Reference Dependence in the Housing Market

Online Appendix

(For Online Publication)

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

August 23, 2022

*Andersen: Copenhagen Business School, Email: san.fi@cbs.dk. Badarinza: National University of Singapore, Email: cristian.badarinza@nus.edu.sg. Liu: The Wharton School, University of Pennsylvania, Email: lliu1@wharton.upenn.edu. 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

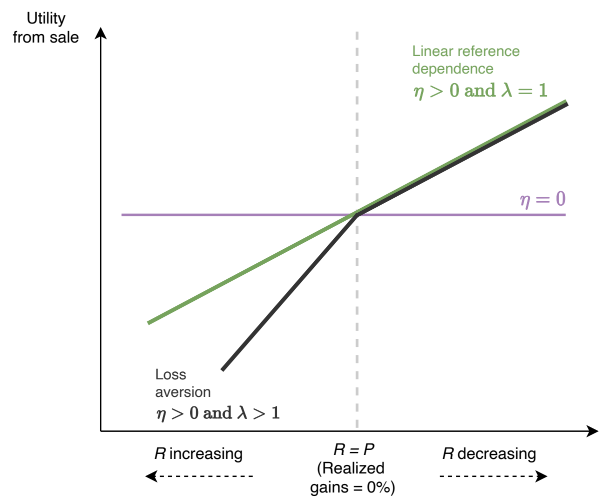
A Additional Tables and Figures for the Paper

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

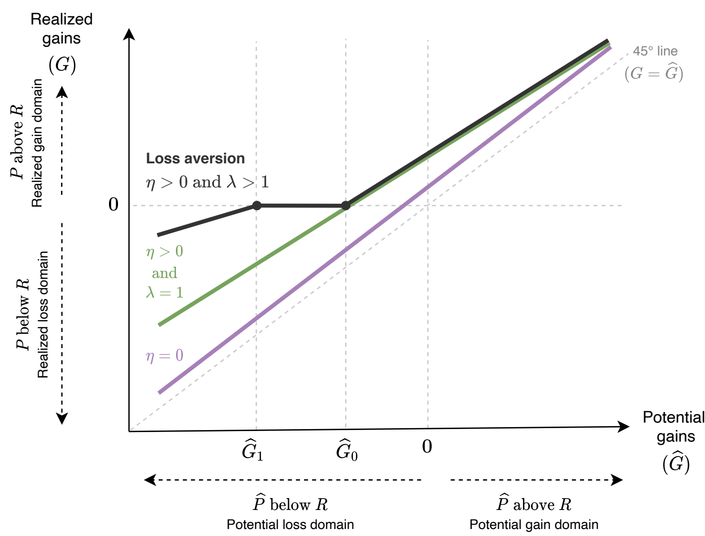


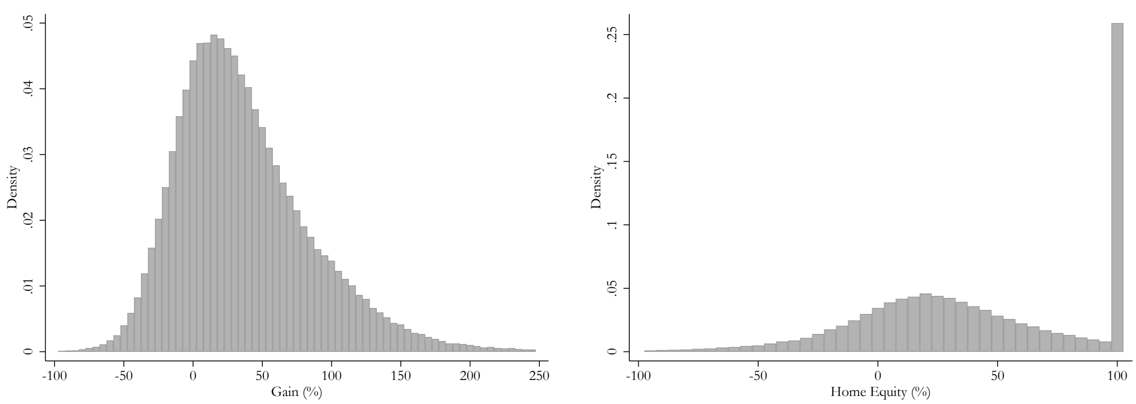
Figure A.2

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 winsorized 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 winsorized at 200 weeks.

Panel A



Panel B

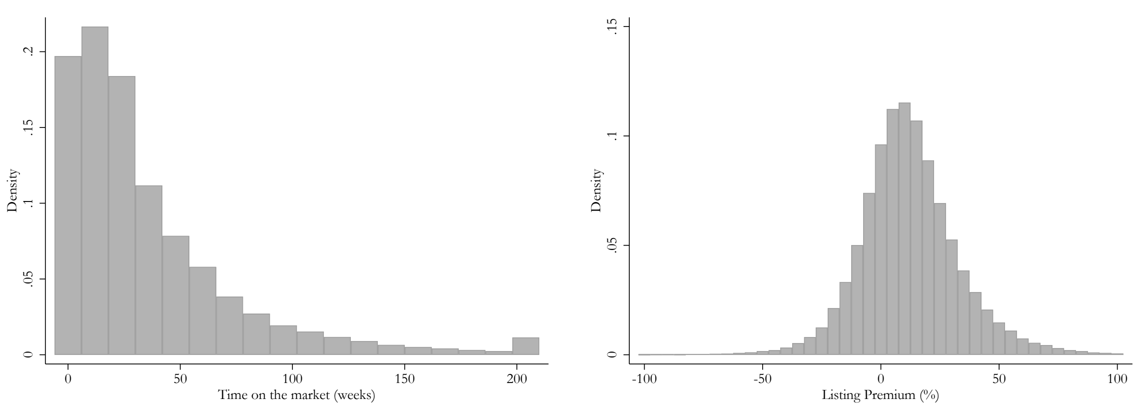


Figure A.3

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.

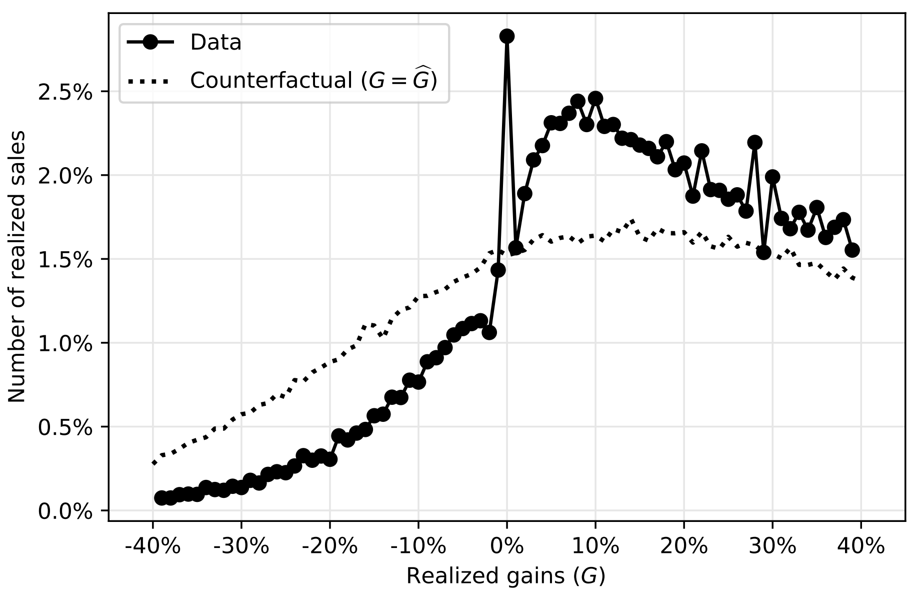


Figure A.4

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.

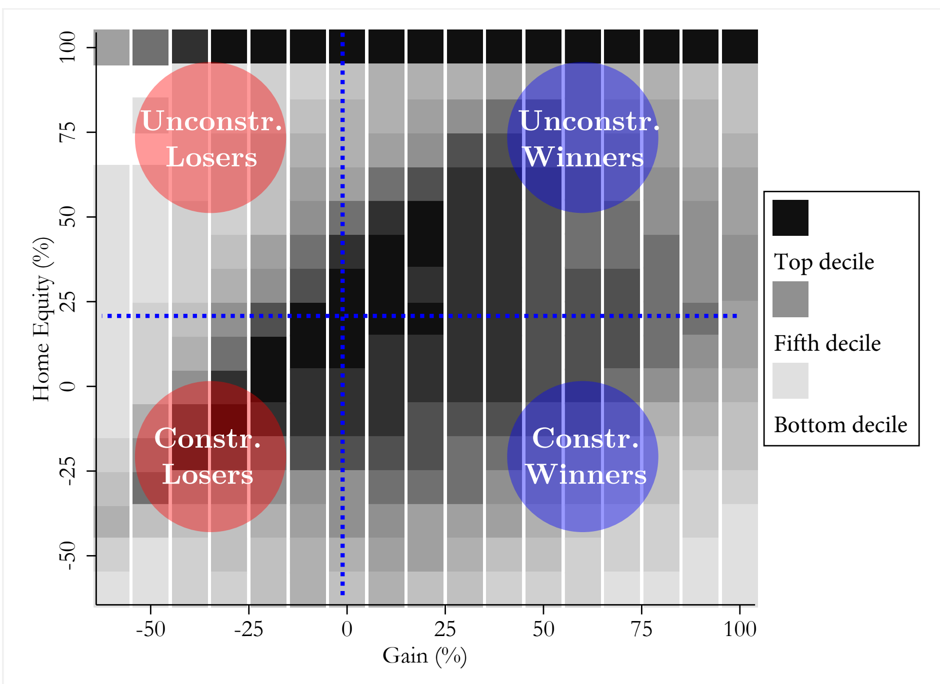


Figure A.5

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).

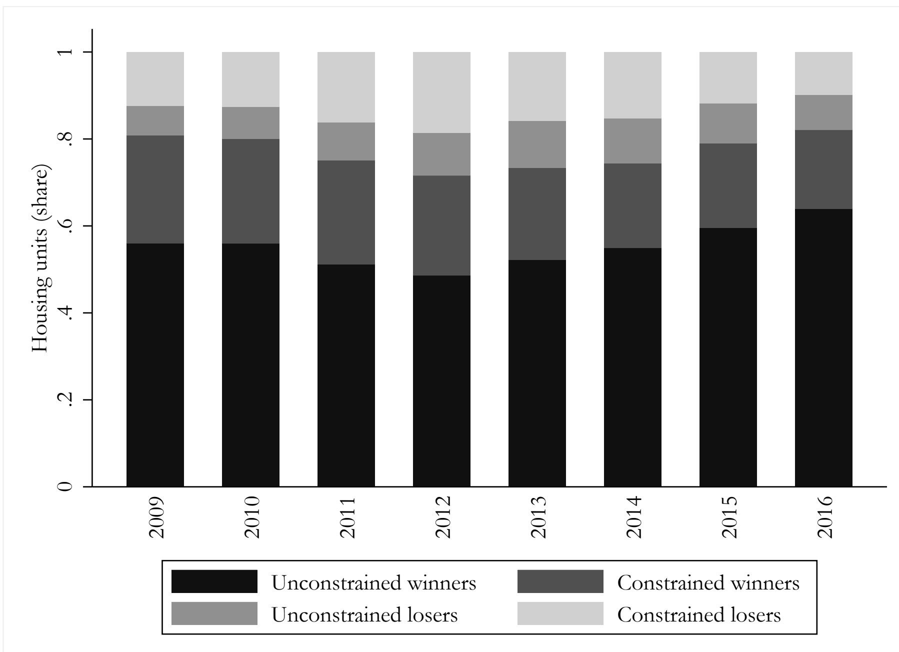


Figure A.6

Realized Gains vs. 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. We report a rolling average of the frequencies, with a bandwidth of 10%. 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.

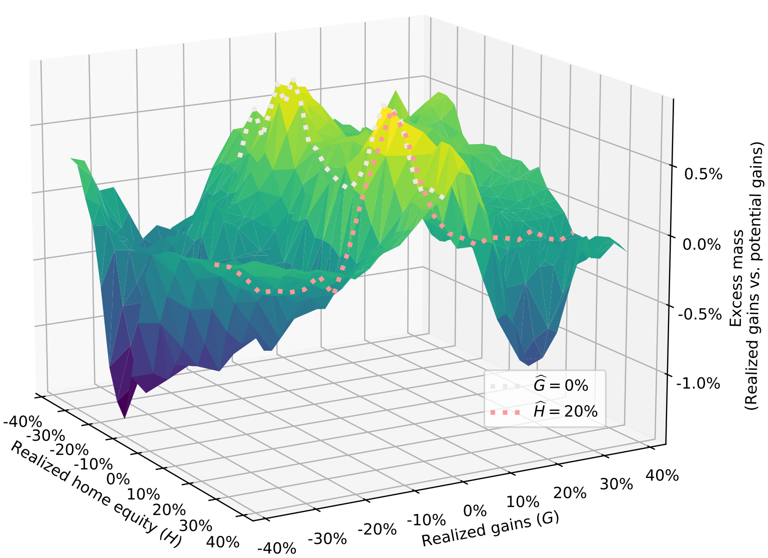


Figure A.7
Bunching around Realized Gains of Zero:
Polynomial Counterfactual

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%].

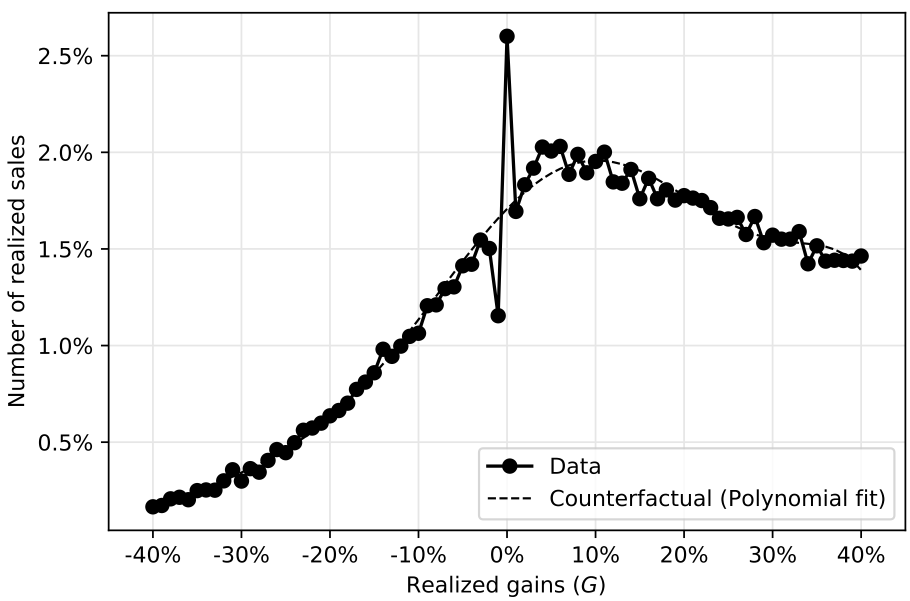


Figure A.8

Price-Volume Correlation

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

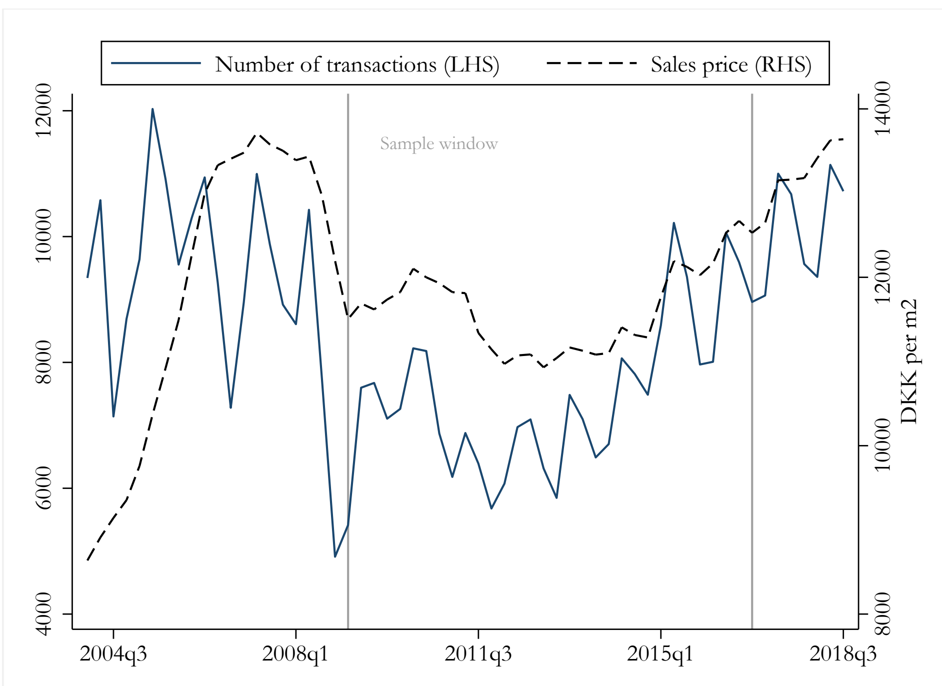


Figure A.9

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.

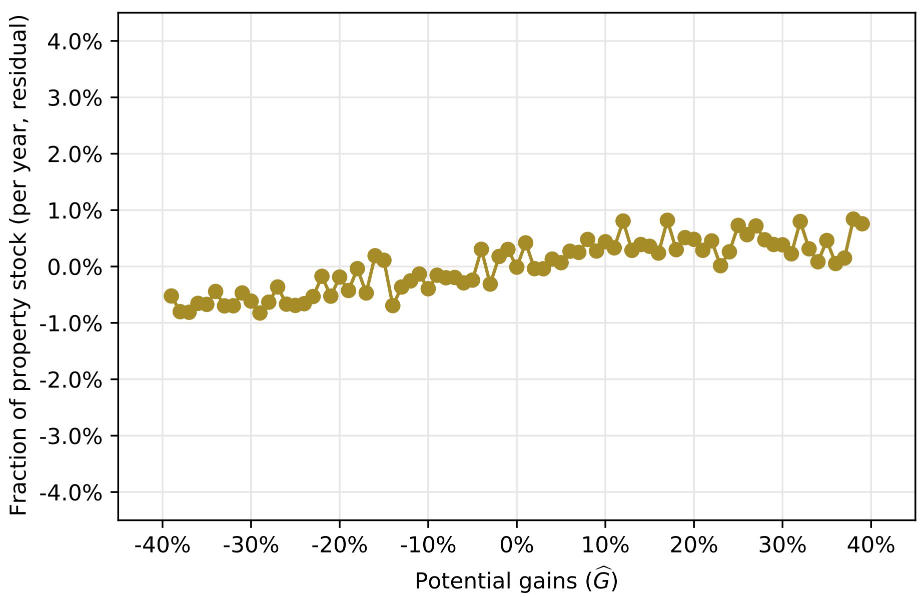


Table A.1

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

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 $\geq$ 80 is the current LTV of the property if the LTV is greater equal to 80 and 0 otherwise. Estimated value is the value of the property implied by the hedonic model, and estimated market index captures 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	0.565***	0.471***	0.519***	0.350***	0.587***	0.494***
	(0.016)	(0.015)	(0.026)	(0.024)	(0.016)	(0.015)
LOSS (squared)			0.001**	0.003***		
			(0.001)	(0.000)		
LTV if $\geq$ 80	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Estimated 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)
Estimated price index	0.991***	0.988***	0.991***	0.987***		
	(0.003)	(0.003)	(0.003)	(0.003)		
Residual from last sales price		-0.096***		-0.098***		-0.093***
		(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	0.410***	0.444***	0.411***	0.448***	76.406***	76.169***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.179)	(0.180)
Year-Quarter FE					✓	✓
Observations	192665	192665	192665	192665	192665	192665
R^2	0.888	0.889	0.888	0.889	0.891	0.892

Table A.3
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. A bankrupt mortgage-credit institution can now adjust administration fees (bidragssats) paid by borrowers, but only if justified by market terms and if at the same time further resources for administration of the bankrupcy estate is required. Changes must be announced in writing at least three month in advance of implementation.
June 2010	Amends how a mortgage-credit institution that has filed for bankruptcy (or is under suspension of payments) can fund payments to mortgage bond owners. Allows a mortgage-credit institution that has filed for bankruptcy (or is under suspension of payments) to, under specific circumstances, transfer series of bonds to other financial institutions. Introduces the option for the FSA to provide dispensation from certain requirements when a bankruptcy estate is converting covered mortgage bonds into uncovered.
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.

Table A.4

Goodness of fit

The table reports root mean squared differences between the level of each moment in the model and the data, relative to the level in the data. This prediction error is interpretable in percent terms, as reported below for each moment separately, as well as jointly for the full set of moments.

Hockey stick	Bunching	Home equity	Extensive margin	Full model
4.29%	7.74%	13.06%	2.73%	6.95%

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 utility is assumed to feature reference dependence and no loss aversion, and the demand functions are assumed to be linear: $\alpha(\ell) = \alpha_0 - \alpha_1\ell$ and $\beta(\ell) = \beta_0 + \beta_1\ell$.

In this case, the maximization problem is given by:

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

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

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

which implies the optimal solution:

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

For a model with loss aversion and reference dependence, the maximization problem is given by:

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

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 with a higher degree of overall reference dependence ($\lambda\eta$):

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

The expression of the optimal listing premium, which is piecewise linear, is then given by:

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

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:

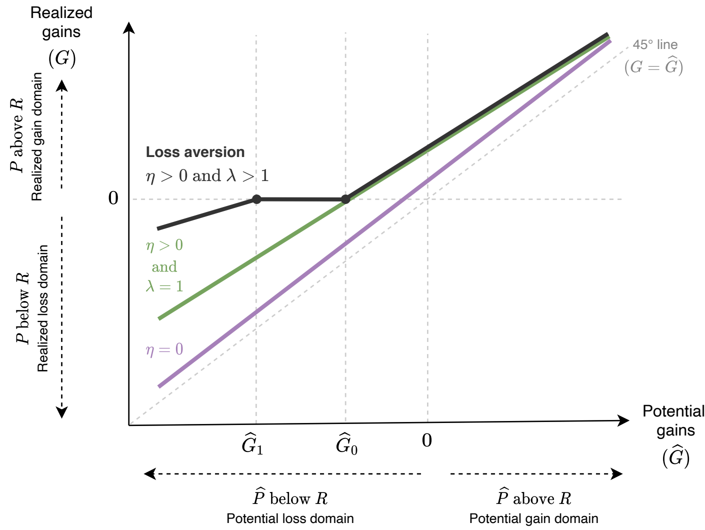
$$G(\widehat{G}) = \begin{cases} \gamma_0 + \gamma_1\widehat{G} & \text{if } \widehat{G} > \widehat{G}_0, \\ 0 & \text{if } \widehat{G} \in [\widehat{G}_1, \widehat{G}_0], \\ \gamma_{\lambda,0} + \gamma_{\lambda,1}\widehat{G} & \text{if } \widehat{G} < \widehat{G}_1. \end{cases} \tag{9}$$

Here, we have:

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

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

with $\gamma_{\lambda,0}$ and $\gamma_{\lambda,1}$ defined analogously to γ_0 and γ_1. The plot below shows the realized gains given the optimal choice of listing premia:



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:

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

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

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 Realization Utility

We assume that households do not receive utility from simply living in a house that has appreciated relative to their reference point R. They exhibit "realization utility", 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. Table L.11 validates these observations in the data.

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), 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 is possible 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

[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: $P = \widehat{P} + \beta_0 + \beta_1(L - \widehat{P}) = \beta_0 + \beta_1 L + (1-\beta_1)\widehat{P}$.

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, which are inconsistent with the data), (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 $\widehat{P}$ corresponds to one of the valuation models that we consider, namely 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 overweighting 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, ζ
Distribution of moving shocks	θ_m, θ_σ
Financial constraints	μ
Listing/search cost	ϕ
Fraction of perfect targeters	π
Calibrated parameters ($\mathbf{x}^c$)	
Down-payment constraint	$\gamma = 20\%$
Exogenous inputs from the data ($\mathbf{x}^d$)	
Density of observations	$f_{stock}(\widehat{G}, \widehat{H}), f_{list}(\widehat{G}, \widehat{H})$
Demand	$\alpha(\ell), \beta(\ell)$
Cross-sectional variation of demand	$\alpha_{k\in\{High,Med,Low\}}(\ell), \beta_{k\in\{High,Med,Low\}}(\ell)$
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	$\ell^*_{k\in\{High,Med,Low\}}(\widehat{G}, \mathbf{x})$

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: $\widehat{\mathbf{x}}^s = \kappa\left(F(\mathbf{x}^s), \mathbf{x}_0^s\right)$. To avoid a situation in which ad hoc initial starting values $\mathbf{x}_0^s$ influence the convergence point, we start with a grid search approach that allows us to solve an exact system of 8 equations in 8 unknowns.

We find the set of parameters:

$$(\eta, \lambda, \theta_m, \theta_\sigma, \mu, \pi, \phi, \zeta)$$

to match the following moments:

1. The average level of the listing premium in the interval $\widehat{G} \in [-40\%, 20\%]$, equal to 9.6% in the data.
2. The average level of the listing premium in the interval $\widehat{G} \in [20\%, 40\%]$, equal to 28% in the data.
3. The slope of the home equity listing premium profile in the interval $\widehat{H} \in [-20\%, 20\%]$, equal to -0.42 in the data.
4. The magnitude of missing mass at $G = -1\%$. In the model, the missing mass below zero has an upper bound equal to π. This is achieved under the assumption that the missing mass at $G = -1\%$ calculated just in the sample of precise targeters is equal to -100%. Our identifying assumption is therefore that the missing mass is equal to -100% for a seller that precisely targets the final price.
5. The magnitude of excess mass at $G = 0$, equal to 69.6% in the data.
6. The magnitude of the spike excess mass relative to the total diffuse mass in the interval $G \in [0, 40\%]$, equal to 22% in the data.
7. Expected utility of a seller with potential gains equal to $\widehat{G}_+ = 40\%$. The identifying assumption here is that this expected utility is equal to zero.
8. The slope of the extensive margin listing decision by potential gains across the domain $G \in [-40\%, 40\%]$, equal to 0.003 in the data.

In Figure L.32, we report a decomposition of the magnitudes of these model-implied moments by the set of parameters.

Finally, the local optimization algorithm takes the form of a gradient search method, which starts from the 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. (In a previous version of the paper, we have also approximated an annealing procedure by first running a Monte Carlo technique, with a set of $N = 50,000$ draws of parameters $\mathbf{x}_{i=1,\ldots,N}^s$ and evaluating the function $F(\mathbf{x}_{i=1,\ldots,N}^s)$ 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.)

To assess the empirical fit quantitatively, we compute the root mean squared difference between the level of each moment in the model and the data, relative to the level in the data. In this way, the prediction error is interpretable in percent terms, as reported in Table A.4 for each moment separately, and for the full model jointly.

B.9 Magnitudes of Financial Constraints

Consider a loan with size M, maturity T, and an effective interest rate i. The monthly payment for this loan is:

$$A(M,i) = M \times \frac{i}{1-(1+i)^{-T}}$$

For an individual with discount rate r, this monthly payment has the following present value:

$$\begin{aligned} NPV(M,i) &= A(M,i) \times \frac{1-(1+r)^{-T}}{r} = M \times \frac{i}{1-(1+i)^{-T}} \times \frac{1-(1+r)^{-T}}{r} \\ &= M \times \frac{1-(1+r)^{-T}}{1-(1+i)^{-T}} \times \frac{i}{r} \end{aligned}$$

To capture the utility penalty in the data, let $P = 1$ be the value of the house. M expressed in units of the house is therefore:

$$M = P \times LTV = LTV$$

We can then have an expression for the additional NPV cost of borrowing, for a general LTV level:

$$\begin{aligned} \kappa^{data}(LTV) &= NPV(LTV, i_1) - NPV(LTV, i_0) \\ &= LTV \times \left(\frac{1-(1+r)^{-T}}{1-(1+i_1)^{-T}} \times \frac{i_1}{r} - \frac{1-(1+r)^{-T}}{1-(1+i_0)^{-T}} \times \frac{i_0}{r} \right) \end{aligned}$$

where i_0 is the not-penalized interest rate on the loan, and i_1 is the penalized one. Figure L.2 shows the average interest rate profile in the data, for different levels of the LTV ratio. Figure L.33 reports the utility penalty calculated in the data based on this interest rate profile, alongside the utility penalty in the model, corresponding to the estimated value of μ in row 8 of Table 2 in the main text:

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 A.8 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), 79.56 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 20% 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 10% 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 ("skuffesalg") 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 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.

We start from 615,040 observations in the raw listings data and once all filters are applied, the sample comprises 214,103 listings of Danish owner-occupied housing in the period between 2009 and 2016, for both sold (70.4%) and retracted (29.6%) properties, matched to mortgages and other household financial and demographic information.[10] These listings correspond to a total of $191,507$ unique households, and $178,933$ 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 $5,538,052$ observations of $807,345$ 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.

C.6 Data Citations

- FinansDanmark (2019a). "Boligmarkedsstatistikken, Ejendomspriser på boligmarkedet, BM010, 2002-2018". https://rkr.statistikbank.dk/201 (accessed in 2019)

[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.5 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 4,663 properties that Statistics Denmark mark as extremely priced, 207 properties with a listing or selling price below 100K DKK, and 629 properties with a listing or selling price above 20M DKK.

[10]The data comprises $172,225$ listings that have a mortgage, and $41,878$ listings with no associated mortgage (i.e., owned entirely by the seller).

- FinansDanmark (2019b). ”Boligmarkedsstatistikken, Bevægelser på boligmarkedet ,BM020, 2002-2018”. https://rkr.statistikbank.dk/201 (accessed in 2019)
- FinansDanmark (2019c). ”Boligmarkedsstatistikken, Salgstider på boligmarkedet,BM030, 2002-2018”. https://rkr.statistikbank.dk/201 (accessed in 2019)
- RealView (2018). Listing data as received by RealView (February 2018)
- Statistics Denmark (2019). “Forbrugerprisindeks (2015=100) efter hovedtal, PRIS112, 1992-2020 [Database]”. https://www.statistikbanken.dk/statbank5a/default.asp?w=1920 (accessed in 2020)
- Statistics Denmark (2021a). ”Befolkningen (BEF, Population Demographics), 1992-2017 [database]”. Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/BEF%20-%20Befolkningen.html (accessed 17/9/2021)
- Statistics Denmark (2021b). ”Boligfinansiering (REAL, Mortgages), 2009-2016 [database]”. Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/REAL%20-%20Boligfinansiering.html (accessed 17/9/2021)
- Statistics Denmark (2021c). ”Boligtællingen (BOL, Housing), 1992-2017 [database]”. Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/BEF%20-%20Befolkningen.html (accessed 17/9/2021)
- Statistics Denmark (2021d). ”Ejendomme salgsoplysninger (EJSA, Property transactions), 1992-2016 [database]”. Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/EJSA%20-%20Ejendomme%20salgsoplysninger.html (accessed 17/9/2021)
- Statistics Denmark (2021e). ”Ejendomsvurderinger (EJVK, Property assessments), 1992-2016 [database]”. Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/EJVK%20-%20Ejendomsvurderinger.html (accessed 17/9/2021)
- Statistics Denmark (2021f). ”Ejere af ejendomme (EJER, Property owners), 1992-2017 [database]”. Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/EJER%20-%20Ejere%20af%20ejendomme.html (accessed 17/9/2021)
- Statistics Denmark (2021g). ”Indkomst (IND, Income from tax returns), 2007-2017 [database]”. Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/IND%20-%20Indkomst.html (accessed 17/9/2021)
- Statistics Denmark (2021h). ”Matrikler (EJMT, Lots), 1992-2016 [database]”. Danmarks Statistiks Forskningsservice. (accessed 17/9/2021)
- Statistics Denmark (2021i). ”Ejendomsvurderinger (EJVK, Property assessments), 1992-2016 [database]”. Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/EJVK%20-%20Ejendomsvurderinger.html (accessed 17/9/2021)
- Statistics Denmark (2021j). ”Uddannelser (UDDA, Education), 2007-2017 [database]”. Danmarks Statistiks Forskningsservice. http://www.dst.dk/extranet/forskningvariabellister/Oversigt%20over%20registre.html (accessed 17/9/2021)

- Statistics Denmark (2021k). "Servicebidragsregister (INDSLUT, Tax deductions), 2011-2017 [database]". Danmarks Statistiks Forskningsservice. https://www.dst.dk/extranet/ForskningVariabellister/INDSLUT%20-%20Servicebidragsregister%20-%20indeholdende%20ligningsm%C3%A6ssige%20fradrag%20mm.%20fra%20Slutligningen.html (accessed 17/9/2021)

D Institutional Background in Denmark

D.1 The Search and Matching Process in the Housing Market

Most Danish homes are sold via a real estate agent and the majority of listings are posted online, although some listings are sold to the real estate agent's network of potential buyers before being posted online ("skuffesalg"). Seller and the real estate agent set an initial listing price. Potential buyers then make offers, which can be both lower or higher than the listing price. By default, bids are subject to two weeks' bank and legal provisos, but they can be waived as part of the negotiation. The seller accepts or rejects the offers. If there are no bids the seller may choose to adjust the price downwards or eventually retract the home from the market.

Sellers employ a real estate agent, who works as the seller representative. The agent advises the seller in setting the listing price and is responsible for marketing, legal work, and getting third-party inspection reports on the house. Third-party inspection reports are mandatory and has to be available before a transaction can take place. Although not as common, buyers can have a buyer agent representing them in the search, negotiation and legal phases.

The average costs of selling a typical home is around 120,000 DKK with about 75,000 DKK paid as broker fees. It also includes inspection reports, marketing, insurance, and official documents.[11] Buying a house cost around 5 to 6 percent of the price. This includes stamp fees (on the deed, on the mortgage, and on a potential bank loan), several different bank fees, legal assistance, cost for construction experts, insurance, moving costs, and renovation costs.[12] For comparison, the typical seller fee in the U.S. ranges from 5 to 6 percent and the typical buyer fee from 2 to 5 percent, also including appraisal, inspection, taxes, insurance and loan-related fees. (Mateen et al, 2021)

D.2 The Taxation Regime for Residential Property

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.3 The Cost of Borrowing

Danish home buyers can take up a mortgage covering up to 80% of the sales price. On top of the interest rate, borrowers are paying fees to the bank, the mortgage bank, and the state. The banks charge administration fees at issuance of the mortgage. They vary, but are in the range

[11]Source: https://www.bolius.dk/saa-meget-koster-det-at-saelge-sit-hus-8664
[12]Source: https://www.bolius.dk/omkostninger-ved-at-koebe-bolig-18145

of 9,000 DKK.[13] The mortgage banks charges administration fees, (bidragssats), which is to be added the interest rate payment each term. The rate increases stepwise with LTV.[14] In addition, mortgage banks charges brokerage for issuing bonds (kurtage) and a spread price (kursskæring) of 0.15% and 0.20% respectively. The Danish state requires stamp fees on the mortgage. They consist of a fixed amount of 1,750 DKK plus 1.45 percent of the loan size, both to be paid out at issuance.

Home buyers are required to provide a down-payment of at least 5%. Bridging the remaining 15%, buyers can take up a bank loan, usually at much higher interest rates than the mortgage. In 2018 interest rates on bank loans varied from 3% to 11% in the 12 largest banks. In addition to higher interest rates, borrowers pay fixed fees to the bank and stamp fees to the state when borrowing from the bank. In 2018 bank fees ranged from 0 to 14,000 DKK. Stamp fees are 1,660 DKK plus 1.5% of the loan value.[15] See Table L.2 for an overview of fees and Figure L.2 for an illustration of average costs over LTV.

D.4 Assumability, Refinancing, and Unsecured Mortgages

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 145bp 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.5 The Foreclosure Process

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

[13]https://www.bolius.dk/omkostninger-ved-at-koebe-bolig-18145
[14]https://www.mybanker.dk/sammenlign/bolig/bidragssatser/
[15]https://www.bolius.dk/boliglaan-i-banken-find-det-billigste-18078

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 section describes the role of the tax-assessted value in the estimation of the baseline hedonic pricing model, and discusses alternative models in more detail.

E.1 Hedonic Model and the Tax-assessed Value

The tax-assessed value 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.

Between 2009 and 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. In 2013, the tax assessments were frozen at 2011 levels in anticipation of a new model of assessment. However, in case of significant value-enhancing adjustments to a house or apartment, a re-assessment took place. 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.

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.

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.3 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 including the third degree polynomial of the tax-assessed value raises the explanatory power to 88 percent.

E.2 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,[16] based on traded properties filtered to match the filtering of the municipality indices provided by Finance Denmark.[17] 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.

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

[17] 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.

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 $ln(P_l) - ln(\hat{P_l})$ 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.3 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.3.1 below. We add $\bar{r}_{it} \equiv \mathbb{1}_{r_{it}} + \mathbb{1}_{r_{it}>0} + r_{it} + \mathbb{1}_{i=f}\mathbb{1}_{r_{it}} + \mathbb{1}_{i=f}\mathbb{1}_{r_{it}>0} + \mathbb{1}_{i=f} \cdot r_{it}$ 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 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.4.

E.3.1 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. 40 percent of claims were at the maximum amount and the average claimed exemption per listing was 14,852 DKK, conditional on claiming a positive amount. To get a sense of magnitudes, 14,852 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[18] (around 6,400 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 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.4 Out-of-Sample Testing

The large number of controls and fixed effects in the hedonic model could give rise to concerns about overfitting. To assess this, we conduct out-of-sample testing of the model. Table L.5 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.[19] 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

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

[19] 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.6. 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.

are preserved when the hedonic price is predicted out-of-sample.

F Functional Form of Down-Payment Constraints

F.1 Alternative Formulation of Penalty Function

In the model, we assume that violating the down-payment constraint leads the seller to incur a monetary penalty for levels of realized home equity below the constraint threshold. Figure 8 in the paper show that the model of down-payment constraints that we use does not perfectly fits the data; in particular, the nonlinearity of the listing premium below potential home equity levels of -20% looks different than the pattern that our model is able to capture. To address this issue, we verify the implications of the model using a concave version of the penalty function:

$$\kappa(P(\ell)) = \begin{cases} \mu(\gamma - H(\ell))^{1/2}, \text{ if } H(\ell) < \gamma \\ 0, \text{ if } H(\ell) \geq \gamma \end{cases}. \tag{23}$$

Figure L.3 shows that the smoothing at the bottom of the home equity profile is better fitted with this formulation, for a set of preference parameters that are very similar to the ones used in the main analysis. However, equation 23 implies a sharp discontinuity and therefore sharp bunching at $\widehat{H} = \gamma$, which is not observed in the data. Addressing these two issues together entails a more heavily parameterised model of home equity constraints, with a significant increase in the computational burden, and without a material impact on the identification of the main structural parameters.

F.2 Downsizing Aversion and Interaction Effects

One possible rationalization for the interaction effects between preferences and constraints in the data 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.12a uses a subsample of the data for which we have information on the households' subsequent down payment ($N = 15,981$). 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 ($N = 36,770$), we show in Figure L.12b 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.

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 36 weeks (24 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.13 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.13, which shows the share of seller *retracted* listings, which also rises with ℓ. Here we also see more "concavity" as $\hat{\ell}$ drops below zero, in that the retraction rate rises the farther $\hat{\ell}$ falls below zero.

In the paper, we simply convert the two plots into a single number, which is the probability of house sale within six months (i.e., $\alpha(\hat{\ell})$) on the y-axis as a function of $\hat{\ell}$ on the x-axis. To smooth the average point estimate at each level of the listing premium, we use a generalized logistic function (Richards, 1959, Zwietering et al., 1990, Mead, 2017) of the form:

$$\alpha(\ell) = A + \frac{K - A}{(C + Qe^{-B\ell})^{1/\nu}}. \tag{24}$$

H Demand Concavity and Housing Stock Homogeneity

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.[20] 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.15 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.7 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,[21] with a baseline level of -0.422. 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 -0.569. With both instruments (i.e., including the distance to the largest cities as well), the overidentified 2SLS specification gives a result of -0.548 without, and -0.428 with controls for average household characteristics in the municipality.

[20] In Figure L.14, we show pictures of typical row houses in Denmark.

[21] 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 Unobserved Quality

Before considering the more general case with loss aversion in section I.4 below, we discuss the role of unobserved quality in a version of our structural model with linear reference dependence. In the context of this model, let L_{ist} denote the listing price chosen by the seller of property i listed for sale in period t, $\widetilde{P}_{it}$ the "true" hedonic value of the property at the time of listing and $\widetilde{\ell}_{ist}$ the "true" listing *premium*, R_{is} the price of the property when initially purchased in period s, and $\widetilde{G}_{ist}$ the "true" potential gain.[22]

This implies the following "true" relationship between listing premia and potential gains:

$$\underbrace{L_{ist} - \widetilde{P}_{it}}_{\widetilde{\ell}_{ist}} = \mu_0 + m(\underbrace{\widetilde{P}_{it} - R_{is}}_{\widetilde{G}_{ist}}) + \varepsilon_{ist}. \tag{25}$$

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:

$$\widetilde{P}_{it} = \widehat{P}_{it} + \xi_{it}. \tag{26}$$

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

$$\begin{aligned} \widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\ &= \widetilde{G}_{ist} - \xi_{it}, \end{aligned} \tag{27}$$

$$\begin{aligned} \widehat{\ell}_{ist} &= L_{ist} - \widehat{P}_{it} \\ &= \widetilde{\ell}_{ist} + \xi_{it}. \end{aligned} \tag{28}$$

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

[22]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:

$$\begin{aligned}\widehat{m} &= \frac{Cov(\widehat{G}_{ist}, \widehat{\ell}_{ist})}{Var(\widehat{G}_{ist})} = \frac{Cov(\widetilde{G}_{it} - \xi_{it}, \widetilde{\ell}_{ist} + \xi_{it})}{Var(\widetilde{G}_{ist} - \xi_{it})} = \frac{Cov(\widetilde{G}_{it} - \xi_{ist}, \mu_0 + m\widetilde{G}_{it} + \varepsilon_{it} + \xi_{it})}{Var(\widetilde{G}_{ist} - \xi_{it})} \\ &= m \underbrace{\frac{Var(\widetilde{G}_{ist})}{Var(\widetilde{G}_{ist}) + Var(\xi_{it})}}_{\substack{\text{Classical measurement} \\ \text{error}}} - \underbrace{\frac{Var(\xi_{it})}{Var(\widetilde{G}_{ist}) + Var(\xi_{it})}}_{\substack{\text{Over-estimation bias} \\ \text{(because } m<0)}} \quad . \end{aligned} \tag{29}$$

Equation (29) 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[23] (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:

$$P_{it} = \underbrace{X_i\beta + \delta_t + \nu_{it}}_{\substack{\text{"True" hedonic value} \\ \text{of property } (=\widetilde{P}_{it})}} + \underbrace{\omega_{it}}_{\substack{\text{Idiosyncratic over-} \\ \text{or under-payment}}} , \tag{30}$$

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:

$$\widetilde{P}_{it} = X_i\beta + \delta_t + \nu_{it}. \tag{31}$$

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:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \nu_{it} = 0, \text{ and } \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \omega_{it} = 0. \tag{32}$$

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:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \omega_{it} = 0. \tag{33}$$

[23] 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 Alternative Estimation Methods

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

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

   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

   $$\widehat{P}_{it} = X_{it}\beta + \delta_t. \tag{35}$$

   Equation 34 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:

   $$\widehat{P}_{it} = X_i\beta + \delta_t + \bar{\nu}_{it_{\tau<t}} + \bar{\omega}_{it_{\tau<t}}, \tag{36}$$

   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

   $$\widehat{P}_{it} = X_i\beta + \delta_t + \nu_{is} + \omega_{is}, \tag{37}$$

   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):

   $$\widehat{P}_{it}^{level} = R_{is}^{level} \cdot \frac{d_t}{d_s}. \tag{38}$$

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

$$\begin{aligned}\widehat{P}_{it} &= R_{is} + \delta_t - \delta_s. \\ &= X_i\beta + \delta_s + \nu_{is} + \omega_{is} + \delta_t - \delta_s \\ &= X_i\beta + \delta_t + \nu_{is} + \omega_{is},\end{aligned} \tag{39}$$

which is equivalent to equation (37).

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

$$P_{it} = X_{it}\beta + \delta_t + \nu_{it} + \omega_{it}. \tag{40}$$

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 $T = 2$:

$$\nu_{is} + \omega_{is} = \underbrace{R_{is} - \widehat{P}_{is}}_{\text{Hedonic model pricing residual at time } s} \tag{41}$$

And more generally:

$$\bar{\nu}_{it_{\tau<T}} + \bar{\omega}_{it_{\tau<T}} = \underbrace{\frac{1}{T-1}\sum_{\tau<T} R_{i\tau} - \widehat{P}_{i\tau}}_{\substack{\text{Average of the hedonic model pricing} \\ \text{residuals across repeat sales for which } \tau < T}} \tag{42}$$

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

$$\widehat{P}_{it} = X_{it}\beta + \delta_t + \frac{\nu_{is} + \nu_{is'}}{2} + \frac{\omega_{is} + \omega_{is'}}{2}, \tag{43}$$

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

$$\begin{aligned}\widehat{P}_{it} = & \underbrace{X_{it}\beta + \delta_t}_{\text{Hedonic model with time-varying observables}} + \frac{1}{T-1}\sum_{\tau<T} R_{i\tau} - \widehat{P}_{i\tau} \\ &= X_{it}\beta + \delta_t + \frac{R_{is} - \widehat{P}_{is} + R'_{is'} - \widehat{P}_{is'}}{2} \\ &= X_{it}\beta + \delta_t + \frac{X_{is}\beta + \delta_s + \nu_{is} + \omega_{is} - (X_{is}\beta + \delta_s) + X_{is'}\beta + \delta_{s'} + \nu_{is'} + \omega_{is'} - (X_{is'}\beta + \delta_{s'})}{2} \\ &= X_{it}\beta + \delta_t + \frac{\nu_{is} + \nu_{is'}}{2} + \frac{\omega_{is} + \omega_{is'}}{2}.\end{aligned} \tag{44}$$

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.3),

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

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:

$$\widehat{P}_{it} = X_i\beta + \delta_t, \tag{46}$$

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

$$\begin{aligned}
\widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\
&= X_i\beta + \delta_t - (X_i\beta + \delta_s + \nu_{is} + \omega_{is}) \\
&= \delta_t - \delta_s - \nu_{is} - \omega_{is} \\
&= \widetilde{G}_{ist} - \nu_{it}.
\end{aligned} \tag{47}$$

When using the observable potential gain $\widehat{G}_{ist}$ to estimate equation (25), 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:

$$\begin{aligned}
\widehat{m} &= \frac{Cov(\widehat{G}_{ist}, \widehat{\ell})}{Var(\widehat{G}_{ist})} \\
&= \frac{Cov(\tilde{G}_{ist} - \nu_{it}, \nu_{it} + \tilde{\ell}_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\
&= \frac{Cov(\tilde{G}_{ist} - \nu_{it}, \nu_{it} + \mu_0 + m\widetilde{G}_{ist} + \epsilon_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\
&= m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} - \underbrace{\frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\substack{\text{Over-estimation bias} \\ \text{(because } m < 0\text{)}}}
\end{aligned} \tag{48}$$

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

$$\widehat{P}_{it} = X_i\beta + \delta_t + \nu_{is} + \omega_{is}. \tag{49}$$

Recall, the true gain is:

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

The potential gain is:

$$\begin{aligned}\widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\ &= \delta_t - \delta_s \\ &= \tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}\end{aligned} \tag{51}$$

Hence the coefficient estimate is:

$$\begin{aligned}\widehat{m} &= \frac{Cov(\widehat{G}_{ist}, \widehat{\ell})}{Var(\widehat{G}_{ist})} \\ &= \frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \nu_{it} + \tilde{\ell}_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\ &= \frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \nu_{it} + \mu_0 + m\tilde{G}_{ist} + \epsilon_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\ &= m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} - \underbrace{\frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\nu_{is}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0 \text{, bias}}\end{aligned} \tag{52}$$

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

$$\widehat{m} = m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \bar{\nu}_{it_{\tau<T}} + \bar{\omega}_{it_{\tau<T}}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} - \underbrace{\frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\bar{\nu}_{it_{\tau<T}}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0 \text{, bias}} \tag{53}$$

   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 (53) 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:

$$\widehat{P}_{it} = X_i\beta + \delta_t + \overline{r}_{it}, \tag{54}$$

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

$$\widehat{G}_{ist} = \widetilde{G} - (\nu_{it} - \overline{r}_{it}), \tag{55}$$

and hence the estimated coefficient is

$$\widehat{m} = m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \bar{r}_{it}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} \underbrace{- \frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\bar{r}_{it}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0 \text{, bias}} \tag{56}$$

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.[24] Table L.4 provides an overview of all the models that we implement. Figures L.16, 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.8 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

[24] 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. $T \geq 2$).

model with renovation expenses and the lagged pricing residual for one repeat sale (IIIa), and the equivalent for any number of repeat sales and average lagged residuals (IIIb). Table L.9 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.8, 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. Figure L.4 documents average prediction errors. The hockey stick of listing premia over negative potential gains is between -0.45 to -0.54 across main models, and between between -0.87 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.9, 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 -0.99, 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.2 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:

$$L_{ist} - \widetilde{P}_{it} = \mu_0 + f(\eta, \lambda, \underbrace{(\widetilde{P}_{it} - R_{is})}_{(\widetilde{G})}) + \varepsilon_{it}, \tag{57}$$

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:

$$\underbrace{L_{ist} - \widetilde{P}}_{\widetilde{\ell}_{ist}} = \mu_0 + m_0 \underbrace{(\widetilde{P}_{it} - R_{is})^-}_{\widetilde{G}^-_{ist}} + m_1 \underbrace{(\widetilde{P}_{it} - R_{is})^+}_{\widetilde{G}^+_{ist}} + \varepsilon_{it}, \tag{58}$$

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:

$$\widehat{m}_0 = \frac{Cov(\widehat{G}^-_{ist}, \widehat{\ell}_{ist})}{Var(\widehat{G}^-_{ist})} = \frac{Cov((\widetilde{G}_{it} - \xi_{it})^-, \widetilde{\ell}_{ist} + \xi_{it})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)} \tag{59}$$

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

$$= m_0 \underbrace{\frac{Cov((\widetilde{G}_{ist} - \xi_{ist})^-, \widetilde{G}^-_{it})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)}}_{\substack{\text{Measurement} \\ \text{error/Smoothing}}} - \underbrace{\frac{Cov((\widetilde{G}_{ist} - \xi_{ist})^-, \xi_{ist})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)}}_{\substack{\text{Over-estimation bias} \\ \text{(because } m<0)}} \tag{60}$$

Equation (60) 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.[25]

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.19).

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

$$\tau = \lim_{v \to \overline{v}_+} \frac{dE[\ell_{it}|V_{it} = v]}{dv}\bigg|_{V_{it}=v} - \lim_{v \to \overline{v}_-} \frac{dE[\ell_{it}|V_{it} = v]}{dv}\bigg|_{V_{it}=v}, \tag{61}$$

[25] 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.

based on the following RKD specification (Landais 2015):

$$E[\ell_{it}|V_{it}=v] = \kappa_m + \kappa_t + \boldsymbol{\xi}\mathbf{X_{it}} + \left[\sum_{p=1}^{\overline{p}} \gamma_p(\nu-\overline{\nu})^p + \nu_p(v-\overline{v})^p \mathbb{1}_{V\geq\overline{v}}\right]. \tag{62}$$

$$\text{where } |v-\overline{v}|<b. \tag{63}$$

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.10 reports results across different bandwidths within which we fit a local linear function on each side of the threshold. Figure L.21 provides further robustness checks on using local quadratic estimation and bandwidth choice.[26]

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

$$\alpha_{ist}(\widetilde{\ell}) = \mu_1 + n_0(\underbrace{L_{ist}-\widetilde{P}_{it}}_{\widetilde{\ell}^-_{ist}})^- + n_1(\underbrace{L_{ist}-\widetilde{P}_{it}}_{\widetilde{\ell}^+_{ist}})^+ + \epsilon_{ist}, \tag{64}$$

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

$$\widehat{\ell}_{ist} = L_{ist} - (\widehat{P}_{it} + \xi_{it}) = \widetilde{\ell}_{ist} + \xi_{it}, \tag{65}$$

the feasible regression is

$$\alpha_{ist}(\widehat{\ell}) = \mu_1 + n_0(\widetilde{\ell}_{ist} + \xi_{it})^- + n_1(\widetilde{\ell}_{ist} + \xi_{it})^+ + \epsilon_{ist}, \tag{66}$$

[26] The precision but not the size of the estimate for unconstrained households depends on the use of a local linear compared to a local quadratic function. Hahn et al. (2001) show that the degree of the polynomial is critical in determining the statistical significance of the estimated effects. In particular, the second-order polynomial needed to identify derivative effects leads to an asymptotic variance of the estimate that is larger by a factor of 10 relative to the first-order polynomial. We verify that the qualitative patterns that we detect are broadly unaffected by the use of either polynomial order, but that the standard errors, consistent with Hahn et al. (2001), are substantially higher for the second-order polynomial, reported in Figure L.22.

with estimated main coefficient of interest

$$
\begin{aligned}
\widehat{n}_1 &= \frac{\alpha_{ist}(\widehat{\ell}), (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+} = \frac{Cov(\mu_1 + n_0\widetilde{\ell}^-_{ist} + n_1\widetilde{\ell}^+_{ist} + \epsilon_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+} \\
&= n_1 \underbrace{\frac{Cov(\widetilde{\ell}^+_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+}}_{\text{Measurement error}} + \underbrace{\frac{Cov(\epsilon_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+}}_{0 \text{ if } \xi_{it} \perp \epsilon_{it}}
\end{aligned}
\tag{67}
$$

Equation 67 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.

I.7 Simulation Approach to Bounding Unobserved Quality Effects

To show model robustness to unobserved quality that is correlated with the list price, we first assume that a portion of the listing premium can be attributed to unobserved quality, i.e., the "true" listing premium $\widetilde{\ell}_{ist}$ is equal to:

$$\widetilde{\ell} = \ell - \zeta, \tag{68}$$

with $\zeta \sim \mathcal{N}(0, \sigma_\zeta^2)$. Second, under the assumption that the same error affects the estimation of $\widehat{P}$, we de-bias $\widehat{P}$, and all variables affected by it, by the same amount:

$$\widetilde{P} = \widehat{P} - \zeta, \widetilde{G} = \widehat{G} - \zeta, \widetilde{H} = \widehat{H} - \zeta. \tag{69}$$

Third, we re-construct the demand function, the distribution of final price realizations and the probability of listing based on the de-biased value of $\widehat{P}$. Finally, we re-estimate values of structural parameters using the set of adjusted empirical moments.

The reason why we opt for a simulation approach here, as opposed to a rescaling of the observed level of the listing premium, is that we want to avoid the assumption that the listing premia for all properties are subject to exactly the same fixed adjustment factor. Instead, it seems more likely that the prediction error (i.e., the degree to which the seller has an informational advantage relative to the econometrician) is idiosyncratically distributed across the set of properties.

One decision that we needed to take is whether the prediction error is symmetric around zero, and we don't see a strong reason to believe otherwise. We also thought about whether we should simulate a correlation between the prediction error and the level of potential gains $\widehat{G}$, for example, because we underpredict prices of low-priced properties and overpredict high-price properties, which would lead to an artificial negative slope of the "hockey stick". Again, we don't see a justification for this, because in the data the effect goes in the opposite direction, i.e., we slightly overpredict prices of lower-value properties and underpredict those at the top (see online appendix section E.1 and Figure L.4.)

Finally, what is a reasonable value of σ_ζ for the calibration of the magnitude of unobserved quality? To answer this question, we thought harder about the nature of the prediction error at work here. By the logic described earlier, if the seller has a true informational advantage over the econometrician, this should be evident in the final transaction price. By estimating the marginal predictive power of listing prices for final prices beyond the hedonic model, we can recover one possible upper bound for this informational advantage.

More formally, let $\varepsilon = P - \widehat{P}$ be the estimated residual from the hedonic model and ε_L the estimated residual from a hedonic model augmented with the listing price as an additional explanatory variable. A natural upper bound for the variance σ_ζ^2 is then given by the variance $\sigma_{\Delta\varepsilon}^2$ of $\varepsilon - \varepsilon_L$. Now, $\sigma_{\Delta\varepsilon}^2$ can stem from two sources: the first is unobserved quality, as discussed, and the second is the equilibrium link between listing prices and realised outcomes stemming from negotiations and bargaining between buyers and sellers. Since there is a strong relationship between observed listing premia and the probability of sale, we infer that this second component has a material effect. This seems reasonable, since assuming that all such variation comes from unobserved quality is tantamount to shutting down the possibility of any "fishing" and subsequent negotiations. Put differently, if all is unobserved quality, de-biasing generates flat $\beta(\ell)$, sellers have no incentive to vary listing premia at all, and this totally shuts down the intensive margin decision in reality. We therefore consider three possible cases in our simulation, allowing unobserved quality to account for 10%, 25% and 50% of the variation of $\sigma^2_{\Delta\varepsilon}$.

The plots in Figure L.34 give an overview of the effects of the de-biasing procedure, adjusting the empirical moments as indicated in equations (68) and (69) above, with ζ drawn from a normal distribution with mean 0 and variance σ^2_ζ. We distinguish between three cases $\sigma_\zeta \in \{0.044, 0.075, 0.105\}$, labeled as 'Conservative', 'Moderate' and 'Aggressive'. In each plot, we report the average value of each binned moment, across the set of simulated samples.

The structural parameters implied by the three alternative magnitudes of unobserved quality variation are reported in Table 3 in the paper.

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.23. 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.24), respectively. We further show that bunching is present across all quintiles of the previous sales price (Figure L.25) and when splitting into quintiles by holding period (Figure L.26), except for the sub-sample with holding periods of greater than 12 years (top quintile).

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.

Overall, the degree of bunching seems to be declining with the holding period. However, a primary reason for this observation is the fact that property prices have trended upwards over the sample period, and longer holding periods are associated with valuations that are increasingly farther away from the reference point. This decreases the mass of all realized prices at or close to the reference point, even in the counterfactual distribution.

When we take this into account, we find that the magnitude of the relative mass (i.e., the excess mass relative to the counterfactual, which is the basis for the identification of reference dependence and loss aversion) remains remarkably stable, even for holding periods up to 12 years.

In the left-hand plot of Figure L.27, we show the robustness of excess mass for different holding periods, measured exactly at the reference point (i.e., the "spike", which is equal to 69% in the full sample). The error bands indicate 95% confidence intervals, calculated across bootstrap samples. Analogously, in the right-hand plot, we show the robustness of the missing mass immediately below the reference point (i.e., the "notch", equal to -24% in the full sample).

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 88,000 DKK and the mean in the sample is 327,000 DKK. When we divide *gross* financial assets by mortgage size, we find that households, at the median, could relax their constraints by around 6.22 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 -86,000 DKK and the mean is -99,000 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 7.9% 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 7.9%, rather than be able to use liquid financial wealth to augment their down payments.

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. In Figure L.30 and Table L.10 we therefore control for the amount of net financial assets, age, and education, to ensure that we accurately measure the impact of these constraints on household decisions. Figure L.31 shows the listing premium "hockey stick" in the sample of sellers with no mortgage outstanding. This part of the analysis addresses measurement concerns that have affected prior work in this area.

Appendix 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, Allan Davids, Benjamin B. Lockwood, and Tarun Ramadorai. 2022. "Diffuse Bunching with Optimisation Frictions."

Anagol, Santosh, Vimal Balasubramaniam, and Tarun Ramadorai. 2018. "Endowment effects in the field: Evidence from India's IPO lotteries." 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).

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

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

Anenberg, Elliot, and Patrick Bayer. 2020. "Endogenous Sources Of Volatility In Housing Markets: The Joint Buyer–Seller Problem." International Economic Review, 61(3): 1195–1228.

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

Backus, Matthew, Thomas Blake, Brad Larsen, and Steven Tadelis. 2020. "Sequential bargaining in the field: Evidence from millions of online bargaining interactions." Quarterly Journal of Economics, 135(3): 1319–1361.

Badarinza, Cristian, John Y Campbell, and Tarun Ramadorai. 2016. "International Comparative Household Finance." Annual Review of Economics, 8(1).

Barberis, Nicholas, and Wei Xiong. 2012. "Realization utility." Journal of Financial Economics, 104(2): 251–271.

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." Quarterly Journal of Economics, 112(2): 407–441.

Campbell, John Y. 2006. "Household finance." 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. 2015a. "The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in Missouri, 2003-2013." American Economic Review, 105(5): 126–30.

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

Card, David, David S Lee, Zhuan Pei, and Andrea Weber. 2017. "Regression kink design: Theory and practice." In Regression discontinuity designs: Theory and applications. 341–382. Emerald Publishing Limited.

Case, Karl E, and Robert J Shiller. 1987. "Prices of single family homes since 1970: New indexes for four cities." National Bureau of Economic Research Working Paper.

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

Chetty, Raj. 2012. "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." Econometrica, 80(3): 969–1018.

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." 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." 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.

El-Gamal, Mahmoud A, and David M Grether. 1995. "Are people Bayesian? Uncovering behavioral strategies." Journal of the American statistical Association, 90(432): 1137–1145.

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." 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.

Gelber, Alexander M, Damon Jones, and Daniel W Sacks. 2020. "Estimating adjustment frictions using nonlinear budget sets: Method and evidence from the earnings test." American Economic Journal: Applied Economics, 12(1): 1–31.

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

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

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

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai. 2021. "Household finance." Journal of Economic Literature, 59(3): 919–1000.

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.

Han, Lu, and William C Strange. 2014. "Bidding wars for houses." Real Estate Economics, 42(1): 1–32.

Han, Lu, and William C Strange. 2016. "What is the role of the asking price for a house?" Journal of Urban Economics, 93: 115–130.

Harrison, Glenn W, and E Elisabet Rutström. 2009. "Expected utility theory and prospect theory: One wedding and a decent funeral." Experimental economics, 12(2): 133–158.

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.

Khaw, Mel Win, Ziang Li, and Michael Woodford. 2021. "Cognitive imprecision and small-stakes risk aversion." The review of economic studies, 88(4): 1979–2013.

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

Kleven, Henrik J, and Mazhar Waseem. 2013. "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan." Quarterly Journal of Economics, 128(2): 669–723.

Kopczuk, Wojciech. 2005. "Tax bases, tax rates and the elasticity of reported income." Journal of Public Economics, 89(11-12): 2093–2119.

Kőszegi, Botond, and Matthew Rabin. 2006. "A model of reference-dependent preferences." 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.

Levitt, Steven D, and John A List. 2009. "Field experiments in economics: The past, the present, and the future." European Economic Review, 53(1): 1–18.

List, John A. 2003. "Does market experience eliminate market anomalies?" 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.

Mateen, Haaris, Franklin Qian, and Ye Zhang. 2021. "The Microstructure of the US Housing Market: Evidence from Millions of Bargaining Interactions."

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.

Newey, Whitney K, and Daniel McFadden. 1994. "Handbook of Econometrics." Elsevier, 4.

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

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.

Pearson, Karl. 1894. "Contributions to the mathematical theory of evolution." Philosophical Transactions of the Royal Society of London. A, 185: 71–110.

Peel, David, and Geoffrey J McLachlan. 2000. "Robust mixture modelling using the t distribution." Statistics and computing, 10(4): 339–348.

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

Pope, Devin G, and Justin R Sydnor. 2015. "Behavioral economics: Economics as a psychological discipline." The Wiley Blackwell Handbook of Judgment and Decision Making, 2: 800–827.

Pope, Devin G, and Maurice E Schweitzer. 2011. "Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes." American Economic Review, 101(1): 129–57.

Rees-Jones, Alex. 2018. "Quantifying loss-averse tax manipulation." 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.

Saez, Emmanuel. 1999. "Do taxpayers bunch at kink points? (NBER Working Paper)."

Saez, Emmanuel. 2010. "Do taxpayers bunch at kink points?" American Economic Journal: Economic Policy, 2(3): 180–212.

Saez, Emmanuel, Joel Slemrod, and Seth H Giertz. 2012. "The elasticity of taxable income with respect to marginal tax rates: A critical review." Journal of economic literature, 50(1): 3–50.

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

Stein, Jeremy 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.

Stock, James H, and Motohiro Yogo. 2002. "Testing for weak instruments in linear IV regression."

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

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

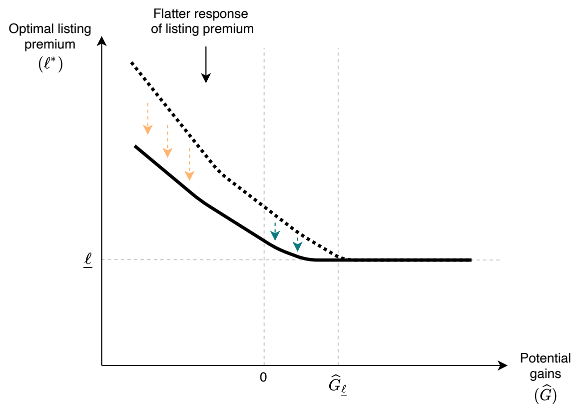
Zhou, Tingyu, John M Clapp, and Ran Lu-Andrews. 2020. "Is the Behavior of Sellers with Expected Gains and Losses Relevant to Cycles in House Prices?" Unpublished working paper.

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.

L Additional Appendix Material

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.



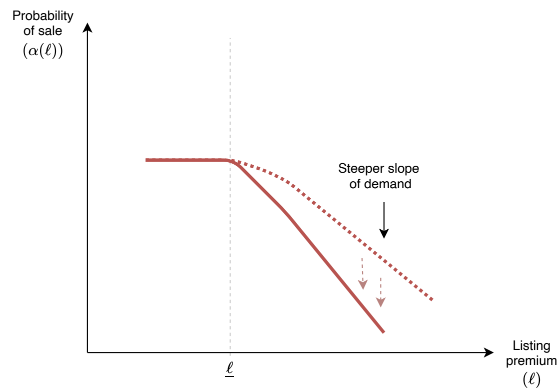


Figure L.2

Cost of borrowing

This figure shows the average cost of borrowing as percentage of loan size for each level of loan-to-value for a house price of 3 million DKK. The range from 0 % to 80% LTV is covered by a 2% mortgage. The solid line shows the cost when bridging the 80% to 95% with a 6% bank loan and the dashed line shows the hypothetical case where the full 95% is funded by a mortgage. Costs include interest rates, fees and other payments to the bank, mortgage bank, and the state.

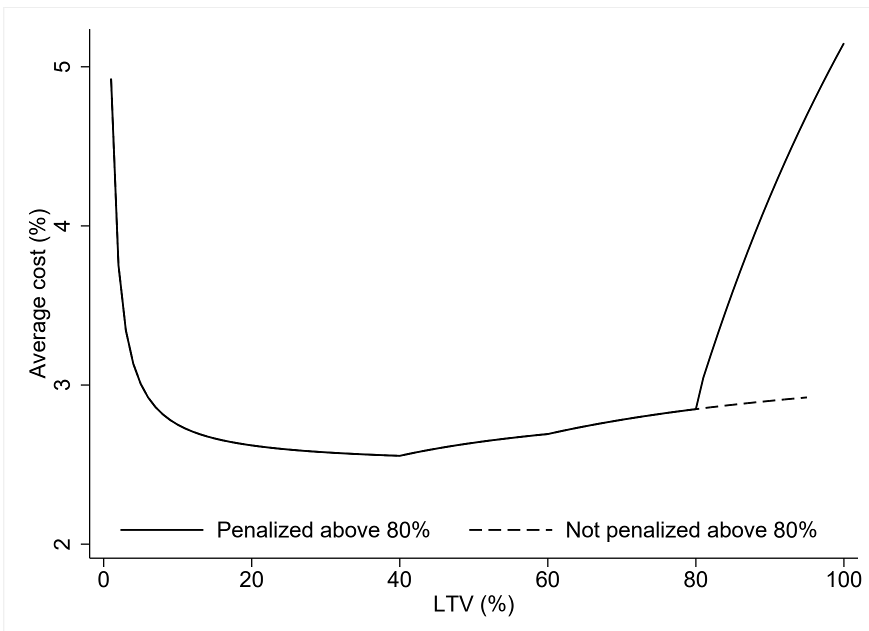


Figure L.3

Listing premia by home equity: Alternative parameterization

This figure shows the fitted listing premium profile for a version of the model with a concave penalty function for violating the down-payment constraint. The model is evaluated at the same set of parameters as in row 8 of Table 2, with $\theta_{mu} = 1.475$ and $\mu = 1.125$.

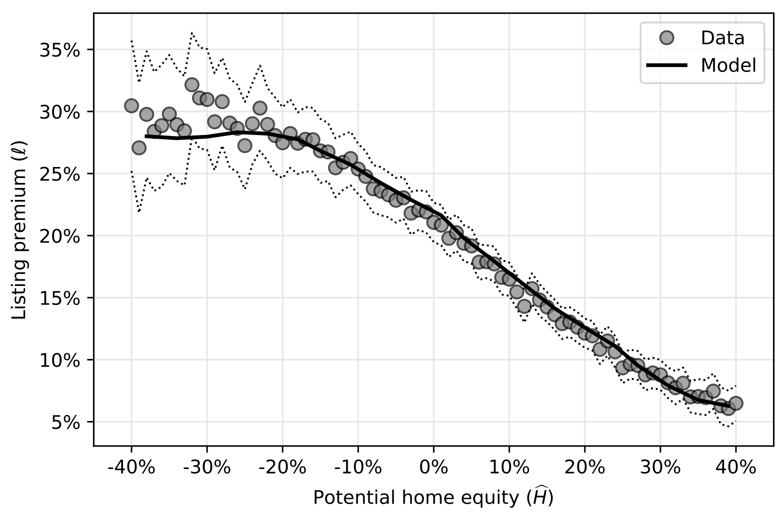


Figure L.4

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:
$ln(P_{it}) = \xi_{tm} + \beta_{ft}\mathbb{1}_{i=f} + \boldsymbol{\beta_x}\mathbf{X_{it}} + \boldsymbol{\beta_{fx}}\mathbb{1}_{i=f}\mathbf{X_{it}} + \Phi(v_{it}) + \mathbb{1}_{i=f}\Phi(v_{it}) + \varepsilon_{it}$, 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). $\mathbb{1}_{i=f}$ 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.

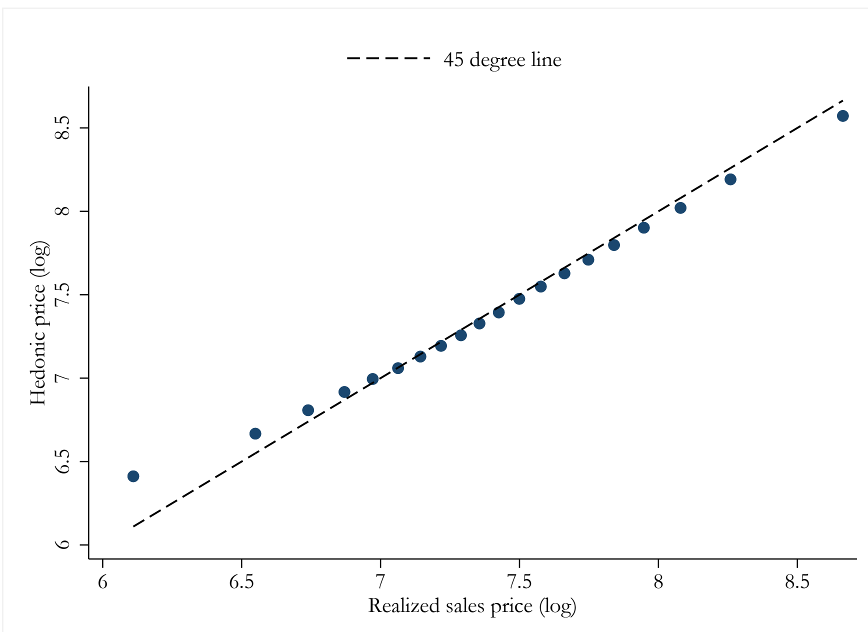
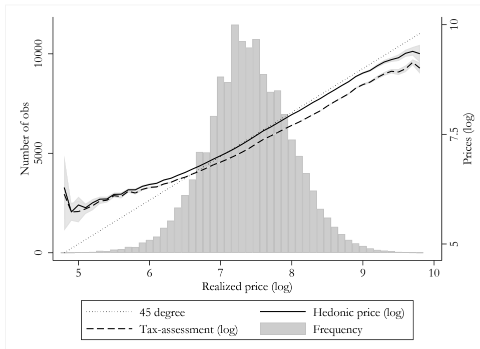


Figure L.5

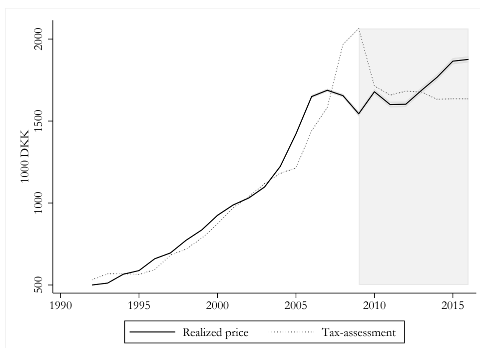
Accuracy of Tax-Assessed Value

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

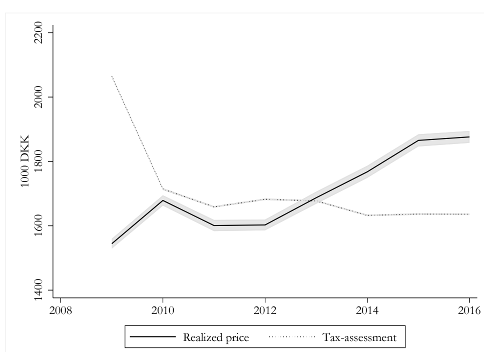


Figure L.6

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

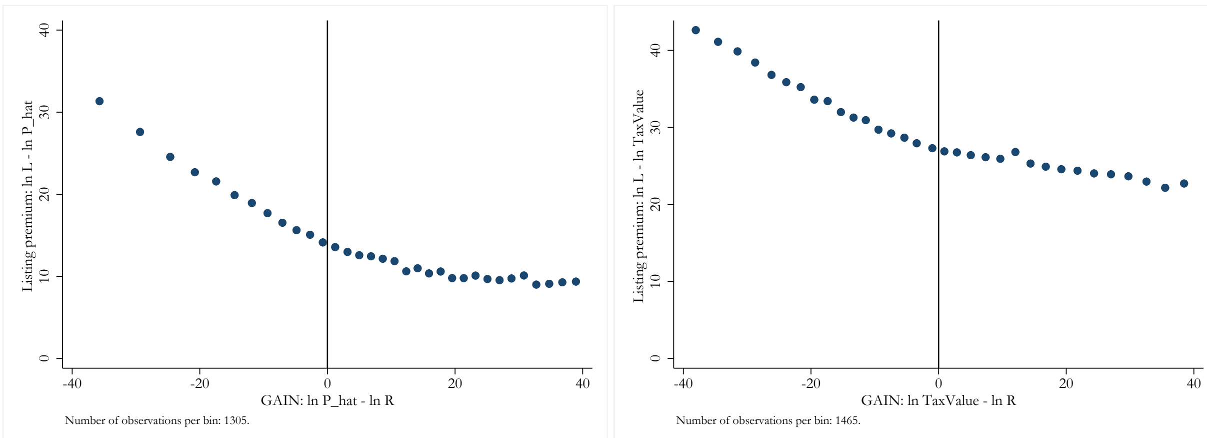


Figure L.7

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

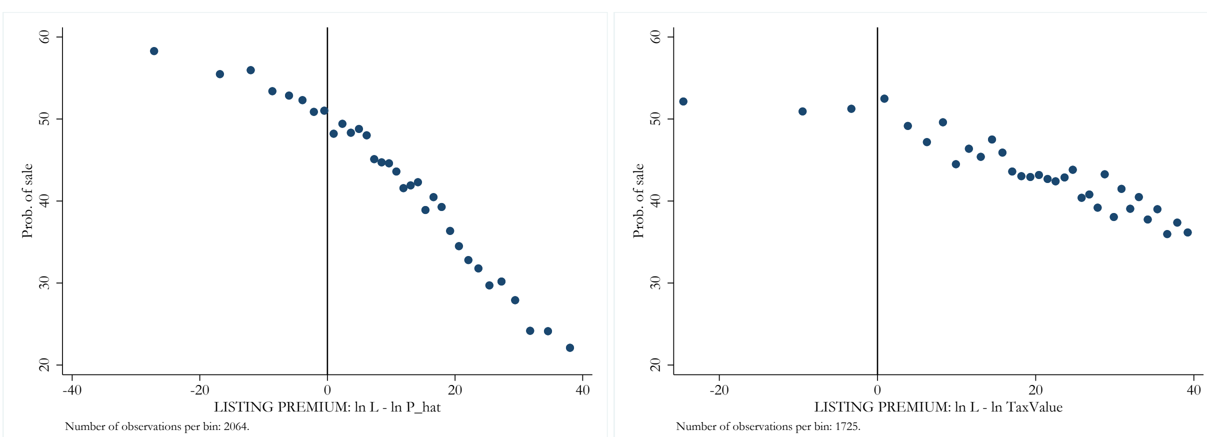
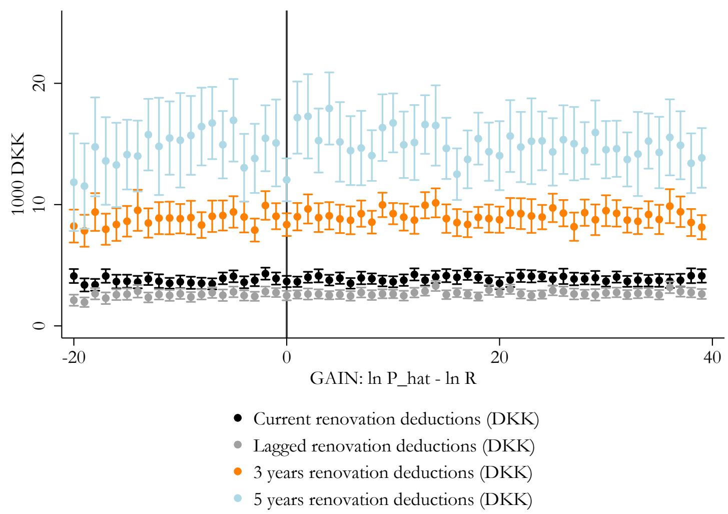


Figure L.8

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 $\widehat{\ell}$

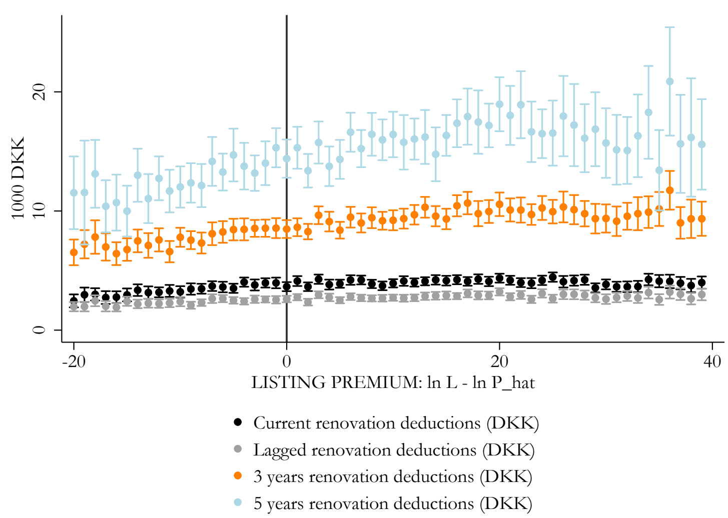
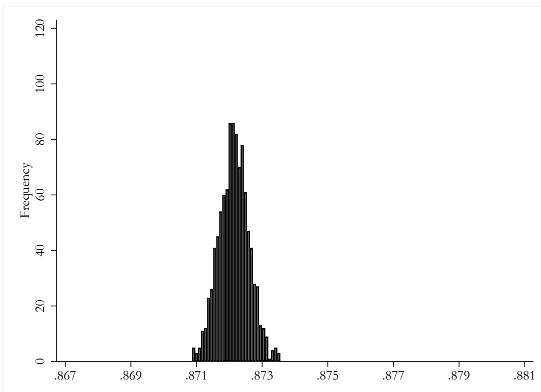


Figure L.9

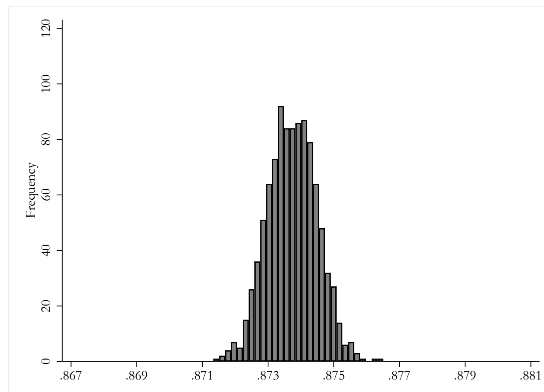
Distribution of R^2s from Out-of-Sample Estimation of the Hedonic Model

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

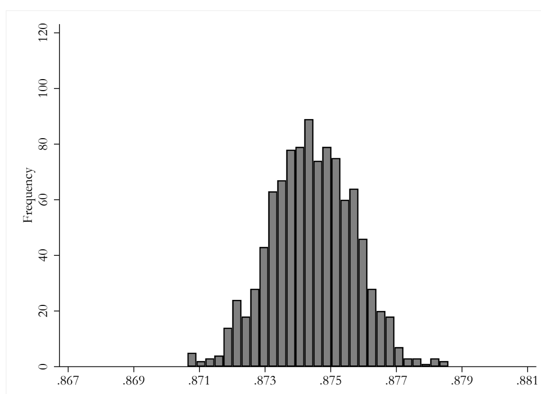
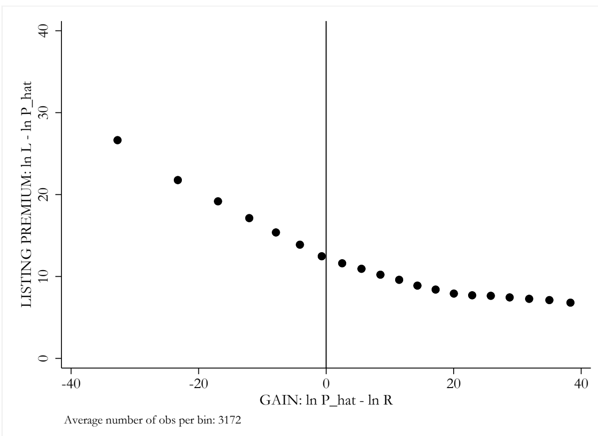


Figure L.10

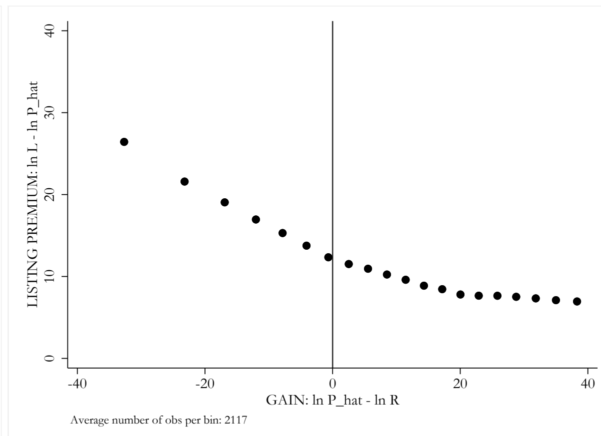
Listing Premia across Potential Gains - Out-of-Sample Predictions

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

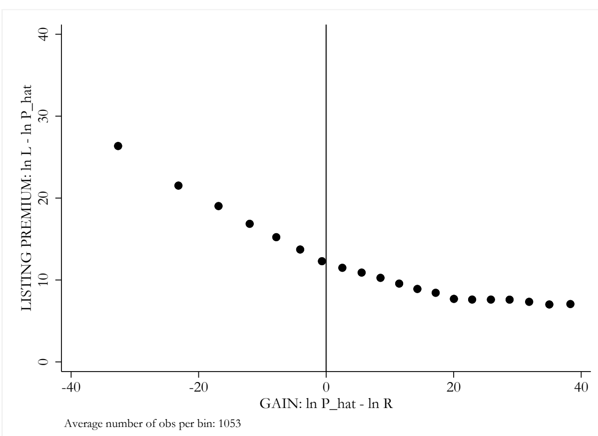
(a) 25 percent out of sample



(b) 50 percent out of sample



(c) 75 percent out of sample



(d) Main data, only sold properties

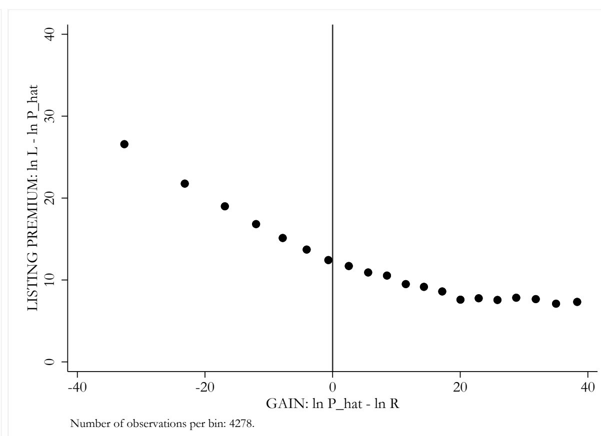
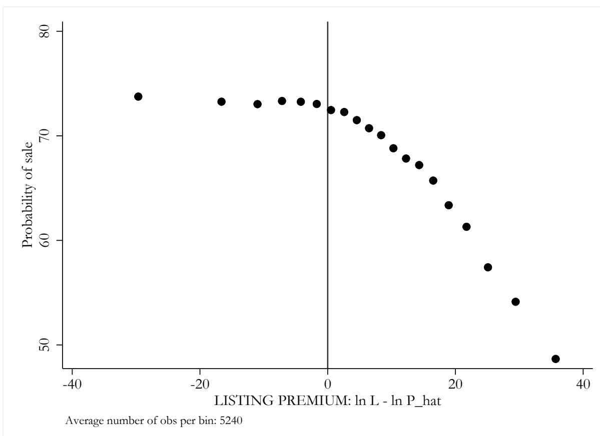


Figure L.11

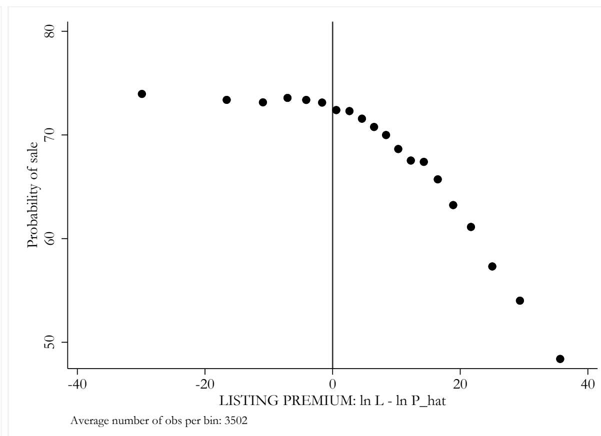
Probability of Sale by Listing Premium (Concave Demand) - Out-of-Sample Predictions

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.

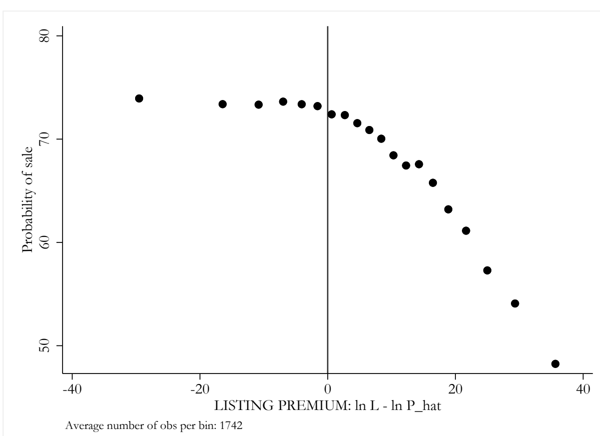
(a) 25 percent out of sample



(b) 50 percent out of sample



(c) 75 percent out of sample



(d) Main data, only sold properties

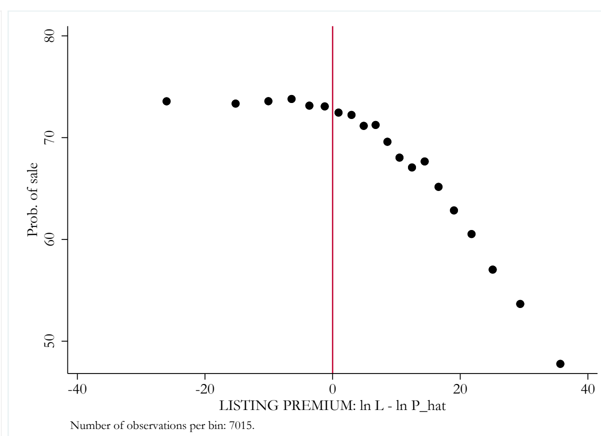
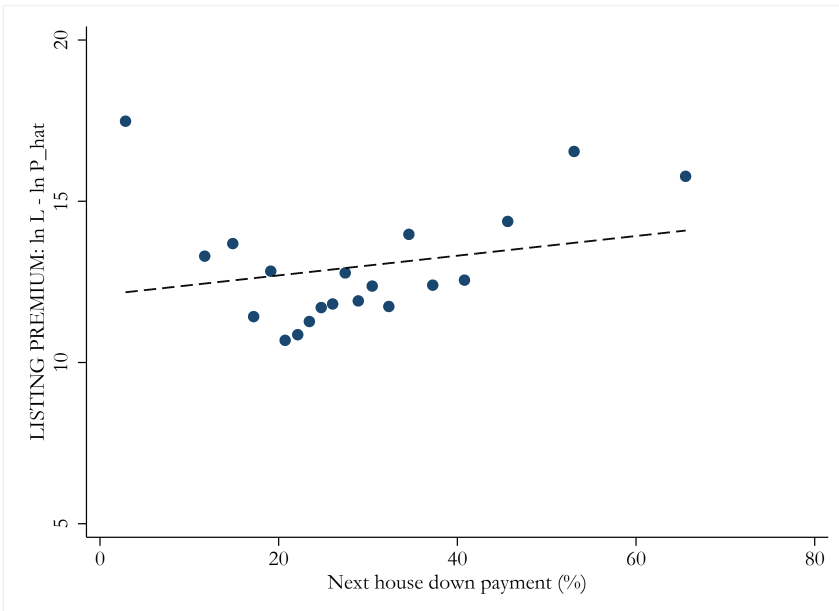


Figure L.12

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

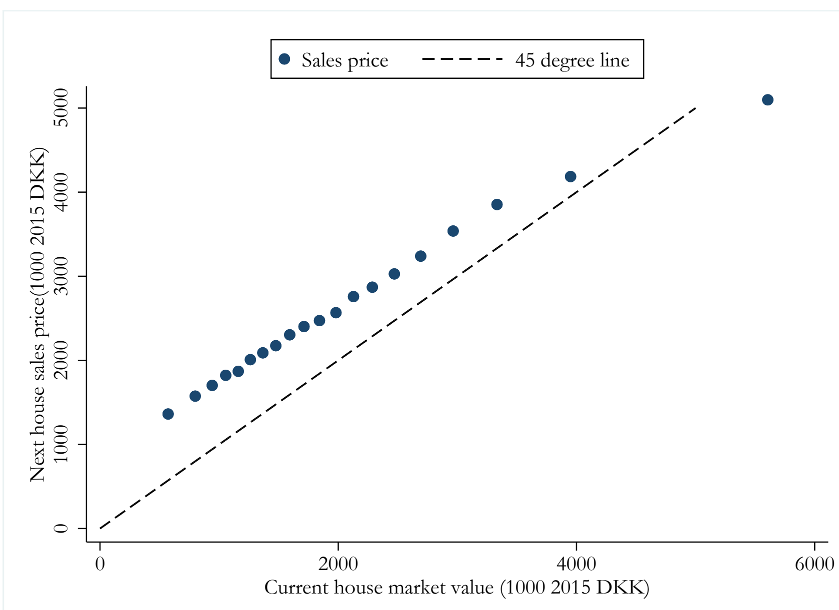


Figure L.13
Time-On-the-Market and Retraction Rate

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

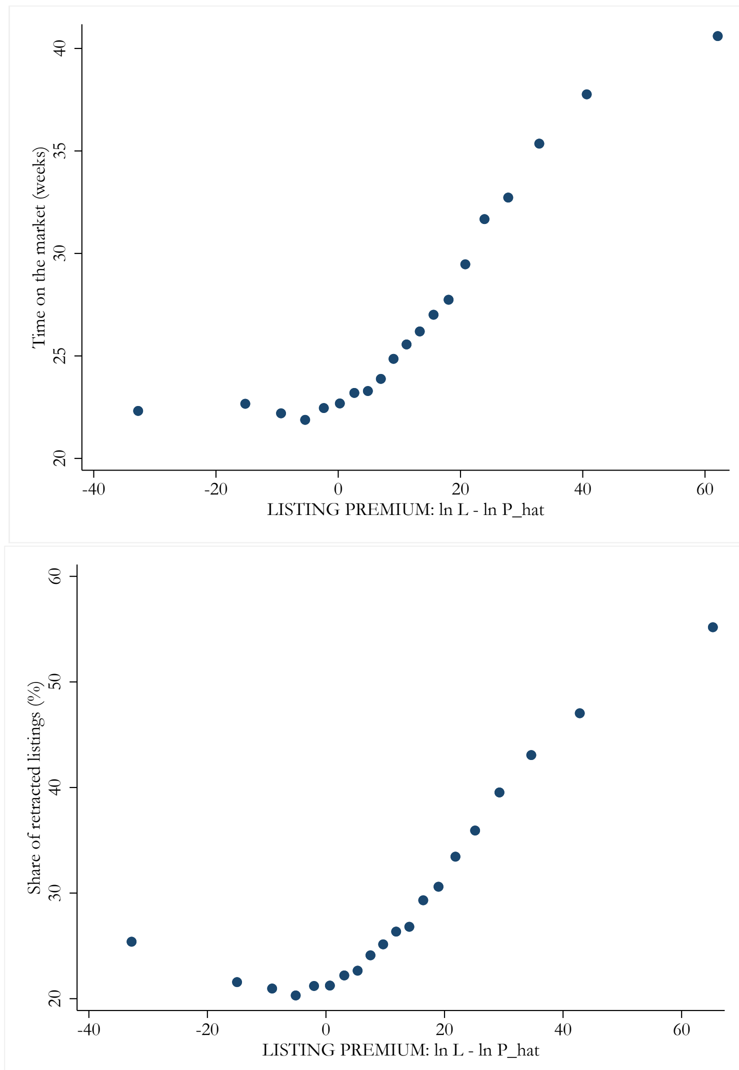


Figure L.14

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). The authors own the copyright of this digital picture. The right-hand side shows a screenshot of property outlines for houses that are part of a row house unit. In contrast, Panel B shows property outlines for detached single family houses, which have visibly different features from other surrounding houses and are less homogeneous than the ones in the row house unit. The mapping plots are reproduced with permission from Styrelsen for Dataforsyning og Effektivisering (https://kort.bbr.dk),

Panel A

Panel B

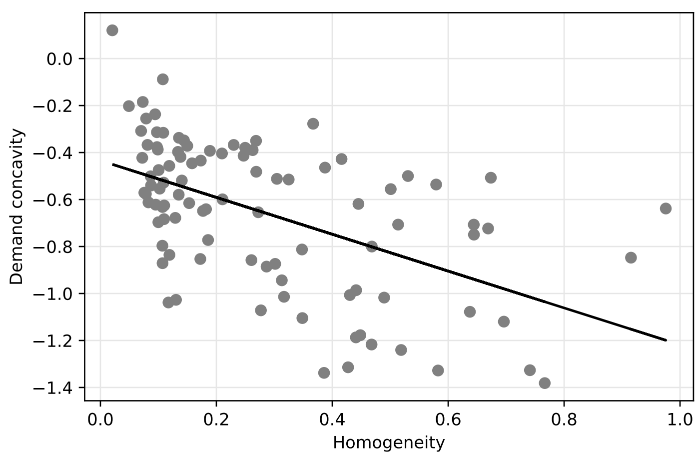
Figure L.15

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

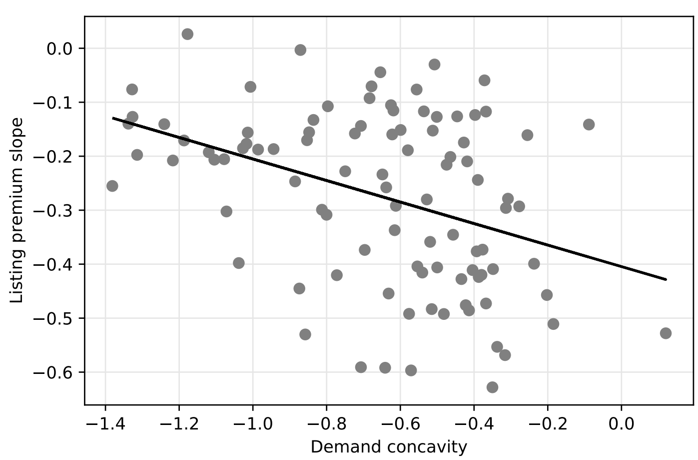
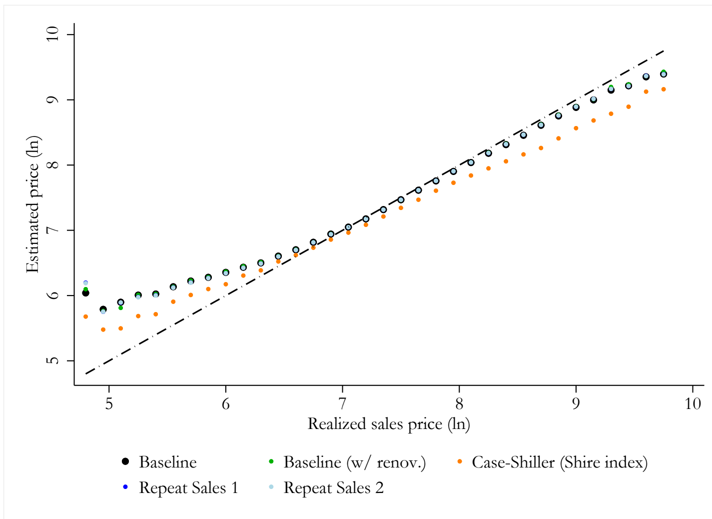


Figure L.16

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

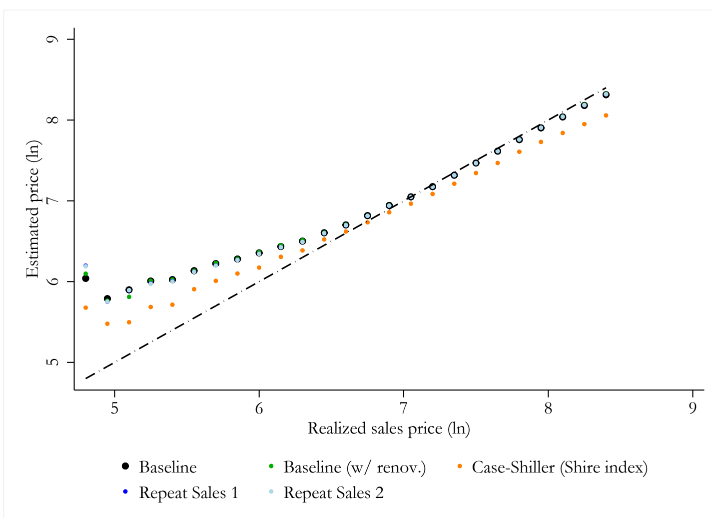
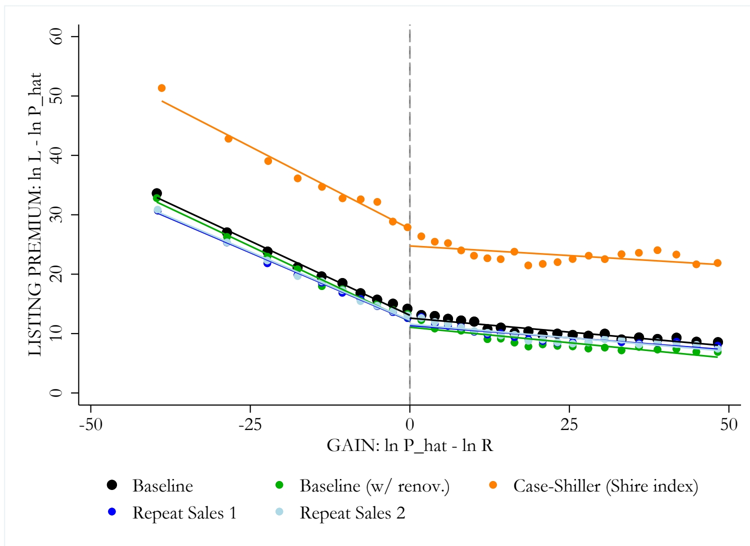


Figure L.17
Hockey Stick and Demand Concavity Across Main Models

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

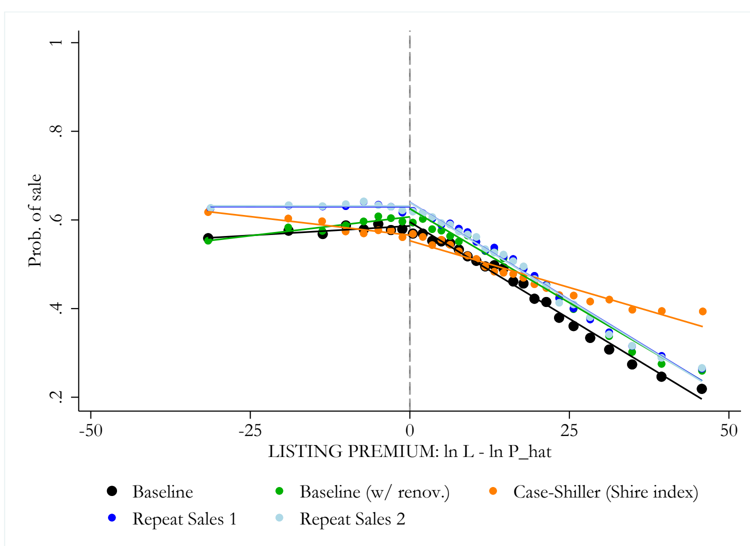
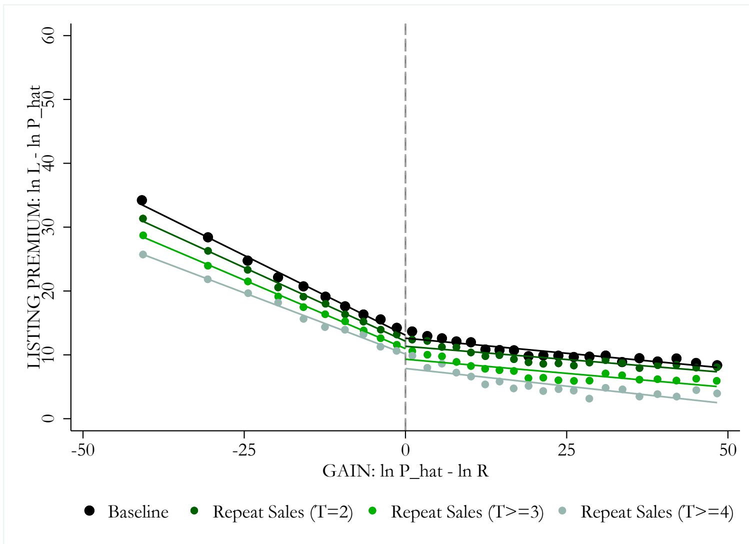


Figure L.18

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

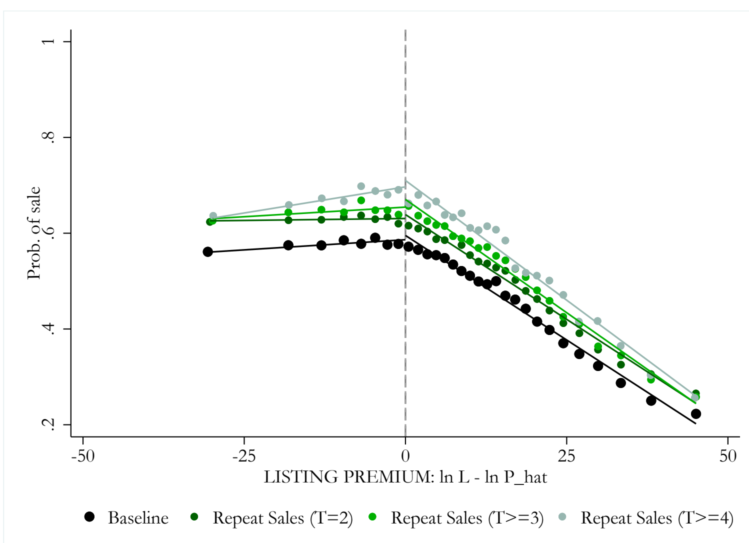


Figure L.19
RKD Validation: Smooth Density of Assignment Variable

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.[27]

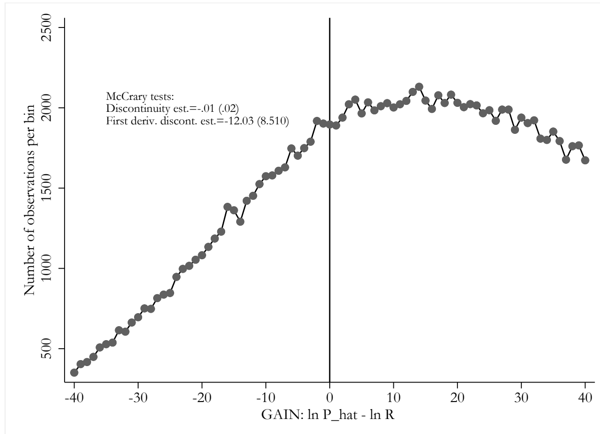


Figure L.20
RKD Validation: Covariates Smooth around Cutoff

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.

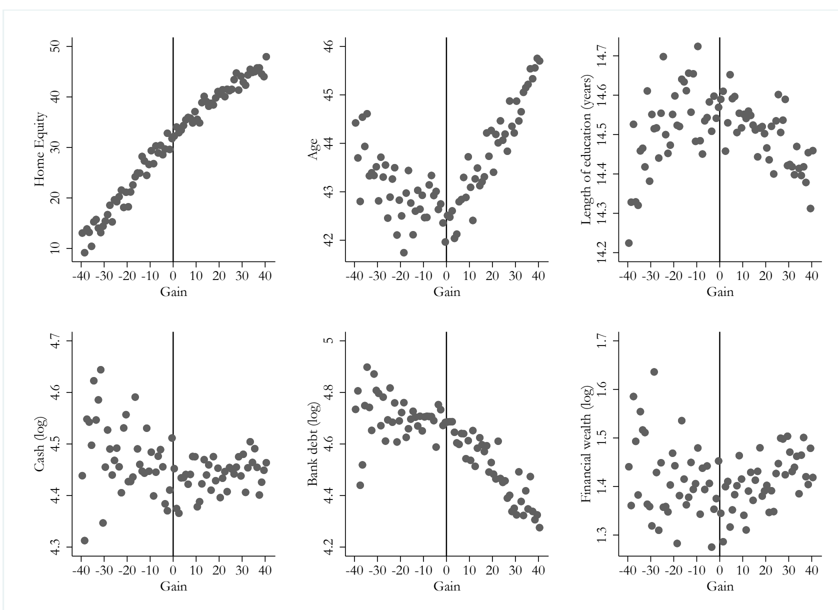


Figure L.21

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).

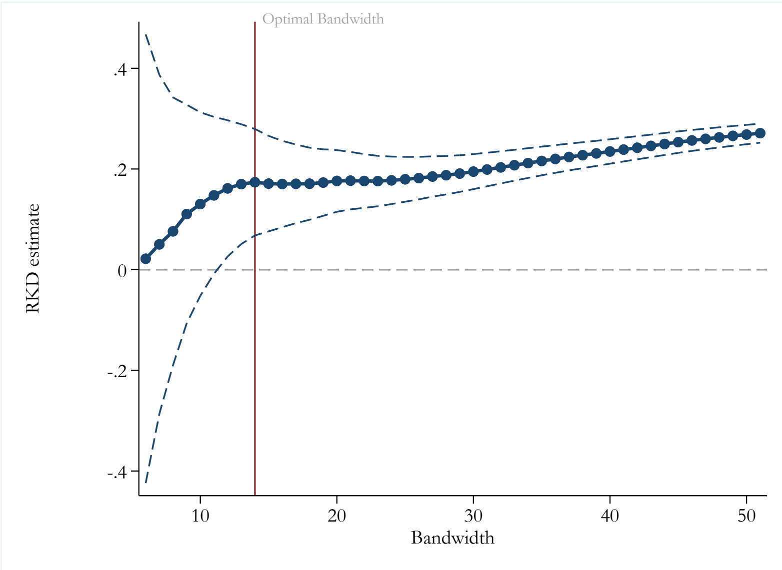


Figure L.22

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).

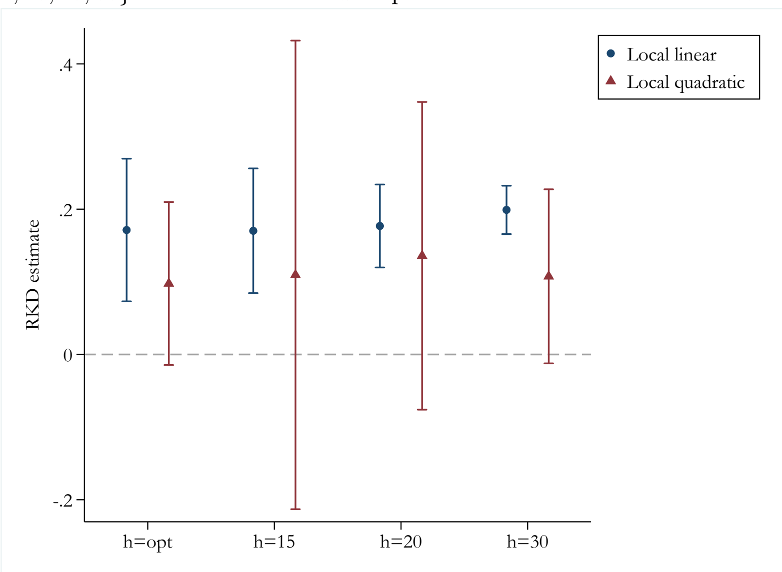


Figure L.23

Incidence of Round Numbers by Rounding Multiple

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

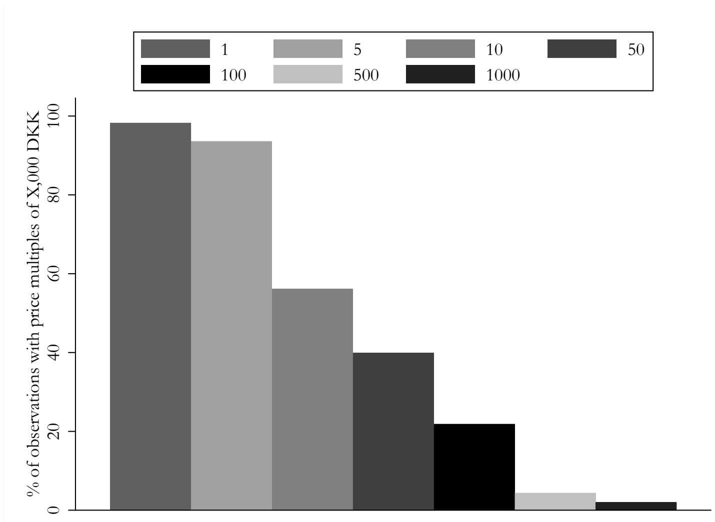
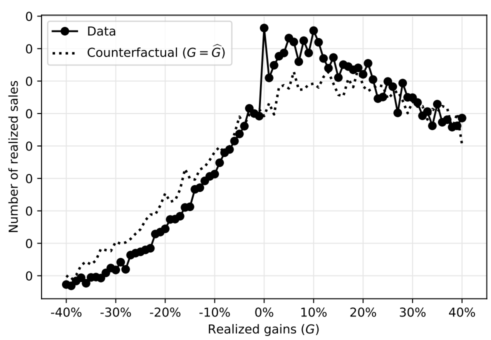


Figure L.24

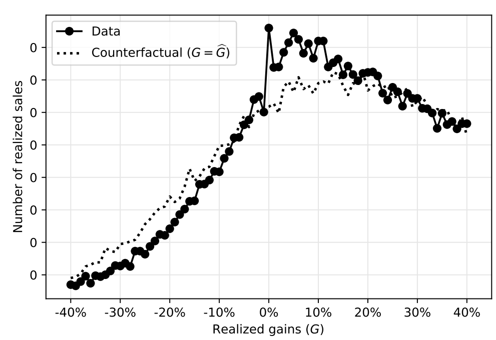
Bunching Robustness: Excluding Sales at Rounded Prices

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.

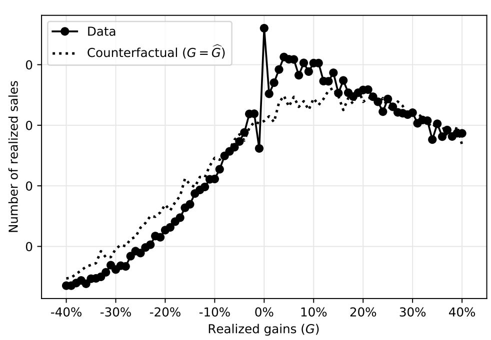
10,000 DKK



50,000 DKK



100,000 DKK



500,000 DKK

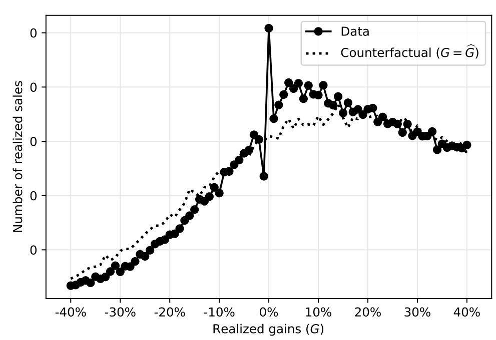
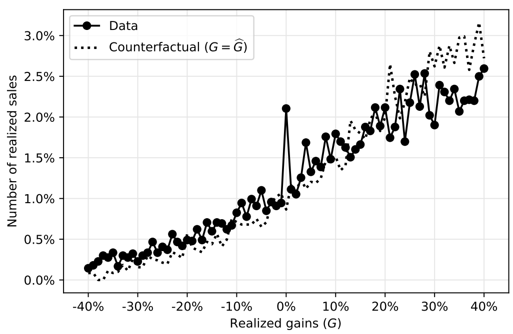


Figure L.25

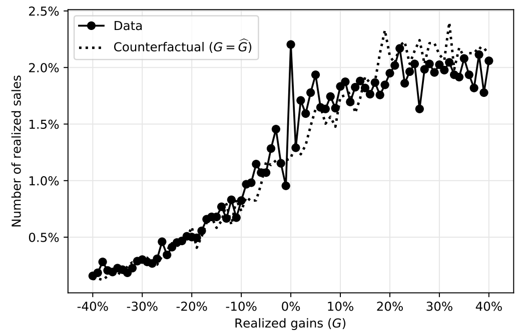
Bunching Robustness: Across Previous Sales Price

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.

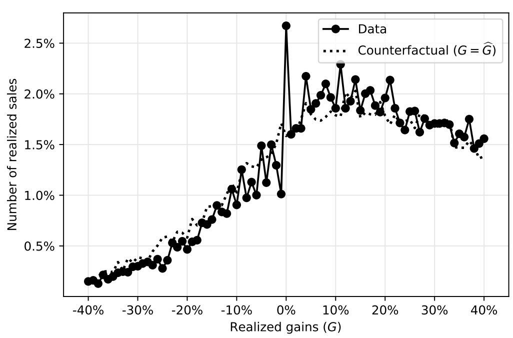
Below DKK 659,000



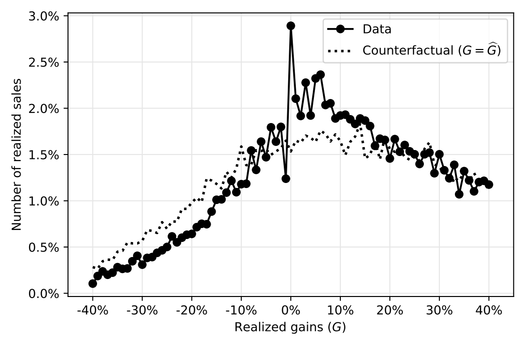
DKK 659,000 – DKK 953,000



DKK 953,000 – DKK 1,313,000



DKK 1,313,000 – DKK 1,901,000



Above DKK 1,901,000

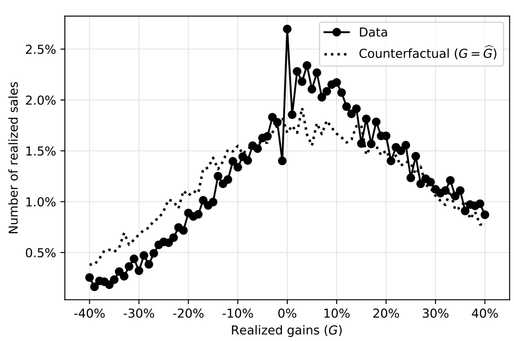
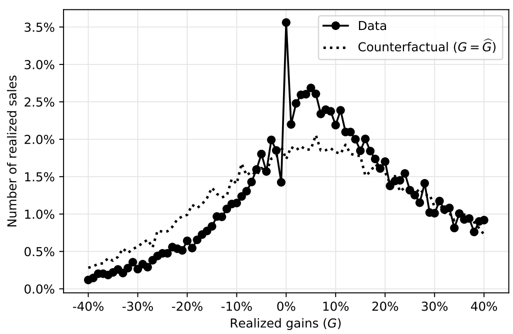


Figure L.26

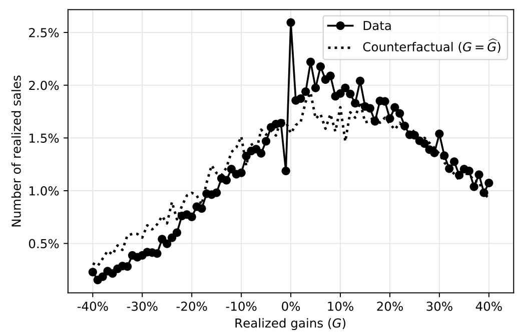
Bunching Robustness: Across Holding Periods

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.

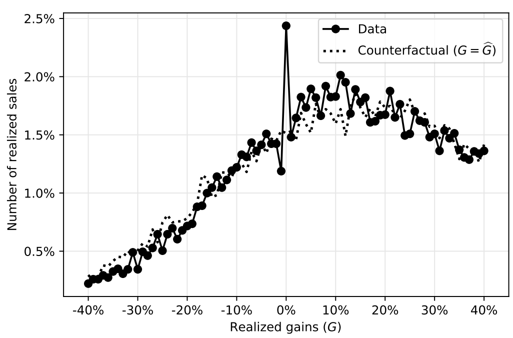
Below 3 years



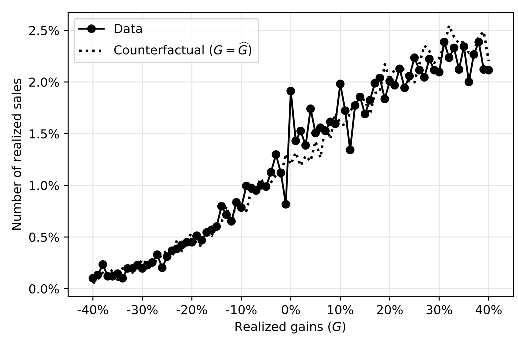
3–6 years



6–9 years



9–12 years



Above 12 years

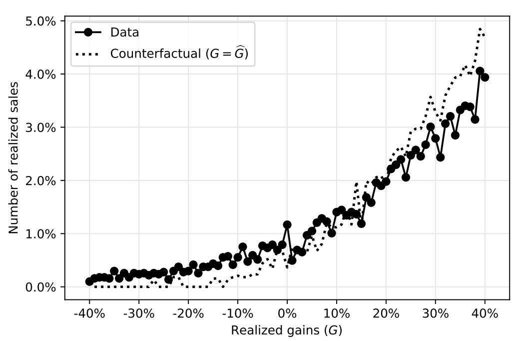


Figure L.27

Bunching Robustness: Excess Mass across Holding Periods

This figure shows robustness for the frequency of sales across gains at the realized price. The dots represent excess mass measures as the frequency of observations in each percentage point bin of realized gains, relative to the frequency of observations in the same percentage point bin, coresponding to potential gains. Error bars indicate 95% confidence intervals, based on bootstrap standard errors.

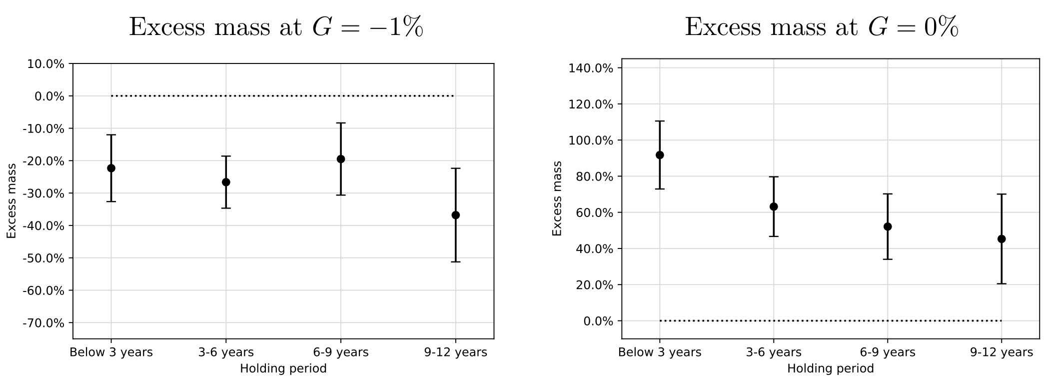


Figure L.28
Bunching Robustness: Model with Cohort Fixed Effects

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.

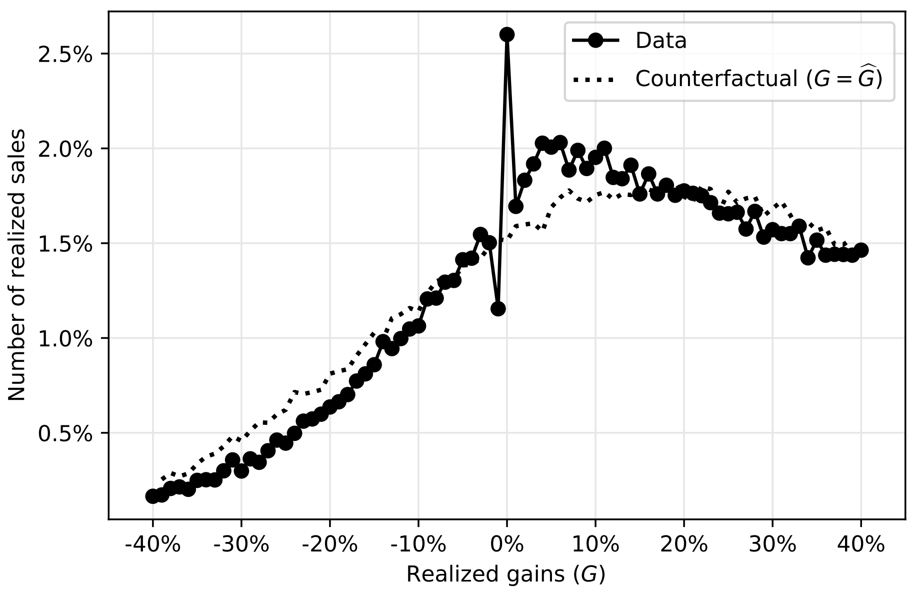
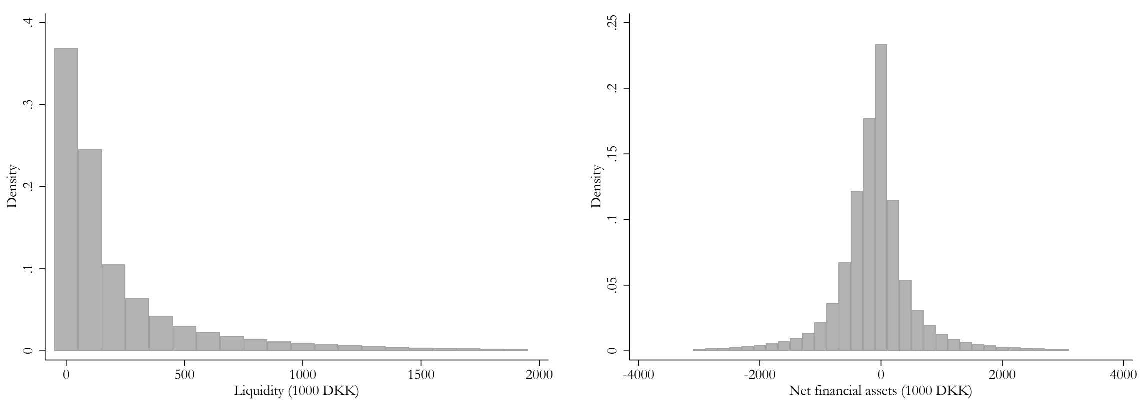


Figure L.29

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

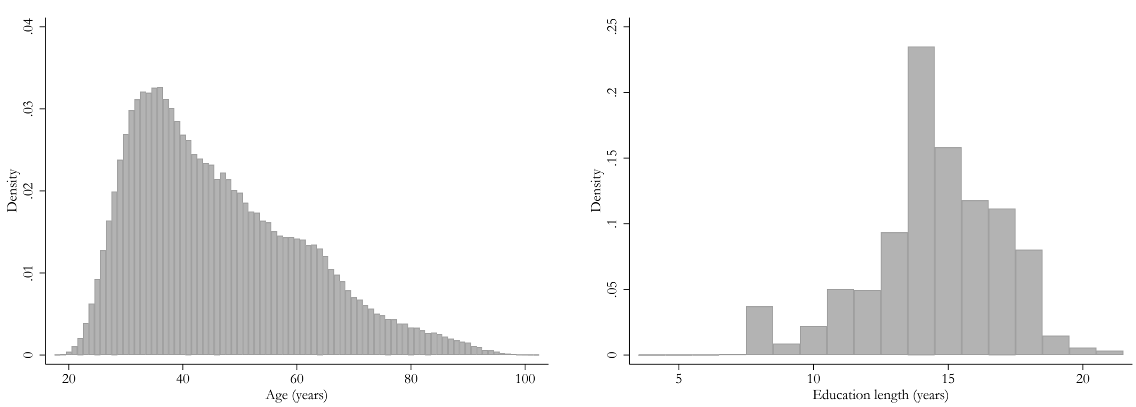


Figure L.30

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.

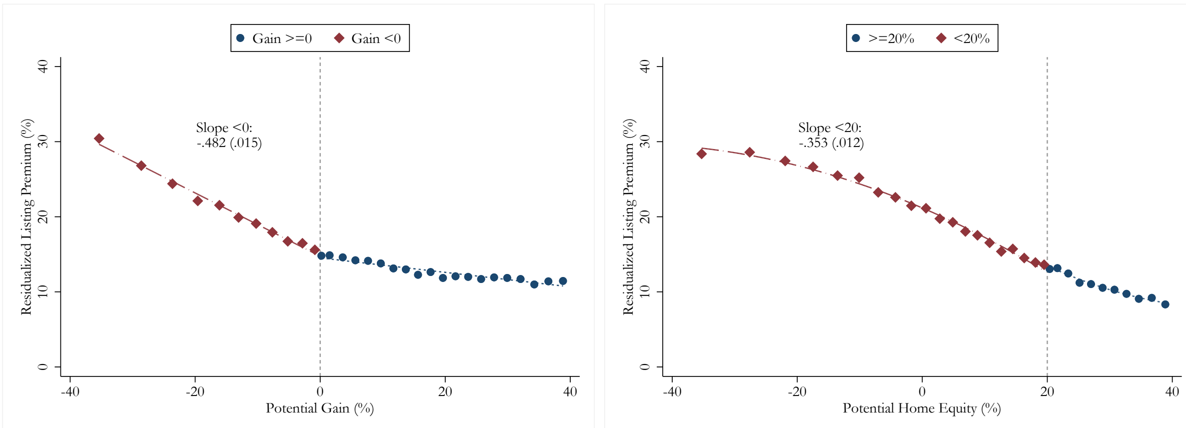


Figure L.31

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

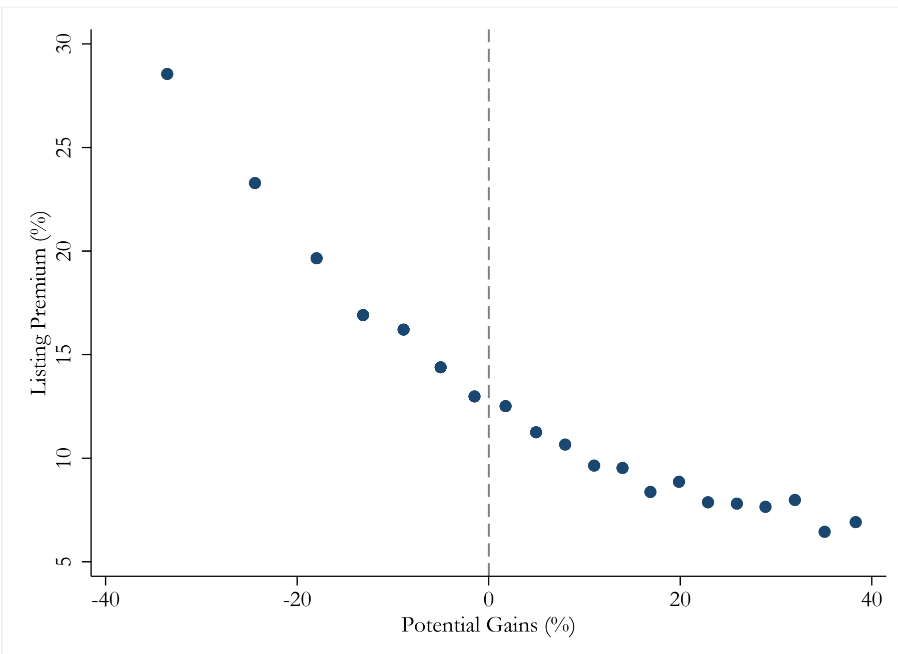


Figure L.32
Decomposition of model-implied moments

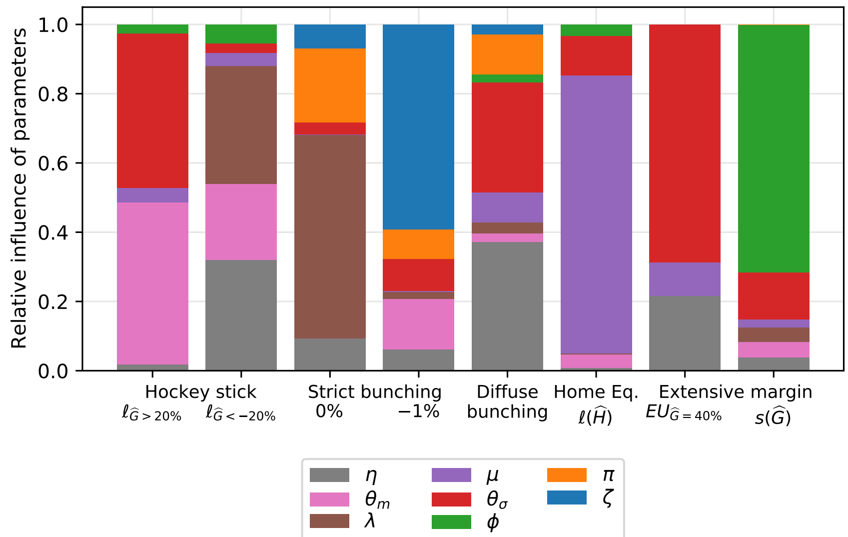


Figure L.33
Utility penalty calculated in the data and the model

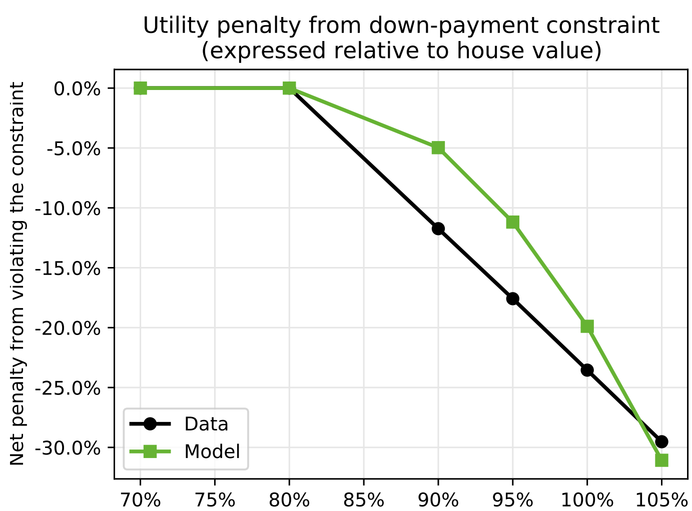


Figure L.34
Simulation results: Unobserved quality

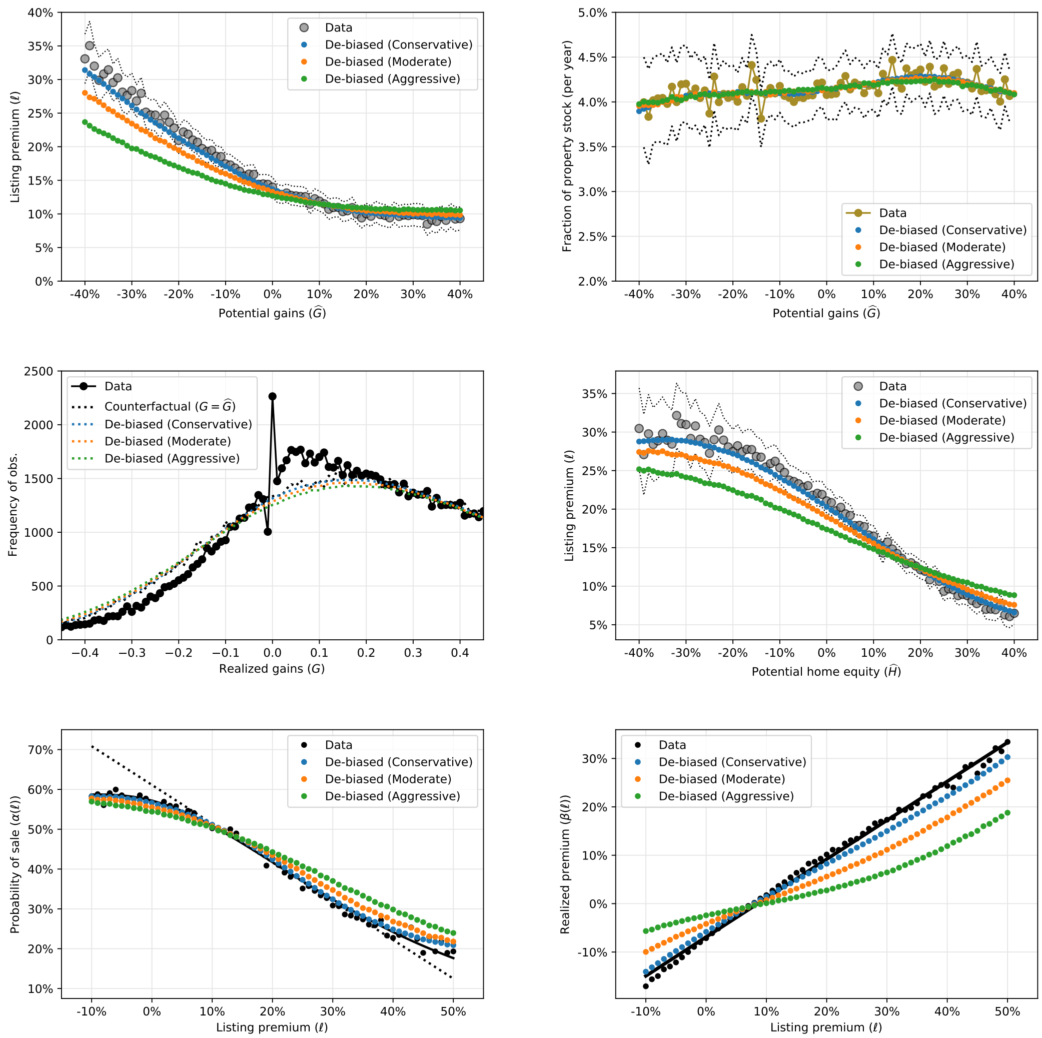


Table L.1

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]	**615,040**
Unmatched in registers[b]	-107,679
	507,361
Cleaning	
No reference price[c]	-144,962
Owner ID not uniquely determined[d]	-71,876
Non-household buyer	-10,382
Foreclosures	-6,416
Extreme price[e]	-5,499
Owner ID not found[f]	-3,987
Missing lot size	-2,823
Error in listing or previous purchase date[g]	-1,915
Intra-family sale and other special circumstances	-2,101
No listing price	-879
Missing hedonic characteristics	-8
	256,513
Sample selection	
Summer house	-24,098
Professional investor[h]	-18,312
Final data	**214,103**
Of which with a mortgage	172,225
Of which without a mortgage	41,878

[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

Cost of borrowing

[28]he table shows approximated costs of funding a home through a mortgage and a bank loan. Prices are based on small surveys of banks and mortgage bank, conducted by bolius.dk and mybanker.dk.

Mortgage (up to 80% LTV)	
Bank fee[29]	∼9,000 DKK
Stamp fee (fixed amount)	1,750 DKK
Stamp fee (percentage)	1.45%
Brokerage	∼0.15%
Spread	∼0.20%
Bidragssats $< 40\% LTV$	∼0.38%
Bidragssats 40-60% LTV	∼0.83%
Bidragssats 60-80% LTV	∼1.11%
Bank loan (80-95% LTV)	
Bank fee	0-14,000 DKK
Stamp fee (fixed)	1,660 DKK
Stamp fee (percentage)	1.5%

Sources:

https://www.bolius.dk/boliglaan-i-banken-find-det-billigste-18078

https://www.mybanker.dk/sammenlign/bolig/bidragssatser/

https://www.bolius.dk/omkostninger-ved-at-koebe-bolig-18145

Table L.3

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	0.550	0.550	0.544	0.544
2) Municipality-year FEs	0.477	0.768	0.478	0.758	0.468	0.775
3) Tax-assessment	0.800	0.876	0.804	0.864	0.834	0.881
4) Renovation exemptions	0.026	0.876	0.009	0.865	0.027	0.882

Table L.4

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.5
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.

	(1)
	Mean
50 pct out-of-sample	0.874
	(0.0000238)
25 pct out-of-sample	0.874
	(0.0000402)
100 pct in-sample	0.876
	(.)
Observations	1000

Standard errors in parentheses

Table L.6
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.

	(1)
	Mean
50 pct out-of-sample	0.764
	(0.0000383)
25 pct out-of-sample	0.765
	(0.0000676)
100 pct in-sample	0.768
	(.)
Observations	1000

Standard errors in parentheses

Table L.7

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.[30] 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	-0.422***	-0.569***	-0.548***	-0.428***
	(0.048)	(0.102)	(0.098)	(0.137)
Household controls				✓
Observations	96	96	96	96
R^2	0.367	-	-	-
First-stage F-stat	-	30.65	30.49	13.36

Table L.8

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 $\widehat{\ell}$ ($\widehat{G} \in (-1, 1]$)	14.054	13.432	13.914	27.406	12.484	13.082
	(0.412)	(0.460)	(0.372)	(0.802)	(0.385)	(0.405)
(2) Slope $\widehat{\ell} - \widehat{G}$ ($\widehat{G} < 0$)	-0.482	-0.485	-0.487	-0.537	-0.452	-0.453
	(0.015)	(0.019)	(0.009)	(0.035)	(0.018)	(0.018)
(3) Slope $\widehat{\ell} - \widehat{G}$ ($\widehat{G} \geq 0$)	-0.111	-0.132	-0.117	-0.081	-0.104	-0.109
	(0.010)	(0.012)	(0.004)	(0.022)	(0.009)	(0.010)
(4) Slope $\widehat{\ell} - \widehat{H}$ ($\widehat{H} < 20$)	-0.353	-0.346	-0.356	-0.632	-0.299	-0.296
	(0.012)	(0.018)	(0.007)	(0.011)	(0.012)	(0.012)
(5) Slope P(sale)-$\widehat{\ell}$ ($\widehat{\ell} < 0$)	0.026	0.147	0.004	-0.214	-0.042	-0.047
	(0.043)	(0.051)	(0.042)	(0.049)	(0.045)	(0.045)
(6) Slope P(sale)-$\widehat{\ell}$ ($\widehat{\ell} \geq 0$)	-0.904	-0.866	-0.890	-0.494	-0.922	-0.922
	(0.013)	(0.016)	(0.013)	(0.014)	(0.014)	(0.014)
Model R^2	0.876	0.881	0.873	0.566	0.881	0.881
Number of observations	214,103	136,717	107,052	202,652	180,545	180,556
(7) P(listing)-$\widehat{G}$	.003	.002	0.021	.003	.037	.036
	(0.002)	(0.002)	(0.000)	(0.004)	(0.003)	(0.003)
Number of observations (ext.)	5,538,052	5,538,052	2,769,026	5,109,438	2,705,243	2,706,078

Table L.9

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 $\widehat{\ell}$ ($\widehat{G} \in (-1,1]$)	12.639	10.873	10.792	12.853	13.432	11.815	12.487
	(0.393)	(0.424)	(0.507)	(0.389)	(0.460)	(0.524)	(0.849)
(2) Slope $\widehat{\ell} - \widehat{G}$ ($\widehat{G} < 0$)	-0.451	-0.421	-0.387	-0.454	-0.485	-0.503	-0.536
	(0.018)	(0.020)	(0.025)	(0.017)	(0.019)	(0.021)	(0.032)
(3) Slope $\widehat{\ell} - \widehat{G}$ ($\widehat{G} \geq 0$)	-0.102	-0.114	-0.145	-0.111	-0.132	-0.151	-0.176
	(0.009)	(0.011)	(0.012)	(0.009)	(0.012)	(0.012)	(0.017)
(4) Slope $\widehat{\ell} - \widehat{H}$ ($\widehat{H} < 20$)	-0.302	-0.233	-0.186	-0.299	-0.346	-0.327	-0.314
	(0.012)	(0.014)	(0.018)	(0.012)	(0.018)	(0.024)	(0.044)
(5) Slope P(sale)-$\widehat{\ell}$ ($\widehat{\ell} < 0$)	0.001	0.065	0.174	-0.029	0.147	0.259	0.244
	(0.045)	(0.061)	(0.085)	(0.045)	(0.051)	(0.061)	(0.102)
(6) Slope P(sale)-$\widehat{\ell}$ ($\widehat{\ell} \geq 0$)	-0.906	-0.969	-0.989	-0.913	-0.866	-0.799	-0.822
	(0.014)	(0.019)	(0.029)	(0.014)	(0.016)	(0.021)	(0.037)
Model R^2	0.880	0.887	0.894	0.881	0.881	0.885	0.890
Number of observations	180,545	95,080	43,894	180,556	136,717	85,483	29,073
(7) P(listing)-$\widehat{G}$	.038	.054	.075	.037	.002	.004	.009
	(0.003)	(0.004)	(0.007)	(0.003)	(0.003)	(0.004)	(0.007)
Number of observations (ext.)	2,705,243	1,150,804	440,439	2,706,078	3,489,967	2,050,917	671,547

Table L.10

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	0.171***	0.177***	0.199***	0.239***
	(0.050)	(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	48,682	48,682	48,682	48,682
N above cutoff	165,421	165,421	165,421	165,421

Table L.11

Alternative Estimation of "Hockey Stick" Pattern

The table reports estimated coefficients from the following regression specifications:

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

$$L_i = a_0 + b_1\widehat{P} + b_2R + \varepsilon_i,$$

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 $\times$ 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