geq 4$ ($\geq$ three repeat sales)	✓	✓	✓	×
IVd	Repeat Sales ($T\geq 2$)	Baseline with average past pricing residuals for any number of repeat sales	✓	✓	✓	×
Va	Renovations (1yr)	Subset of Ib, where renovation expenses over the past year are available	✓	✓	×	✓
Vb	Renovations (3yr)	Va, but where cumulative 3-year renovation expenses are available	✓	✓	×	✓
Vc	Renovations (5yr)	Va, but where cumulative 5-year renovation expenses are available	✓	✓	×	✓

Table L.4

Out-of-Sample Test of Hedonic Model

This table shows the mean R^2 from 1000 regressions of realized price on three different in-sample estimation shares and accompanying predicted prices from the baseline hedonic model. Standard errors of the mean are in parentheses.

	Mean
50 pct out-of-sample	0.875
	(0.00002)
25 pct out-of-sample	0.876
	(0.00004)
100 pct in-sample	0.878
	(.)
Observations	1,000

Table L.5

Out-of-Sample Test of Hedonic Model without Tax-Assessed Value

This table shows the mean R^2 from 1000 regressions of realized price on realized price on three different in-sample estimation shares and accompanying predicted prices from the baseline hedonic model, without controlling for the tax-assessed value. Standard errors of the mean are in parentheses.

	Mean
50 pct out-of-sample	0.765
	(0.00004)
25 pct out-of-sample	0.766
	(0.00007)
100 pct in-sample	0.769
	(.)
Observations	1,000

Table L.6

Regional Variation in Demand Concavity and Hockey Stick - OLS and IV 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.[23] 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	-0.322***	-0.534***	-0.483***	-0.500**
	(0.061)	(0.099)	(0.085)	(0.253)
Household controls				✓
Observations	96	96	96	96
R^2	0.385	-	-	-
First-stage F-stat	-	37.25	34.15	20.05

Table L.7

Comparison of Moment Summary Metrics Across Models of $\widehat{P}$ - Main Models

This table estimates simple linear coefficients to summarize and compare the information contained in the non-parametric moments we use for the structural estimation. (1) is the average listing premium ($\widehat{\ell}$) around zero potential gains ($\widehat{G} \in (-1, 1]$). (2) is the piecewise-linear slope of the hockey stick in listing premia, over negative potential gains ($\widehat{G} \in [-40, 0)$). (3) is the piecewise-linear slope of the hockey stick in listing premia, over positive potential gains ($\widehat{G} \in [0, 40]$). (4) is the piecewise-linear slope of the hockey stick in listing premia, over home equity in the constrained range ($\widehat{H} \in [-40, 20)$). (5) is the piecewise-linear slope of the probability of sale with respect to negative listing premia ($\widehat{\ell} \in [-20, 0)$). (6) is the piecewise-linear slope of the probability of sale with respect to positive listing premia ($\widehat{\ell} \in [0, 40]$). We refer to (5) and (6) as summarizing "concave demand". (7) is the slope in the probability of listing with respect to potential gains, estimated in the data comprising the full housing stock. The underlying number of observations vary slightly for models using repeat sales and the shire-level price index, as we cannot compute past pricing residuals or the price index, respectively, due to data limitations for some observations. *Model Ib is estimated by randomly sampling 50% of the data to estimate the baseline model (Ia), and only using the remaining 50% of the data to compute the summary metrics out of sample, based on 100 random draws from the full housing stock data.

	Ia Baseline	Ib Baseline (w/ renov.)	Ic Baseline (OOS)*	II Simple Repeat (Shire index)	IIIa Repeat Sales 1 ($T = 2$)	IIIb Repeat Sales 2 ($T \geq 2$)
(1) Level of $\widehat{\ell}$ ($\widehat{G} \in (-1, 1]$)	13.879	14.433	13.814	26.698	12.403	12.889
	(0.405)	(0.382)	(0.390)	(0.848)	(0.402)	(0.428)
(2) Slope $\widehat{\ell} - \widehat{G}$ ($\widehat{G} < 0$)	-0.486	-0.478	-0.485	-0.535	-0.452	-0.450
	(0.016)	(0.016)	(0.010)	(0.034)	(0.018)	(0.018)
(3) Slope $\widehat{\ell} - \widehat{G}$ ($\widehat{G} \geq 0$)	-0.113	-0.113	-0.115	-0.085	-0.107	-0.111
	(0.010)	(0.010)	(0.007)	(0.021)	(0.009)	(0.009)
(4) Slope $\widehat{\ell} - \widehat{H}$ ($\widehat{H} < 20$)	-0.351	-0.348	-0.357	-0.627	-0.300	-0.297
	(0.012)	(0.012)	(0.006)	(0.011)	(0.012)	(0.012)
(5) Slope P(sale)-$\widehat{\ell}$ ($\widehat{\ell} < 0$)	0.021	0.017	0.012	-0.180	-0.025	-0.029
	(0.043)	(0.043)	(0.043)	(0.048)	(0.045)	(0.045)
(6) Slope P(sale)-$\widehat{\ell}$ ($\widehat{\ell} \geq 0$)	-0.905	-0.912	-0.892	-0.491	-0.922	-0.927
	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)
Model R^2	0.876	0.876	0.873	0.570	0.881	0.881
Number of observations	214,523	214,523	107,254	203,051	181,179	181,191
(7) P(listing)-$\widehat{G}$	0.003	0.002	0.021	0.012	0.037	0.036
	(0.002)	(0.002)	(0.000)	(0.004)	(0.003)	(0.003)
Number of observations (ext.)	5,540,349	5,540,349	2,770,188	5,113,712	2,712,310	2,713,136

Table L.8

Comparison of Moment Summary Metrics Across Models of $\widehat{P}$ - Additional Models

This table estimates simple linear coefficients to summarize and compare the information contained in the non-parametric moments we use for the structural estimation. (1) is the average listing premium ($\widehat{\ell}$) around zero potential gains ($\widehat{G} \in (-1, 1]$). (2) is the piecewise-linear slope of the hockey stick in listing premia, over negative potential gains ($\widehat{G} \in [-40, 0)$). (3) is the piecewise-linear slope of the hockey stick in listing premia, over positive potential gains ($\widehat{G} \in [0, 40]$). (4) is the piecewise-linear slope of the hockey stick in listing premia, over home equity in the constrained range ($\widehat{H} \in [-40, 20)$). (5) is the piecewise-linear slope of the probability of sale with respect to negative listing premia ($\widehat{\ell} \in [-20, 0)$). (6) is the piecewise-linear slope of the probability of sale with respect to positive listing premia ($\widehat{\ell} \in [0, 40]$). We refer to (5) and (6) as summarizing "concave demand". (7) is the slope in the probability of listing with respect to potential gains, estimated in the data comprising the full housing stock.

	IVa	IVb	IVc	IVd	Va	Vb	Vc
	Repeat sales only				With renovations data only		
	($T = 2$)	($T \geq 3$)	($T \geq 4$)	($T \geq 2$)	(1yr)	(3yr)	(5yr)
(1) Level of $\widehat{\ell}$ ($\widehat{G} \in (-1, 1]$)	12.27	10.726	10.781	12.77	13.511	12.264	12.649
	(0.386)	(0.401)	(0.514)	(0.394)	(0.451)	(0.539)	(0.803)
(2) Slope $\widehat{\ell} - \widehat{G}$ ($\widehat{G} < 0$)	-0.450	-0.418	-0.388	-0.452	-0.488	-0.504	-0.540
	(0.018)	(0.021)	(0.025)	(0.017)	(0.019)	(0.021)	(0.032)
(3) Slope $\widehat{\ell} - \widehat{G}$ ($\widehat{G} \geq 0$)	-0.105	-0.114	-0.143	-0.112	-0.135	-0.153	-0.179
	(0.009)	(0.011)	(0.012)	(0.009)	(0.011)	(0.011)	(0.017)
(4) Slope $\widehat{\ell} - \widehat{H}$ ($\widehat{H} < 20$)	-0.301	-0.232	-0.183	-0.299	-0.344	-0.327	-0.310
	(0.013)	(0.014)	(0.018)	(0.012)	(0.018)	(0.024)	(0.043)
(5) Slope P(sale)-$\widehat{\ell}$ ($\widehat{\ell} < 0$)	-0.007	0.074	0.113	-0.017	0.150	0.259	0.289
	(0.045)	(0.061)	(0.085)	(0.045)	(0.051)	(0.061)	(0.101)
(6) Slope P(sale)-$\widehat{\ell}$ ($\widehat{\ell} \geq 0$)	-0.910	-0.969	-1.005	-0.918	-0.870	-0.804	-0.810
	(0.014)	(0.019)	(0.029)	(0.014)	(0.016)	(0.021)	(0.037)
Model R^2	0.881	0.888	0.895	0.881	0.881	0.884	0.890
Number of observations	181,179	95,589	44,387	181,191	137,026	85,717	29,177
(7) P(listing)-$\widehat{G}$	.038	.054	.076	.036	.002	.004	.009
	(0.003)	(0.005)	(0.007)	(0.003)	(0.003)	(0.004)	(0.008)
Number of observations (ext.)	2,712,310	1,155,737	445,126	2,713,136	3,491,461	2,051,816	671,837

Table L.9

Regression Kink Design

The table shows results from sharp regression kink tests of a discontinuous increase in the listing premia slope over potentials gains, at the 0% potential gain cutoff, for varying bandwidths $b \in \{b^*, 20, 30, 40\}$. b^* refers to the optimally chosen bandwidth using a MSE-optimal bandwidth selector from Calonico et al. (2014). All estimations include the following control variables: 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) h=opt	(2) h=20	(3) h=30	(4) h=40
RD_Estimate	0.180***	0.184***	0.204***	0.241***
	(0.047)	(0.029)	(0.017)	(0.012)
Cutoff	0.00	0.00	0.00	0.00
Bandwidth	14	20	30	40
Polynomial order	1	1	1	1
N below cutoff	49130	49130	49130	49130
N above cutoff	165393	165393	165393	165393

Table L.10

Alternative Estimation of "Hockey Stick" Pattern

The table reports estimated coefficients from the following regression specifications:

$$\ell_i = a_0 + b_0 \widehat{G}_i + \varepsilon_i,$$

$$L_i = a_0 + b_1 \widehat{P} + b_2 R + \varepsilon_i,$$

with all variables defined as in the paper. In Panel B, we interact terms with an indicator variable which takes the value of 1 if potential gains are positive, and zero otherwise. For consistency with the binned moments used in structural estimation along the potential gains dimension, we restrict the support to potential gains domain between -40% and +40%. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively, based on standard errors clustered at the municipality × year level.

Panel A

	Listing premium ($\ell = L - \widehat{P}$)	Listing price (L)
Potential gains ($\widehat{G} = \widehat{P} - R$)	-0.269***	
	(0.007)	
Hedonic valuation ($\widehat{P}$)		0.709***
		(0.007)
Reference point (R)		0.258***
		(0.006)
Number of obs.	122,916	122,916
R^2	0.073	0.083

Panel B

		Listing premium ($\ell = L - \widehat{P}$)	Listing price (L)
Potential gains ($\widehat{G} = \widehat{P} - R$)	- Loss domain ($\widehat{G} < 0$)	-0.494***	
		(0.015)	
	- Gain domain ($\widehat{G} \geq 0$)	-0.118***	
		(0.009)	
Hedonic valuation ($\widehat{P}$)	- Loss domain ($\widehat{P} < R$)		0.496***
			(0.016)
	- Gain domain ($\widehat{P} \geq R$)		0.861***
			(0.010)
Reference point (R)	- Loss domain ($\widehat{P} < R$)		0.472***
			(0.015)
	- Gain domain ($\widehat{P} \geq R$)		0.107***
			(0.009)
Number of obs.		122,916	122,916
R^2		0.840	0.885