Imperial College London
Imperial College Business School, Department of Finance

Essays on Mortgage and Housing Markets

Lu Liu

15 January 2023

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Business Research

Abstract

This thesis studies household behavior in housing and mortgage markets. The thesis uses micro-level data and structural models to highlight novel features of the choice environment and trade-offs that households face when they make high-stakes financial decisions in these markets. The thesis consists of three chapters.

In the first chapter, I show that the choice of mortgage fixation length tracks the life-cycle dimension of credit risk in the mortgage market, where the loan-to-value (LTV) ratio declines over time. High-LTV borrowers face a trade-off between their demand to insure against repricing, and obtaining a lower credit spread over time using shorter-term contracts. To quantify demand for longer-term contracts, I develop a life-cycle model of optimal mortgage fixation choice. With baseline house price growth and interest rate risk, high-LTV households prefer shorter-term contracts, in line with the data.

In the second chapter, I use comprehensive data on issued and offered mortgages in the UK to document that lenders differ substantially in the fees they charge, and that borrowers appear less overall cost-sensitive to products with fees. I show that lenders pass on firm-specific funding cost shocks via fees, but not interest rates, consistent with strategic pricing of fees, and maintaining competitive prices in the salient price dimension, interest rates.

The third chapter quantifies reference dependence and loss aversion in the housing market using rich Danish administrative data. The structural model matches key nonparametric moments, including a “hockey stick" in listing prices with nominal gains, and bunching at zero realized nominal gains. Households derive substantial utility from gains over the original house purchase price; losses affect households roughly 2.5 times more than gains. The model helps explain the positive correlation between aggregate house prices and turnover, but cannot explain visible attenuation in reference dependence when households are more financially constrained.

Acknowledgements

I am incredibly grateful to Tarun Ramadorai, Marcin Kacperczyk and Antoinette Schoar for their generosity, advice and mentorship throughout the PhD and beyond.

Financial support from the Imperial College President's PhD Scholarship, Economic and Social Research Council (LISS-DTP) and Brevan Howard Centre for Financial Analysis is gratefully acknowledged. I thank Antoinette Schoar and the finance department at MIT Sloan for hosting my visit.

I am grateful to Franklin Allen, Chris Hansman, Alex Michaelides and other members of the finance department at Imperial College for their continued support. Thanks to my co-authors, from whom I continue to learn. I thank Matt Waldron and Jagdish Tripathy for their support at the Bank of England.

I thank my PhD cohort, and Heidi, Jamie, Daniel, Will and Lydia for the shared PhD journey. Thanks to Pari, Olivia, Peter, Kim and Alex for their help and feedback.

Lastly, I thank my family, friends – in particular Peggy, Sarah, Stefanie, Sylvia and Lauren – and community for believing in me and supporting me on my journey. Very special thanks to Rufus, for everything.

Statement of Originality

This is to certify that the content of this thesis is my own work. Any other assistance received in preparing this thesis and sources have been appropriately referenced.

Disclaimer

Chapter 1 and 2 of this thesis use data provided to the Bank of England by the Financial Conduct Authority. The views expressed are those of the author and do not necessarily reflect the views of the Bank of England or any of its policy committees.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Contents

Abstract i

Acknowledgements ii

Statement of Originality iv

Disclaimer iv

Copyright Declaration iv

1 The Demand for Long-Term Mortgage Contracts and the Role of Collateral 3

1.1 Introduction . . . 3

1.1.1 Related Literature . . . 9

1.2 Institutional Setting and Data . . . 11

1.2.1 UK Fixed-Rate Mortgages and Institutional Setting . . . 11

1.2.2 Dataset Construction . . . 14

1.3 Mortgage Contract Choice and the Life Cycle of Credit Risk . . . 15

1.3.1 Mortgage Fixation Length Choice and the Role of LTV . . . 16

1.3.2 Mortgage Pricing . . . 19

1.3.3 Credit Repricing and the Life Cycle of Credit Risk in Mortgages . . . 20

1.3.4 Credit Repricing And The Relative Cost of Long-Term Mortgages . . . 22

1.3.5 Discussion of Mechanisms . . . 23

1.4 Life-Cycle Model of Optimal Mortgage Fixation Choice . . . 25

1.4.1 Model Setup . . . 26

1.4.2 Calibration and Solution . . . 31
1.4.3 Results . . . 33
1.4.4 Discussion of Results . . . 37
1.5 Conclusion . . . 39

2 Non-Salient Fees in the Mortgage Market 50
2.1 Introduction . . . 50
2.2 Background and Data . . . 58
2.2.1 Background on the UK Mortgage Market . . . 58
2.2.2 Data . . . 59
2.3 Stylized Facts and Mortgage Pricing . . . 60
2.4 Borrower-Level Analysis . . . 64
2.4.1 Discrete Choice Model Setup . . . 64
2.4.2 Discrete Choice Model Results . . . 66
2.5 Lender-Level Analysis . . . 67
2.5.1 Pricing Incentives With Non-Salient Fees . . . 67
2.5.2 Identification Strategy . . . 68
2.5.3 Results . . . 71
2.6 Interpretation and Discussion . . . 73
2.6.1 The Role of Fees . . . 73
2.6.2 Pricing Fees in a Specialization Equilibrium . . . 74
2.6.3 Evidence for Specialization Equilibrium . . . 75
2.6.4 Discussion of Alternative Channels and Robustness . . . 77
2.7 Conclusion . . . 79

3 Reference Dependence in the Housing Market 97
3.1 A Model of Household Listing Behavior . . . 103
3.1.1 Demand Functions . . . 104
3.1.2 Reference Dependence and Loss Aversion . . . 105
3.1.3 Down-Payment Constraints . . . 106

3.1.4 Structural Parameters . . . 108
3.2 Data . . . 113
3.2.1 Property Transactions and the Housing Stock . . . 113
3.2.2 Property Listings Data . . . 114
3.2.3 Mortgage Data . . . 114
3.2.4 Owner/Seller Demographics . . . 115
3.2.5 Final Merged Data . . . 115
3.2.6 Hedonic Pricing Model . . . 116
3.3 Reduced-Form Facts . . . 117
3.3.1 Fact 1: Listing Premia and Potential Gains ("Hockey Stick") . . . 117
3.3.2 Fact 2: "Hockey Stick" and Optimization Frictions . . . 119
3.3.3 Fact 3: Bunching in Realized Sales . . . 122
3.3.4 Fact 4: Extensive Margin: Probability of Listing . . . 123
3.3.5 Robustness . . . 124
3.4 Structural Estimation . . . 126
3.4.1 Moments . . . 126
3.4.2 Moments in the Model . . . 127
3.4.3 Estimation . . . 128
3.4.4 Results . . . 129
3.5 Validating the Model . . . 133
3.5.1 Measurement Error and Robustness . . . 133
3.5.2 Interactions Between Preferences and Constraints . . . 134
3.5.3 Price-Volume Correlation . . . 135
3.6 Conclusion . . . 138

References **149**

Appendices **162**

1 Appendix for "The Demand for Long-Term Mortgage Contracts and the Role of Collateral" **163**

A Further Background on Mortgage Market Contract Structure, Pricing and Choice 164
1 Mortgage Contract Structure and Payment Profile . . . 164
2 Mortgage Pricing and the Role of LTV . . . 164
3 Mortgage Contract Choice And The Role of LTV . . . 165
4 Summary Statistics By Contract Choice . . . 166
B Mortgage Fixation Period Across Countries . . . 168
C Dataset Construction . . . 169
1 Merge (PSD007 and PSD001) for Borrower Panel . . . 169
2 Data Cleaning in PSD007 . . . 172
D Evidence On Alternative Choice Mechanisms . . . 174
1 Within-Household Contract Choice . . . 174
2 Ex Post Default by Fixation Choice and LTV . . . 175
3 Household Refinancing Behavior with Prepayment Penalties . . . 176
E Decomposing the Relative Cost of Longer-Term Mortgages . . . 178
1 Decomposition of Mortgage Yield Difference . . . 178
2 Mapping Decomposition of Mortgage Yield Difference to the Data . . . 180
F Model Appendix . . . 184
1 Value Function for $\theta^{LT} = 5$ . . . 184
2 Numerical Solution . . . 185
G Additional Model Results - Robustness . . . 187

2 Appendix for "Non-Salient Fees in the Mortgage Market" 189
A Further Data Description . . . 189
B Additional Figures and Tables . . . 191

3 Appendix for "Reference Dependence in the Housing Market" 201
A Additional Tables and Figures for the Paper . . . 201
Figures . . . 201
Tables . . . 208
B Details on Model Framework . . . 211

1 Derivation of $\widehat{G}_0$ and $\widehat{G}_1$ . . . 211
2 Mapping Between Potential and Realized Gains . . . 213
3 Extensive Margin Decision . . . 213
4 Realization Utility . . . 214
5 The Role of Concave Demand . . . 215
6 State Variables: Listing Premia and Potential Gains . . . 216
7 The Role of the Outside Option . . . 217
8 Structural Estimation . . . 220
9 Magnitudes of Financial Constraints . . . 222
C Detailed Data Description . . . 224
1 Property Transactions and Other Property Data . . . 224
2 Property Listings Data . . . 225
3 Mortgage Data . . . 226
4 Owner/Seller Demographics . . . 226
5 Final Merged Data . . . 227
6 Data Citations . . . 228
D Institutional Background in Denmark . . . 230
1 The Search and Matching Process in the Housing Market . . . 230
2 The Taxation Regime for Residential Property . . . 231
3 The Cost of Borrowing . . . 231
4 Assumability, Refinancing, and Unsecured Mortgages . . . 232
5 The Foreclosure Process . . . 233
E Hedonic Pricing Model and Alternative Models of $\widehat{P}$ . . . 235
1 Hedonic Model and the Tax-assessed Value . . . 235
2 Repeat Sales Models . . . 236
3 Repeat Sales Model with Renovations Data . . . 237
4 Out-of-Sample Testing . . . 239
F Functional Form of Down-Payment Constraints . . . 239
1 Alternative Formulation of Penalty Function . . . 239

2 Downsizing Aversion and Interaction Effects . . . 240
G Functional Form of Measured Concave Demand . . . 241
H Demand Concavity and Housing Stock Homogeneity . . . 241
I Unobserved Quality . . . 244
1 A General Formulation of Unobserved Quality . . . 244
2 Sources of Estimation Error . . . 245
3 Alternative Estimation Methods . . . 247
4 Reference Dependence with Loss Aversion . . . 256
5 Regression Kink Design (RKD) . . . 257
6 Demand Concavity . . . 259
7 Simulation Approach to Bounding Unobserved Quality Effects . . . 261
J Bunching Estimation . . . 263
1 Robustness . . . 263
2 Bunching of Listing Prices around Nominal Purchase Price . . . 264
K Household Demographics . . . 264
1 Liquid Financial Wealth . . . 264
2 Age and Education . . . 265
L Additional Appendix Material . . . 266
Figures . . . 266
Tables . . . 289

List of Tables

1.1 5-Year Contract Choice Regressions 45
1.2 Baseline Calibration of Model Parameters 46
1.3 Long-Term Contract Shares 47
1.4 Value of Long-Term Contract 48

2.1 Example of a mortgage price table 54
2.2 Descriptive Statistics 90
2.3 Proportion of Cost-Dominated Products by Fee Category (in %) 91
2.4 Logit Model Estimation Results 92
2.5 Changes in Pricing and Total Cost 93
2.6 Changes in Pricing and Total Cost, By Lender Pricing Type 94
2.7 Lender Characteristics By Pricing Type 95

3.1 Variables and parameters in the model 145
3.2 Structural Parameters 146
3.3 Alternative Model Specifications 147
3.4 Price-Volume Correlation in the Data and the Model 148

A.1 Summary statistics for 2-year and 5-year fixed-rate borrowers 167
C.1 Matching to 2015H2 mortgage stock 171
C.2 Balance of matched observations 171
C.3 Sample selection balance (fixation length observed) 172
D.1 5-Year Fixation Length Contract Choice Regressions - Panel 174
G.1 Long-Term Contract Shares (Robustness) 187

G.2 Value of Long-Term Contract (Robustness) . . . 188

B.1 Example of Moneyfacts Data Structure . . . 196
B.2 Proportion of Cost-Dominated Products By Fee Category (in %) (High Loan Value) . . . 197
B.3 Descriptive Statistics For Discrete Choice Analysis . . . 197
B.4 Balance Tests For Funding Shock ϕ . . . 198
B.5 Changes In Price Components And Total Cost (High LTV) . . . 199
B.6 Fee-Rate Trade-Off . . . 200

A.1 Literature Overview . . . 208
A.2 Genesove and Mayer (2001) Replication . . . 209
A.3 Amendments to the Danish Mortgage-Credit Loans and Mortgage-Credit Bonds Act Between 2009 and 2016 . . . 210
A.4 Goodness Of Fit . . . 211
L.1 Construction of Main Dataset . . . 288
L.2 Cost of borrowing . . . 289
L.3 R^2 of Hedonic Model - Contributions . . . 290
L.4 Overview and Description of Models of $\widehat{P}$ . . . 291
L.5 Out-of-Sample Test of Hedonic Model . . . 292
L.6 Out-of-Sample Test of Hedonic Model without Tax-Assessed Value . . . 292
L.7 Regional Variation in Demand Concavity and Hockey Stick - OLS and IV Regressions . . . 293
L.8 Comparison of Moment Summary Metrics Across Models of $\widehat{P}$ - Main Models . . . 294
L.9 Comparison of Moment Summary Metrics Across Models of $\widehat{P}$ - Additional Models 295
L.10 Regression Kink Design . . . 296
L.11 Alternative Estimation of "Hockey Stick" Pattern . . . 297

List of Figures

1.1 Fixation Length Contract Choice by LTV . . . 41
1.2 LTV Pricing . . . 41
1.3 LTV and Interest Rates Over the Life Cycle, by Cohort . . . 42
1.4 Simulation of LTV and Interest Rate Over The Life Cycle . . . 43
1.5 Optimal Mortgage Choice Over the Life Cycle . . . 44

2.1 Example Price Table from MoneySupermarket.com . . . 81
2.2 Distribution of Fees . . . 82
2.3 Example of Products on Offer . . . 83
2.4 Within-Bank Price Dispersion . . . 84
2.5 Logit Model Prediction Error (By Product Type) . . . 85
2.6 Illustration of pricing with non-salient fees . . . 86
2.7 Average Interest Rates and Fees and Branch Density . . . 87
2.8 Excess Cost Dispersion by Fee Category . . . 88
2.9 Relative market shares by lender excess cost level . . . 89

3.1 Reference Dependence and Loss Aversion . . . 139
3.2 Listing Premia and Potential Gains . . . 140
3.3 Listing Premia Across Gains and Home Equity . . . 141
3.4 Optimization Frictions in the Data . . . 141
3.5 Geographic Variation in Concave Demand and Housing Stock Homogeneity . . . 142
3.6 Bunching Around Realized Gains of Zero . . . 142
3.7 Extensive Margin . . . 142

3.8 Overview of Model Fit . . . 142
3.9 "Hockey Stick" across Municipalities . . . 143
3.10 Out-of-Sample Fit: Independent and Interactive Variation . . . 144

A.1 Illustration of Payment Profile for UK Mortgage Contracts with Initial Fixation Period . . . 164
A.2 Mortgage Pricing Regression R^2 . . . 165
A.3 Contract Choice by LTV band . . . 165
A.4 Explaining Fixation Length Choice - LTV and Other Channels . . . 166
B.1 Mortgage Fixation Period Across Countries . . . 168
D.1 Ex Post Default by Fixation Choice and LTV . . . 175
D.2 Refinancing behavior . . . 176
D.3 Ex Post Repricing . . . 177
E.1 Observed and Counterfactual Credit Spread Pricing . . . 182
E.2 Decomposition of Mortgage Yield Difference . . . 182
F.1 Illustration of Value Function and Repricing States . . . 185

B.1 Products By Interest and Fee Quintile . . . 191
B.2 Total Cost Dispersion by Fee Category . . . 192
B.3 Loan-to-Deposit Ratio (2008) . . . 193
B.4 Illustration of Mortgage Price Components . . . 194
B.5 Loan Value Histogram and Fee Distribution . . . 194
B.6 Cost Differential For Average Product In Interest And Fee Quintile (High Loan Value) . . . 195
B.7 Mortgage Payment Profile Over Time . . . 195

A.1 Reference Dependence and Loss Aversion . . . 201
A.2 Graphical Summary Statistics: Potential Gains, Potential Home Equity, Listing Premium and Time-on-the-Market . . . 202
A.3 Bunching of Listing Prices around Reference Point . . . 203
A.4 Joint Distribution of Gains and Home Equity and Regions with $\widehat{G} \lesseqgtr 0$ and $\widehat{H} \lesseqgtr 20$ 204

A.5 Seller Groups - Listed (Relative Shares) . . . 204
A.6 Realized Gains vs. Realized Home Equity: Bunching . . . 205
A.7 Bunching around Realized Gains of Zero: Polynomial Counterfactual . . . 206
A.8 Price-Volume Correlation . . . 206
A.9 Extensive Margin - Residualized . . . 207
L.1 Concave Demand . . . 266
L.2 Cost Of Borrowing . . . 267
L.3 Listing premia by home equity: Alternative parameterization . . . 267
L.4 Actual vs. Predicted Price of Sold Properties . . . 267
L.5 Accuracy of Tax-Assessed Value . . . 268
L.6 Listing Premia across Potential Gains and Tax-Assessed Value . . . 269
L.7 Probability of Sale by Listing Premia (Concave Demand) and Tax-Assessed Value 269
L.8 Renovation Expenses across Potential Gains and Listing Premia . . . 270
L.9 Distribution of R^2s from Out-of-Sample Estimation of the Hedonic Model . . . 271
L.10 Listing Premia across Potential Gains - Out-of-Sample Predictions . . . 271
L.11 Probability of Sale by Listing Premium (Concave Demand) - Out-of-Sample Predictions . . . 271
L.12 Listing Premia and Down Payment, and Current and Next House Price . . . 272
L.13 Time-On-the-Market and Retraction Rate . . . 272
L.14 Illustration of Homogeneity of Housing Stock for IV Estimation . . . 273
L.15 Regional Variation in Demand Concavity, Listing Premium-Gain Slope and Housing Stock Homogeneity . . . 274
L.16 Estimated vs. Realized ln(price) Across Main Models . . . 275
L.17 Hockey Stick and Demand Concavity Across Main Models . . . 275
L.18 Hockey Stick and Demand Concavity Across Repeat Sales Models . . . 275
L.19 RKD Validation: Smooth Density of Assignment Variable . . . 276
L.20 RKD Validation: Covariates Smooth around Cutoff . . . 276
L.21 RKD Robustness: Estimates for Different Bandwidths (Gain) . . . 277
L.22 RKD Estimation: Local Linear vs. Local Quadratic Estimation Results . . . 277

L.23 Incidence of Round Numbers by Rounding Multiple 278
L.24 Bunching Robustness: Excluding Sales at Rounded Prices 279
L.25 Bunching Robustness: Across Previous Sales Price 280
L.26 Bunching Robustness: Across Holding Periods 281
L.27 Bunching Robustness: Excess Mass across Holding Periods 282
L.28 Bunching Robustness: Model with Cohort Fixed Effects 283
L.29 Summary Statistics: Household Demographics 284
L.30 Residualized Listing Premium and Gains and Home Equity 285
L.31 Listing Premium "Hockey Stick" for Sellers Without Mortgage 286
L.32 Decomposition Of Model-Implied Moments 287
L.33 Utility Penalty Calculated In The Data And The Model 287
L.34 Simulation Results: Unobserved Quality 287

The Demand for Long-Term Mortgage Contracts and the Role of Collateral

Lu Liu, Imperial College London

Abstract

Long-term fixed-rate mortgage contracts protect households against interest rate risk, yet most countries have relatively short interest rate fixation lengths. Using administrative data from the UK, the paper finds that the choice of fixation length tracks the life-cycle decline of credit risk in the mortgage market: the loan-to-value (LTV) ratio decreases and collateral coverage improves over the life of the loan due to principal repayment and house price appreciation. High-LTV borrowers, who pay large initial credit spreads, trade off their insurance motive with reducing credit spreads over time using shorter-term contracts. To quantify demand for long-term contracts, I develop a life-cycle model of optimal mortgage fixation choice. With baseline house price growth and interest rate risk, households prefer shorter-term contracts at high LTV levels, and longer-term contracts once LTV is sufficiently low, in line with the data. The findings help explain reduced and heterogeneous demand for long-term mortgage contracts.

JEL classification: D15, E43, G21, G22, G5, G52
Keywords: mortgage choice, house prices, collateral, credit risk, interest rate risk, household risk management, household finance

Chapter 1

The Demand for Long-Term Mortgage Contracts and the Role of Collateral

1.1. Introduction

Long-term contracts offer households protection against repricing when fundamentals change (**??**). The most important financial contract based on its weight in household balance sheets is the mortgage, with loan repayment over around 30 years. Yet the period over which households fix their mortgage rate is typically much lower, between two to five years in Canada, Australia and most European countries including the UK. Such short fixation periods imply frequent repricing and exposure to different sources of risk when the fixed rate expires, most prominently the risk of increasing aggregate interest rates.[1] The length over which mortgage rates stay fixed is hence an important dimension of household contract choice, but has not been studied explicitly thus far. The paper aims to fill this gap.

To provide intuition, a basic insurance framework suggests that risk-averse households prefer one long-term mortgage contract with no repricing risk, to rolling over two short-term mortgage contracts with a zero-mean risk in mortgage payments, with the same expected cost. In order

[1]The paper abstracts from mortgage contract choice trade-offs with inflation risk, which is studied in **?**. Fixed-rate mortgages are implicitly treated as inflation-indexed, as in **?**.

to evaluate this prediction empirically, I employ granular UK administrative data. The paper generates three main findings. First, it documents a novel fact: the share of relatively long-term mortgages is decreasing in the loan-to-value (LTV) ratio, a measure of credit risk, meaning that riskier borrowers with smaller down payments insure less against interest rate risk. Second, the paper proposes a mechanism to explain this fact, by taking into account the life-cycle dimension of credit risk in the mortgage market: over the life of the loan, the loan-to-value ratio typically declines and thus collateral coverage improves, due to principal repayment and house price appreciation. When considering a longer fixation length, borrowers trade off their regular demand to insure against repricing, to obtaining a lower credit spread over time by repricing more frequently. This raises the opportunity cost of insurance against interest rate risk for high-LTV borrowers, who would lock in large initial credit spreads.

Can the relative insurance cost justify the lack of insurance take-up at high levels of LTV? As a third step, to quantitatively evaluate this trade-off, I build a life-cycle model of optimal mortgage fixation choice. The model allows me to evaluate the net insurance benefit of longer-term contracts by varying households' available contract choice sets. The model suggests that borrowers at high LTV prefer shorter-term contracts under standard calibrations for house price growth and risk, and income and interest rate risk. They prefer longer-term contracts once their LTV is sufficiently low. The insurance value of longer-term contracts, measured as a consumption certainty equivalent, is around 2 to 3 times smaller for borrowers at high LTV, compared to borrowers at low LTV. The paper hence proposes a mechanism for demand heterogeneity and reduced demand for longer-term contracts in high-LTV segments of the mortgage market, which may help explain the prevalence of relatively short mortgage fixation lengths across most countries.

I exploit the UK mortgage market setting as an ideal laboratory for mortgage fixation length choice. Mortgage contracts in the UK allow households to explicitly choose the length over which the mortgage rate stays fixed, separately from the period over which the mortgage is repaid. The choice of fixation length is difficult to isolate in the frequently-studied US market because it is confounded by the simultaneous choice of repayment window, e.g. a 15-year fixed-rate mortgage has a fixed rate for 15 years, but is also repaid over 15 years, giving rise to other

choice factors over and above the insurance choice against repricing. Second, the UK allows me to study credit risk as an important pricing factor in mortgage contracts. While high-LTV mortgage issuance is common in both the US and UK, prices reflect public credit risk guarantees provided by government-sponsored entities in the US (**?**), rather than market prices of credit risk. Lastly, the UK market structure is representative of most of the world's largest mortgage markets,[2] where the fixed rate resets to a more expensive floating rate at the end of the fixation period. These rate resets provide regular economic incentives to refinance into new fixed-rate contracts. UK mortgage rates are typically fixed for two to five years, similar to countries such as Canada, Australia and Ireland, and allow me to study frequent household fixation choices.[3]

To study contract pricing and household behavior over time, the paper utilizes two datasets provided by the Financial Conduct Authority (FCA), comprising the universe of UK residential mortgage originations, and stock of all outstanding mortgages. The full origination data is used to study contract pricing and choice. In addition, I build a panel dataset for first-time borrower cohorts between 2013 and 2017 to track contract choice and loan performance for these borrowers over time.

In simple descriptive analysis, the loan-to-value ratio plays an important role for fixation length choice. A borrower at 95% LTV is between two to three times less likely than a 70%-LTV borrower to take out a 5-year fixed-rate contract compared to a 2-year contract. LTV remains the strongest cross-sectional predictor of 5-year fixed-rate contract choice when controlling for other characteristics such as the loan-to-income ratio, borrower age, loan size and loan maturity.[4] In the UK, LTV is the main dimension along which credit risk is priced.[5] The loan-to-value ratio is an inverse measure of collateralization of the mortgage. The higher the

[2]Including Canada, Australia, Germany, Italy, Spain, Sweden, Denmark and Ireland.

[3]Throughout the paper, I refer to a long-"term" contract as a contract with a relatively longer fixation period, to indicate the interval between repricing. I focus on the most prevalent mortgage fixation lengths of two and five years, which make up around 90% of the UK market. UK variable-rate mortgages also feature rate resets, but take-up is very low, around 4%, over the sample period.

[4]There are no restrictions on the supply side to offer certain fixation lengths, and there are 3.6% of mortgages with fixation lengths greater than 5 years over the sample period, but which are mostly concentrated at low levels of LTV, shown in the online appendix. The pattern is also replicated in the sub-sample of borrowers who originate their loan directly, rather than via a broker.

[5]Lenders have "full recourse" in the UK, meaning they can recover losses from defaulted borrowers though their assets and incomes for up to seven years, until the debt is paid (**?**), which may help explain why measures of household-specific creditworthiness such as the FICO score are only accounted for via a minimum threshold at loan application, but result in no price variation conditional on LTV (as shown by **?**).

LTV, the greater the loss in case of default, as the value of the house which can be seized by the lender, relative to the outstanding loan amount, decreases. Lenders charge a credit spread that is increasing and convex in the LTV ratio in the region between 70 and 95% LTV, i.e. mortgage rates are increasingly "collateral-sensitive" for an LTV above 70%. This credit spread is sizeable: For instance, the mortgage rate at an LTV of 95% is on average 220 basis points higher than at 70% LTV over the sample period.

Tracking borrower outcomes over time, I show that there is a trend decline in the loan-to-value ratio over the life of a typical mortgage loan. The numerator, the loan balance, decreases due to principal repayment, while the denominator, the value of the house, increases given positive house price growth. This implies a life-cycle dimension to credit risk in mortgage contracts, as expected losses from the perspective of the lender, and hence credit spreads, decrease over time. When households choose how long to fix their current mortgage rate, the credit spread component plays a larger role for borrowers at high LTV. High-LTV borrowers can price in lower levels of LTV over time by repricing via shorter-term contracts with successively lower credit spreads. I find that households lock in similar credit spreads regardless of the fixation length, meaning that households cannot obtain these credit spread reductions ex ante. This raises the opportunity cost of the longer-term contract for high-LTV borrowers.

I illustrate the credit repricing effect on the relative cost of longer-term contracts by borrowing from the literature on government bonds (**?**). As a direct comparison, households can compute the expected yield difference between the longer-term mortgage contract, compared to rolling over a sequence of short-term contracts. For low-LTV borrowers, the yield difference reflects the standard bond term premium, the term premium pertaining to the riskless interest rate.[6] For high-LTV borrowers, the life-cycle dimension of credit risk raises the relative cost of the longer-term contract. Given a standard loan repayment path and calibrated expected house price growth of 2.6 percent per annum, I find that the rate on a 5-year 95% LTV contract held over 5 years would have to be 69 basis points lower to be equivalent in expected cost to a 2-year contract sequence. Market prices for long-term mortgage contracts are hence relatively

[6]I indeed find that this measure tracks banks' funding cost spread between longer and shorter maturity interest rate swap rates.

expensive for high-LTV borrowers.

Existing work has demonstrated that long-term contracting can be affected by market imperfections, across a range of different markets including long-term care insurance and unsecured credit markets (e.g. **??**). This paper proposes a novel explanation that reduces long-term contracting in the mortgage market, where key risks are systematic.

Due to the life-cycle trend in LTV, lenders would have to price in a forward-looking path of collateral, and implicitly bear the risk of future house price developments over the fixation horizon of the longer-term contract, in order to make the longer-term contract attractive to high-LTV borrowers. The findings are consistent with lenders requiring compensation for a systematic source of risk, and the lack of financial instruments available to hedge aggregate house price risk, as observed by **?** and **?**. The findings also provide a general intuition for why the initial insurance benchmark does not hold, as long-term contracts in markets with systematic risks may command a premium above the expected cost of the short-term contract sequence.

I find less direct evidence for information frictions. Borrowers may strategically select into fixation lengths (**???**) if they have private information about future repricing risks, which could be more severe at higher LTV bands. I find limited evidence for net adverse selection into longer-term contracts, in particular not at high levels of LTV. I find that ex ante measures of risk such as local house price betas are weakly negatively correlated with 5-year take-up. Ex post default rates within a given LTV band are similar across contract types, with the caveat that the sample window reflects a time period with relatively stable house price growth and low overall rates of default. A related long-term contracting problem is selective household attrition over time (**???**): Households who receive better shocks ex post can leave the borrower pool over time, such that lenders retain an adversely selected pool. In contrast, I find that attrition is minimal over the initial fixation length, due to significant prepayment penalties that penalize early contract termination within the fixation window.[7]

[7]Mortgages in the UK have prepayment penalties of about 3 to 5% of the loan value throughout the initial fixation period (see e.g. https://moneyfacts.co.uk/mortgages/).

In order to quantitatively evaluate contract choices and quantify the insurance value of longer-term contracts net of cost, I build a life-cycle model of optimal mortgage fixation choice. Throughout the life of the loan, households optimally choose between two contract fixation lengths, a novel contribution to existing models of mortgage choice.[8] Depending on the fixation length chosen, households face repricing based on realized loan-to-value ratios, driven by shocks to house prices and regular loan repayment, and shocks to aggregate interest rates. Households also optimally choose consumption over the life cycle, and face income risk. In the model, households trade off their regular demand to insure against repricing, against obtaining a lower credit spread over time by repricing more frequently.

I then use the model to quantify the marginal welfare benefit of adding a longer-term contract to the choice set, as a standard consumption certainty equivalent. I find that high-LTV households, evaluated at 90% LTV, have a reduced willingness-to-pay for 5-year fixed-rate contracts, compared to low-LTV households at 70% LTV. Under a baseline calibration for income, interest rates and house prices, the marginal insurance value of longer-term contracts is 0.36% of annual consumption for high-LTV borrowers, around half of that for low-LTV borrowers. I also evaluate a counterfactual 10-year fixed-rate contract, where the relative willingness to pay is lower, 0.74% for high-LTV borrowers, amounting to about a third of the 2.03% for low-LTV borrowers, meaning that take-up of high-LTV borrowers is relatively lower compared to low-LTV borrowers. When simulating contract choices, I find a life-cycle pattern in mortgage fixation choice, as optimal fixation length is decreasing in LTV. Households are increasingly likely to take out 5-year fixed-rate contracts over the life of the mortgage as LTV decreases over time, in line with the data.

The model suggests that there would be substantial welfare gains for high-LTV households if they could access longer-term mortgages that lock in the base interest rate while adjusting for the trend in credit spreads.

The model findings also imply that heterogeneity in initial LTV levels alone can generate differential long-term contract take-up, for households who are otherwise identical. The observed

[8] These have focused on the 30-year fixed-rate contract in the US and hence do not feature contemporaneous choice throughout the 30-year repayment window, see e.g. **??**.

credit spread levels can plausibly generate reduced demand for long-term contracts for high-LTV borrowers, under a baseline calibration of house prices and interest rates. Thus the mechanism may help explain the missing (or very small) markets for even longer mortgage fixation lengths beyond 5 years with high levels of LTV.

The paper proposes a mechanism that reduces demand for longer-term mortgage contracts and hence risk-sharing in high-LTV segments of the mortgage market, which is relevant for the continuing policy debate on optimal mortgage contract and market design (**?????**). In the US, such risk-sharing is done via public credit risk guarantees by government-sponsored entities, and may help explain why it is one of the few mortgage markets in the world where high-LTV borrowers take out 30-year fixed-rate contracts.

The results are further important from a monetary policy perspective. The relative cost of long-term contracts for high-LTV borrowers influences the length over which their mortgage rates are locked in, and hence the monetary transmission mechanism (**???**). The paper suggests that any mortgage policy interventions should be state-dependent: Having high-LTV borrowers take out shorter-term fixation lengths may improve pass-through when interest rates decrease, but may be more concerning from a financial stability perspective at a time of rising interest rates, as these high-LTV borrowers are less insured against interest rate rises.

1.1.1. Related Literature

The paper contributes to several strands of literature. Household choice of mortgage fixation length is an important household risk management problem. This paper adds to existing work which has focused on the US institutional framework (**?**) and the choice between 30-year fixed and adjustable-rate mortgages (**??**), and the trade-offs related to interest rate and inflation risks, but not interest rate and credit risk. The most conceptually related papers are by **??** and **?** who study the risk-sharing effects of enforcing commitment with long-term mortgage contracts via prepayment penalties in the US context. My findings suggest that pooling in longer-term contracts in the high-LTV segment may be difficult to sustain under market pricing of credit risk, even with binding prepayment penalties, as long as the expected cost of the longer-term

contract is above that of a sequence of shorter-term contracts.

The paper highlights the role of collateral and credit repricing on mortgage contract choice, where expected house price growth and reductions in LTV can override household insurance demand against interest rate risk. The findings support research emphasizing the role of house prices for household behavior in the mortgage market (**???**). In the paper, house prices affect credit repricing risk and mortgage borrowing cost, consistent with a collateral channel.[9] House price growth matters most for high-LTV borrowers, whose borrowing cost are the most sensitive to collateral.

The paper evaluates the willingness-to-pay for longer-term mortgage contracts using comprehensive micro-data and a life-cycle model of mortgage fixation choice. The goal of the model is to quantify why households may not have insurance demand given prevailing prices, and is thus complementary to papers that study optimal mortgage contract features when prices are endogenous (**???**). I use the model to estimate counterfactual demand for non-traded contracts, building on influential work on mortgage choice by **?**, and similar approaches in other insurance markets (see e.g., **?**), to helps explain the lack of take-up at even longer mortgage fixation lengths.

Lastly, the findings relate to a broader literature on long-term contracting and contract choice given dynamic repricing risks (**???????**).[10] Previous papers have studied the front-loaded nature of pricing (e.g. **?**) to overcome dynamic contracting problems, as well as the pricing of callable bonds, i.e. bonds that can be prepaid (**?**). In the mortgage market setting with prepayment penalties, I show that the pricing of house price risk raises long-term contract cost for high-LTV borrowers, despite effective commitment over the fixation horizon. This likely reflects the importance of house price risk and lender willingness to bear a source of systematic risk (**??**) in this market. As a result, households appear to bear much of the house price risk in the market, pointing to novel trade-offs in mortgage market design (**?**).

[9] **?** find that the collateral channel matters most for housing wealth effects: they find that principal reductions that leave households underwater, i.e. in the collateral-insensitive pricing region, do not reduce default, while short-term payment reductions do.

[10] Also referred to as rollover risk in corporate finance (e.g., **???**), and reclassification risk in insurance markets (e.g., **??**).

This paper is organized as follows. Section **??** introduces the institutional framework and data. Section **??** outlines the empirical analysis. Section **??** develops the model and discusses results, and Section **??** concludes.

1.2. Institutional Setting and Data

This section provides background on the UK mortgage market and the typical mortgage contract structure, and provides a brief summary of the data.

1.2.1. UK Fixed-Rate Mortgages and Institutional Setting

Fixed-Rate Mortgage Contract Structure. The dominant mortgage product in the UK is a fixed-rate contract[11] that automatically resets to a so-called "revert rate" at the end of the initial fixation period, for the remainder of the loan maturity, unless the borrower refinances into a new contract. Figure **??** in the online appendix illustrates the payment profile, further explained in the following. The initial fixed rate is in place for a period of typically two or five years – 2 and 5-year fixation lengths together account for 87% of all fixed-rate contracts. The revert rate can be thought of as a penalty rate and is priced at a relatively large spread to the floating base rate, the Bank of England's Bank Rate. The revert rate spread is around 300 to 400 basis points between 2009 and 2017. The rate reset provides a regular economic incentive to refinance into another fixed rate contract,[12] in which case the contract is repriced.[13] 80 to 90% of first-time buyers refinance within six months of the reset date, consistent with findings by **?**, as the loan balance in the initial years since loan origination is usually sufficiently high to warrant refinancing. As a result, the paper focuses on typical refinancing behavior where

[11]The paper focuses on fixed-rate as opposed to floating or adjustable-rate mortgages, but the contract structure is analogous in adjustable-rate mortgages which reset at regular time periods. In the UK, adjustable-rate mortgages feature an initial spread over a floating base rate, which can reset to a larger spread after an initial discounted period and hence provides similar incentives to reprice intermittently. Over the sample window, the share of floating-rate mortgages is very low, at about 4% of all mortgages originated.

[12]Depending on other factors that affect optimal refinancing such as loan size, the interest rate incentive, and the cost of refinancing (**??**). **?** show that the expected option value of staying on the revert rate is only positive under a counterfactually volatile interest rate process.

[13]In the US mortgage market, this type of product would typically be referred to as a "hybrid" adjustable-rate mortgage (ARM).

households choose a new fixed-rate contract after the existing fixed-rate contract expires.

Mortgage Contract Characteristics. Mortgage contracts feature the following characteristics: the maturity over which the loan is repaid, most commonly between 25 and 35 years; the mortgage interest rate; initial fixation period; the rate type over the initial period (fixed or floating); and prepayment penalty due if households prepay and terminate the contract within the initial fixation period. The UK contract structure allows households to choose the fixation length separately from the loan repayment term, which is not possible for the 30-year fixed-rate contract (which is repaid over 30 years), or 15-year fixed-rate contract (which is repaid over 15 years) in the US. The UK market thus provides an ideal setting to study household fixation length choice, separately from the repayment term, which is likely affected by other considerations over and above the marginal insurance choice against repricing. Mortgage interest rates differ by rate type, loan-to-value ratio, and rarely borrower type (first-time borrower, home mover or refinance), but not other borrower characteristics. Borrowers go through an approval process where lenders screen applications using minimum criteria related to the current age, age at loan repayment, loan maturity, loan-to-income ratio, credit score and credit history. Subject to passing these lender criteria, risk-based pricing is focused on the LTV dimension (as shown by **?**), in contrast to the US, which features variation in mortgage guarantee fees along the LTV and FICO score dimensions (**?**). UK mortgage rates are priced across LTV bands in steps of five percentage points, starting from an LTV pricing threshold of 60 to 70%, up to an upper bound of 95% LTV.[14] Offered rates apply across the UK and are not further personalized (**?**).

Repricing. At each refinance, mortgages are repriced in two dimensions: the prevailing mortgage rate at 60 to 70% LTV (which is affected by prevailing aggregate interest rates), and a credit spread adjustment depending on the current level of LTV. Most households do not face repricing due to household-specific creditworthiness and income, as lenders typically do not carry out new credit or affordability checks for their existing customers[15] (**?**). LTV is typically

[14]The resulting step function pricing schedule can be verified in posted prices on offer, as well as realized interest rates (see e.g. **?**).

[15]Around two thirds of mortgages are refinanced with their existing lender, while the remaining third change

adjusted based on changes to local property prices, rather than external re-appraisals.[16]

Prepayment Penalties. Mortgage contracts feature prepayment penalties in case of early repayment (so-called "early repayment charges"), but which only apply for prepayment within the initial fixation period, and vary from around 3 to 5% of the outstanding loan balance. Prepayment terms are not collected as part of the administrative mortgage data, but research using complementary data by a private data provider on the universe of mortgage contracts on offer shows that they are fairly stable and do not vary systematically over time or across lenders (**?**). I verify in the online appendix (section **??**) that the resulting contract structure incentivizes households to commit to the initial fixed rate over the initial fixation period, with a negligible share of households refinancing before the end of the initial fixation period. In addition, most mortgages in the UK are portable, meaning households can transfer a similar-sized mortgage to a different house.[17] Portability implies that a longer fixation length does not lock households in (and require them to pay the prepayment penalty) in case they are moving.

1.2.2. Dataset Construction

Data Overview. This subsection describes the data and provides a brief overview of the main dataset construction. A more detailed description is provided in the online appendix (section **??**). The main data source is the Product Sales Database (PSD), a comprehensive loan-level dataset on residential mortgages in the UK, collected by the Financial Conduct Authority (FCA), and accessed via a data-sharing agreement with the Bank of England. The data comprise the universe of new loan originations at quarterly frequency (PSD001), and also track the stock of all outstanding mortgage loans issued by all regulated financial institutions in the UK at semi-annual frequency (PSD007). The datasets have been used in a range of academic studies (e.g. **??????**). I use both the PSD001 loan origination data from January 2013 to December 2017, and a merged subset of the data that combines the information at

lenders, as further studied by **?**.

[16] Exceptions could be made for instance if households intend to extract home equity (see e.g. **?**).

[17] See e.g. https://www.forbes.com/uk/advisor/mortgages/porting-a-mortgage/.

loan origination with the stock of all outstanding mortgages between 2015 and 2017 in the PSD007 data, which further includes information on refinancing status, interest rate paid and loan performance reported in semi-annual snapshots. The merged data forms a borrower-level panel that is tracked at semi-annual frequency.

Data on New Mortgage Originations (PSD001). The dataset collects detailed borrower characteristics such as income, age, address, loan amount, property value, and detailed loan characteristics such as the loan maturity, interest rate, fixed-rate window, and which lender originated the mortgage. I use the origination data between 2013Q1 to 2017Q4 for the pricing analysis and results that do not require the borrower panel dimension, containing around 2.9 million loans. I further use the data to identify first-time buyer cohorts who newly originate their mortgage between 2013H2 to 2015H1 to create the borrower panel. The origination data prior to 2015H1 does not require to report the fixed-rate window, so I do not observe the fixed-rate window for about 40% of first-time borrowers in this time period. The sample for which fixed-rate windows are observed appears to be a highly balanced sample compared to where it is not observed, as noted by **?** and demonstrated in Table **??** in the online appendix. The sample for which the fixed-rate window is observed contains 414,643 first-time borrowers.

Data on the Stock of All Outstanding Mortgages (PSD007). The stock data contains information on the current interest rate type, current interest rate paid, current loan amount, current lender, and whether the loan is in arrears. I create the borrower panel by merging first-time buyer cohorts from the origination data with stock data waves 2015H1 to 2017H2, to track refinancing behavior and outcomes over time.

Data Merge. The origination and stock datasets do not have unique borrower identifiers, but a borrower can be identified up to an (anonymized) date of birth and six-digit postcode, which is approximately the building block in which a UK household resides. The merge using this borrower identification is almost comprehensive, with only 1.8% of first-time borrowers in 2015H2

not matched to the stock data in 2015H2, which provides an estimate of unmatched observations driven by pure merging error. This results in a panel of around 2.8 million borrower-half-year observations. Lastly, I supplement the merged dataset with administrative data on UK house prices from HM Land Registry, using house price indices at local-authority administrative unit-level (with data going back to 1995), and merging these at the local-authority level based on borrower location.

1.3. Mortgage Contract Choice and the Life Cycle of Credit Risk

This section illustrates empirical patterns in household fixation length choice and the role of the loan-to-value ratio. It introduces three findings. First, the probability of choosing a 5-year, relative to a 2-year fixed-rate mortgage contract is decreasing in LTV. Second, there is a life-cycle dimension of credit risk in the mortgage market due to loan amortization and positive house price growth, reducing the LTV ratio, and improving the collateralization of the loan over time. Third, the downward trend in expected credit risk and hence credit spreads raises the expected cost of a 5-year fixed-rate contract, compared to a sequence of 2-year fixed-rate contracts, where credit spreads are repriced over time.

1.3.1. Mortgage Fixation Length Choice and the Role of LTV

Which households choose a 5-year, rather than a 2-year fixed-rate contract? I estimate a linear probability model for choosing a 5-year fixed-rate contract relative to a 2-year contract, based on a range of household and loan characteristics. The dependent variable is an indicator that takes the value 1 if a household chooses a 5-year contract, and 0 if the household chooses a 2-year contract.[18] The regressors comprise the LTV ratio, the LTV interacted with an indicator variable if the LTV is greater equal to 80%, the loan-to-income (LTI) ratio, borrower age

[18]These account for 87% of all contracts. The excluded contracts have a fixation length of 1 year or less (2% of contracts), 3 to 4 years (7%), and 6 years or more (4%), but results are very similar for an analysis of less than and greater equal 5-year fixed-rate contracts and including these contracts.

(linear and square term), loan term, local house price growth, house price volatility and house price beta, and are converted to standard deviations of the variable.[19] Results are reported in Table **??**. Column (1) reports the baseline results, controlling for time (year-month), local-authority×time, lender×time, and borrower-type (first-time buyer, home mover, refinance), loan decile, and sales channel fixed effects (direct/online or intermediated via a broker).

Households who choose a 5-year contract relative to a 2-year contract tend to have a lower LTV, and lower mortgage maturity. Measures of ex ante risk are ambiguous, households who choose 5-year contracts experienced somewhat lower local house price growth, but also live in places with lower house price beta. 5-year contract choice features an inverse-u-shaped relationship with age, as the linear coefficient on age is positive (but statistically indistinguishable from zero), but the squared coefficient is negative.

LTV emerges as the most quantitatively important predictor of cross-sectional choice in the data, particularly at high levels of LTV. For a borrower with an LTV greater than 80%, a one standard deviation increase in LTV (around 10 percentage points) reduces the probability of choosing a 5-year fixed-rate contract by around 18%, or more than half of the average probability of choosing a 5-year fixed-rate contract. The results hold in each time period or controlling for year-month fixed effects, meaning that high-LTV borrowers are always less likely to choose 5-year fixed-rate contracts compared to low-LTV borrowers, even while average 5-year share levels fluctuate over time due to variation in expectations of the path of future interest rates.

While LTV could be correlated with the degree of household financial constraints, the effect of the loan-to-income ratio, a direct measure of the life-time borrowing and monthly payment constraint, goes to zero when controlling for high levels of LTV. Another way to illustrate the importance of LTV is to compare the adjusted R^2 for univariate predictive regressions of contract choice, reported in online appendix Figure **??**, with LTV having the highest univariate adjusted R^2. The overall adjusted R^2 for the choice regressions is between 0.14 and 0.15, which

[19]Local house price growth is measured as growth two years prior to the choice of contract, local house price volatility is computed as the rolling 10-year volatility in log house price returns, while house price beta is computed as a rolling 10-year beta of local house price returns with respect to aggregate UK house price returns. "Local" refers to a local-authority-level of aggregation which is a typical administrative unit in the UK, with 415 local authorities with an average population of around 200,000, similar to counties in the US.

is moderate but large compared to existing analysis on mortgage choice (see e.g. **?**).

The results also suggest ambiguous ex ante selection patterns. While households who choose a 5-year fixed-rate contract experience slightly lower house price growth, which could be a measure of lower expected house price growth in the future, other ex ante measure of risk such as local house price volatility and local house price beta prior to contract choice are insignificant or even negative. The predictive effect of local house price growth on contract choice is also relatively small, as reported in Figure **??**. Ex post measures of risk such as realized LTV and default are studied further below.

Zooming in on the effect of LTV, Figure **??** shows the coefficient of LTV band on 5-year fixed-rate contract choice, with aforementioned covariates partialled out.[20] At an LTV of 70% or lower, contract choice is roughly split, with 58% of borrowers choosing a 2-year contract, and 42% of borrowers choosing a contract with a fixation period of 5 years. The 5-year contract share decreases consistently across LTV bands, with only 18% of borrowers at an LTV of 90-95% choosing a 5-year fixation window, with the raw numbers corresponding to 44% and 16%, respectively. A borrower at 95% LTV is thus around two to three times less likely to take out a 5-year fixed-rate contract, compared to a borrower at 70% LTV.

A concern may be that the relationship between LTV and contract choice is driven by an unobserved variable that determines both the choice of fixation length, and household LTV, such as variation in risk aversion. Less risk-averse households may have greater LTV ratios, and less demand for insurance against repricing via 5-year fixed-rate contracts. I can use the panel of first-time borrowers to study the effect in repeated contract choice. Table **??** reports the linear probability analysis, focusing on the effect of LTV. The coefficients on LTV and LTV interacted with an indicator for an LTV greater equal to 80% are similar, but a bit smaller than in the full origination data.[21] Column 2 estimates the specification with borrower fixed effects, meaning that identification comes from changes in LTV between repeat choices, differencing

[20] The pattern in the raw data is very similar and reported in online appendix figure **??**.

[21] This could be due to the fact that the indicator captures an initial LTV of greater equal to 80%, and so subsequent contract choices are done at lower levels of LTV. In addition, current LTV is reported at origination, but is a derived variable for repeat choices and so may attenuate coefficient values, as loan balances are reported in the panel data, but house prices are based on local house price changes.

out any borrower-specific time-invariant unobservable factors. Column 3 conditions on the set of households for which the first choice of fixation length is 2 years, and measures the effect of LTV changes on subsequent choices. Both specifications suggest that an increase in LTV significantly reduces 5-year contract choice probabilities for high-LTV borrowers.

Another potentially unobserved variable is future income growth. Using UK survey data **?** finds that future income growth is most closely correlated with the current loan-to-income ratio, which I can control for. In addition, **?** use a longer panel of repeat refinancers in the UK and show that realized income growth for these borrowers is not correlated with fixation length choice.

Lastly, there is limited evidence that the effect of LTV on contract choice is driven by selective supply by lenders or steering. I can split the sample by sales channel, whether the mortgage was sold directly or via a broker, and the decreasing pattern in 5-year fixed-rate choice across the LTV distribution persists in both. Market concentration as measured by a Herfindahl-Hirschman index is similar across fixation length and within a given LTV band. While it is difficult to observe the true choice set from which households can choose, I submitted queries through the UK regulator (the Prudential Regulation Authority, or the PRA) to the six largest lenders to confirm that lenders do not offer different fixation lengths to borrowers in different LTV bands.

To summarize:

Finding 1. *The probability of choosing a 5-year relative to a 2-year fixed-rate mortgage contract is decreasing in LTV. The effect is robust to controlling for a range of other variables including the loan-to-income ratio, loan size, loan term and sales channel.*

In order to quantify the choice mechanism separately from any selection effects, I will turn to a model in section **??**.

1.3.2. Mortgage Pricing

Why does LTV matter for fixation length choice? In the UK, LTV is the main pricing measure for credit risk and so determines the credit spread that is locked in as part of the mortgage rate.[22]

Mortgage rates are increasing and convex in LTV bands. Figure **??** illustrates the pricing of LTV by showing the credit spread paid on 5-year fixed-rate mortgages across loan-to-value (LTV) bands, extracted as LTV-band fixed effects from a regression of interest rates on LTV bands, controlling for a range of fixed effects, with further detail provided in the online appendix.[23] Below the lower LTV threshold of 70% LTV, interest rates typically do not vary with changes in LTV, which I refer to as the "collateral-insensitive" pricing region. Starting from an LTV of 70% LTV, mortgage rates become increasingly sensitive to the level of LTV and rise in LTV bands of 5 percentage points, up to the highest LTV band of 90-95% LTV, above which very few mortgages are originated and households pay the revert rate. The region between 70% to 95% LTV can be thought of as the "collateral-sensitive" mortgage pricing region. The LTV credit spread is sizeable, and reaches on average 220 basis points for a borrower at 95% LTV, compared to a borrower at 70% LTV over the sample period. Importantly, credit spreads for 2-year fixed-rate contracts are very similar to 5-year fixed-rate contracts (reported in the online appendix), meaning that households essentially lock in the LTV credit spread at contract origination, regardless of the subsequent fixation length, which makes up a larger proportion of the mortgage rate for high-LTV borrowers.

[22] To illustrate the main drivers of cross-sectional variation in UK mortgage rates, I regress observed mortgage rates on a range of fixed effects, including time, lender, buyer type, fixation length, and all interaction effects. Figure **??** in the online appendix reports the adjusted R^2 values from these regressions. When comparing the inclusion of different household covariates, the marginal increase in R^2 is highest when including the LTV band fixed effects, with the adjusted R^2 rising from about 55% to around 85%, while the inclusion of income and age deciles only leads to an increase of a few percentage points, consistent with earlier analysis by **??**, who also confirms that mortgage rates do not differ across FICO scores. FICO scores are only used as a minimum threshold for loan approval, in contrast to the US market setting.

[23] Raw averages of mortgage rates across LTV bands yield very similar results, and results are robust to different fixed-effect specifications.

1.3.3. Credit Repricing and the Life Cycle of Credit Risk in Mortgages

Long-term contracts protect against repricing in general, including along the credit dimension. The risk of credit repricing has been studied in the corporate finance literature. **?** shows that firms who are already very risky may have a higher chance to improve their credit risk rather than deteriorate further over time. The empirical findings are more ambiguous. **?** and **?** show that credit spreads of riskier firms decrease over time, while **?** show that they increase over time. In comparison, a residential mortgage loan to households is a specific type of collateralized loan, with the value of the house serving as collateral that can be seized in case of default, and full loan repayment over time. This implies that there is a downward trend in the measure of credit risk, the LTV ratio, as the loan balance, the numerator, decreases, and the value of the house, the denominator, increases with some positive house price growth. A mortgage loan hence becomes typically less risky over time from the perspective of the lender.

I find evidence for this life-cycle dimension of credit risk in LTV ratios and mortgage rates in the data. Figure **??** shows LTV ratios (Panel A and B) and mortgage rates (Panel C and D) for different cohorts of first-time borrowers (between 2013 and 2015), tracking their outcomes over time, between 2015 to 2017. Panel A illustrates the average decline in LTV for borrowers who start with an LTV of 75%, who reduce their LTV by around 10 percentage points over two years. Since mortgage rates are virtually collateral-insensitive below a threshold of 70%, there is only a small associated decrease in credit spreads and interest rates, as shown in Panel C. In contrast, for borrowers who start with an LTV of 90%, the same percentage point reduction in LTV (Panel B) is associated with large decreases in credit spreads and mortgage rates (Panel D), as LTV changes are in the collateral-sensitive pricing region.[24] While each cohort captures different aggregate interest rate levels, there is a significant reduction in credit spreads for each cohort after two years, as many borrowers refinance and reprice at that point.

The data reflect one particular realized house price path, with relatively strong house price growth of 4-5% per annum over the sample period, and so provides an ex post view of realized credit repricing. In order to get an ex ante measure of risk which is relevant for household in-

[24] These findings are consistent across the LTV distribution.

surance choices, I can also simulate the distribution of LTV and credit spreads using a baseline calibration for house prices. As described in more detail in the model calibration, I calibrate a house price process with average house price growth $\mu_h = 0.0258$ and standard deviation $\sigma_h = 0.077$. The distribution of LTV and credit spreads over time is shown in Figure **??**, for simulated households with a starting LTV of 90%. While there is some upward repricing risk that LTV increases above 90%, there is a strong downward trend such that the probability of upward repricing is less than 5% after 10 years, and the median household is in the collateral-insensitive pricing region below 70% LTV after approximately 7 years. Note that I can also simulate alternative house price scenarios and loan repayment patterns. With lower house price growth and higher house price volatility, the risk that credit spreads rise is more material, and I can evaluate these alternative scenarios as part of the model analysis in section **??**.

Finding 2.1. *There is a life-cycle dimension of credit risk in the mortgage market. Under a baseline calibration of house prices and loan repayment, the median household who starts with a 90% LTV at origination reaches an LTV below 70% after approximately 7 years, leading to a decline in credit spreads over that time.*

1.3.4. Credit Repricing And The Relative Cost of Long-Term Mortgages

The expected decrease in credit risk over the life of the mortgage raises the opportunity cost of longer-term mortgages, as these lock in initial credit spreads for longer. I show that the relative cost can be written as the expected yield difference between the longer-term mortgage and sequence of shorter-term contracts, and can be decomposed into two parts (which is summarized here, and outlined in more detail in the online appendix section **??**): a standard bond term premium, and a premium arising from the expected credit repricing benefit of shorter-term contracts. Intuitively, the latter can be interpreted as an implicit term premium in the credit dimension, in this case the initial LTV level, which raises the cost of longer-term mortgage contracts for riskier borrowers, over and above the bond term premium.

The decomposition can be thought of as an extension of the expectations hypothesis of the

term structure (**?**) for mortgage rates, which depend on LTV, in addition to aggregate interest rates. Under the assumption that mortgages with an LTV below the lower pricing threshold are essentially credit-risk free, the expected mortgage yield difference measured at the lowest LTV band of 70% reflects the standard bond term premium. In the data, I indeed find that this measure tracks the funding cost spread between longer and shorter maturity interest swap rates.

Reductions in credit spreads over time imply that longer-term contracts carry an opportunity cost over and above the bond term premium, for mortgages with an LTV above 70%.[25] As an illustrative example, I can calculate the average effect using an assumption for average house price growth, calibrated to be 2.6 percent per annum (using UK data from 1987-2017), and fully-amortizing payments for a loan over 30 years. I can then compute the expected 2-year rate path, using the expected LTV at the time of repricing and associated credit spreads. To offset the decreasing expected 2-year rate path over time, the rate on a 5-year 95% LTV contract held over 5 years would have to be 69 basis points lower than the 2-year rate at 95% LTV.

Finding 2.2. *The downward-trending life-cycle profile of credit risk raises the opportunity cost of a long-term mortgage contract, compared to a sequence of shorter-term mortgage contracts.*

1.3.5. Discussion of Mechanisms

Why is the longer-term mortgage contract not cost-equivalent to a sequence of shorter-term contracts? This subsection discusses why the relative cost of a longer-term contract could be increasing in LTV. Current pricing seems consistent with systematic house price risks being reflected in LTV. There is less direct evidence for information and other contracting frictions.

Pricing of House Price Growth. In order to make longer-term contracts attractive to high-LTV borrowers, lenders would have to price in a forward-looking LTV path, including

[25] There could also be a pure pricing difference which I discuss further in the appendix, e.g. the 5-year 80% LTV mortgage rate could be higher than the 2-year 80% LTV rate.

a forward-looking path for house prices. Thus lenders would implicitly bear risk of future house price developments over the fixation horizon of the longer-term contract. A longer-term contract that is cost-equivalent to a shorter-term contract sequence would insure households against house price risk, in addition to interest rate risk. The expected decline in LTV would be more sensitive to the rate of expected house price growth at higher LTV bands. The fact that longer-term contracts are relatively more costly at high levels of LTV may be consistent with the lack of financial instruments available to hedge house price risk as observed by **?** and **?**, and lenders requiring compensation for exposure to house price risk as a systematic source of risk.

Selection and Screening. Borrowers may strategically select into longer-term contracts if they have private information about future repricing risks, and those with worse future risks may adversely select into longer-term contracts. Lenders may charge a premium on 5-year contracts to screen for that type of selection. In order to test this channel, I evaluate both ex ante and ex post measures of borrower risk. Reviewing the covariates that correlate with 5-year contract choice in the choice regressions in Table **??** suggest that there is limited adverse selection into longer-term contracts based on ex ante observables. Local house price beta as a measure of local house price risk is slightly negatively correlated with 5-year contract take-up.

As a measure of realized risk, I assess ex post default rates that I track in the borrower panel data. Ex post default rates over the sample period are fairly similar across contract types conditional on a given LTV band, shown in Figure **??**. If anything, 2-year borrowers at higher LTV who stay with their lender have a slightly higher ex post default probability compared to 5-year borrowers, with the caveat that the sample window reflects a time period with relatively stable house price growth and low overall rates of default.[26] The finding is consistent with the intuition that one would expect less asymmetric information in a market where the main measure of credit risk is the value of house price collateral, which is considered largely observable due to

[26]Note that ex post default outcomes reflect net selection effects: other factors that affect contract choice, such as financial constraints, could be positively correlated with default, inducing "advantageous" selection into 5-year contracts that may offset adverse selection incentives. Lenders could also use historical data to price default that differs from current default rates.

observable changes in local house price indices.[27] This is in contrast to unsecured credit markets where information asymmetries have been shown to affect contract maturity choice (see e.g., **?**).

Selective Early Prepayment and Adverse Retention. Another canonical long-term contracting problem is selective household attrition over time (**???**): households who receive better shocks ex post can leave the borrower pool over time, such that lenders retain an adversely selected pool, potentially leading to market unravelling à la **?** in a dynamic sense. 5-year fixed-rate contracts could price in this adverse retention relative to 2-year contracts, and this effect may be more pronounced at higher LTV. In contrast, I find that attrition is limited over the initial fixation window (shown in online appendix section **??**), likely due to significant prepayment penalties. Mortgages in the UK have prepayment penalties of about 3 to 5% of the loan value. More detail on household refinancing behavior is also provided in the online appendix.

Summary and Motivation for Model

So far, the analysis has established that the cost of insurance when taking out a 5-year fixed-rate contract is increasing in LTV due the downward-trending life-cycle profile of credit risk. Can the credit risk profile justify reduced longer-term contract take-up? In order to evaluate household contract choices quantitatively, and in order to assess the *net* benefit of longer-term contracts to households given cost, risk aversion and the joint distribution and evolution of risks, I develop a model of optimal mortgage fixation choice in the following section.

[27] UK lenders in fact use changes in local house prices to re-evaluate collateral values for refinances with existing customers, without new credit or affordability checks (see FCA Mortgage Market Study, Interim Report 2018).

1.4. Life-Cycle Model of Optimal Mortgage Fixation Choice

In this section, I develop a partial equilibrium life-cycle model of household consumption and mortgage contract choice. The goal of the model is to quantify household insurance demand for longer-term contracts given prevailing prices. The model allows me to evaluate the net insurance benefit of longer-term contracts, taking into account realistic features of the household choice problem such as income, interest and house price risk, and how these risks evolve over the life cycle. The model also provides a way to separate the effect of selection on contract choice, by evaluating household choices in a setting where households are ex ante identical.

The model features household choice of interest rate fixation length over the life of the loan, and frequent repricing based on realized loan-to-value ratios and aggregate interest rates, as an extension to existing models of optimal mortgage choice. The model matches the mortgage contract structure in the UK which is common in many countries. It hence differs from influential work by **?**, who evaluate mortgage choice in the US market context with 30-year fixed-rate and adjustable-rate mortgages that are held over the life of the loan, with the option to refinance, by introducing frequent repricing and allowing for a flexible contract choice path between two fixation lengths throughout the life of the loan.

1.4.1. Model Setup

Overview. In the model, households optimally choose consumption and mortgage contracts given two different fixation periods until the loan is repaid, and only consumption thereafter. Households have a finite time horizon with a working life, after which they retire and die. For simplicity, I focus on homeowners, and assume the size of the house is fixed.[28] Households buy the house at the beginning of their working life with a mortgage and repay it over the

[28]The model does not endogenize the decision to buy a house or rent, and the choice of the size of the house, which is assumed to be fixed. Hence households are assumed to strictly prefer buying a house to renting and cannot adjust their house size in response to shocks, as in **?**.

maturity of the loan T. Mortgage rates depend on the relative value between the outstanding loan balance and the value of the house price as collateral, i.e. the loan-to-value (LTV) ratio, and the aggregate interest rate (r), at the time when the loan was last repriced. Since the utility derived from the house is fixed, it can be omitted from the household optimization problem.[29]

Utility. Households maximize expected utility with time discount rate δ and discount factor $\beta = \frac{1}{1+\delta}$. Households have constant relative risk aversion (CRRA) utility over consumption:

$$U(C) = \frac{C^{1-\gamma} - 1}{1-\gamma}. \tag{1.1}$$

Dynamic Budget Constraint. Households face idiosyncratic income risk. At each time period, households pay mortgage payment $M_t = L_t \cdot \frac{r_t^m}{1-(1+r_t^m)^{-T}}$, where L_t is the remaining loan balance outstanding at time t. The mortgage interest rate r_t^m depends on the aggregate interest rate r_t^τ plus a time-invariant premium over the base rate ρ^m, which compensates the lender for the cost of issuing a given loan independent of LTV, and the LTV ratio LTV_t^τ locked in at the last instance of repricing (at time $t = \tau$), tracked in superscript:

$$r_t^m = \rho^m + r_t^\tau + f(LTV_t^\tau, \theta^\tau).$$

$f(\cdot)$ is the lender credit pricing function which is increasing and convex in LTV, and which may differ across contract fixation length θ. This component can be thought of as the credit spread for a given level of LTV. The LTV ratio at time t is determined by the outstanding loan value L_t relative to the current value of the house:

$$LTV_t = \frac{L_t}{H_t}$$

[29] This assumption is justified when households have separable utility between housing and consumption (**?**) or CES utility with a unitary elasticity of substitution (**?**).

House prices H_t follow a lognormal distribution, and the change in log house prices is given by

$$\Delta \log H_t = g + \eta_t \tag{1.2}$$

with constant g and an i.i.d normally distributed shock with mean zero and variance σ_η^2. The expected log real return on a one-period bond $r_t = \log(1 + R_t)$ follows an AR(1) process:

$$r_t = (1 - \rho_r)\mu_r + \rho_r r_{t-1} + \xi_t, \tag{1.3}$$

where ξ_t is a normally distributed white noise shock with mean zero and variance σ_ξ^2. Household net wealth X_t evolves according to the following dynamic budget constraint:

$$X_{t+1} = (1 + r_t)(X_t - C_t) - M_t(LTV_t^\tau, r_t^\tau) + Y_{t+1},$$

subject to the borrowing constraint $(1 + r_t)(X_t - C_t) - M_t \geq \bar{B}$.[30] Next period net wealth is net savings compensated at the risk-free rate, less mortgage payments, plus income. Log income $\ln(Y_t)$ has a deterministic component $f(t)$ that is a function of time t, and is subject to transitory shocks ϵ_t. ϵ_t is an i.i.d normal shock with mean 0 and standard deviation σ_ϵ.[31]

Mortgage Contract Choice. Households choose the fixation length θ over which they lock in the current mortgage rate, and hence the point in time at which they get repriced next, i.e. θ periods from when the contract is chosen. A longer θ exposes household less frequently to repricing risk, but overall house price and aggregate interest rate changes accrue over the repricing window and are repriced at the end of the repricing window. Since house price shocks are i.i.d., multi-period house price risk over the duration of the fixation period can be expressed

[30] The model abstracts from the ability to extract home equity, and housing wealth does not enter household utility directly. This would introduce additional variation in the cost of borrowing across the LTV distribution, as this would be captured in the mortgage rate that is increasing in LTV. This may also understate an additional benefit of shorter-term mortgages, as they could provide borrowers with greater flexibility to cash out at shorter time intervals (given costly prepayment penalties over the initial fixation period).

[31] Income shocks are assumed to be i.i.d in order to simplify the problem and economize on state variables.

with mean and variance:

$$E\left(\eta_{t,t+\theta}\right) = E\left(\sum_{i=1}^{\theta} \eta_{t+i,t+i-1}\right) = \sum_{i=1}^{\theta} E\left(\eta_{t+i,t+i-1}\right) = \theta\mu_h$$
$$\text{Var}\left(\eta_{t,t+\theta}\right) = \text{Var}\left(\sum_{i=1}^{\theta} \eta_{t+i,t+i-1}\right) = \sum_{i=1}^{\theta} \text{Var}\left(\eta_{t+i,t+i-1}\right) = \theta\sigma_h^2.$$

In the baseline model, mortgages are assumed to be fully amortizing, i.e. households repay both capital and interest over the life of the loan, and so the loan value L_t decreases over time, i.e. $\Delta L_{t,t+\theta} \leq 0$. Households can choose between a relatively longer-term fixation period θ^{LT}, and relatively shorter-term fixation period θ^{ST}, i.e. $\theta^{LT} > \theta^{ST}$. Once they make a choice, the mortgage rate is locked in over the fixation length chosen, and a new contract can be chosen at the end of the fixation period. In order to economize on state variables, the model assumes that the loan balance can be tracked using time t alone.[32] The model tracks LTV as a state variable, which is then sufficient to track the house price evolution over time.[33]

Value Function and Repricing States. In order to determine optimal mortgage choice (stored in policy function $\mathcal{R}$), the household value function tracks two auxiliary value functions, the value function if the household chooses the short-term contract V^{ST}, which implies repricing in θ^{ST} periods, and new choice of fixation length thereafter; and the value function if the long-term contract is chosen, V^{LT}, with repricing and new choice in θ^{LT} periods. Both take into account that the current rate is locked in and repriced at the end of the chosen fixation window. Rather than tracking the time at which repricing next takes place (τ), repricing depending on contract choice is tracked more parsimoniously by repricing state variables $\mathcal{S}_t^{\theta}$ which take θ states defined as follows: for each fixation window, interest rates are locked in at a given LTV level for θ periods ($\mathcal{S}_t^{\theta} = \theta$), locked in for $\theta - 1$ periods ($\mathcal{S}_t^{\theta} = \theta - 1$), up until when the remaining fixation window reaches 1 period, at the end of which there is θ-period repricing

[32] This is a common and quantitatively small approximation that abstracts from small variations in the loan amortization path due to differences in interest rates (**??**).

[33] The model abstracts from an explicit strategic default decision given the full recourse regime of the UK. Default behaviour is implicitly captured by the household utility maximization problem that avoids states with high mortgage payments (and hence low consumption). In the robustness check with high revert rates, this rate could be considered a penalty rate that serves as a proxy for the expected cost of default, which becomes more likely at an LTV exceeding 95%.

($\mathcal{S}_t^\theta = 1$), for $\theta \in \{\theta^{LT}, \theta^{ST}\}$. V^{LT} and V^{ST} are further defined in the following. The vector of state variables is $\Omega = \{X, t, LTV, r, \mathcal{S}^{\theta^{LT}}, \mathcal{S}^{\theta^{ST}}\}$, representing household net wealth, time, LTV, aggregate interest rate, repricing state for the longer-term contract, and repricing state for the shorter-term contract, respectively. To simplify notation, the θ-superscripts for the repricing state variables are omitted if the information is not required. First, in order to capture the temporal dependence of repricing states for the value function when choosing a contract with fixation window θ, it is useful to note that in the θth-period from choosing the contract, the value function is

$$\begin{aligned} &V_{t+\theta-1}\left(X_{t+\theta-1}, LTV_{t+\theta-1}, r_{t+\theta-1}, \mathcal{S}_{t+\theta-1} | \mathcal{S}_{t+\theta-1} = 1\right) \\ &= \max_{C_{t+\theta-1}, R_{t+\theta-1}} U(C_{t+\theta-1}) + \beta E_{t+\theta-1}\left[V_{t+\theta}^*\left(X_{t+\theta}, LTV_{t+\theta}, r_{t+\theta}\right)\right], \end{aligned} \tag{1.4}$$

with repricing at the end of the period, indicated by $\mathcal{S}_{t+\theta-1} = 1$. Note that the continuation value, with $V_t^* = \max\left\{V_t^{ST}, V_t^{LT}\right\}$, takes into account that the household can optimally choose a short- or long-term contract after this period, and does not depend on the repricing state. Over the fixation length of the contract, the household is protected from repricing, indicated by $\mathcal{S} \in \{2, ...\theta\}$. In $\theta - 1$ periods, the value function is

$$\begin{aligned} &V_{t+\theta-2}\left(X_{t+\theta-2}, LTV_{t+\theta-2}, r_{t+\theta-2}, \mathcal{S}_{t+\theta-2} | \mathcal{S}_{t+\theta-2} = 2\right) \\ &= \max_{C_{t+\theta-2}, R_{t+\theta-2}} U(C_{t+\theta-2}) \\ &+ \beta E_{t+\theta-2}\left[V_{t+\theta-1}^{LT}\left(X_{t+\theta-1}, LTV_{t+\theta-1}, r_{t+\theta-1}, \mathcal{S}_{t+\theta-1} \mid \mathcal{S}_{t+\theta-1} = 1\right)\right], \end{aligned} \tag{1.5}$$

which can be extended analogously for each period up until (and including) the current period. In the current period, the value function for choosing the long-term contract is

$$\begin{aligned} &V_t\left(X_t, LTV_t, r_t, \mathcal{S}_t \mid \mathcal{S}_t = \theta\right) \\ &= \max_{C_t, R_t} U(C_t) + \beta E_t\left[V_{t+1}^{LT}\left(X_{t+1}, LTV_{t+1}, r_{t+1}, \mathcal{S}_{t+1} \mid \mathcal{S}_{t+1} = \theta - 1\right)\right]. \end{aligned} \tag{1.6}$$

For notational simplicity, the value functions for choice of the short-term and long-term contract

are respectively defined as

$$V_t^{ST} \equiv V_t\left(X_t, LTV_t, r_t, \mathcal{S}_t^{\theta^{ST}} \mid \mathcal{S}_t^{\theta^{ST}} = \theta^{ST}\right),$$
$$V_t^{LT} \equiv V_t\left(X_t, LTV_t, r_t, \mathcal{S}_t^{\theta^{LT}} \mid \mathcal{S}_t^{\theta^{LT}} = \theta^{LT}\right). \tag{1.7}$$

Online appendix section **??** further outlines the dependencies across time and repricing states for V^{LT} with $\theta^{LT} = 5$.

The dynamic budget constraint can then be rewritten without τ, and using the repricing state variable instead:

$$X_{t+1} = (1 + r_t)(X_t - C_t) - M_t(LTV_t, r_t, \mathcal{S}_t) + Y_{t+1}. \tag{1.8}$$

Policy Functions. Households choose optimal consumption and the optimal mortgage contract in the model. Let policy function C denote household optimal consumption in state $\Omega = \{X, t, LTV, r, \mathcal{S}^{\theta^{LT}}, \mathcal{S}^{\theta^{ST}}\}$ where $C : \Omega \to [0, \infty)$, and $\mathcal{R}$ denote optimal mortgage choice of either the long-term ($\mathcal{R} = 1$) or short-term contract ($\mathcal{R} = 2$) where $\mathcal{R} : \Omega \to \{1, 2\}$.

Bellman Equation. The resulting Bellman equation for the household problem is

$$\begin{aligned}
V_t\left(X_t, LTV_t, r_t, \mathcal{S}_t^{\theta^{ST}}, \mathcal{S}_t^{\theta^{LT}}\right) &= \max_{C_t, R_t} U(C_t) + \beta E_t\left[V_{t+1}^*\left(X_{t+1}, LTV_{t+1}, r_{t+1}\right)\right] \text{ with} \\
V_t^* &= \max\left\{V_t^{ST}, V_t^{LT}\right\} \\
V_t^{ST} &= \max_{C_t, R_t} U(C_t) + \beta E_t\left[V_{t+1}^{ST}\left(X_{t+1}, LTV_{t+1}, r_{t+1}, \mathcal{S}_{t+1}^{\theta^{ST}}\right)\right] \\
V_t^{LT} &= \max_{C_t, R_t} U(C_t) + \beta E_t\left[V_{t+1}^{LT}\left(X_{t+1}, LTV_{t+1}, r_{t+1}, \mathcal{S}_{t+1}^{\theta^{LT}}\right)\right] \\
\text{s.t.} \quad X_{t+1} &= (1 + r_t)(X_t - C_t) - M_t(LTV_t, r_t, \mathcal{S}_t) + Y_{t+1}, \\
(1 + r_t)(X_t - C_t) - M_t &\geq \bar{B}.
\end{aligned} \tag{1.9}$$

1.4.2. Calibration and Solution

Table **??** provides an overview of the parameters used for the baseline calibration of the model.

House Prices. The house price process is calibrated using aggregate UK house price from 1987 to 2017, with nominal house prices deflated using RPI, yielding an average log house price growth of 0.0258 and standard deviation $\sigma_h = 0.0770$.[34] The initial house price level is set to fit the average loan-to-income (LTI) ratio of borrowers with an initial LTV of 85% in the data, yielding an LTI ratio of 3.56.[35]

Interest Rate. The real log interest rate is calibrated using UK data from 1987 to 2017, with mean $\mu_r = 0.0164$, standard deviation $\sigma_r = 0.0193$ and autocorrelation coefficient $\rho_r = 0.95$. Real rates are calibrated using 5-year UK inflation-indexed gilts, and supplemented using 1-year nominal rates deflated by 1-year ahead survey-based household expectations of inflation.[36]

Income and Borrowing Constraint. Working age is set to 30 to 60, after which households retire and die at age 80. The deterministic hump-shaped component of income over the life cycle is adopted from **?**, following standard life-cycle models (**?**).[37] The standard deviation of the transitory income shock σ_ϵ is set to 0.1 based on the literature.[38] The borrowing constraint $\bar{B}$ is set to 0 and households cannot extract home equity, which likely understates the benefit of shorter-term contracts, as they may give borrowers greater flexibility to cash out at shorter time intervals given costly prepayment penalties over the initial fixation period.

[34] As an alternative, local authority-level house price indices are used to capture cross-sectional variation in house price risk which yields similar magnitudes.

[35] Because all values are standardized in terms of units of permanent income, the loan-to-after-tax-income ratio of 4.82 is used after applying a tax rate of 35.5%, based on 2017/2018 effective UK tax rates.

[36] The autocorrelation coefficient is calibrated using 1-year rates as the preferred implied 5-year real rates are highly persistent with autocorrelation ≈ 1.

[37] The deterministic income profile is based on a simple average of households with college education and households with high school education in **?**, to approximate the population with a mortgage.

[38] See e.g. **??** for the UK, and **??** for a more general range of estimates and alternative specifications.

Mortgage Contract. The maturity of the loan T is set to 30 years. The fixation windows that households can choose from are set to 5 (θ^{LT}) and 2 (θ^{ST}) years, respectively, matching the UK institutional setting and the two most common types of contracts available. For some counterfactuals, θ^{LT} is set to 10, reflecting a 10-year fixed-rate contract.

Mortgage Pricing. The mortgage rate premium ρ^m is set to the difference between the average 2-year mortgage rate at an LTV of 60% or lower and the real rate, and uses data between 2013 and 2017 to match the LTV premia derived from the loan-level data. I use the empirical LTV credit spreads introduced earlier, estimated as LTV-band fixed effects in steps of 5 percentage points from 70% to 95% LTV, controlling for time (year-month), lender, region, time×lender and buyer-type fixed effects, to calibrate the credit pricing function $f(LTV, \theta)$.[39] The revert rate premium is obtained from the Bank of England database, as the difference between the average revert rate and the average 2-year mortgage rate at an LTV of 60%. Based on this calibration, the real mortgage rate with an aggregate rate of 1% for a 2-year 80% LTV contract is 2.24%, while it is 2.96% for a 2-year 90% LTV, and 4.07% for a 2-year 95% LTV contract.

Summary and Model Solution. To summarize, the household decision problem is tracked using the state variable vector $\Omega = \{X, t, LTV, r, \mathcal{S}^{\theta^{LT}}, \mathcal{S}^{\theta^{ST}}\}$ and contains household net wealth, time/age, LTV, aggregate interest rate, repricing state for the longer-term contract, and repricing state for the shorter-term contract. The model solution is briefly outlined in the following, with more detail provided in appendix section **??**. The state space is discretized in an equal-spaced grid for the continuous state variables X and LTV, interest rates are discretized using five states, and the model is solved recursively by setting $V_T = C_T$ in the last period, i.e. assuming households consume all wealth in the last period. Household consumption, also discretized, and mortgage choice functions are obtained as optimal choices across each combi-

[39]The calibration assumes that the relative differences in credit spreads across the LTV distribution are preserved in real terms. The opportunity cost of longer-term contracts would likely be larger if credit spreads apply to nominal LTV, as nominal LTV may decrease more quickly over time given nominal rather than real house price growth.

nation of the discretized points on the state space using a grid search. The policy functions are then used to simulate consumption and mortgage choice given simulated realizations of income, house price and interest rate shocks for 10,000 households.

1.4.3. Results

This subsection provides an overview of the main findings from the model. After solving for optimal mortgage fixation choice and simulating household choices, I can vary the choice set to evaluate the marginal welfare benefit of adding a longer-term contract. I find that high-LTV households have a reduced willingness to pay for adding a longer-term contract to their choice set compared to low-LTV borrowers, amounting to 0.5% less in annual consumption for 5-year, and 1.3% less for 10-year fixed-rate contracts. Optimal mortgage fixation length is decreasing in LTV, and increasing over the life of the loan, in line with the data.

Optimal Mortgage Fixation Policy. To provide intuition on the model results, the following illustrates patterns in the policy function for optimal mortgage fixation choice. Under the baseline calibration, given an intermediate aggregate interest rate state and holding net wealth constant, borrowers prefer the short-term contract (with $\theta^{LT} = 2$) to the long-term contract (with $\theta^{ST} = 5$) whenever the mortgage rate is in the collateral-sensitive pricing region between 70% to 95% LTV. For an LTV below or equal to 70%, the mortgage rate is in the collateral-insensitive pricing region, i.e. there are no further interest rate reductions for an LTV lower than 70%, and the optimal choice is to lock in this rate for longer. Households hence trade off their insurance motive against upward interest rate (and downward house price) risk, and expected cost reductions in credit spreads using shorter-term contracts.

The model can also be solved under different configurations of risk and house price growth, to evaluate how optimal choice depends on the risk environment. In a scenario with no house price growth, and higher interest rate risk ($\sigma_r = 0.0193 \times 2$), households choose long-term contracts at higher levels of LTV, up to 90%, and closer to the initial loan origination date, i.e. within the first 5 to 10 years, as the risk of upward repricing in the credit dimension is largest given

larger loan balances. In a scenario with greater interest rate risk, households choose long-term contracts at higher levels of LTV throughout.

Household Valuation of Long-Term Contract. I quantify the marginal welfare benefit of adding a longer-term contract to the choice set as a standard consumption certainty equivalent. This requires solving the model twice, first under the restriction that households can only choose a short-term contract, and second when households are allowed to choose between the short-and longer-term contract. The consumption certainty equivalent is computed as the percentage increase in consumption across states that a household would require to reach the same life-time expected utility when a longer-term contract is available. Table **??** shows results from a comparison of consumption certainty equivalents under different scenarios. Each row represents a different scenario, while columns 1 and 2 show results for low (70%) and high (90%) LTV contracts under observed pricing, and columns 3 and 4 show the equivalent under an assumption when the longer-term contract is priced at the expected cost of rolling over the short-term contract sequence ("expected cost equivalence").[40] In the baseline calibration, households' marginal benefit of adding a 5-year contract to their choice set is 0.85%, but this declines to 0.36% for households with an LTV of 90% (Panel A).

I can also use the model to evaluate the value of a counterfactual 10-year fixed-rate contract,[41] with results shown in Panel B. The gap in the welfare benefit between low- and high-LTV borrowers is even greater: while the value is 2.03% of lifetime consumption for low-LTV borrowers, this value is only a third, 0.74%, for high-LTV borrowers, a larger proportional decrease compared to the 5-year contract. The intuition is that the opportunity cost of not repricing in the credit dimension over 10 years is even greater than not repricing over 5 years, and so the credit repricing effect is exacerbated at longer fixation lengths.

[40] Households' initial LTV level is taken as given, based on the empirical findings that control for other variables such as the LTI ratio and past house price growth which could be a proxy for house price beliefs (**?**) and finds that the effects on fixation choice are small.

[41] This assumes effective commitment over the initial fixation window under a binding prepayment penalty, and using credit risk pricing for the 5-year contract.

Finding 3.1. *Under the baseline calibration, high-LTV households' willingness to pay for 5-year fixed-rate contracts is around half that of low-LTV households.*

The results in alternative scenarios in the following rows suggest that house price growth plays an important role in both raising the expected credit repricing benefit, as well as reducing repricing risk to the upside. High-LTV borrowers almost reach the same insurance value from longer-term contracts in a scenario without house price growth. In addition, in a scenario with higher interest risk the insurance value rises for both low- and high-LTV borrowers, to 1.47% and 1.44%, respectively. The expected cost pricing columns estimate the insurance value if high-LTV households were able to pay the expected cost-equivalent of rolling over a short-term contract sequence, while taking out the longer term contract, which gives a sense of how valuable it is to insure households against both interest rate and credit repricing risk, especially in a more risky environment.

Finding 3.2. *The insurance value of 5-year fixed-rate contracts only equalizes across low- and high-LTV borrowers in scenarios with less house price growth, and/or higher interest rate and house price risk.*

Life-Cycle Pattern in Simulated Mortgage Choice. Figure **??** shows the simulated mortgage choices for an initial LTV of 70% (Panel A) and 90% (Panel B) for the baseline calibration. In line with the policy function intuition and results above, borrowers at 70% LTV start off with a 5-year contract, but those who receive negative house price or positive interest rate shocks move onto a short-term contract. Borrowers at 90% LTV start off with a 2-year contract, but the majority of borrowers moves onto a 5-year contract after about seven years, as the LTV reaches the long-term choice boundary at 70%.

Table **??** summarizes long-term contract take-up over the initial first 10 years of the loan (Panel A) and over the entire loan maturity of 30 years (Panel B). Take-up is computed as the share of borrower-year observations under the 5-year fixed-rate contract, relative to the 2-year fixed-

rate contract. Long-term contract take-up for low-LTV borrowers is close to 100% throughout. High-LTV take-up under the baseline is about half that of low LTV borrowers, and rises in particular for combinations of no house price growth and higher house price risk (second row), and higher house price and interest rate risk (fifth row).

In the model and simulation, households are homogeneous other than their initial starting LTV. I show that this dimension of heterogeneity alone generates a substantial decrease in 5-year contract take-up.[42] This is consistent with the empirical findings from the multivariate linear probability model of predicting 5-year contract choice – 5-year contract choice is strongly decreasing in LTV, holding fixed other covariates such as loan-to-income, age and maturity.

Finding 3.3. *Optimal mortgage fixation length is decreasing in LTV, and increasing over the life cycle of the mortgage, as LTV decreases over time.*

Robustness and Alternative Scenarios. I can also vary household discount rate and risk aversion parameters, with results shown in Tables **??** and **??** in the online appendix. Take-up for high-LTV borrowers compared to the baseline is lower in the scenario with a higher time discount rate, and a higher revert rate, but is unchanged with greater risk aversion, suggesting that the cost reduction motive weighs strongly against households' insurance motive. Relative value-added compared to the baseline is lower with a higher revert rate and discount rate. The insurance value is raised with greater risk aversion for low-LTV borrowers, but unchanged for high-LTV borrowers.

1.4.4. Discussion of Results

The model results confirm that under a baseline scenario for house prices, interest rates and income, optimal mortgage fixation length is decreasing in LTV, and increasing over the life

[42]There is no other source of heterogeneity in the model, which is best interpreted as a model for a representative household, and so contract shares over time arise due to different realized house price, interest rate and income shocks.

cycle, in line with the data. As a result, high-LTV borrowers insure less against interest rate risk compared to low-LTV borrowers.[43] The paper highlights a tension between households' insurance motive, and a cost savings motive, as repricing more frequently allows households to reduce the credit spread while LTV decreases over time. The findings further suggest that the welfare benefits of alternative contract designs, such as allowing high-LTV borrowers to lock in the base rate, but reprice their credit spreads, is potentially large.

The model quantifies potential welfare gains and distributional effects of long-term contract pricing for representative low- and high-LTV borrowers, focusing on how credit repricing affects optimal contract choice throughout the life of the loan. It is hence complementary to models that study optimal mortgage contract design and house prices in general equilibrium, and which emphasize the benefits of state-contingent contract elements (**?**), such as the option to convert a 30-year fixed-rate mortgage to an adjustable-rate mortgage which can enhance the stabilizing effects of monetary policy in downturns (**??**), or indexing mortgage payments to house prices (**?**). The quantification of household willingness-to-pay for long-term contracts follows approaches in the public finance and insurance literatures (**??**) to estimate counterfactual demand for non-traded or missing markets and help explain the lack of longer mortgage fixation lengths in the market.

Long-term contracting can be difficult to sustain due to a number of market imperfections. In the mortgage market, key dimensions of risk are systematic. The paper highlights the role of the loan-to-value ratio as a measure of collateralization and important determinant of mortgage credit spreads. Lenders would have to price in a forward-looking path of collateral, and hence provide insurance against house price risk, in order to make long-term contracts attractive to high-LTV borrowers who pay high initial credit spreads. While I do not find strong evidence for ex ante adverse selection, if households were to have private information about their future repricing risks, this would likely further impede long-term contracting in the mortgage market.

From a policy perspective, one way to sustain long-term contracts is government insurance

[43]The model simulates a scenario where households are ex ante identical and only vary in the LTV they are endowed with, such that the variation in contract choice is caused by resulting variation in credit spreads faced, overcoming potential household selection effects that may be correlated with LTV levels in the data.

of house price risk, as seen in the US market. Centralized pricing by government-sponsored entities has been shown to have regionally redistributive effects (**?**). A flattening of the credit pricing curve across the LTV distribution in longer-term contracts could be interpreted as an additional dimension of redistribution: cross-subsidization from low to high-LTV borrowers, or younger and older cohorts of borrowers, as the credit spread is spread out over time and across LTV groups.[44] Of course, the unintended cost of government insurance of house price risk and securitization could involve an increase in moral hazard and systemic risk (**????**). The paper hence highlights novel trade-offs for mortgage contract design and market reform.

The paper findings are consistent with cross-country evidence of mortgage fixation lengths being relatively short, and the US 30-year fixed-rate mortgage being an exception. Figure **??** in the online appendix shows average fixation lengths across a range of advanced economies. The average fixation period is around 2 to 2.5 years in the UK, Greece and Spain, around five years in the Netherlands and Italy, around ten years in Denmark and Germany.[45] The right-hand axis also plots a measure of average household indebtedness relative to average house price levels. Household indebtedness is somewhat negatively correlated with average mortgage fixation length, and the US has similar levels of indebtedness, but much longer average fixation lengths of 25 years, suggesting that this outcome may be sustained by policy interventions in the mortgage market, and may not arise in markets without public credit risk guarantees.

1.5. Conclusion

This paper studies household mortgage fixation choice in a setting with frequent repricing and market pricing of credit risk, the UK mortgage market. It documents a life-cycle dimension of credit risk in the mortgage market: over the life of the loan, the loan-to-value ratio typically

[44] An alternative policy could be repricing credit spreads, while locking in the base rate, as seen in the Netherlands, or structuring mortgages into a non-risky, i.e. LTV-insensitive, tranche, and a risky tranche, as done in Denmark, where collateralized mortgage loans are only available up to an LTV of 80%. Borrowers can then borrow an unsecured loan to raise their LTV up to 95%. This way, base rates can be locked in at origination, while the risky portion of the loan can be repaid separately. One disadvantage is that these unsecured loans are potentially quite expensive.

[45] The data is taken from **?**, which does not include averages for Canada and Australia, so the most common range of products is shown, which is between 2 and 5 years.

declines and thus collateral coverage improves, due to principal repayment and house price appreciation. When considering a longer fixation length, high-LTV borrowers face a trade-off between their regular demand to insure against repricing, and obtaining a lower credit spread over time by repricing more frequently, raising the opportunity cost of the longer-term contract.

In order to quantify whether this mechanism can explain reduced demand for longer-term contracts, the paper further proposes a model of optimal mortgage fixation choice. I use the model to evaluate the marginal insurance benefit of longer-term contracts by varying households' available contract choice sets. I find that high-LTV households' willingness to pay for 5-year fixed-rate contracts is around half that of low-LTV households, suggesting that the insurance benefits of longer-term contracts are unequally distributed across households due to credit risk. A baseline calibration of risks is not sufficient to generate demand for 5-year fixed-rate contracts for high-LTV borrowers. The model helps explain the missing, or very small, markets for even longer-term contracts such as 10-year fixed-rate contracts in the UK, and the prevalence of relatively short mortgage fixation lengths across most countries.

The paper proposes a mechanism that reduces pooling and hence risk-sharing in longer-term contracts in high-LTV segments of the mortgage market. The lack of long-term contract take-up in the mortgage market was noted by **?** in a comprehensive review of the UK mortgage market, and appears common in many other mortgage markets. In the US mortgage market context, mortgage contracts have historically evolved from short-term balloon-mortgages with substantial repricing risk in the 1930s, to the institutional framework surrounding the 30-year fixed-rate mortgage today (**?**). This has remained a subject of on-going discussion on mortgage market reform (**?**) given the large public cost externalities posed by public credit risk guarantees following the 2008/09 financial crisis. One way to interpret the role of public credit risk guarantees is that they allow for risk-sharing with riskier households and lower the cost of insurance against interest rate risk for these borrowers. The paper also raises questions regarding the incidence of house price risk. In a market where risky households frequently reprice, house price risk is mostly borne by households, whereas with public credit risk guarantees, that risk is borne by the government. Without such subsidized credit risk pricing, market pricing incentivizes higher-LTV borrowers towards shorter-term contracts, rendering the combination of

high-LTV borrowing and long-term contracting particularly challenging, and may help explain continued government interventions observed in mortgage markets.

Figure 1.1: Fixation Length Contract Choice by LTV

This figure shows the residualized probability of choosing a 5-year fixed-rate contract across LTV bands, by regressing an indicator variable that takes the value 1 if households chooce a 5-year fixed-rate contract, and 0 if households choose a 2-year fixed-rate contract on LTV band indicators, with the effect of covariates (LTI ratio, age, age squared, loan term) and time (year-month), local authority, time×local authority, lender, lender×local authority, borrower type, loan deciles, and sales channel (intermediated or direct sale) fixed effects partialled out. The LTV-band coefficients are added to the base category (LTV≤ 70%) level. The non-residualized raw probabilities are reported in online appendix Figure **??**.

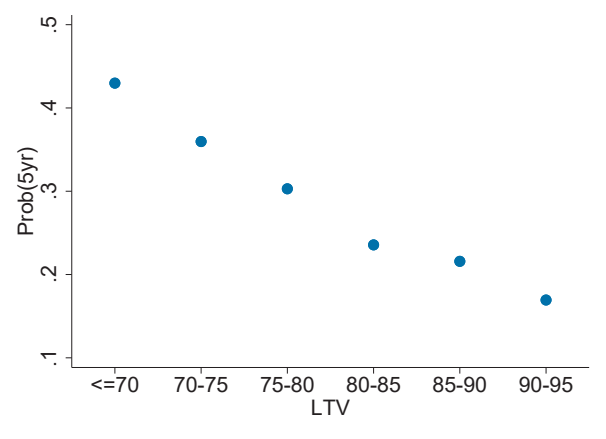


Figure 1.2: LTV Pricing

This figure plots the credit spread paid on 5-year fixed-rate mortgages across LTV bands (≤70%, (70-75]%, (75-80]%, (80-85]%, (85-90]%, and (90-95]%), by extracting LTV-band fixed effects from a regression of interest rates on LTV bands and fixation length, controlling for year-month, lender, buyer-type, year-month×lender fixed effects, using data from 2013 to 2017. The fixed-effect point estimates are plotted as LTV band levels, with confidence bands based on standard errors clustered by time and lender.

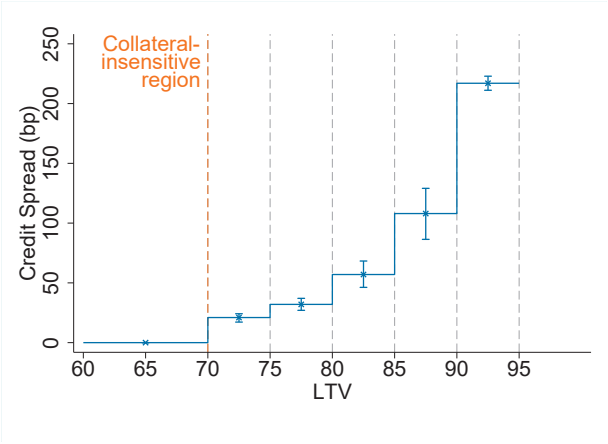
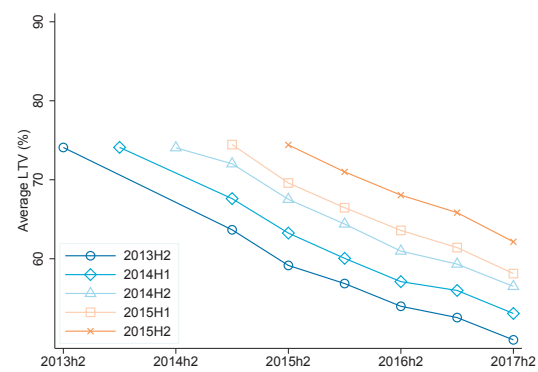


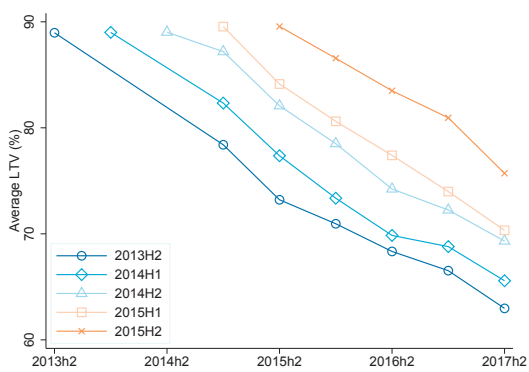
Figure 1.3: LTV and Interest Rates Over the Life Cycle, by Cohort

This figure shows the average loan-to-value (LTV) ratio and mortgage interest rates for different first-time borrower cohorts (originating their loans between 2013H2 and 2015H2), tracked over time at half-yearly frequency using the borrower panel data between 2015H1 to 2017H2. The current LTV ratio is computed as the reported current loan balance, divided by the current house price, obtained by scaling the initial house price with the local-authority-specific house price index. The current mortgage rate is directly reported in the panel data. Panel A and C show cohorts of borrowers that originated with an initial LTV between 70 to 75%, Panel B and D show cohorts of borrowers with an initial LTV between 90 and 95% LTV.

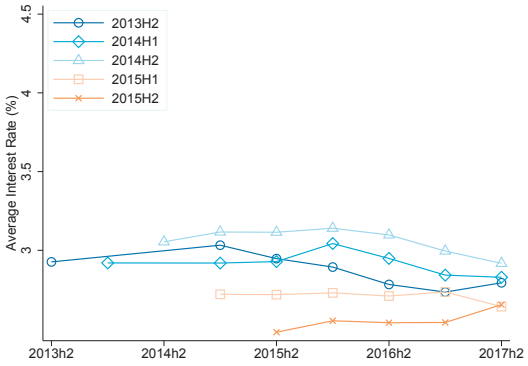
(a) LTV (75%)



(b) LTV (90%)



(c) Mortgage Rate (75% LTV)



(d) Mortgage Rate (90% LTV)

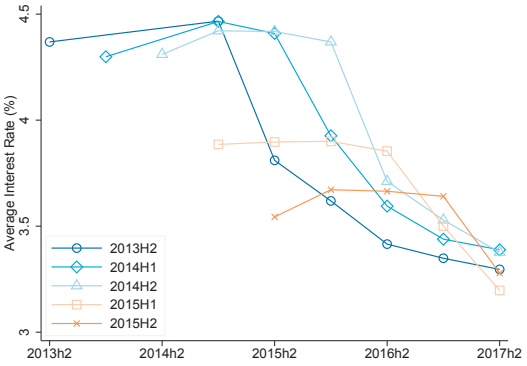
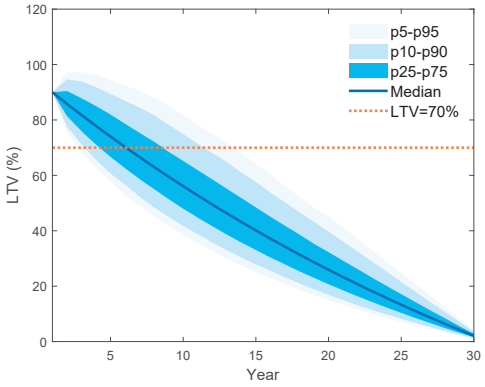


Figure 1.4: Simulation of LTV and Interest Rate Over The Life Cycle

These figures illustrate the distribution of repricing risks over the life of the loan. The simulation is based on a fully-amortizing loan, repaid over 30 years, using LTV credit pricing estimated from the data, and a calibrated house price process as described in Table **??**. Panel A shows the distribution of LTV and Panel B shows the mortgage rate distribution, respectively, with an initial LTV of 90%. The dark blue line indicates the median (50th percentile) of the distribution, the dark blue swathe indicates the interquartile range (25th to 75th percentile), the light blue swathe indicates the interdecile range (10th to 90th percentile), and the grey swathe the 5th to 95th percentile range. The dotted orange line indicates the LTV pricing boundary at 70% LTV, and the interest rate associated with the LTV pricing boundary, respectively.

(a) Simulated LTV (90% LTV)



(b) Simulated Mortgage Rate (90% LTV)

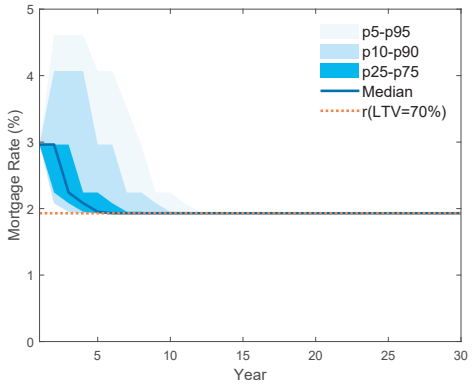
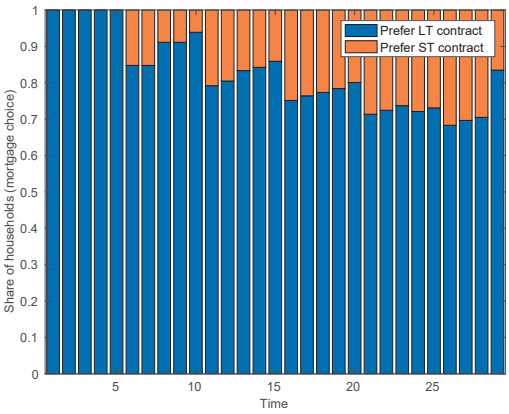


Figure 1.5: Optimal Mortgage Choice Over the Life Cycle

These figures show simulated optimal mortgage choice for a relatively long-term contract with fixation length $\theta^{LT} = 5$ against a shorter-term contract with $\theta^{ST} = 2$, under the baseline calibration for households with an initial LTV of 70% (Panel A) and 90% (Panel B), respectively, over the life of the loan.

(a) Low LTV



(b) High LTV

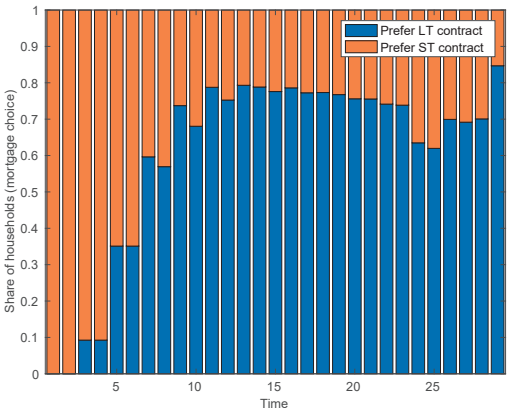


Table 1.1: 5-Year Contract Choice Regressions

This table shows three regressions of the probability of taking out a 5-year contract, compared to taking out a 2-year contract, on a set of covariates that could be related to 5-year contract choice. The dependent variable is an indicator that takes the value 1 if the contract fixation length is 5 years, and 0 if the fixation length is 2 years. All independent variables are expressed in standard deviations of the variable. Column (1) includes local-area×time fixed effects. Column (2) includes past rolling (10-year) local quarterly house price volatility, while column (3) uses a rolling (10-year) beta of the local house price index with respect to the aggregate house price index, with region×time fixed effects. Local refers to local authority districts, while region refers to 12 administrative regions in the UK. Borrower types are first-time borrowers, second-time borrowers and remortgagors.

	(1) I[5yr]	(2) I[5yr]	(3) I[5yr]
LTV	-0.048***	-0.046***	-0.046***
	(0.002)	(0.002)	(0.002)
I[LTV≥80%]=1 × LTV	-0.133***	-0.134***	-0.134***
	(0.003)	(0.003)	(0.003)
LTI	-0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Age	0.004	0.002	0.002
	(0.003)	(0.003)	(0.003)
Age (sq.)	-0.011***	-0.009**	-0.009**
	(0.004)	(0.004)	(0.004)
Loan Term	-0.029***	-0.028***	-0.028***
	(0.001)	(0.001)	(0.001)
Local HP Growth		-0.005*	-0.005*
		(0.003)	(0.003)
Local HP Volatility		0.001	
		(0.002)	
Local HP Beta			-0.002*
			(0.001)
Year-Month FE	✓	✓	✓
Local-Authority×Time FE	✓		
Region×Time FE		✓	✓
Lender×Time FE	✓	✓	✓
Borrower-Type FE	✓	✓	✓
Loan-Decile FE	✓	✓	✓
Sales-Channel FE	✓	✓	✓
Observations	2,865,661	2,865,661	2,865,661
R^2	0.15	0.14	0.14

Table 1.2: Baseline Calibration of Model Parameters

This table provides an overview of calibrated model parameters for the baseline lifecycle model. Real interest rate parameters are estimated using UK average annual rates on 5-year inflation-indexed gilts between 1987 to 2017, and 1-year real rates deflated using 1-year ahead inflation expectations. House price parameters are estimated using the UK annual house price index between 1987 to 2017. The average loan-to-income ratio is estimated based on PSD data between 2013 and 2017 of first-time buyers within the LTV band of 80-85%, and is converted to an loan-to-after-tax-income ratio using the 2017 effective UK tax rate. The long-term and short-term contract fixation lengths are based on the two most common types of products in the PSD data. The 5-year to 2-year swap rate premium is based on average UK swap rates between 2013 and 2017 using yield curve data from the Bank of England. The mortgage rate premium is computed as the difference between the 2-year 60% LTV mortgage rate and the 2-year UK swap rate over the same time period. The interest rate premia are extracted from a regression (see specification in Figure **??**) using PSD loan-level data between 2013 and 2017. The revert rate premium is computed as the difference between the average revert rate and the 2-year 60% LTV mortgage rate.

Parameter		**Value**	**Source**
Panel A: Household preferences			
Risk aversion	γ	3	Literature
Time discount rate	δ	0.02	Literature
Panel B: Interest rates			
Mean of log real rate	μ_r	0.0164	UK (1987-2017)
Standard deviation of log real rate	σ_r	0.0193	UK (1987-2017)
Autoregression coefficient of log real rate	ρ_r	0.95	UK (1987-2017)
Panel C: House prices			
Mean of house price shock	μ_h	0.0258	UK (1987-2017)
Standard deviation of house price shock	σ_h	0.077	UK (1987-2017)
Loan-to-income ratio (85% LTV)	LTI	3.56	PSD (2013-2017)
Loan-to-after-tax-income (85% LTV)		4.82	Computed (tax rate of 35.5%)
Panel D: Income			
Standard deviation of transitory income shock	σ_ϵ	0.1	Literature
Panel E: Mortgage rates and fixation length			
Fixation length of long-term contract (years)	θ_{LT}	5	PSD
Fixation length of short-term contract (years)	θ_{ST}	2	PSD
Mortgage rate premium (bp)	ρ^m	93	Bank of England database, 2yr 60% LTV rate (2013-2017)
2yr 70-75% LTV premium (bp)	$\rho^{2,75-75}$	2	PSD (2013-2017)
2yr 75-80% LTV premium (bp)	$\rho^{2,75-80}$	15	PSD (2013-2017)
2yr 80-85% LTV premium (bp)	$\rho^{2,80-85}$	31	PSD (2013-2017)
2yr 85-90% LTV premium (bp)	$\rho^{2,85-90}$	103	PSD (2013-2017)
2yr 90-95% LTV premium (bp)	$\rho^{2,90-95}$	214	PSD (2013-2017)
5yr 70-75% LTV premium (bp)	$\rho^{5,75-75}$	21	PSD (2013-2017)
5yr 75-80% LTV premium (bp)	$\rho^{5,75-80}$	32	PSD (2013-2017)
5yr 80-85% LTV premium (bp)	$\rho^{5,80-85}$	57	PSD (2013-2017)
5yr 85-90% LTV premium (bp)	$\rho^{5,85-90}$	108	PSD (2013-2017)
5yr 90-95% LTV premium (bp)	$\rho^{5,90-95}$	217	PSD (2013-2017)
Revert rate premium	ρ^{REV}	269	Bank of England database (2013-2017)

Table 1.3: Long-Term Contract Shares

This table shows the simulated long-term contract shares given optimal household choice under different scenarios. The columns show results taking pricing as given for low- (70%) and high-(90%) LTV borrowers, and under pricing that equalizes the expected cost of the long-term contract and short-term sequence (shown in online appendix Figure **??**), respectively. The rows show different scenarios for house price growth and risk, and interest rate risk. The long-term contract shares are computed as the share of household-year observations that are under a long-term, compared to a short-term contract, over the first ten years since loan origination (Panel A) and over the entire maturity of the loan, 30 years (Panel B). The simulation tracks household optimal behavior given realization of shocks, based on 10,000 households for each scenario and LTV band.

	Baseline		Expected Cost Equivalence	
	Low LTV	High LTV	Low LTV	High LTV
Panel A: Share on Long-Term Contract Over Initial 10 Years				
Baseline	0.95	0.35	0.96	0.93
No house price growth ($\mu_h = 0$)	0.89	0.66	0.95	0.84
No house price growth and higher risk ($\mu_h = 0$, $\sigma_h = 0.0770 * 2$)	0.87	0.80	0.86	0.81
Higher interest rate risk ($\sigma_r = 0.0193 * 2$)	0.97	0.93	0.97	0.96
No house price growth & higher interest rate risk	0.96	0.82	0.97	0.93
Panel B: Share on Long-Term Contract Over 30 Years				
Baseline	0.83	0.62	0.84	0.83
No house price growth ($\mu_h = 0$)	0.80	0.66	0.83	0.77
No house price growth and high vol ($\mu_h = 0$, $\sigma_h = 0.0770 * 2$)	0.77	0.71	0.78	0.72
Higher interest rate risk ($\sigma_r = 0.0193 * 2$)	0.91	0.89	0.89	0.89
Higher house price & interest rate risk	0.89	0.82	0.89	0.87

Table 1.4: Value of Long-Term Contract

This table shows the consumption-equivalent of introducing a long-term contract to an existing short-term contract under different scenarios. The columns show results taking pricing as given for low- (70%) and high-(90%) LTV borrowers, and under pricing that equalizes the expected cost of the long-term contract and short-term sequence (shown in online appendix Figure **??**), respectively. The rows show different scenarios for house price growth and risk, and interest rate risk. The consumption certainty equivalent is computed as the percentage increase in consumption across states that a household would require to reach the same life-time expected utility when a longer-term contract is available, in addition to the shorter-term contract. Life-time expected utility is simulated by tracking household optimal behavior given realization of shocks, based on 10,000 households for each scenario and LTV band.

	Baseline		Expected Cost Equivalence	
	Low LTV	High LTV	Low LTV	High LTV
Panel A: Value of 5-year Contract				
Baseline	0.85	0.36	0.85	0.86
No house price growth ($\mu_h = 0$)	0.84	0.73	0.86	0.85
No house price growth and higher risk ($\mu_h = 0$, $\sigma_h = 0.0770 * 2$)	0.96	0.87	1.01	0.96
Higher interest rate risk ($\sigma_r = 0.0193 * 2$)	1.47	1.44	2.60	3.46
Higher house price & interest rate risk	2.60	3.41	2.61	3.43
Panel B: Value of 10-year Contract				
Baseline	2.03	0.74	2.03	2.04

Non-Salient Fees in the Mortgage Market

Lu Liu, Imperial College London

Abstract

This paper studies supply-side product pricing when consumers underreact to non-salient fees. Using comprehensive data on issued and offered mortgages in the UK, I document that lenders differ substantially in the fees they charge, and that borrowers appear less overall cost-sensitive to products with fees. In order to distinguish from demand factors such as unobservable preferences or product characteristics, I show that lenders pass on firm-specific funding cost shocks via fees, but not interest rates, consistent with strategic pricing of fees, and maintaining competitive prices in the salient price dimension, interest rates. I further find heterogeneity in pricing across lenders: those who rely on high fees tend to have higher funding cost, lower return on equity and larger branch networks, in line with a specialization equilibrium in which high-cost lenders are able to match with less cost-sensitive consumers.

JEL classification: G21, G41, D12, D14, D18
Keywords: price dispersion, salience, strategic pricing, mortgages

Chapter 2

Non-Salient Fees in the Mortgage Market

2.1. Introduction

The "law of one price" fails to hold across a range of seemingly homogeneous consumer goods markets.[1] Existing work suggests that this may be due to demand-side factors, taking prices as given: consumers may have unobserved preferences, e.g. for specific brands that increase their willingness to pay; or alternatively, search and cognitive frictions may prevent them from finding the cost-minimizing product (**?**, **?**). One such friction is how readily available and visible information on cost is to consumers. In many markets, one cost component is naturally more *salient* than others, e.g. used as reference category in price comparison websites, such as basic airline fares and hotel rates, with less visible cost components such as fees for luggage and usage of amenities, respectively, added on (**?**, **?**). In order to find the cost-minimizing product, consumers have to correctly compute total cost as the sum of all relevant cost components. However, if one cost dimension is more salient to consumers and if consumers underreact to non-salient cost components (**?**, **?**), recent theoretical work implies that firms have an incentive to adjust pricing of these cost dimensions differentially (**?**, **?**), suggesting a *supply-driven* mechanism behind price, or total cost dispersion, given demand-side frictions.

[1] As noted by **?**, see **?** for a review.

I study this mechanism in the context of the UK mortgage market, which provides a useful laboratory of a market with one highly salient cost component, the interest rate, and a less visible and non-optional cost component,[2] the origination fee.[3] The market in itself is interesting as mortgage costs have significant implications for household finances (**?**), while frictions in search, as documented by **?** and **?** for the US, are important for macroeconomic policy (**?**).

Most mortgages in the UK have a limited set of product characteristics, namely the product type (fixed vs. adjustable rate), fixation period and loan-to-value (LTV) ratio, and are hence fairly standardized.[4] However, even when holding fixed a given set of product characteristics, such as a 2-year fixed rate product with a maximum 75% LTV,[5] there exists a range of products that differ purely in price terms, most commonly a high interest rate and low fee, and low interest rate and high fee variant of the same product by the same lender. Using a novel product-level dataset of the universe of mortgages on offer between 2009 and 2016, I document that more than half of all products on offer are significantly cost-dominated and more than £1000 more expensive than the cheapest alternative.[6] In addition, I find substantial heterogeneity in fees charged across lenders which is much greater than the relative dispersion in interest rates.[7] I hence find reduced-form evidence for substantially dominated offers and fee heterogeneity. However, these facts alone are insufficient to pin down the mechanism. The evidence could be fully driven by demand motives, if fees are for instance compensating for unobserved brand preferences or services. Furthermore, there could be multiple potential supply mechanisms behind differential fees and cost, such as screening or liquidity motives, in addition to strategic pricing.

In order to disentangle these mechanisms more formally, I propose a two-part analysis. First, I estimate a simple discrete choice framework on administrative data on mortgages originated

[2]The non-optionality is important as it allows for a simple and unique total cost comparison across consumers. In many of the examples given, some cost components are optional such that variation in realized cost may arise due to unobserved heterogeneity in preferences across optional components.

[3]In the UK context, fees are a pure price component required at origination in contrast to the US "points" system. Early repayment penalties are allowed and charged separately to fees.

[4]The homogeneous product benchmark is more likely to hold when consumers are sufficiently homogeneous and creditworthy, which is why the paper focuses on low risk mortgages, i.e. mortgages with an LTV smaller than 80%.

[5]Meaning *after* the optimal choice of mortgage type such as fixed vs. adjustable rate (**?**) and loan-to-value (LTV) (**?**) have been made.

[6]Assuming an average loan size of £150,000, where £1000 is about 10 to 20% of the total cost of the mortgage over 2 years and approximately 4% of the median household's annual disposable income in 2016.

[7]For instance, the cross-sectional standard deviation has risen from around £500 to £700 for fees, and declined from around 80 to 50 basis points for interest rates between 2009 and 2016, making cross-sectional fee dispersion much larger than interest rate dispersion relative to the respective average levels.

in the UK, to test the cost-minimization benchmark on the demand side while controlling for observable demand factors. The model assumes that consumers treat both interest cost and fees as equally salient, i.e. they weigh both cost components in £s equally, and allows the consumer to choose across lenders and price-variants within lenders, for a given loan value and product category. Based on observed loan choices, I find that the total cost sensitivity is lower for products with fees than without, which is partially driven by borrowers who benefit from borrowing the fee to bunch just at the LTV threshold, as documented by **?**.[8] However, the high-fee product fixed effect remains significant and sizeable even when excluding borrowers who bunch. The model hence systematically under-predicts market shares for high fee products relative to the data, and vice versa, even when controlling for lender-level fixed effects. This suggests that there may be "unexplained" demand for products with low rates, but high fees, which is difficult to match in a model where consumers minimize total cost and dislike interest cost and fees equally. To the extent that unobserved demand factors such as preferences for other services or convenience vary across lenders and not across price-variants within a given lender, borrowers seem to indeed neglect fees.[9]

However, possible concerns that remain are potentially time-varying confounding factors and that the demand analysis provides only indirect evidence for a supply-driven mechanism.

Hence as a second step, I study lender pricing directly. The UK market is particularly suited to study mortgage pricing as borrowers can directly obtain the interest rates and fees that are quoted and advertised. There is no ex post price discrimination based on borrower-specific characteristics such as credit scores or bargaining, and lenders operate fully flexible interest rate schedules instead of fixed tick sizes, in contrast to the US and other countries. Using the mortgage offer data, I build a quarterly panel from 2009 to 2016 of prices set by lenders across products and time. I then study price adjustment of lenders in response to firm-specific, time-varying shocks to wholesale funding cost, which I construct as a "shift-share" shock (**?**, **?**) based on cross-sectional predetermined variation in loan-to-deposit ratios across lenders and aggregate instrumented variation in wholesale funding cost over time. I find that lenders respond to these firm-specific cost shocks by leaving interest rates broadly unchanged, but

[8]This could be thought of as a "soft" borrowing constraint as interest rates rise discretely in 5 to 10% LTV steps (**?**). In contrast, in this setting it is unlikely that borrowers demand different price-variants to achieve different payment profiles over time, as borrowers have the option to add the fee to the loan value and repay it over time, at no additional cost.

[9]There are no known differences in eligibility or services across price types (low fee or high fee), which is confirmed in conversations with industry participants.

increasing fees for their lowest interest rate product. The increase in fees is economically large: given a one standard deviation increase in the funding cost shock, lenders raise fees for their lowest interest rate product by around £100, which is about 15 per cent of the average level of fees and corresponds to a 0.5 standard deviation change in fees. For this to be evidence for a supply-driven mechanism, the identifying assumption is that time-varying unobserved demand factors are uncorrelated with lender-specific cost shocks over time.[10] This finding cannot be rationalized by time-varying preferences of lenders for fee income compared to interest income or a boost to current earnings alone, as they seem to increase total cost, instead of decreasing interest rates in line with the increase in fees. Instead, it would require a demand pattern that is somewhat less elastic to increases in fees, in line with a more strategic motive: in a market where (some) consumers narrowly focus on the salient cost dimension, interest rates, lenders may be able to use fees to improve their position in the interest cost ranking and appear cheap to consumers who neglect fees. This may affect search outcomes. In originated loans, I indeed find lower realized excess cost dispersion (accounting for borrower, product and regional characteristics) for products without fees, compared to products with fees, i.e. borrowers seem to come closer to the cost-minimizing benchmark in the class of products without fees than those with fees.

To illustrate the pricing incentives with non-salient fees, Table **??** shows a stylized example of a pricing table.[11] It illustrates two features in the data: first, lender L offers the cost-minimizing product 4 in the market, priced at 2% interest and zero fees, which I refer to as a low interest rate, high fee price type ($\{r_h, f_l\}$). Lender H who is more expensive than lender L can only attract less cost-sensitive consumers, and reprices its product at 1.95% interest and £495 fees, instead of a 2.23% interest and zero fee product,[12] which is dominated in the fee dimension, but not in the salient interest rate dimension (product 3). The same holds for lender L's pricing of the low interest rate, high fee price type ($\{r_l, f_h\}$), and lender H's product 1. Second, the interest-sorted price table shows low interest rate, high fee products more prominently than high interest rate, low fee products. In this example, the cheapest product is hence least visible on the interest rate ranking. The zero-fee interest column converts interest cost and fees into a

[10] This provides an alternative way to rule out that fees reflect compensation for unobserved demand factors, as it accounts for time-invariant unobservables such as branch distance, and assumes that variation in e.g. unobserved service quality does not vary at the frequency of cost shocks.

[11] It provides a close approximation of popular price comparison tables such as MoneySupermarket.com, as shown in Figure **??**.

[12] Assuming that the mortgage is repaid over an initial 2-year fixed rate period with a £150,000 loan value and 25 year amortization period, and is subsequently refinanced.

zero-fee interest rate, which reveals the total cost ranking and makes more visible that product 4 would be the cost-minimizing choice.

Table 2.1: Example of a mortgage price table

Product	Lender	Interest (%)	Fees (£)	Zero-fee interest*	Price type
1	H	1.34	1,700	2.32	$\{r_l, f_h\}$
2	L	1.44	1,000	2.02	$\{r_l, f_h\}$
3	H	1.95	495	2.23	$\{r_h, f_l\}$
4	L	2.00	0	2.00	$\{r_h, f_l\}$

**Assuming a loan value of £150,000, 2-year fixation and 25-year amortization period.*

The findings suggest that a lender can use fees as an active margin of price adjustment to maintain its competitive position in terms of interest rates, in response to a relative (i.e. firm-specific) cost shock vis-à-vis its competitors. For this to be an optimal pricing strategy with fee heterogeneity in equilibrium requires two assumptions, and I discuss in turn why they appear valid in this context. First, there needs to be a significant fraction of the borrower population that does neglect fees. This does not seem far-fetched in the UK, where mortgage prices are commonly framed as part of "best buy" tables ranked by interest rates. Figure **??**, a snapshot from one of the biggest price comparison websites, MoneySupermarket.com, illustrates that interest rates are the salient price dimension that price comparison tables coordinate on, while fees are shown separately, or, in this case, only specified in the footnotes. If borrowers are fully cost-minimizing, however, they should be indifferent between cost paid in fees and interest rates, and consider a total cost ranking. Offering cost-dominated products with high fees but low interest rates hence attracts borrowers who neglect fees, but not total-cost-minimizing borrowers.

Second, (some) lenders find it profitable to pass-through relative cost shocks via fees rather than interest rates, suggesting that demand is (locally) less elastic with respect to fees. I find that there is substantial heterogeneity in fees across lenders, in contrast to interest rates, pointing to heterogeneity across lenders where some lenders rely more on fees than others. For this to be an equilibrium requires that the marginal demand sensitivity to fees also varies across lenders,[13] such that there is specialization across lenders. Hence one way to rationalize the results is that lenders with higher fees and overall cost need to be able to attract less cost-sensitive borrowers, while lenders with low overall cost attract cost-sensitive borrowers (**?**, **?**).

[13]This is also supported in the discrete choice analysis, where controlling for borrower-characteristics substantially increases the estimated cost sensitivity.

I provide further evidence for this type of specialization equilibrium by splitting lenders into pricing types based on terciles of the difference between their minimum total cost and interest rate ranking, as a measure of how much they skew their interest rate ranking using higher than prevailing average fees. The "high" pricing type, defined as the upper tercile of lenders, have a positive ranking difference of 6 on average, i.e. they have the $x+6$th-highest position in the total cost ranking, but move up the interest rate ranking to the xth-highest position. As expected, they charge similar interest rates as their competitors, but higher fees, both for the high fee and low fee product. While they have similar size and funding structures as other lenders, they earn substantially lower returns on assets and equity and have higher funding costs. They also tend to have a higher branch density, measured as the number of households for whom a given lender operates one of the closest branches, and higher market share, suggesting that existing forms of market power based on branch networks (**?**) interact with those based on consumer mistakes and may make it easier for these lenders to acquire less cost-sensitive borrowers. Across lenders, fees do not seem to compensate directly for a greater branch density, however, meaning that both the financial necessity and ability to attract less cost-sensitive borrowers seem to affect lenders' pricing decisions. I further find that high-skew lenders also adjust their fees by more in response to firm-specific cost shocks, which could be interpreted as an additional dynamic source of market power in line with anchoring effects (**?**) where consumers become less sensitive to bigger absolute fee changes if they are used to higher equilibrium levels of fees.

Next, I consider other supply motives such as screening that could drive pricing, and could be correlated with cost shocks. I find little evidence that fees are used to screen for borrower risk types. In contrast to the US, lenders frequently charge early repayment penalties in the UK, which are a more direct measure to screen for prepayment risk than fees.[14] For almost 90% of products in the sample, prepayment penalties do not vary across products within a given lender.[15] This seeming lack of screening for heterogeneous prepayment risk could be explained by the relatively short initial fixation periods prevalent in the UK, with most borrowers refinancing at the end of a 2 to 5 year fixation period (**?**).[16] The pricing model by lenders implies,

[14]In the US mortgage market, borrowers typically have the option to pay "points" (fees) upfront to obtain a lower interest rate, which decreases the refinancing incentive and signals lower prepayment risk (**?**).

[15]This extends to a regression setup, where prepayment penalties do not significantly affect the interest rate-fee trade-off within and across lenders.

[16]As another screening mechanism, lenders may use high fees to screen for liquidity risk, as liquidity-constrained borrowers may be less able to pay an upfront fee, or any other unobservable characteristics that may be correlated with default probabilities. The institutional framework in the market makes this less likely as borrowers are allowed to add the fee to the loan balance and repay it over the duration of the mortgage, at

however, a consistent link between fees and loan size, as the benefit of a lower interest rate, the interest cost reduction, is greater for larger loan balances.[17] For a given lender's product offering of $\{r_l, f_h\}$, $\{r_h, f_l\}$ there exists a unique loan value at which a borrower should be indifferent between choosing the low or high fee product. From the perspective of a given borrower with a fixed loan size, she should only consider a single product by a given lender, such that price dispersion purely arises from choice across lenders. This does not seem to be the case in realized borrowing outcomes: I find substantial realized fee dispersion across the loan size distribution and even *within*-lender price dispersion, suggesting that consumers indeed make mistakes in the fee dimension, as unobservable preferences seem more likely to affect choice across lenders, and less so choice within.

These findings are important for two reasons. First, the paper provides empirical evidence that lenders price strategically in response to demand-side biases and document a supply-driven mechanism behind price dispersion. The framework provides testable predictions of firms' pricing incentives in the presence of non-salient fees. This work is the first to empirically identify this supply-side channel using cost shocks, to the best of my knowledge. And second, fee dispersion may make total cost comparisons more difficult for consumers, pointing to an interaction mechanism with existing demand-side frictions.

The paper makes three contributions to the existing literature. First, recent work in household finance provides evidence of consumer mistakes (**?**, **?**, **?**), and firm behavior that exploits these (**?**, **?**, **?**, **?**). The first part of the analysis, that consumers seem to be less cost-sensitive to non-salient price components, is consistent with evidence from the credit card market (**?**), mutual fund (**?**, **?**) and social security markets (**?**), as well as consumers underreacting to non-salient taxes (**?**) and neglecting expected out-of-pocket costs in health insurance plan choice (**?**). As a second step, I provide direct evidence that firms take into account these consumer frictions and set the non-salient price component jointly in order to compete in the salient price dimension.

Second, the identification strategy also allows me to highlight the active role of the supply side for household finance problems (**?**), as firms dynamically respond to cost shocks to adjust their prices optimally, in response to a change in their competitive position. It thereby reveals a pricing strategy in line with search and cognitive frictions on the demand side, while making a

no additional cost, which is observed in the data.

[17] In that sense, fees can be thought of as very coarsely screening for loan size.

preference-driven mechanism less likely, hence adding to the previous literature that has used experimental (**?**) and model-based approaches (**?**, **?**) to distinguish between these two channels. The findings further show that supply-side incentives play a role in understanding the drivers behind price dispersion in markets for homogeneous goods (**?**, **?**, **?**, **?**) which could exacerbate existing search frictions. Alternatively, the degree to which this supply channel matters may also depend on the degree of existing frictions. The UK mortgage market appears to have relatively low search frictions as a market where prices are posted and can be directly obtained. In closely related work, **?** find substantial price dispersion in US mortgage rates that varies with borrower types, highlighting the role of ex post price discrimination in a market where lenders post initial prices, but where final prices are subject to negotiation. This suggests that the market structure and ease of search may affect the relative role of supply and demand behind price dispersion.

Third, I provide an empirical framework with testable predictions for pricing to link to theoretical work on price obfuscation (**?**, **?**, **?**, **?**)[18] and show evidence for a specialization equilibrium (**?**, **?**, **?**) in which high-cost lenders attract less fee-sensitive consumers, as observed in other retail finance settings (**?**).[19]

The remainder of this paper is organized as follows. Section **??** provides background on the UK mortgage market and the data used. Section **??** presents stylized facts and the mortgage pricing structure. Section **??** provides the borrower-level discrete choice analysis, and Section **??** describes the identification strategy and results based on the lender-level analysis. Section **??** discusses the mechanism and Section **??** concludes.

2.2. Background and Data

2.2.1. Background on the UK Mortgage Market

Mortgage borrowing accounts for around half of the median household's liabilities in the UK, which is similar to the US and one of the highest levels across developed economies (**?**). Mort-

[18] See **?** for a review.

[19] The findings are also consistent with the salience theory of **?** in the sense that differences in interest rates are quoted in percentage points, while differences in fees are quoted in £s, which consumers may be less sensitive to due to diminishing sensitivity to price differences at higher price levels.

gage product choice in the UK can be summarized using four product characteristics: mortgage type (fixed vs. floating), fixation period, LTV band, and mortgage borrower type. The large majority of UK mortgages are short duration fixed-rate products, with 2 and 5 years the most common fixation periods. These revert to so-called Standard Variable Rates that tend to be substantially higher than prevailing rates in the market, such that most borrowers refinance at the end of the fixation period. Lenders have "full recourse",[20] default rates are comparably low and default risk pricing takes place through a discrete interest rate schedule at maximum LTV bands in steps of 5 to 10% (**?**). Most mortgages are available to first-time buyers, second-time buyers and refinancing borrower types, with some only available to a particular borrower type.

In contrast to the US (**?**) and Canada (**?**), UK borrowers can directly obtain the interest rates and fees that are quoted and advertised, i.e. there is no ex post price discrimination based on borrower-specific characteristics such as credit scores or bargaining. Prices are set UK-wide and hence not differentiated by branch location or region. Lenders also operate fully flexible interest rate schedules, in contrast to fixed tick sizes in the US, such that both interest rates and fees can be adjusted flexibly. These factors make the UK mortgage market an ideal laboratory to study lenders' pricing strategies. I observe the full universe of mortgages and product characteristics that a borrower can shop from, and the prices reflect the final interest rates and fees that can be obtained.[21]

Similar to other countries, the market is relatively concentrated. The largest six UK lenders together account for around 75% of the stock and new flows of mortgage lending.[22] The largest 27 lenders that I study together account for approximately 95% of new mortgage lending.

[20]Meaning lenders can recover losses from defaulted borrowers though their assets and incomes for up to seven years, until the debt is paid (**?**).

[21]In an earlier step, lenders will accept and reject loan applications based on a borrower's credit history, such that prices are implicitly conditional on approval. The approval mechanism depends on lender-specific internal credit models, but these do not differentiate between a borrower who takes out a high fee product, compared to a low fee product, and so should not confound the analysis. This is likely to play a greater role for more risky mortgages, as demonstrated by **?**.

[22]Between 2010 and 2015, see former quarterly "Trends in Lending" reports from the Bank of England: http://www.bankofengland .co.uk/publications/Pages/other/monetary/trendsinlending.aspx.

2.2.2. Data

I combine three datasets. First, my main data source is Moneyfacts which is one of the most commonly used financial price comparison websites in the UK,[23] and is accessed through the Bank of England. It comprises the universe of mortgage products on offer, with detailed product characteristics, since June 2008, by lender and at monthly frequency. The data used covers the time period from January 2009 to December 2016.

Since the focus of this paper is to study the consumer choice problem for a given set of product characteristics, the analysis focuses on 2-year fixed rate products that are low risk (maximum 70-75% LTV) and available for first-time buyers, and I assume that borrowers refinance at the end of the fixation period, such that they need to search anew after each fixation spell.[24] Fixed-rate mortgages available to first-time borrowers account for around 80% and 70% of the mortgages on offer, respectively. They also reflect respectively around 80% and 30% of mortgage issuance in the UK.

Table **??** illustrates the product characteristics and a representative menu structure based on four products by Halifax, one of the largest UK mortgage lenders, as observed in April 2013. It shows that a borrower with a maximum 75% LTV ratio can choose between an annual interest rate of 3.39% and a total arrangement fee of £295, or "trade down" the interest rate to 2.69% by paying a higher fee, £1290. Further descriptive statistics are provided in Table **??**, Panel 1.

Second, I augment the Moneyfacts dataset with data on lender characteristics and funding cost. Data on lender characteristics for the 27 largest lenders in the UK is obtained from SNL Financial. These contain lender characteristics from balance sheet and income statement data (Table **??**, Panel 2). Data on wholesale funding cost are based on daily 2-year LIBOR and USD swap rates and CDS premia averaged by quarter for the largest six lenders, and are obtained from Bloomberg and the Bank of England.

Third, in order to study choice outcomes, I use the Financial Conduct Authority's Product Sales Database (PSD) which collects data on all regulated mortgage originations in the UK, and is

[23] Recommended by the formerly government-led Money Advice Service on its "Mortgage comparison checklist", see https://www.moneyadviceservice.org.uk/en/articles/your-mortgage-comparison- checklist

[24] I abstract from dynamic dependencies that may give lenders different pricing incentives - I do not find a relation between fees and Standard Variable Rates, which is likely explained by the fact that borrowers tend to refinance consistently at the end of their fixation period, as documented by **?**.

accessed through the Bank of England via a data sharing agreement. The data used covers mortgage originations between 2015 and 2016. Each loan contains detailed loan and borrower characteristics such as the mortgage type, fixation period, LTV, interest rate, fees paid, LTI, age, income and postcode of the borrower. I further use a 2016 snapshot from SNL Financial on all UK bank branch locations to compute the distance from a given lender's nearest branch to a given borrower.

I use loan-level data from the PSD for the borrower-level analysis. For the lender-level analysis, I construct a panel with lender characteristics and pricing statistics, in particular changes in interest rates and fees for the low rate, high fee, and high rate, low fee product on offer at the lender level, and a lender-specific funding shock, at quarterly frequency between 2009Q1 to 2016Q4. Data selection and variable construction are further described in the Appendix **??**.

2.3. Stylized Facts and Mortgage Pricing

This section documents two stylized facts in the data: first that there is evidence for substantial price dispersion in the UK mortgage market, and second that this is accompanied by large heterogeneity across fees. It then describes how supply-side mortgage pricing could be related to price dispersion, and why fees could be interpreted as a non-salient cost component.

Price dispersion in the mortgage market. From the perspective of a borrower with an approximately average loan demand of $L = £150,000$, who refinances the mortgage after the end of the fixation period d (2 years, i.e. 24 months) and amortizes the mortgage over 25 years ($T = 300$ months), the total cost C of the mortgage with monthly interest r and origination fee f over time period d is

$$C_d = \underbrace{\frac{r}{1 - \frac{1}{(1+r)^T}} \cdot L \cdot d}_{\text{interest cost}} + f. \tag{2.1}$$

When the borrower searches for a mortgage in the universe of products on offer in a given month, there is substantial dispersion in total cost and the majority of products on offer at any one point in time is cost-dominated by more than £1000.[25] Table **??** shows the share of

[25]The total cost measure abstracts from time discounting as the fee can be amortized over the duration of

cost-dominated products by fee categories. Overall, only 14.5% of products are within £500 of the cheapest product in a given month. While about half of all products with a low to medium (up to £1000) fee are strongly cost-dominated, this is true for more than 90% of all products with higher fees.[26]

Fee heterogeneity. Next, I examine the pricing patterns that underlie the cost dispersion. Figure **??** shows the share of products on offer, by double-sorting all products by interest and fee quintile in a given month. It gives a sense of the most common type of products on offer. Similar to the illustrative example above, there are two product clusters in general: one with low interest rates (lowest interest quintile, bottom row), but with medium to high fees (third to fifth fee quintile), and another with relatively low fees (lowest two fee quintiles), and medium-level interest rates. The prices show a pattern of horizontal differentiation along the fee dimension, reflecting products with similar interest rates, but higher fees.[27] Figure **??** illustrates the heterogeneity across fees based on a histogram of fees for 2-year fixed rate, 75% LTV products, with clusters at £0 and £1000 and substantial variation in between and beyond £1000, with the largest fees at around £3000 to £4000. Overall, this illustrates that mortgage products on offer tend to have similar interest rates, but appear differentiated along the fee dimension.

Mortgage pricing. The above pattern reflects that in the UK market, each lender typically offers product variants with different price types based on an interest rate-fee trade-off, e.g. a high interest rate (r_h) and low fee (f_l), compared to a low interest rate (r_l) and high fee (f_h) product. The average and median lender offers around 2 of these products, and the analysis focuses on the case of two price types for simplicity, but can be easily generalized.[28] For a given borrower with a fixed loan size, the pricing scheme hence implicitly defines loan value cut-offs L^* at which a borrower should be indifferent between the $\{r_h, f_l\}$ and $\{r_l, f_h\}$ product, i.e. is

the loan, and does otherwise not affect the preference ranking across products.

[26]Figure **??** in the appendix demonstrates that this pattern holds across the loan size distribution. And Table **??** shows that a large share of products with fees remains cost-dominated even for borrowers with a high loan value of £250,000 (around the 90th percentile of the loan size distribution). One notable difference is that the share of low fee cost-dominated products also increases for high loan values, as the interest rate differential gets magnified at higher loan values.

[27]Figure **??** further gives a sense of how expensive these products are compared to the cheapest product in a given month. The cost differential is naturally lowest close to the left lower corner where both interest rates and fees are low, and increases most visibly along the fee dimension. High fee products command a £3000 to £4000 premium on overage, compared to the cheapest product.

[28]The minimum lender-quarter observation has only one product on offer, while the maximum is 5 products.

indifferent between paying the higher fee to obtain the interest cost reduction, which is greater for larger loan balances, or paying a lower fee but higher interest rate. L^* is obtained by solving for the loan value for which $C^{(r_h,f_l)}$ is equal to $C^{(r_l,f_h)}$ using the total cost equation (**??**) from above, which gives

$$L^* = \frac{-(f_h - f_l)}{d \cdot \left(\frac{r_l}{1-\frac{1}{1-(1+r_l)^T}} - \frac{r_h}{1-\frac{1}{1-(1+r_h)^T}} \right)}.$$

This suggests that a given borrower with a fixed loan size should only consider one product per lender, and that the products available for a given borrower should be segmented according to loan size. However, this does not seem to be reflected in the way the product market is structured in practice. The implied loan value cut-offs vary strongly across lender and over time, and products appear marketed with an overall interest rate ranking in mind. That is, prices in the UK mortgage market tend to be prominently framed in terms of the cheapest interest rates available in the market, e.g. via "best buy" tables, and many price comparison websites sort products by interest rate cost by default, as shown in Figure **??**. So while interest rates are made visible as the most salient price dimension, fees are often shown as separate price category, and in this example, are only specified in the footnotes.[29] The emphasis on the overall interest rate ranking leads to products with higher fees being more visible in the market overall, as illustrated in the following. Figure **??** shows the set of available products in a particular month, first as a pure interest rate ranking, and then in interest rate and fee-space (Figure **??**). The top 10 products with the lowest interest rates are clustered very tightly together when considering just the interest rate dimension, but all belong to the group of products with "higher fees", with some having fees up to £2000.

Hence the pricing structure may allow cost-dominated products (with fees) to compete with cost-minimizing products in the salient price dimension, interest rates, even though these products would not be intended to compete if borrowers correctly ruled out dominated products by applying the cut-off rule based on their loan size. Figure **??** also illustrates the interest rate-fee combinations that yield the same total cost for the borrower (with average loan value $\bar{L}$) as isocost curves $\bar{C}^{\bar{L}}$. It shows that for an average borrower, focusing narrowly on the interest rate

[29] Composite cost measures are not necessarily readily available. The APR, for instance, is measured over the full amortization period (usually 25 to 30 years), which is often not representative of a mortgage that is refinanced after the end of the initial fixation period.

ranking tends to be misleading as the cost-minimizing choice tends to be a product without fees (green squares). The excess total cost paid for a given product compared to the cost-minimizing product can be read as the distance from the cost-minimizing isocost curve, which is at least £500 and up to £2000 for most of the interest-rate minimizing products in this example from the data.[30] Over the sample period more generally, only a relatively small sample of borrowers with large loan values should consider products with fees, in particular high fees, in the first place, if they minimize total cost rather than interest rate cost.

Within-lender price dispersion. One immediate consequence of the pricing structure is that the product alternatives by price types can be a source of price dispersion even *within* a given lender. Figure **??** shows the log-distance of the total cost of the product chosen, and the cheapest alternative, for the same lender at the same point in time. If price dispersion only arose from choice across lenders, this distribution would be degenerate and collapse to zero. In the data, the average within-lender price dispersion is around £400, with substantial variation across lenders. This confirms the idea that borrowers with a fixed loan size are free to pick any of the products by a given lender, and may overpay in that process. Within-lender price dispersion reflects variation in prices paid while holding constant any lender-specific unobservable preferences as there are no known services or other unobserved product characteristics and eligibility criteria that differ between a product with a low or high fee within a given lender. This provides some suggestive evidence that consumers do not pick across lenders according to the loan value threshold rule, and may not fully consider total cost as the sum of interest cost and fees. I test this idea more formally in the following two sections.

2.4. Borrower-Level Analysis

This section develops the first part of the analysis, which is to study consumer choice as a cost-minimization problem based on observed interest rates and fees (taking choice of product characteristics and loan demand as given), and controlling for other lender and borrower-specific factors. I use a discrete choice model to create a simple empirical benchmark for predicted choice

[30] I verify that this tends to hold more generally and over time, the loan value cut-off from which it is worthwhile to pay a medium fee (up to £1000) is at around the 75th percentile of the actual loan value distribution, while a loan value needs to be at around the 90th percentile for it to be worthwhile to pay a high fee (greater than £1000).

which I compare to observed market shares. The approach has recently increased in prominence to study mortgage market demand (**?**, **?**).

2.4.1. Discrete Choice Model Setup

Suppose borrower i's utility from a mortgage that is homogeneous in terms of observable product characteristics, but differentiated by lender and borrower-level factors, by lender j with mortgage pricing type k is

$$U_{ijk} = \underbrace{-\gamma C_{jk} + \beta X_{ij} + \xi_j + \eta_{ij} + \epsilon_{ijk}}_{\delta_{ijk}},$$

where $k \in \{1, 2\}$ denotes the type of mortgage: $k = 1$ if the consumer chooses the low rate, high fee combination $\{r_l, f_h\}$, and $k = 2$ if the consumer chooses $\{r_h, f_l\}$. Utility decreases in total cost C_{jk} which is the sum of interest rate cost and fees, and increases in observable borrower-lender-specific variables captured by X_{ij}, most notably the (negative) distance to the nearest branch of lender j based on borrower i's location, and ξ_j, η_{ij} and ϵ_{ijk} are respectively lender, borrower-lender and borrower-lender-type-specific unobservables. As is common in the literature, the random utility shock ϵ_{ijk} is assumed to be i.i.d and follows a type 1 extreme value distribution (**?**).[31] I define choice indicators as

$$y_{ijk} = \begin{cases} 1, & \text{if } i \text{ chooses product } k \text{ by lender } j, \\ 0, & \text{otherwise.} \end{cases}$$

This yields a closed-form solution for choice probabilities and market shares of logit form for each product type by lender:

$$s_{jk} = \frac{\exp \delta_{jk}}{\sum_{m=1}^{M} \exp \delta_m}.$$

Borrowers choose from the largest lenders J, where I group the 27 largest lenders that I study as the big six and three other lender groups, and the remaining lenders as the outside option.

[31] This implicitly assumes that the odds ratio between any two product types does not depend on the number of products available and poses restrictive assumptions on cross-product substitution patterns. These could be alleviated using a nested logit or random coefficients logit approach.

I aggregate the choice set and matched observed choices such that each lender offers a high ($\{r_l, f_h\}_j$) and low fee ($\{r_h, f_l\}_j$) product, which is a simplification for some lenders, but matches the average and median menu. Each borrower chooses the utility-maximizing product from the resulting choice set M in a given quarter, with $M = 2 \times J = 20$. Note that I am mainly interested in choice across pricing types, so I assume for identification that any unobserved variation in product characteristics that may jointly affect pricing and choice is at the lender level (ξ_j), such as an unobserved brand preference for a given lender that I can control for, and that unobserved borrower-lender-specific unobservables (η_{ij}) are i.i.d. mean zero when controlling for observed borrower-lender-specific nearest branch distance.[32]

I estimate the model on a main sample of all borrowers with a 2-year, maximum 70 to 75% LTV loan between 2015 to 2016, who fulfil a set of standard characteristics (age 25 to 45, loan-to-income-ratio ≤ 4.5, amortization period ≤ 30 years) in order to come as close to the homogeneous product choice benchmark as possible. I also do robustness checks on a sub-sample of borrowers who do not bunch at the 75% LTV threshold, as that may be an additional demand factor behind products with fees in order to borrow the additional fee amount, and results are also robust to both using an unrestricted or more restricted sample.

2.4.2. Discrete Choice Model Results

I estimate alternative-specific logit models with results shown in Table **??**. Columns (1) to (5) show results based on the main sample. Column (1) estimates a simple model only including total cost, which yields a coefficient of -0.20. This can be understood as a baseline cost sensitivity, indicating that borrowers indeed dislike total cost. Columns (2) to (5) estimate alternative model specifications. Column (2) estimates a model that accounts for the borrower-specific distance to the nearest branch of a given lender directly, which is significantly negative and picks up part of the total cost sensitivity. This suggests that borrowers indeed have a preference for convenience or familiarity. Column (3) includes lender fixed effects, with results similar to the baseline cost sensitivity, suggesting that cost sensitivity for choice within a given lender is not more sensitive than when choosing across lenders. Column (4) includes an indicator for the low interest, high fee product type which is significantly positive and the coefficient

[32]In addition, while it would be possible to include interest rate cost and fee components separately to estimate demand elasticities for both price components, the price components are not set independently in this context in contrast to other settings, e.g. comparing salient and non-salient taxes as in **?**.

on total cost increases (in absolute terms) to -0.26, suggesting that borrowers have significant additional demand for high-fee products. Lastly, column (5) includes borrower-specific controls for age and income, which increases the cost sensitivity coefficient to -0.32, consistent with the idea that there is substantial heterogeneity in cost sensitivity across borrowers. One additional demand-driven reason why borrowers may have a preference for high fee products is that they can borrow the fee, allowing them to bunch just at the LTV threshold as the fee is not counted towards the LTV balance (as discussed in **?**). The intuition is confirmed by re-running the analysis with a subset of borrowers who do not bunch at the LTV threshold (Columns (6) to (10)).[33] But while the positive demand coefficient for high-fee products decreases by a factor of around 2/3, it remains sizeable and significantly positive.

One way to illustrate the results for demand across price types is to compare the fit of model-predicted market shares with realized market shares, by price types. I hence compute the model prediction error e_{jk} as the difference between the actual and predicted market shares $(s_{jk} - \hat{s}_{jk})$ using different model estimates of $\hat{s}_{jk}$, as discussed above. I then average these across lenders for the low fee and high fee products ($k = 1$ and $k = 2$) in Figure **??**. In the Figure, Model (1) to (4) correspond respectively to columns (1) to (4) in Table **??**. The model seems to systematically over-predict low-fee market shares compared to the data (negative prediction error), and under-predict high-fee market shares (positive prediction error). It shows that borrower-lender specific variables such as distance to nearest branch (2) and lender-specific fixed effects (3) cannot explain the systematic prediction errors across lenders. The average difference in errors only disappears once explicitly controlling for product-type fixed effects (4). The findings seem indicative of an "unexplained" demand for products with high fees that cannot be explained in the data if consumers weigh interest costs and fees equally, as is the assumption in the model. The caveat is that these results should be considered in the spirit of a calibration exercise, as it does not yet take into account any further challenges to identification, for instance if fees are differentially related to unobserved preferences and product characteristics than interest rates. However, to the extent that these unobservables do not vary within a given lender, the specification with lender fixed effects remains a puzzling result.

Overall, the demand analysis provides indirect evidence for fees being a less salient price component than interest rates, and does not account for more complex challenges to identification,

[33] Excluding borrowers who bunch at the LTV threshold which are about 23% of borrowers in the main sample.

such as time-varying variation in unobserved preferences. In order to address these concerns, the following section studies lender pricing directly.

2.5. Lender-Level Analysis

There are a number of competing explanations behind lender fee pricing. In order to distinguish supply-side pricing strategies from demand-driven mechanisms such as unobservable preferences or product characteristics, and to differentiate behind different supply-driven explanations, this section studies lender pricing in response to lender-specific time-varying funding cost shocks. I provide evidence that a lender-specific cost shock, i.e. a relative deterioration in the competitive position of the lender, is associated with significantly higher fees, in particular for the high fee, low interest rate product, while interest rates remain broadly unchanged.

2.5.1. Pricing Incentives With Non-Salient Fees

If consumers underreact to fees but emphasize a lender's position in the interest cost ranking, lenders have an incentive to adjust pricing accordingly. The pricing incentives are illustrated in Figure **??**. More specifically, suppose lender j receives a shock to its funding cost, which rise from isocost level $\bar{C}^1$ to $\bar{C}^2$, where $\bar{C}^2 > \bar{C}^1$, while prevailing average market interest rates and fees $\{r_l^{avg}, f_h^{avg}\}$ remain at $\bar{C}^1$, meaning it is a firm-specific shock. It decides how to reprice its products following the shock from $\bar{C}^1$ to $\bar{C}^2$. If consumers emphasize the interest cost ranking, its lowest interest rate product (r_l, f_h) is particularly relevant. It can reprice the product by raising interest rates, raising fees, or a combination of both. However, if consumers neglect fees, it is as if they narrowly consider "iso-interest-cost" curves $\bar{R}$, where $\bar{R}^2 > \bar{R}^1$, which are flat across the fee dimension, instead of considering total cost and hence iso-(total)-cost curves. The lender hence has an incentive to reprice its products to dominate in the non-salient price dimension fees (f_h'), while maintaining a competitive low interest rate r_l^{avg} at iso-interest-cost level $\bar{R}^1$, if sufficiently many consumers neglect fees.

This strategic pricing prediction in response to firm-specific cost shocks is further useful to differentiate it from a different plausible mechanism, which is that lenders may have time-varying preferences for fee income, for instance to boost their reported current earnings. This

would predict that lenders move "along" the same isocost curve to increase fee income, while decreasing interest income, all other things equal. Since total cost is held constant, this would attract the same borrower base and increase current fee income while decreasing interest income, for all products, not just the $\{r_l, f_h\}$ product, and is hence distinct from the prediction for strategic pricing.

2.5.2. Identification Strategy

In order to test these predictions and study pricing in response to supply shocks, I require a supply-side cost shifter. Work by **?** illustrates that the largest marginal cost component for mortgages in the UK is funding cost, where long-term wholesale debt is typically considered the marginal source of funding due to its more elastic supply compared to retail deposits. Based on this, I construct a *lender-specific* funding shock using a lender's pre-determined past loan-to-deposit ratio as a measure of its dependence on wholesale funding,[34] interacted with aggregate (exogenous) changes in wholesale funding costs.[35]

The exogeneity of this so-called "shift-share" shock can be assured in two ways: first, as suggested by **?**, lender-specific exposures to wholesale funding need to be relatively sticky and as-good as randomly assigned after controlling for observables, i.e. loan-to-deposit ratios need to be uncorrelated with lender-specific characteristics conditional on controls. One way to do such a balance test is to regress the funding shock on lagged levels and changes of lender characteristics, which is reported in the appendix.[36] None of the lender characteristics (including size, return on assets, net interest margin and leverage) seem systematically correlated with the funding shock, especially not when measured in changes.[37] Second, recent work by **?** suggests that this assumption can be relaxed as long as the aggregate shock is exogenous, with exposure sufficiently dispersed. Based on this idea, I further instrument aggregate variation in

[34] As a related application, **?** use pre-crisis variation across lenders in the loan-to-deposit ratio in a difference-in-differences setup to compare banks which are relatively more exposed to the wholesale funding shock of the 2007-2008 financial crisis to those that are relatively less dependent on wholesale funding.

[35] This is akin to a **?** shock, originally using local industry employment shares × national industry employment growth rates as an instrument for labour demand (**?**), commonly used in the trade and labor literatures.

[36] Table **??** shows results for this exercise, based on individual years.

[37] In addition, another concern could be that lenders who are dependent on wholesale funding differ systematically from lenders who are not, in terms of both depositor and borrower characteristics. The results show substantial variation even within lender types (e.g. large universal banks with similar business models and funding structures), suggesting that the pre-determined difference in loan-to-deposit ratios is not the primary driver of the main findings, further discussed in Section **??**.

UK long-term wholesale funding cost, 2-year LIBOR swap rates (ρ_t^{UK}), using 2-year US dollar swap rates (ρ_t^{US}), and construct the aggregate funding cost shock using the fitted values from a regression of ρ_t^{UK} on ρ_t^{US}, denoted $\hat{\rho}_t^{UK}$. This ensures that aggregate variation comes from changes in US wholesale funding markets that are plausibly exogenous to UK mortgage market dynamics. Following **?**, long-term wholesale funding cost are then approximated as 2-year swap rates plus senior CDS spreads (s_{jt}) where I use lender-specific CDS spreads for the largest six lenders, and for all other lenders, I use the average CDS spread ($\bar{s}_t$).

The lender j, time t-specific funding cost shock, denoted ϕ_{jt}, is then constructed as the lender-specific loan-to-deposit ratio in 2008 (one year prior to the start of my analysis), $ltd_{j,2008}$, based on annual balance sheet data, interacted with instrumented wholesale funding cost:

$$\phi_{jt} = ltd_{j,2008} \times \left(\hat{\rho}_t^{UK} + \bar{s}_t\right) \tag{2.2}$$

Identifying variation comes from cross-sectional variation in wholesale funding shares, cross-sectional variation in wholesale funding cost for the largest six lenders, and variation in instrumented aggregate wholesale funding cost over time.

One possible concern is that CDS spreads may not be fully exogenous to contemporaneous mortgage pricing strategies. As explained in **?**, banks' operations may usefully alleviate links between funding and mortgage markets within a given bank. Lenders usually centralize their funding operations within a treasury department across the institution, which makes funding available to other business units, who further decide on business-specific lending margins (known as "transfer pricing"). That makes it more likely that for instance the risk strategy chosen for the mortgage market is at the most an outcome of shocks to funding cost, but not the other way round.[38] This is in line with the idea that banks can be considered as "price takers" in wholesale funding markets, in particular from the perspective of the mortgage business unit over the main sample period.

[38]In addition, my analysis focuses on analysis within homogeneous mortgage product categories, such as 70-75% LTV, where default risk is low. Within LTV band variation in default risk is explicitly not priced (as per the discrete pricing scheme commonly used), making the risk adjustment channel within LTV band in response to changes in CDS spreads likely to be small in the first place. And in most of the sample period from 2010, bank CDS spreads appear to be driven by banks' exposure to systemic factors such as the Euro Area sovereign debt crisis, that have limited links to the domestic mortgage market.

The identifying assumption for the overall identification strategy is

$$E\left[\epsilon_{jt} \mid \phi_{jt}, \gamma_t, \theta_j\right] = 0, \tag{2.3}$$

i.e. time-varying unobservables such as lender-time-specific demand shocks should not be correlated with the funding shock.[39]

I construct the dependent variables as follows. I track the minimum fee ($\{r_h, f_l\}$) and minimum rate ($\{r_l, f_h\}$) products for 2-year fixed rate, 70-75% LTV mortgages for first-time borrowers, per lender over time, and compute the average rate, fee and total cost (over one and two years) for each lender and quarter. I then merge these pricing statistics to lender-specific funding cost shocks to create a quarterly lender panel from 2009Q1 to 2016Q4. The main specification regresses changes in the outcome variables on changes in the funding cost shock ϕ, and γ_t and δ_j are time and lender fixed effects, respectively:

$$\Delta y_{jt} = \alpha + \beta \cdot \Delta\phi_{jt} + \gamma_t + \theta_j + \epsilon_{jt}, \tag{2.4}$$

where the outcomes are $y \in \{r, f, C_{1yr}, C_{2yr}\}$, for each price type and for the average across all products. The specification studies how lenders respond to firm-specific shocks to their competitive position, as aggregate shocks and lender-specific levels are absorbed in the fixed effects. While aggregate shocks such as changes in aggregate financial conditions are expected to be passed through, the pricing incentives with non-salient fees generate predictions as to how lenders respond to shocks that change their relative competitive position, in particular by matching prevailing interest rates while increasing fees when funding costs increase.

2.5.3. Results

Pricing in response to funding cost shocks. Table **??** reports results for the main regression analysis. Column (1) to (4) show results for average interest rates, fees and total cost across products. In response to a one standard deviation change in the funding cost shock, interest rates remain unchanged, while fees and total cost (at the one year horizon) go up significantly, by around £40 and £50, respectively. This seems largely driven by pricing changes for the

[39] By construction, the lender-specific loan-to-deposit share is predetermined, and the aggregate funding shock is not driven by firm-specific decisions or should be exogenous to mortgage pricing decisions.

low rate, high fee product: while r_l (Column (5)) remains unchanged or if at all decreases a little, f_h (Column (6)) increases significantly by £110, which is also economically significant at around 15% of the average level of fees and a 0.5 standard deviation increase in fees. Total cost over one and two years increase by £102 and £94, respectively (Column (7) and (8)). This is in contrast to the high rate, low fee product: while total cost does seem to increase for a similar amount as the low rate, high fee product, albeit not significantly so, this seems driven by increases in r_h. The results suggest that lenders maintain the relative pricing of their low interest rate, but increase fees in response to a deterioration in their competitive position. The fact that the overall increase in fees is driven by the highest fee product is also intuitively consistent with the idea that competing for the lowest interest rate in the market is important - which can be partly achieved by increasing fees. I later also provide evidence that high-cost lenders appear more dependent on the high fee product type.

Pricing of low rate, high fee product by lender pricing type. Next I test if the dynamic pricing strategy is related to the heterogeneity in fee levels across lenders, i.e. if lenders who rely on high fees also adjust their fees more in response to a relative cost shock. I define lender pricing types based on how much prices are skewed towards the interest rate dimension via fees that are higher than the prevailing average, using terciles of the ranking difference in the total cost minus the interest rate ranking position. Lenders in the "low" category have a negative skew and hence relatively low fees (Columns (1) to (4)), "medium" lenders' position in the total cost ranking is similar to that in the interest rate ranking (Columns (5) to (8)), and "high" lenders have a substantially better (i.e. lower) position in the interest rate ranking compared to the total cost ranking, i.e. have a high skew in their prices using high fees (Columns (9) to (12)). I find that the price adjustment in the low rate, high fee product seems indeed strongly driven by lenders with skewed price levels. While there is no significant adjustment in interest rates or fees in response to a relative funding cost shock for "low" skew types, fees respond marginally significantly for "medium" pricing types, and are strongly significant for lenders with highly skewed prices. The latter raise f_h by £201 (Column (10)) and total cost over two years by £144 (Column (12)) in response to a one standard deviation cost shock, while r_l remains relatively unchanged.

Robustness. The results are robust to using UK swap rates, different pricing statistics (e.g.

median, minimum and maximum rates and fees of lender menus) and similar when omitting lender fixed effects. In addition, the analysis so far has focused on low risk mortgages for which the homogeneous product market and limited heterogeneity in unobserved borrower characteristics assumptions seem more valid. In contrast, I do not find cost pass-through via fees for riskier mortgages (reported in Table **??** in the appendix). Average total costs across products increase substantially in response to a funding cost shock, by between £80 to £200, but this is almost entirely driven by increases in interest rates. This may suggest that pricing strategies differ across less risky and riskier LTV markets and may depend on the competitive structure and borrower population of a given LTV market. An intuitive explanation could be that if the adverse selection problem is much worse for high LTV loans, lenders may not want to attract borrowers based on a low interest rate with high fees.[40]

Overall, I find evidence that lenders maintain their relative pricing of interest rates following a cost shock and increase fees for low risk mortgages, driven by changes in the pricing of the low interest rate, high fee product, and lenders who have high average levels of fees and hence more skewed prices relative to others.

2.6. Interpretation and Discussion

2.6.1. The Role of Fees

There are two conventional views of the function of mortgage origination fees. On the one hand, they could compensate for a fixed cost component of originating a mortgage such as paper work and processing cost. That implies that they should not be related to higher frequency changes in marginal cost such as funding cost. Alternatively, fees could reflect a variable cost of originating larger mortgages. For instance in Denmark, consumers pay a percentage of the loan value in administration fees that depends only on loan characteristics,[41] meaning that a given borrower does not have to compare fees for her cost minimization problem.

[40] There is existing evidence that firms choose rent-extraction strategies differentially across borrower groups, for instance **?** finds that US lenders target existing clients who have high credit scores but seem less likely to switch banks to increase credit card rates, while this strategy is not employed for low credit score borrowers where default risk is the main pricing factor.

[41] E.g. collateral and period of interest rate fixation, see Danmarks Nationalbank, Statistics on Banking and Mortgage Lending, Interests, April 2018.

In the UK, while the interest rate-fee trade-offs imply optimal cut-offs for different loan values, the market is not clearly segmented or standardized by loan size and all product variants appear marketed as pooled, with an emphasis on the overall interest rate ranking. The evidence further suggests that fees serve as an active margin and additional degree of freedom when setting mortgage prices. This affects the direct comparability of total cost across mortgages: instead of comparing mortgage prices using a scalar, where the interest rate is a sufficient statistic for the total interest rate cost and can be compared using a general best-buy table, consumers face a price vector of interest rates and fees. In order to compare total cost across products, borrowers need to add fees to the loan-specific interest rate cost, which depends on the loan amount borrowed, and would require loan-amount-specific best-buy tables. This separation of pricing components (**?**) and limiting of comparability across products (**?**, **?**) may decrease borrowers' total price sensitivity and make search more difficult. In addition, borrowers may be less able to find the cheapest product if both fees and interest rates vary, compared to if they were confronted with a composite (scalar) price measure. This is similar to findings by **?** in an online shopping environment for a homogeneous consumer electronic good, who document a range of case study practices to make search more difficult, including shrouding shipping cost and competing on additional quality dimensions. One interesting implication of their findings could be that without fees, the market would be extremely price-sensitive given the ease of price search if there is a unique price ranking by interest rates. While this counterfactual is unobserved, I provide some supportive evidence by looking at the sample of borrowers who choose a zero fee product, who indeed generate less price dispersion in choice outcomes (Figure **??**).[42]

2.6.2. Pricing Fees in a Specialization Equilibrium

The empirical results point to lenders adjusting interest rates and fees separately in a way that may allow (some) lenders to extract rents from consumers who neglect fees.

The findings are consistent with a standard search framework that yields a specialization equilibrium, in which high-cost lenders are able to attract "uninformed", that is less cost-sensitive consumers, but not informed consumers, who go to low-cost lenders, yielding a separating equi-

[42]This may point to trade-offs between the volatility of bank profits and improved consumer search for policy makers that are outside the scope of this paper.

librium or equilibrium price dispersion (**?**, **?**, **?**). In the following, I discuss how this can be interpreted in the context of a differentiated product market model in line with the borrower-level discrete choice analysis. Suppose lender j maximizes profits

$$\Pi_j = (z_j - \kappa_j) \cdot s_j,$$

where z_j is a fee-inclusive price ("zero-fee interest"), κ_j is marginal cost and s_j is the market share based on borrower demand defined in Equation **??** (ignoring product types k and default risk for simplicity). The first-order condition pins down the standard optimal pricing rule for prices as the sum of marginal cost and mark-up:

$$\underbrace{z_j}_{\text{price}} = \underbrace{\kappa_j}_{\text{marginal cost}} + \underbrace{\frac{1}{\gamma^z(1-s_j)}}_{\text{mark-up}},$$

where γ^z is proportional to the total-cost elasticity from the demand estimation. It does not, however, pin down the optimal split of price components. If lenders are indifferent between interest and fee income, and if borrowers indeed exhibit lower demand elasticities with respect to fees than interest rates, then lenders would choose a corner solution and receive their entire income through fees, with 0% interest rates, which is highly counterfactual. Rather than fees being unconditionally non-salient, it seems intuitive of thinking of the fee elasticity as an (exponentially) increasing function of the level of fees, meaning that consumers become more attentive to fees, the higher they are. If different lenders are able to attract differentially cost-sensitive consumers (with γ_j^z), the marginal benefit of increasing the fee equals the marginal cost at different levels of fees, generating fee heterogeneity.[43] In that sense, lenders with higher equilibrium fees are able to attract less cost-sensitive consumers, while lenders who do not skew their pricing are likely to attract more cost-sensitive consumers, generating a specialization pattern in line with the intuition from search models.

[43] This is related but slightly different to the approach by **?** for the US credit card market, who assume that non-salient fees are capped at an exogenous maximum level as their market setup does not require matching fee heterogeneity. An alternative way to preserve non-degenerate interest pricing as observed in the data is to allow for a fixed price frame r_l, r_h, as outlined in a previous version of this paper.

2.6.3. Evidence for Specialization Equilibrium

Heterogeneity in pricing across lenders. I provide further evidence for this type of specialization equilibrium by splitting lenders into terciles of the difference between their minimum total cost and interest rate ranking, as a measure of how much they skew their interest rate ranking using higher fees compared to other lenders. The results are shown in Table **??**. The upper tercile of lenders have a positive difference of 6 on average, i.e. they have the $x + 6$th-highest position in the total cost ranking, but only the xth highest interest rate ranking. As expected, they charge similar interest rates as their competitors, but higher fees, both for the high fee and low fee product. While they have similar size and funding structures as other lenders, they earn substantially lower returns on assets and equity and have higher funding costs. They also tend to have a higher branch density, measured as the number of households that have a branch nearby, and higher market share, suggesting that existing forms of market power based on branch networks (**?**) interact with those based on consumer mistakes and may make it easier for these lenders to acquire less cost-sensitive borrowers. Across lenders, fees do not seem to compensate directly for a greater branch density, however. Figure **??** and **??** show that there is a somewhat positive association between branch density and interest rates, but not fees, meaning that both the financial necessity and ability to attract less cost-sensitive borrowers seem to affect lenders' pricing decisions.

Dependence on zero-fee and high-fee products. Another way of showing variation in pricing across low and high-cost lenders is to look at their dependence on zero-fee compared to high-fee products. The true cost ranking across lenders should be fully and uniquely revealed (i.e. valid across the loan size distribution) when looking at the interest rate ranking across products without fees. So low cost lenders should have a relatively high market share in products without fees, while high-cost lenders may try to avoid products without fees, as they cannot improve their interest cost ranking by increasing fees. Hence high-cost lenders should also be relatively reliant on products with fees, in particular high fees. This seems to hold out in the data. I measure within-lender market share as the share of products for a given lender in a given quarter that has zero or high fees, out of all products issued by that lender in that quarter (based on mortgage origination data from PSD). As a proxy of a lender's cost-level, I compute the average total cost distance to the cost-minimizing product across all the products

offered by a given lender in a given quarter[44] which I refer to as lender-time-specific excess cost. Figure **??** shows that there is indeed a negative relationship between the within-lender market share of zero fee products and lender-time-specific excess cost, while Figure **??** shows that there is a positive relationship for high-fee products.[45] This provides suggestive evidence that low-cost lenders are relatively more competitive in the zero-fee product market, and that high-cost lenders are relatively more reliant on high-fee products.[46]

Excess cost dispersion in borrowing outcomes. In borrowing outcomes, I compute excess cost dispersion following **?**, as the residual of a regression of total cost on borrower, product and regional characteristics. The excess cost residuals are plotted by fee categories in Figure **??**. I find higher excess cost dispersion for products without fees, compared to products with fees, i.e. borrowers seem to come closer to the cost-minimizing benchmark in the class of products without fees than those with fees. It shows that borrowers with zero fee products indeed seem to have a narrower distribution and hence lower price dispersion, while the distributions for products with fees are more dispersed, and in particular high fees seem to come with a larger right tail of excess cost. This is further suggestive evidence that borrowers appear to underreact to fees and seem better at cost minimization if there is no fee involved.[47]

2.6.4. Discussion of Alternative Channels and Robustness

This section considers additional alternative channels, most notably screening motives using fees which could be correlated with lender-specific cost shocks. I find limited evidence for lenders' use of fees to screen for risk types.

Screening for borrower risk. First, the market setting makes it unlikely that lenders in

[44] Based on the Moneyfacts product data and an average loan value of £150,000 as a reference value for comparison. As shown in Figure **??** in the appendix, the cost ranking does not vary substantially across the loan size distribution, suggesting that some products are unambiguously more expensive than others.

[45] This relationship is not mechanical, as there is an omitted category with medium fee products (0,1000].

[46] Note that the proxy for high and low-cost lender is measured as time-varying, reflecting the idea that lenders' position in the total cost ranking fluctuates over time.

[47] Related work by **?** using a strictly dominated choice criterion finds that borrowers tend to take out mortgages that are strictly dominated in the interest rate dimension, which is a more narrow framework compared to the discrete choice model that allows borrowers to have lender- and borrower-lender-specific demand. The finding could be consistent with the evidence that preference factors, such as the branch network, appear more compensated in the interest rate dimension, while fees may still distort the subjective cost ranking within and across (interest-cost dominated) lenders.

the UK screen for prepayment risk using fees. In contrast to the US, lenders commonly set early repayment penalties, providing them with a more direct way to screen for prepayment risk than by using fees, also known as points (**?**) in the US. There is also limited variation in prepayment penalty terms for products within a given lender. Only 7 out of 27 lenders have any variation across products, and only around 10% of product observations offer different prepayment penalty terms within a given lender at the same point in time. I further confirm that prepayment penalty terms do not significantly alter the interest rate-fee trade-off pricing of a given lender in a regression framework, outlined below. This seeming lack of screening and pricing of heterogeneous repayment risk could be explained by the relatively short initial fixation periods of 2 to 5 years prevalent in the UK, at the end of which most borrowers refinance, as documented by **?**.

Second, lenders could price in potential borrower selection on unobservable characteristics that affect default probabilities, for those who choose a relatively more expensive fee-interest rate alternative. For instance, **?** show that demand for high-fee index funds seems to be primarily driven by mistakes due to financial illiteracy, which could be correlated with unobservable default risk. Lenders may also use products with fees to screen for liquidity risk which may be correlated with default, as highly liquidity-constrained borrowers may be unable to pay an upfront fee. But as seen in the main results, the cost pass-through occurs via fees only for low LTV products, indicating that the mechanism may play more of a role when default risk is low and selection on unobservables may play less of a role.

Complementary factors. Recent work has established advertising (**?**, **?**) and advice (**?**, **?**) as important factors behind dominated consumer choice. In my analysis, these factors are likely to reinforce the mechanism at hand. Interest rates are made salient in the way prices are framed and advertised in the market, while differences in fees are noted much less prominently or simplified as indicator variable in many price comparison tables. Brokers can steer borrowers towards more expensive mortgages due to incentive problems, as they may earn higher commission payments for such products. Lender pricing using high fees can be seen as a strategic complement to any of these channels, as the choice of high fee product could be promoted via advertising campaigns and commissions, and may affect consumers' choice and choice sets even when they process their mortgage through a broker.[48] There does not seem

[48]In the UK, about half of borrowing is processed via brokers (**?**). Changes to advisory requirements man-

to be strong direct link between lender fees and broker commissions in the data, however, as the correlation in the loan sample is positive but small (3%), pointing to a largely independent channel.

Screening for loan size. Lastly, a given lender's pricing scheme does imply that borrowers with different loan sizes should choose different products, i.e. it should serve as a coarse way of screening for loan sizes. Based on offered prices and the loan value distribution over the sample period, only a relatively small sample of borrowers should consider products with fees at all, in particular high fees, compared to picking a product alternative with no or low fees. For the average borrower (borrowing at the average of the loan size distribution at a given point in time), the isocost curve with cost-minimizing fee-rate combinations tends to be steeper than the price terms that are on offer (as seen in Figure **??**), meaning that the cost-minimizing choice tends to be a product without fees or low fees. I can show that the interest rate cost reduction for the average borrower tends to be lower than the fee required to get the lower interest rate more formally in a product-level regression of interest rate cost on fees shown in Table **??**. If a borrower is indifferent between the low fee, high interest rate and the high fee, low interest rate product, the slope coefficient β on fees would be -1, i.e. a £1 paid in fees would yield a £1 interest rate cost reduction. The baseline correlation in column (1) with time and product (LTV) fixed effects is negative, but not significant. Column (2) and (3) saturate the regression further with lender and lender-time fixed effects, in order to measure the interest cost-fee trade-off within a given lender that is on offer at a given point in time. The coefficient is strongly negatively significant, but is only around -0.4. Columns (4) to (6) add additional product-level control variables including the early prepayment penalty, a cashback indicator as an example of measurable additional incentives and the length of the incentive description text. Only the incentive length shows up as positively significant within lender, tentatively suggesting that lenders provide a worse trade-off when the description of incentives is more extensive, consistent with obfuscation motives that make the product more complex (**?**). Lastly, column (7) introduces a square term for fees which is strongly positively significant, capturing the fact that the trade-off becomes worse for products with high fees, as expected.

dated in the FCA Mortgage Market Review in 2014 were found to have had a limited impact on cost (**?**), suggesting that the impact of brokers on chosen product cost may be relatively neutral.

2.7. Conclusion

This paper provides evidence for supply-driven fee dispersion in two steps: first, by showing that borrowers are less cost-sensitive towards products with high fees, in line with a market environment that emphasizes interest rates and leaves fees as a non-salient, but financially significant price component; and second, by showing that lenders internalize this demand-side friction and maintain competitive interest rates, while increasing fees in response to lender-specific shocks to funding cost. The empirical strategy makes it less likely that fees compensate for (time-varying) variation in unobserved preferences or product characteristics. The paper further illustrates heterogeneity in lender pricing strategies: lenders who skew their prices towards low interest rates using higher-than-average fees indeed rely more on business from high fee products, and tend to have higher funding cost, lower return on equity and a bigger branch network, suggesting that the pricing strategy is related to both competitive pressure, as well as the ability to attract less cost-sensitive consumers.

To put these findings into a bigger perspective, they illustrate that incentives of financial intermediaries are important for household financial outcomes (**?**), which could become a more prominent issue with the rise in data availability to single out consumers and their biases (**?**). Pricing of non-salient cost components further relates to the changing nature of competition and market power in consumer goods and retail financial markets, away from physical branch networks (**?**) towards online-based distribution channels and hence an increasing importance of the information available to consumers, and what information consumers pay attention to, with potentially important macroeconomic consequences (**?**, **?**). The work shows that competitive pressure may exacerbate firms' reliance on non-salient price components (**?**) and generate incentives for financial innovation (**?**).

Any policy responses[49] will have to assess the cost and benefits of decreasing price dispersion in the market, such as consumer cost savings through standardizing products against the risk of limiting (potentially welfare-increasing) product innovation, and indirect effects on demand and supply. Policies aimed at informing consumers may be beneficial to the extent that consumers for instance believe that fees are less dispersed than they are in reality, and hence correct beliefs

[49] See **?** for a review.

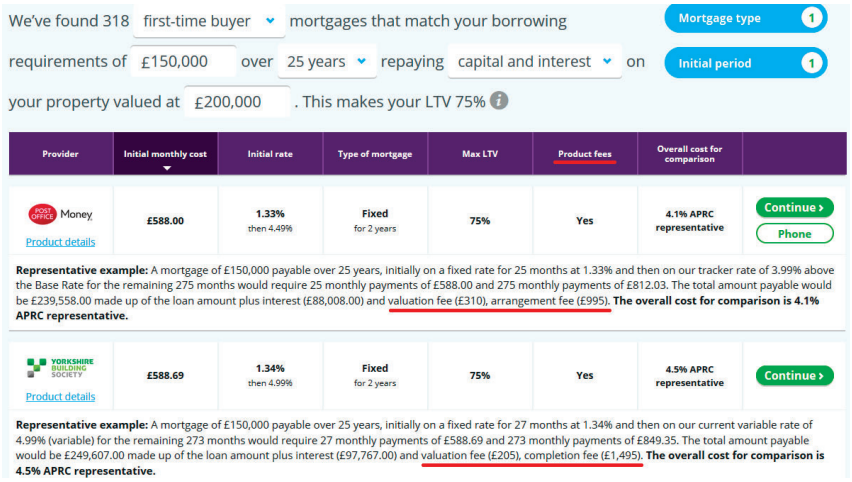
about the marginal benefits of search or induce more efficient search.[50] The framing of cost may also affect the relative salience of cost components, in line with **?**.[51] However, it also risks making certain features salient at the expense of others (**?**), e.g. by introducing a composite price, which firms have an incentive to deviate from.

Overall, these findings pose new questions relating to competition and consumer search in the presence of behavioral biases, and the importance of both supply and demand factors as well as their interaction effects, which seem fruitful topics for future work.

[50] The effectiveness of such a policy may depend on what motivates consumers to search and what type of information would improve their search behavior (**?**).

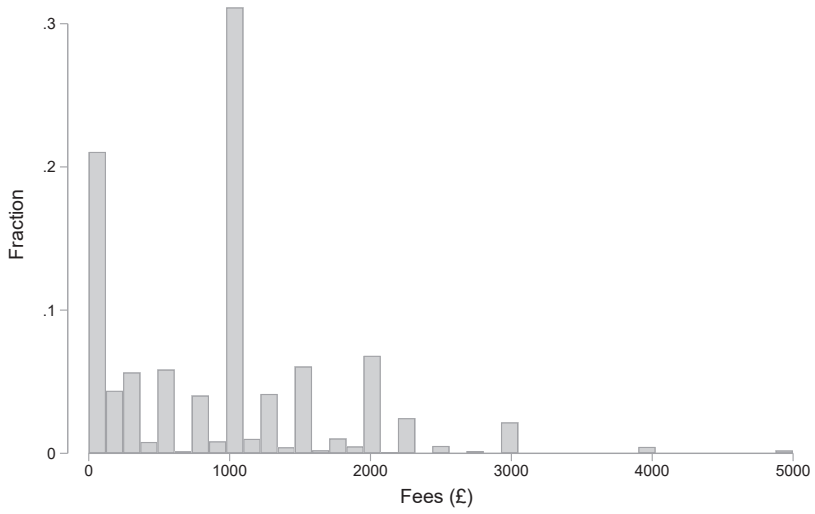
[51] For instance, (relative) differences in fees may be less salient than differences in interest rates, which may be reversed if interest rates were quoted in interest cost terms.

FIGURE 2.1: EXAMPLE PRICE TABLE FROM MONEYSUPERMARKET.COM



Notes: This figure shows a typical price table for mortgage products, based on the search result for a 2-year fixed rate mortgage product, with maximum 75% LTV, for first-time buyers and a loan size of £150,000, from MoneySupermarket.com. The default ranking is by interest rates, with a fee indicator and the value of fees in the footnotes.

Figure 2.2: Distribution of Fees



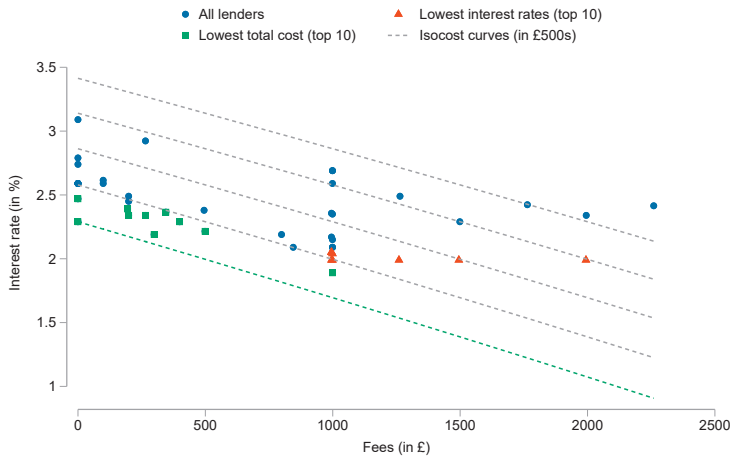
Notes: This figure shows a histogram of fees, based on all products on offer for 2yr fixed rate, 70-75% LTV products over the sample period between January 2009 and December 2016.

Figure 2.3: Example of Products on Offer

(a) Interest rate ranking (w/o fees)

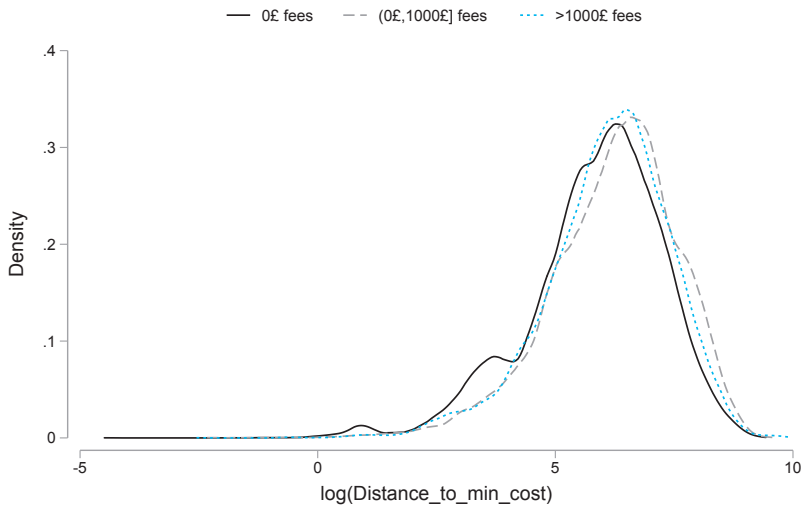


(b) Interest rates, fees and total cost curves



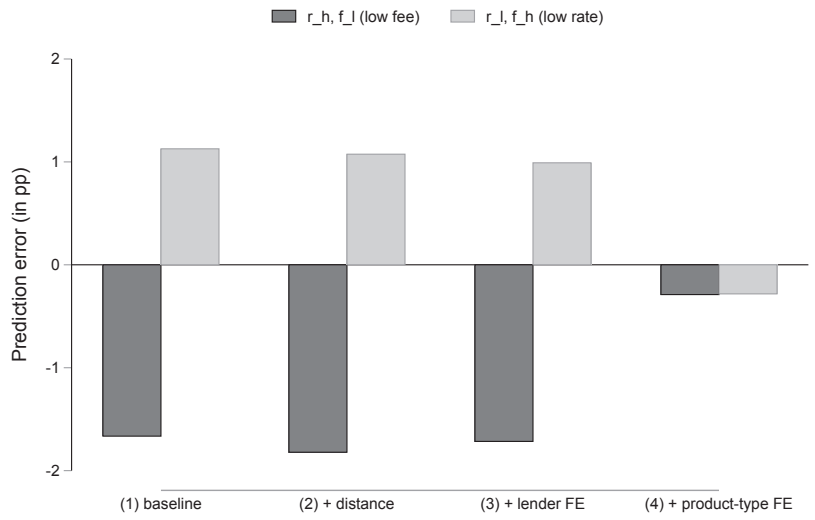
Notes: Figure **??** shows the set of 2yr 70-75% LTV products on offer in February 2014 as a pure interest rate ranking, when ignoring fees. Figure **??** shows the same set of products in interest rate-fee space, together with isocost-curves based on the total cost of a given product over two years and the average loan size in that month, amortized over 25 years.

Figure 2.4: Within-Bank Price Dispersion



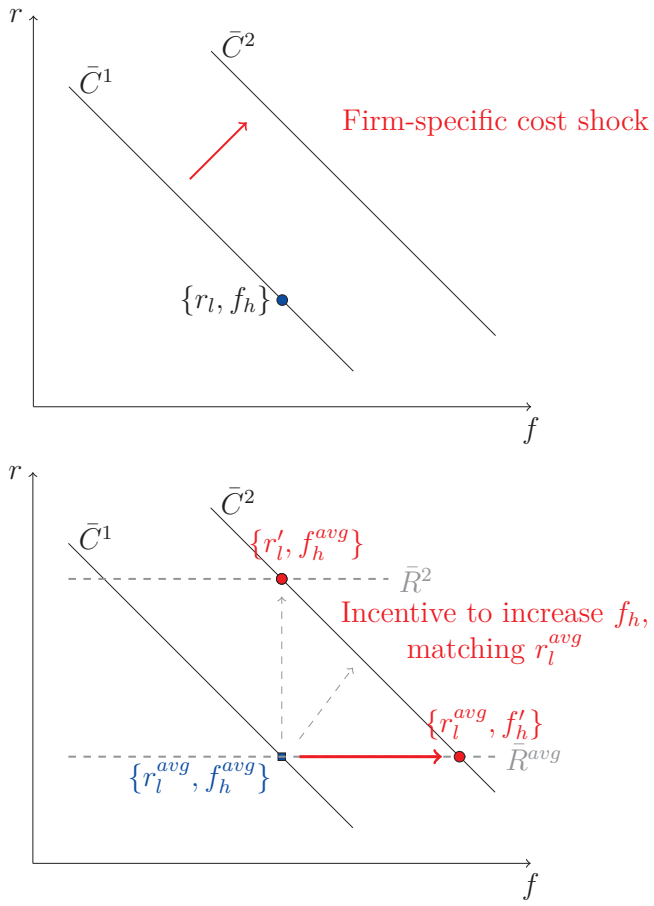
Notes: This figure shows the distribution of the log distance of the cost of the chosen product type (no fees, medium and high fees), compared to the cheapest product alternative available, for a given lender at a given point in time. It hence captures price dispersion "within" a given lender, which would be a degenerative distribution centered at 0 if borrowers always chose the cost-minimizing product for a given lender and point in time. The figure is based on realised 2-year fixed rate, low LTV loan outcomes from 2015 to 2016.

FIGURE 2.5: LOGIT MODEL PREDICTION ERROR (BY PRODUCT TYPE)



Notes: This figure shows the average logit model prediction error (e_k), defined as the difference between the realized and predicted market shares (where $e_{jk} \equiv s_{jk} - \hat{s}_j k$), across all lenders, by product type k ($\{r_h, f_l\}$ or $\{r_l, f_h\}$), for four model specifications as estimated in Table **??**. Models (1) to (4) correspond to column (1) to (4) in Table **??**, respectively. It shows that borrower-lender specific variables such as distance to nearest branch (2) and lender-specific fixed effects cannot explain the systematic prediction errors for low fee-type products ($\{r_h, f_l\}$), compared to low rate but high fee-type products ($\{r_l, f_h\}$). The average difference in prediction errors only disappears once explicitly controlling for product-type fixed effects (4).

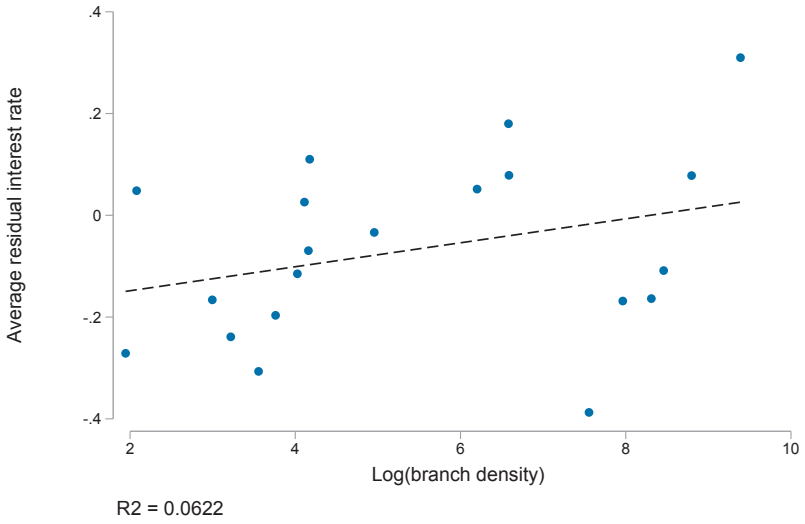
FIGURE 2.6: ILLUSTRATION OF PRICING WITH NON-SALIENT FEES



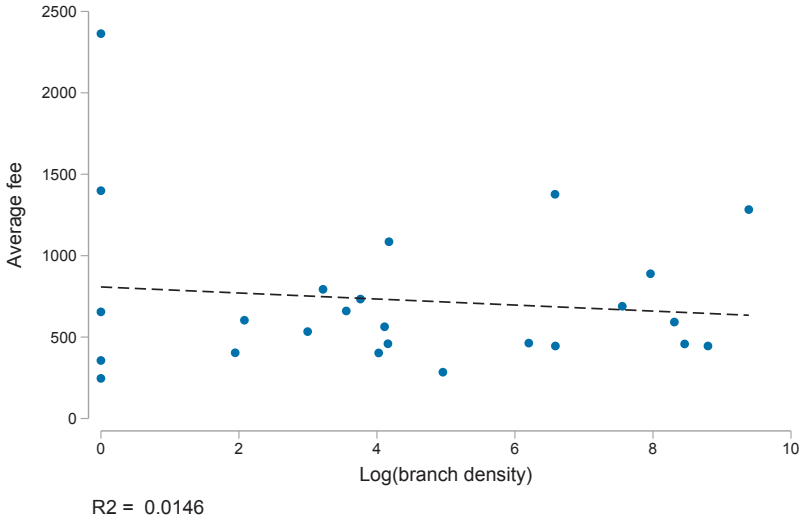
Notes: This figures illustrates the pricing incentives of a lender following a firm-specific shock to funding cost, from isocost level $\bar{C}^1$ to $\bar{C}^2$, where $\bar{C}^2 > \bar{C}^1$. The now cost-dominated product (relative to prevailing average interest rates and fees $\{r_l^{avg}, f_h^{avg}\}$) is repriced to dominate in the non-salient price dimension, fees (f_h'), while maintaining a competitive interest rate (r_l^{avg}), if consumers neglect fees. Non-salient fees can be interpreted as consumers narrowly considering "iso-interest-cost" curves $\bar{R}$, where $\bar{R}^2 > \bar{R}^1$, which are flat across the fee dimension, compared to when consumers consider total cost.

Figure 2.7: Average Interest Rates and Fees and Branch Density

(a) Interest rates

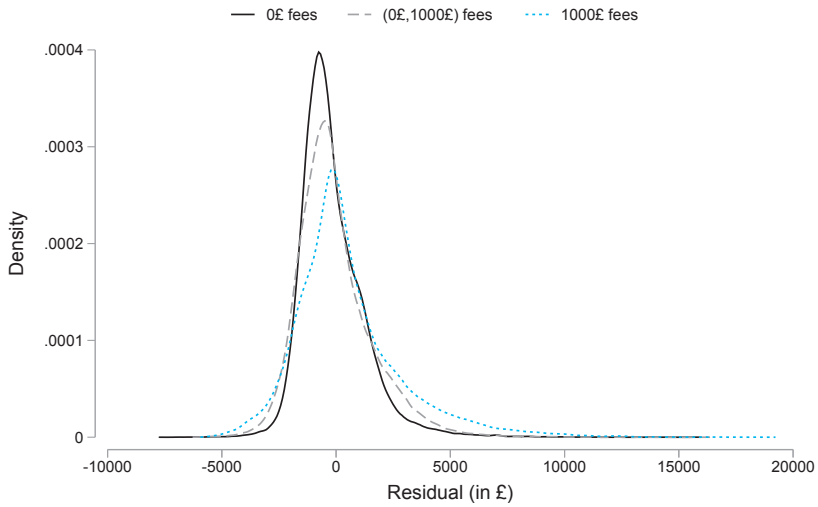


(b) Fees



Notes: This figure shows a scatter plot of the average residual interest rate (with monthly average levels partialled out) and fee across the full sample period from 2009 to 2016, respectively, and log branch density. Branch density is measured as the number of borrowers for whom a given lender operates the closest nearby branch (yielding similar distributions when using up to second, third etc. closest branches). It suggests that higher branch density is partially compensated via higher interest rates, but less so higher fees, where the relationship is flatter.

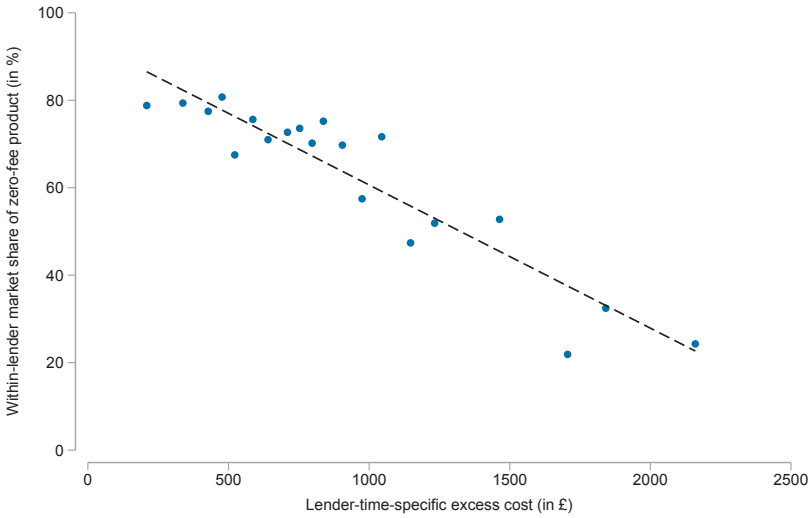
FIGURE 2.8: EXCESS COST DISPERSION BY FEE CATEGORY



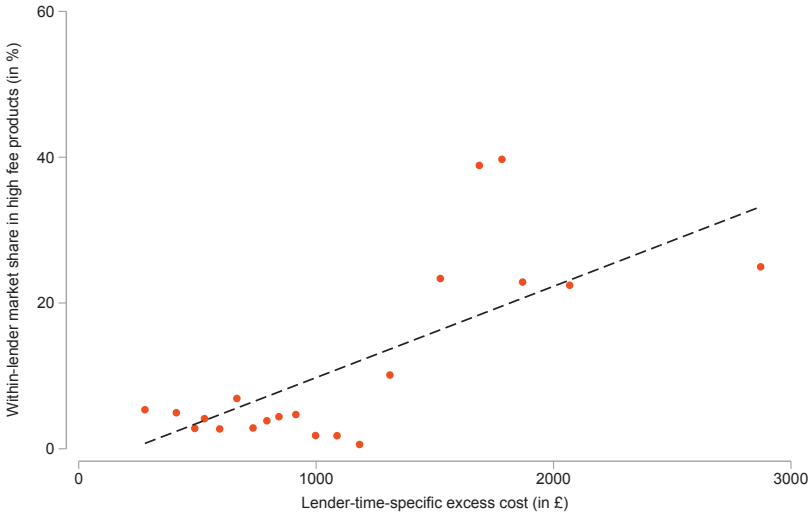
Notes: This figure shows the densities of excess cost by fee categories (low, medium and high fees). Excess cost is measured as the residual e_{ijt} from a regression of loan-level realized total cost on LTV band, year-month and region fixed effects, using loan-level data from 2015-2016: $C_{ijt} = \alpha + \beta X_{it} + \gamma_t + \theta_j + \epsilon_{ijt}$. It measures the dispersion in cost paid when partialling out other observable characteristics, which seems to be larger, with a bigger right tail for high fee products (>£1000).

Figure 2.9: Relative market shares by lender excess cost level

(a) Relative Market Share In Zero-Fee Product



(b) Relative market share in high-fee product



Notes: Figure **??** shows a binned scatter plot of within-lender market share (share of products in a given fee category out of all products issued by the lender in a given quarter) in the zero-fee product by lender-time-specific excess cost, measured as the difference between a given lenders' average total cost of mortgages on offer and the cost-minimizing product in a given quarter. Figure **??** shows the equivalent binned scatter plot for high-fee products where high fees are defined as fees $>$ £1000.

Table 2.2: Descriptive Statistics

Panel 1: Summary of Moneyfacts variables (monthly)					
Variable	Observations	Mean	Std. Dev.	Min	Max
Interest rate (in %)	26451	3.36	1.24	0.99	7.89
Max LTV, first-time buyer	26451	77.43	10.04	60.00	95.00
No additional product fees (indicator)	26451	0.21	0.40	0.00	1.00
Arrangement fees (in £)	9423	1009.34	449.48	99.00	3995.00
Booking fees (in £)	10987	465.78	520.76	95.00	4999.00
Completion fees (in £)	7286	506.93	362.09	50.00	2400.00
Reservation fees (in £)	735	563.97	298.93	99.00	999.00
Fees (sum of fee components)	26451	708.43	611.46	0.00	4999.00
Panel 2: Lender characteristics (annual)					
Variable name	Observations	Mean	Std. Dev.	Min	Max
log(total assets)	212	17.78	2.07	12.39	21.71
Return on assets (in %)	212	-0.07	1.47	-10.57	4.91
Return on equity (in %)	212	0.10	15.95	-74.39	39.96
Loan to deposit-ratio	212	1.06	0.30	0.28	2.59
Net interest margin (in %)	212	1.41	0.71	0.30	4.57
Panel 3: Lender panel variables (quarterly)					
Variable name	Observations	Mean	Std. Dev.	Min	Max
Levels					
Wholesale funding shock	696	9.51	9.49	0.29	70.94
Avg. interest ($\bar{r}$)	696	3.04	0.92	1.42	5.92
Avg. fees ($\bar{f}$)	696	689.91	385.35	0.00	3665.67
Avg. total cost over 1yr ($\bar{C}_{1yr}$)	696	8351.79	979.46	6480.50	11298.91
Avg. total cost over 2yr ($\bar{C}_{2yr}$)	696	16013.67	1849.75	12865.75	21948.01
Changes					
Wholesale funding shock	671	0.00	1.00	-2.78	3.50
Avg. interest ($\bar{r}$)	671	-0.08	0.26	-1.12	1.27
Avg. fees ($\bar{f}$)	671	4.28	205.87	-800.00	2666.67
Avg. total cost over 1yr ($\bar{C}_{1yr}$)	671	-69.74	313.22	-1365.31	2772.89
Avg. total cost over 2yr ($\bar{C}_{2yr}$)	671	-143.75	535.36	-2533.13	3429.85

Panel 1 reflects the main dataset that contains offers for first-time buyers and 2-year fixed rate contracts only. Data on lender characteristics and funding data (Panel 2 and 3) are obtained from SNL Financial and Bloomberg. Variables in Panel 3 are built from Moneyfacts, lender characteristics and funding data.

Table 2.3: Proportion of Cost-Dominated Products by Fee Category (in %)

Excess cost	Fees					Total
	zero fees	(0,500]	(500,1000]	(1000,1500]	>1500	
<=500	4.6	5.4	4.3	0.1	0.0	14.5
(500,1000]	4.4	5.4	10.0	1.3	0.0	21.1
(1000,2000]	6.6	6.3	13.4	5.5	2.7	34.5
(2000,4000]	2.8	5.1	8.0	4.1	2.8	22.9
>4000	0.7	1.8	2.7	1.2	0.6	7.1
total	19.1	23.9	38.5	12.3	6.2	

Notes: This table shows the proportion of all products (in %) between January 2009 and December 2016 split by excess cost, computed as the total cost less the lowest cost product available in a given month, and fee categories. Total cost is computed for a 2-year fixed rate mortgage over two years, at 70-75% LTV for first-time buyers, for a loan size of £150,000, amortized over 25 years.

Table 2.4: Logit Model Estimation Results

	Main sample					Non-bunching sample				
Dep. Var.: *chosen*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total cost (in £,000)	-0.202***	-0.131***	-0.187***	-0.259***	-0.322***	-0.191***	-0.110***	-0.225***	-0.208***	-0.339***
	(0.009)	(0.010)	(0.013)	(0.010)	(0.017)	(0.011)	(0.011)	(0.015)	(0.011)	(0.021)
Distance to nearest branch (in km)		-0.022***					-0.027***			
		(0.001)					(0.001)			
$I_{[r_l, f_h]}$				0.562***					0.135***	
				(0.014)					(0.015)	
Lender FE			✓							✓
Borrower-specific controls					✓			✓		
Observations	450297	450297	450297	450297	450297	347147	347147	347147	347147	347147
Cases	23067	23067	23067	23067	23067	17767	17767	17769	17767	17768

Notes: This table displays estimates of equation **??** for two samples and five specifications each. The main sample contains borrowers with 2yr fixed rate, 70-75% LTV products and is restricted to standard characteristics (age 25 to 45, loan-to-income ratio ≤ 4.5, and amortization period of ≤ 30 years). The non-bunching sample contains only borrowers whose original LTV is below 75 and who are hence not reliant on fees to borrow within the 75% LTV band, in order to avoid higher interest rates (which jump discretely at 5-10% LTV bands). Column (1) shows the baseline specification using only total cost as explanatory variable for the random-utility model and is estimated as an alternative-specific conditional logit, using the main sample. Columns (2) to (5) estimate alternative specifications. Column (2) includes the borrower-lender-specific distance to the closest branch (in km) and hence differs across lenders within the choice options of the same borrower. Column (3) includes lender fixed effects. Column (4) includes an indicator-variable for $k = 2$ (product type $\{r_h, f_l\}$) which is equivalent to a product-type fixed effect. Column (5) is a specification with borrower-specific controls, income and age. Columns (6) to (10) correspond to Columns (1) to (5) for the baseline model, using the non-bunching sample. Standard errors are shown in parentheses, coefficients are denoted with *** for p<0.01, ** p<0.05, * p<0.1.

Table 2.5: Changes in Pricing and Total Cost

Dep. Var:	$\Delta\bar{r}$	$\Delta\bar{f}$	$\Delta\bar{C}_{1yr}$	$\Delta\bar{C}_{2yr}$	Δr_l	Δf_h	$\Delta C_{1yr}^{r_l,f_h}$	$\Delta C_{2yr}^{r_l,f_h}$	Δr_h	Δf_l	$\Delta C_{1yr}^{r_h,f_l}$	$\Delta C_{2yr}^{r_h,f_l}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\Delta\phi$ (std.)	0.00	41.35**	48.74*	56.12	-0.01	109.97**	102.08**	94.20*	0.04	0.04	45.06	92.05
	(0.03)	(19.52)	(24.94)	(52.87)	(0.03)	(44.79)	(40.63)	(50.70)	(0.04)	(0.04)	(37.31)	(73.70)
Year-Quarter FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lender FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	671	671	671	671	671	671	671	671	671	671	671	671
R^2	0.498	0.097	0.451	0.518	0.513	0.071	0.388	0.519	0.443	0.443	0.413	0.477

Notes: This table reports results from 12 different regressions of changes in average interest rates, fees and total cost (over one and two years) on changes in funding shock ϕ (columns (1)-(4)), and the equivalent for $\{r_l, f_h\}$-products (columns (5)-(8)), and $\{r_h, f_l\}$ products (columns (9)-(12)), based on lender-level panel data between 2009 Q1 and 2016 Q4, at quarterly frequency. Prices are for 2-year fixed rates and 70-75% LTV product offers only. All regressions absorb time (year-quarter) and lender fixed effects. Standard errors are reported in parentheses and clustered at the lender level. */**/*** denote p<0.1, p<0.05 and p<0.01, respectively.

Table 2.6: Changes in Pricing and Total Cost, By Lender Pricing Type

	Low				Medium				High			
Dep. Var:	Δr_l	Δf_h	$\Delta C_{1yr}^{r_l,f_h}$	$\Delta C_{2yr}^{r_l,f_h}$	Δr_l	Δf_h	$\Delta C_{1yr}^{r_l,f_h}$	$\Delta C_{2yr}^{r_l,f_h}$	Δr_l	Δf_h	$\Delta C_{1yr}^{r_l,f_h}$	$\Delta C_{2yr}^{r_l,f_h}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\Delta\phi$ (std.)	0.01	13.16	17.66	22.16	-0.06	116.93*	66.56	16.19	-0.03	201.66***	173.22***	144.77**
	(0.02)	(9.02)	(24.16)	(48.35)	(0.08)	(49.46)	(46.33)	(113.06)	(0.03)	(46.97)	(42.32)	(59.38)
Year-Quarter FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lender FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	190	190	190	190	234	234	234	234	217	217	217	217
R^2	0.477	0.196	0.426	0.484	0.631	0.119	0.522	0.645	0.577	0.212	0.424	0.559

Notes: This table reports results from 12 different regressions of pricing characteristics for the low rate, high fee product-type only, for different lender pricing types. Lender pricing types are defined using the tercile of ranking difference (total cost minus interest rate ranking position), as a measure of how much prices are skewed towards the interest rate dimension, using fees that are higher than the prevailing average. Lenders in the "low" category have a negative skew and hence relatively low fees (columns (1)-(4)), "medium" lenders' position in the total cost ranking is similar to that in the interest rate ranking (columns (5)-(8)), and "high" lenders have a substantially better (i.e. lower) position in the interest rate ranking compared to the total cost ranking, i.e. have a high skew in their prices using high fees (columns (9)-(12)). Regressions are based on lender-level panel data between 2009 Q1 and 2016 Q4, at quarterly frequency. Prices are for 2-year fixed rates and 70-75% LTV product offers only. All regressions absorb time (year-quarter) and lender fixed effects. Standard errors are reported in parentheses and clustered at the lender level. */**/*** denote p<0.1, p<0.05 and p<0.01, respectively.

Table 2.7: Lender Characteristics By Pricing Type

Pricing type	Low	Medium	High
Lender characteristics			
Number of lenders	9	9	8
Δ ranking	-7.6	-0.3	5.7
log(assets)	16.6	18.5	17.9
ROA (in %)	0.5	0.3	-0.1
ROE (in %)	8.1	2.8	2.0
Net interest margin (in %)	1.4	1.7	1.5
Loan/dep	1.0	1.0	1.1
Funding cost shock	1.8	2.2	2.3
Lender FE (norm.)	1.1	2.9	3.8
Branch density	715.9	1826.5	2724.2
Market share	0.028	0.030	0.075
Pricing			
Low rate (r_l)	1.92	1.76	1.76
High fee (f_h)	733.33	1083.76	1243.99
High rate (r_h)	2.18	2.30	2.16
Low fee (f_l)	151.43	61.26	373.94
Excess cost	833.06	858.27	1699.96
Within-lender pay dist.	366.40	519.86	631.17
Borrower characteristics			
LTI	3.28	3.53	3.63
LTV	72.35	72.34	72.63
Income (in £)	52900	48707	50021
Age	32.21	32.21	31.39
Loan amount (in £)	170411	166076	172911

Notes: This table displays average lender, pricing and borrower characteristics across "low", "medium" and "high" pricing types, defined using the tercile of ranking difference (total cost minus interest rate ranking position), as a measure of how much prices are skewed towards the interest rate dimension, using fees that are higher than the prevailing average. There are 9 low, 9 medium and 8 high-type lenders. Δ ranking captures the average ranking difference, which is negative for low types (position in total cost ranking is lower than interest rate ranking), around 0 for medium types, and positive for high types (position in total cost ranking is high than interest rate ranking, i.e. is more expensive in total cost terms, indicating high fees). Branch density is computed as the number of borrowers for whom a given lender operates the closest nearby branch (yielding similar distributions when using up to second, third etc. closest branches). Excess cost is computed as the difference between a given product and the minimum cost product in a given month. Within-lender pay distance is the difference between the offer taken by a particular borrower in a given quarter for a given lender, compared to the cheapest alternative by that lender at the same point in time. All values are averaged across lenders by pricing type.

Reference Dependence in the Housing Market

Steffen Andersen, Copenhagen Business School
Cristian Badarinza, National University Singapore
Lu Liu, Imperial College London
Julie Marx, Copenhagen Business School
Tarun Ramadorai, Imperial College London

Abstract

We quantify reference dependence and loss aversion in the housing market using rich Danish administrative data. Our structural model includes loss aversion, reference dependence, financial constraints, and a sale decision, and matches key nonparametric moments, including a "hockey stick" in listing prices with nominal gains, and bunching at zero realized nominal gains. Households derive substantial utility from gains over the original house purchase price; losses affect households roughly 2.5 times more than gains. The model helps explain the positive correlation between aggregate house prices and turnover, but cannot explain visible attenuation in reference dependence when households are more financially constrained.

JEL classification: D15, E43, G21, G22, G5, G52
Keywords: mortgage choice, house prices, collateral, credit risk, interest rate risk, household risk management

Chapter 3

Reference Dependence in the Housing Market

How behavioral are households? In one of the most widely used behavioral economics models of utility, **?** posit that economic agents are reference dependent, measuring utility in terms of gains and losses relative to a reference point, and loss-averse, with losses reducing utility more than gains increase it. It is difficult to assess whether and to what degree this is an accurate representation of human behavior; convincingly answering this question requires detailed data about high-stakes household choices. Correctly interpreting such field evidence also requires taking the richness of households' incentives and the broader features of their decision environment into account.

We quantify reference dependence and loss aversion in the Danish housing market, a unique field setting where a range of household decisions leading up to house sales are tracked with high precision over a long time span. Housing is the largest asset for the vast majority of households (**???**), and selling a house is one of the largest personal economic transactions that households undertake in their lifetimes. These high stakes provide strong incentives for households to think carefully about their choices, making this an important setting to assess whether preferences are non-standard (**?**).

We pin down household preference parameters using a parsimonious structural model of the house selling decision estimated on rich and detailed Danish administrative data. The model

includes loss aversion, reference dependence, household financial constraints, and a sale decision. The data illuminate novel aspects of the choices that households make when they sell houses: a "hockey stick" pattern traced out by list prices as nominal gains vary, visible and sharp bunching of realized prices exactly at the original nominal house purchase price, and lower listing propensities for properties facing nominal losses. The estimated parameters reveal that households are strongly reference dependent, deriving substantial utility from gains and losses relative to the original purchase price of their homes, and that losses hurt households between 2 and 2.5 times more than gains contribute to their utility.[1]

Our paper is not the first to use housing market decisions to study household behavioral biases. In a classic highly-cited paper, **?** compare households that list houses of similar current value, but different original purchase prices. They show that households facing nominal losses set list prices significantly higher, and link this finding to loss aversion. Subsequent literature (see, e.g., **?????**) also investigates other aspects of housing sales through a behavioral lens. However, as we later describe in detail, it is complicated to credibly identify the magnitude of preference parameters from reduced-form empirical evidence in this setting (online appendix Table A.1 documents issues and maps them to a comprehensive review of the literature). To achieve this task, we surmount important conceptual challenges, and bring new data to our analysis.

Selling a house is a multi-layered process involving several choices which are influenced by market conditions. We develop a structural model to capture the richness of this environment. The model maps underlying preference parameters to the main decisions that households make when selling their houses, while taking into account the market factors and frictions that also affect these decisions.

The first decision that we model is whether or not a household lists their property for sale. The second is the listing price that households set if they decide to put the house on the market. When making these decisions, households anticipate the demand conditions they will face in the local housing market, which we model in reduced form. These include the effect of listing prices on the final sale price, and on the probability that the transaction will complete swiftly. Sellers in the model also anticipate the degree of control they have over final price realizations, which may be affected by noise arising from bilateral negotiations with buyers.

[1]This estimate of loss aversion is in line with early estimates which lie between 2 and 2.5 (e.g., **??**), and slightly higher than those seen in more recent literature (e.g., **?** put this multiple at 1.59).

We embed a flexible specification of household preferences into this setup, and predict how different constellations of parameters affect final outcomes. The baseline model considers a "rational" seller, who simply derives utility from the final sale price of the house. A seller who is reference-dependent, in contrast, enjoys an additional boost to utility on realizing a successful sale that delivers gains relative to a reference price, à la **?** (throughout the paper, we assume that the reference price is the original purchase price of the house). And a seller who is both reference dependent and loss averse suffers a greater decrement to utility from a loss than the positive utility boost derived from an equivalent gain.

Importantly, the model also embeds the effect of financial constraints on seller utility. As **?** highlights, a household pays off its outstanding mortgage when selling a house, and faces a mandatory down payment to take on a new mortgage for any subsequent house purchase. If the home equity that they realize post-sale falls short of the mandatory down-payment on the new house, households face either downsizing or costly unsecured borrowing. To capture this, we include a penalty parameter for any shortfall in final realized home equity below the regulatory down-payment threshold.

To understand the model's predictions, consider sellers who derive utility from the final sale price of the house, as well as (symmetrically) from any realized gains or losses relative to the reference price. In this case, sellers optimally set listing price "premia," i.e., markups over a measure of "fair" or hedonic property value, that are linear and downward-sloping in expected gains. The intuition is that sellers compare the potential gain/loss utility they derive if they sell the house (their realization utility) to their reservation utility from staying put. Higher expected realization utility creates incentives to lower the listing price and increase the likelihood of a sale, and lower expected realization utility generates the opposite incentive, to raise the listing price and effect a higher final price. When losses and gains are viewed symmetrically, the marginal impact of this effect is constant, leading to a linear relationship between the listing price and sellers' potential gains.

This linear relationship changes when sellers are loss averse and feel greater disutility from realized losses than the utility enjoyed from realized gains. The asymmetry in preferences in this case translates into an optimal listing premium profile which is also asymmetric, sloping up more sharply when sellers face losses than when they face gains. Moreover, loss averse sellers have an incentive to "fish" for higher prices that avoid the discontinuity of the marginal

loss at the reference point. In this case, the model also predicts sharp bunching of final sales transactions quantities precisely at realized nominal gains of zero.

These insights suggest an easy interpretation of the asymmetric listing price behavior originally detected in **?**, i.e., that it reflects underlying preference asymmetry. However, several key confounds can interfere with such simple inferences. The first issue, mentioned earlier, is that down-payment constraints can manifest similar effects as loss aversion. This creates an empirical confound, because the nominal purchase price of the house (the assumed reference point for loss aversion) can be very close in practice to the outstanding mortgage balance (the point at which down-payment constraints kick in). To cleanly separate these forces, we harness significant independent variation in the Danish data between sellers' home equity position and their gains/losses since purchase.

A second problem comes from the need to assess sellers' gains/losses at the point of their listing decisions. Measuring these "potential gains" requires an estimate of the expected value of houses. As **?**, **?**, **?**, and others note, unobserved property quality can cloud assessments of house values. To deal with this issue, we first extend pre-existing empirical approaches to analyze how different sources of measurement error generate specific biases in inferences about underlying structural parameters. This analysis leads us to modify the standard hedonic model estimates of fair house value, in turn generating different data moments that we use to confirm structural parameter robustness. We also employ a simulation-based approach to assess how inferences are affected when buyers and sellers have information about quality that is invisible to the econometrician. As we later describe, this exercise provides additional theoretical confidence that our setup can accurately recover underlying structural parameters.

A third confound comes from the optimization frictions that sellers face in the market. **?** shows that U.S. housing markets are characterized by "concave demand," where lowering list prices past a point does not boost sale probabilities, but does negatively impact realized sale prices.[2] We confirm this finding in Danish data, which also exhibit strong demand concavity. Our model reveals that optimizing sellers who face concave demand set asymmetric listing premium schedules even if their underlying preferences are symmetric across gains and losses.

[2]The paper attributes this to the search process in the housing market, where potential buyers are less likely to visit a property if the initial listing price is unusually high, but they are not more likely to visit it if the listing price is unusually low. One interpretation of this is that a low price can signal a "lemon," leading to a longer time-on-the-market for such properties.

We confront this challenge by harnessing regional variation across housing markets in Denmark, and calculating moments in the model that reflect regional variation in the strength of demand concavity. We assume that seller preferences are consistent across locations, meaning that variation of listing behavior that is unexplained by differences in demand concavity helps to pin down the degree of loss aversion.[3]

Another important optimization friction arises because final sale price realizations result from post-listing negotiations with buyers. Most sellers have only a limited ability to control final negotiation outcomes, which dilutes the link between the elasticities observed in the data—the observed listing, pricing, and selling decisions associated with particular reference points—and their structural counterparts. To introduce greater realism, we incorporate these optimization frictions into the model, and estimate the fraction of sellers that can only imprecisely target final outcomes as an additional model parameter. This "imprecise targeting" friction bears similarities to the imperfect ability to manipulate income modelled in the public finance literature (see, e.g., **???**).

In Danish data, as in US data, the listing price schedule has the characteristic "hockey stick" shape first identified by **?**, rising substantially as potential losses mount, and virtually flat in potential gains. We find slopes of similar magnitude to the **?** estimates despite the differences in location, sample period, and sample size. We also see a similar "hockey stick" along the dimension of potential home equity controlling for potential gains, consistent with the distinct effect of down-payment constraints on seller listing behavior.

The data also show that the shapes of listing premia schedules vary across regional housing markets in Denmark, mirroring the degree of demand concavity in these regions. Listing premia with sharp responses to losses are evident in regional markets with weaker demand concavity, and muted responses to losses are a feature of regional markets with strong demand concavity, confirming the model-implied link between demand conditions and sellers' optimal listing decisions.

The distribution of final sales in the data exhibits sharp bunching of transactions at realized gains of zero, diffuse bunching mass just to the right of zero, and a significant shift in total

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

mass from realized losses towards realized gains. The spike at precisely zero gains is clear evidence of loss aversion, which is the only force in the model that can generate this pattern. Fitting just the spike requires only a modest degree of loss aversion in a model in which sellers can precisely target desired final realized prices. In a more realistic model with optimization frictions inhibiting precise targeting, loss aversion is larger, to account for the diffuse excess bunching mass visible just to the right of zero gains.

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

Collectively, these moments in the data allow us to structurally estimate sizable magnitudes for both behavioral frictions and down-payment constraints. We converge on a coefficient of reference dependence of 0.629 (*s.e.* 0.043), meaning that realized gains contribute roughly 60% as much as final prices to household utility, and a coefficient of loss aversion $\lambda = 2.473$ (s.e. 0.128), meaning that the disutility of losses is 2.473 times the utility of gains. We also find that these behavioral effects coexist and interact with financial frictions. In particular, sellers who are financially constrained appear "less behavioral" in the sense that reference dependence and loss aversion are visibly less pronounced in the listing prices that they set. We cannot rationalize this intriguing observation with our model; we view this as an important avenue for future research.

Using the model, we structurally decompose the positive correlation between aggregate house prices and transaction volume routinely observed in housing market data. Explaining this correlation was an important motivation for the early literature in this space, including **?** and **?**. In our model, changes in fair house values change the distributions of both potential gains and potential home equity. These changes affect listing behavior, which in turn translates into realized traded volumes through the estimated concave demand mapping from listings to final sales. Attributing this correlation to the different economic channels, we find that the effect of down-payment constraints is large, closely followed by the effects of reference dependence and loss aversion.

The paper is organized as follows. Section 2 introduces our model of household listing behavior.

Section 3 discusses the construction of our merged dataset, and provides descriptive statistics. Section 4 introduces the moments that we use for structural estimation and uncovers new facts on listing prices and listing decisions. Section 5 describes our structural estimation procedure, and reports parameter estimates. Section 6 describes validation exercises, and highlights areas where the model falls short in explaining features of the data. Section 7 concludes.

3.1. A Model of Household Listing Behavior

We develop a model in which a household (the "seller"), optimally decides the price at which they list their house (the "intensive margin" decision of listing price), as well as whether or not to list the house (the "extensive margin"). The model framework flexibly embeds different preferences and constraints commonly used to explain patterns in listing behavior. We describe the intuition of the model here; online appendix B provides a more detailed discussion of model arguments and derivations of equations.

The market comprises a continuum of sellers and buyers of residential property. In period 0 of the two-period model, the property owner receives a shock $\theta \sim N(\theta_m, \theta_\sigma)$. This "moving shock" θ can be thought of as a "gain from trade" (**?**), i.e., a boost to lifetime utility which the seller receives in the event of successfully selling and moving in period 1. It captures a variety of reasons for moving, including moves due to labor market opportunities, or the desire to upsize or downsize.

Let L denote the listing price set by the seller, and $\widehat{P}$ the "fair" or "fundamental" property value. Conditional on the chosen listing premium $\ell = L - \widehat{P}$, the demand function $\alpha(\ell)$ indicates the probability that a willing buyer is found in period 1, and $P(\ell) = \widehat{P} + \beta(\ell)$ is the (uncertain) realization of the final sale price resulting from the negotiation with that buyer. Analogously to the definition of the listing premium, $\beta(\ell)$ is the final realized premium of the price over the hedonic value that the seller is able to negotiate for in period 1. As in **?** and **?**, we assume that the seller takes the distribution of potential outcomes of negotiations as given when optimizing utility. This restricts their action space, and captures the basic tradeoff when deciding on a listing strategy: a larger ℓ can lead to a higher ultimate transaction price, but it decreases the probability of quickly finding a willing buyer.

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

$$\max_{s\in\{0,1\}}\left\{s\times\max_{\ell}\left[\alpha(\ell)\left(E\left[U\left(P(\ell),\cdot\right)\right]+\theta\right)+(1-\alpha(\ell))\underline{u}-\phi\right]+(1-s)\times\underline{u}\right\}. \tag{3.1}$$

The function $U\left(P(\ell),\cdot\right) = u(P(\ell),\cdot) - \kappa(P(\ell),\cdot)$ has two separate components. The first accommodates reference-dependent utility and loss aversion à la **?** and **??**. The second is the penalty function for violating down-payment constraints à la **?**.

The seller decides on the extensive margin, i.e., whether ($s = 1$) or not ($s = 0$) to list, and sets ℓ to maximize the expected utility from the property sale. Once the property is listed, the sale goes through with probability $\alpha(\ell)$. The seller receives utility and pays any penalty associated with violating the down-payment constraint based on the realized final price $P(\ell)$, and also receives θ. The listing fails with probability $1-\alpha(\ell)$, in which case the seller falls back to their outside option level of utility $\underline{u}$. Listing incurs a one-time utility cost ϕ, sunk at the point of listing, which captures a range of frictions including psychological "hassle factors," and search, listing, and transaction fees.

In the model solution and estimation exercise, we normalize $\widehat{P}$ to 1. All model quantities are therefore simply expressed in units of $\widehat{P}$. In this way, the calculated premia in the model are easily mapped to log differences, consistent with the definitions of gains/losses and home equity employed in our empirical work.[4]

3.1.1. Demand Functions

The functional mappings between listing premia, sale probabilities and final sale prices simplify the characterization of the outcome of the bargaining process. In our estimation, we define a period as equal to six months, meaning that $\alpha(\ell)$ captures the probability that the transaction goes through within six months after the initial listing. In the online appendix, we verify robustness to varying the length of this period.

To introduce greater realism into the formation of sale prices, we assume that sellers vary in their ability to target a particular final price. We distinguish between two separate types, according to their deal-making ability: a fraction π of sellers are able to precisely steer the transaction

[4]This relies on the usual approximation $\ln(1+x)\approx x$. For example, the listing premium expressed in units of $\widehat{P}$ equals $L/\widehat{P}-1$ in the model, and is estimated as the log difference between L and $\widehat{P}$ in the data.

towards their desired outcome; the remainder $1-\pi$ target final prices only imprecisely, have limited bargaining power, and expect to settle for a random price realization drawn from a distribution that we estimate in the data.

Both types of sellers solve the optimization problem laid out in equation (**??**), except that precise targeters face a deterministic path of final prices $P(\ell) = \widehat{P} + \beta_0 + \beta_1 \ell$, while imprecise targeters form rational expectations about the distribution of $P(\ell)$:

$$P(\ell) = \widehat{P} + \beta_0 + \beta_1 \ell + \varepsilon, \varepsilon \sim B(0, \sigma(\ell)), \tag{3.2}$$

with β_0, β_1 and $\sigma(\ell)$ estimated in the data.[5] The mixture distribution of precise and imprecise targeters, weighted by the parameter π, allows us to calculate the final set of aggregate model-implied moments.[6]

3.1.2. Reference Dependence and Loss Aversion

We model reference-dependent loss-averse preferences in a standard fashion (**??**). The seller's reference price level is denoted by R, and $u(P(\ell), R)$ is a piecewise linear function of realized nominal gains $G(\ell) = P(\ell) - R$:

$$u(P(\ell), R) = \begin{cases} P(\ell) + \lambda\eta G(\ell), \text{ if } G(\ell) < 0 \\ P(\ell) + \eta G(\ell), \text{ if } G(\ell) \geq 0 \end{cases}. \tag{3.3}$$

The parameter η captures the degree of reference dependence. Sellers derive utility from the final price realized from the sale as usual, but there is an incremental utility "bump" from the realized nominal gain. η measures the extent to which realized gains augment utility over and above the final utility of wealth arising from the sale.[7]

[5] As we later discuss, the right-hand plot of Figure **??** shows that the variance of negotiation outcomes is larger for large listing premia, and it flattens as listing premia approach zero. We capture this phenomenon by explicitly estimating the function $\sigma(\ell)$ in the data; we find that the best fit is a third-degree polynomial of ℓ.

[6] This structure of the population is consistent with previous mixture models of behavioral frictions, which originate in statistical theory with **?**, and have since been used to cluster sets of continuous multivariate data for a wide variety of random phenomena (**????**).

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

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

We set the outside option $\underline{u} = \widehat{P}$, which implies that absent reference dependence ($\eta = 0$) or any additional reasons to move ($\theta = 0$), and if listing is frictionless ($\phi = 0$), the seller is indifferent between staying in their home and receiving the hedonic value in cash. This specification of the outside option is equivalent to the seller not receiving gains from moving, but experiencing θ disutility in the event of a failed sale (i.e., the outside option is then rewritten as $\underline{u} = \widehat{P} - \theta$). We do not place any restriction on θ, simply recovering it as a latent variable in structural estimation. In our model, as sellers only experience utility from house price appreciation if they realize a sale, the case of linear reference dependence with $\lambda = 1$ is essentially equivalent to the "realization utility" framework of **?**. In online appendix section B.7, we discuss this issue in detail.

3.1.3. Down-Payment Constraints

We now describe the financial penalty function $\kappa(P(\ell), \cdot)$ for violating down-payment constraints. Let M denote the household's outstanding mortgage balance. The potential home equity position of the household is then $\widehat{H} = \widehat{P} - M$, and their realized home equity position $H(\ell)$ arises from the potential level $\widehat{H}$ plus the realized price premium $\beta(\ell)$, i.e.:

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

Assume γ to be the required down-payment fraction on a *new* mortgage origination for a property of similar size and quality, with a hedonic value of $\widehat{P}$. Based on the realized home equity level H, we can then distinguish between constrained households, for which $H(\ell) < \gamma\widehat{P}$, and unconstrained ones, for which $H(\ell) \geq \gamma\widehat{P}$.

If down-payment constraints bind, only unconstrained sellers can move to another property of

the same or greater value. However, there are several ways in which Danish households can relax these constraints despite legal restrictions on LTV at mortgage initiation (as we discuss later, the Danish Mortgage Act restrict the down-payment fraction at issuance to a value of 20% or higher). The first way is for households to downsize to a less expensive home than $\widehat{P}$, or indeed, to move to the rental market—either decision might incur a utility cost. The second is that households can engage in non-mortgage borrowing to fill the gap $\gamma\widehat{P} - H(\ell)$. A common approach in Denmark is to borrow from a bank, or occasionally from the seller of the property, to bridge funding gaps between 80% and 95% loan-to-value (LTV); this is typically expensive.

Taking these features into account, we assume that violating the down-payment constraint does not lead the seller to withdraw the sale offer, but instead that the seller incurs a monetary penalty for levels of realized home equity below the constraint threshold. The financial penalty function $\kappa(P(\ell), \cdot)$ captures this monetary value:

$$\kappa(P(\ell)) = \begin{cases} \mu(\gamma\widehat{P} - H(\ell))^2, \text{ if } H(\ell) < \gamma\widehat{P} \\ 0, \text{ if } H(\ell) \geq \gamma\widehat{P} \end{cases}. \tag{3.4}$$

We choose a smooth quadratic penalty function to avoid a discontinuity at the threshold level $H(\ell) = \gamma\widehat{P}$. Such a discontinuity would predict bunching in realized prices at 20% home equity, which we can firmly reject in the data. Online appendix F contains a detailed discussion and additional evidence on downsizing/upsizing, which has a dilutive effect on the exact point at which the down-payment constraint becomes binding.

3.1.4. Structural Parameters

We next discuss selected predictions of the model to build intuition. This guides the choices of the main moments of the data used to structurally estimate the model's parameters. Table **??** provides an overview of all model variables and structural parameters.

Consider a simple version of the model without the extensive margin decision, listing costs, or constraints $\kappa(\cdot)$, and assuming that all sellers can precisely target outcomes.[8] In this case, the maximization problem becomes:

$$\max_{\ell} \left[\alpha(\ell)\left(u(P(\ell),\cdot)+\theta\right)+(1-\alpha(\ell))\underline{u}\right]. \tag{3.5}$$

The first-order condition determining the optimal listing premium in equation (**??**) balances the marginal utility benefit arising from a higher realized premium in the event of a successful sale against the marginal cost associated with an increased chance of the transaction failing, which results in the outside option utility level.

To derive more intuition, we analytically solve this simple model under the additional assumptions that $\alpha(\ell) = \alpha_0 - \alpha_1\ell$ and $\beta(\ell) = \beta_0 + \beta_1\ell$ are linear and deterministic in ℓ. Figure **??** illustrates how the optimal listing premium varies in this version of the model.

In the absence of reference dependence ($\eta = 0$), the figure shows that utility derives purely from the terminal house price P. In this case, the left-hand plot (dashed line) shows that, for any given level of $\widehat{P}$, ℓ^* is unaffected by the reference price R. With "linear reference dependence" (i.e., $\eta > 0$, $\lambda = 1$), holding $\widehat{P}$ constant, the plot shows how different R values affect utility. In this case, there is a negatively-sloped linear relationship between ℓ^* and $\widehat{G} = \widehat{P} - R$, indicated with a dotted line. Here, R does not affect the marginal benefit of raising ℓ^*, but it affects the marginal cost by changing the distance between $\underline{u}$ and the achievable utility level in the event of a successful transaction. If realized gains and losses contribute symmetrically to household utility, optimal ℓ^* decreases linearly with the magnitude of this distance.

In the case of reference dependence plus loss aversion (i.e., $\eta > 0$, $\lambda > 1$), indicated with a solid line in Figure **??**, the kink in the piecewise linear utility function leads to a more complex piecewise linear pattern in ℓ^*. This is because ℓ controls the gains/losses that sellers ultimately realize. The figure shows that there is a range of potential gain values which map to realized gains of precisely zero (recall $G(\ell^*) = \widehat{G} + \beta(\ell^*)$). For all $\widehat{G} \in [\widehat{G}_1, \widehat{G}_0]$, the seller can choose a

[8] We provide a full solution of this model in online appendix sections B.1-B.3, and discuss it graphically below.

listing premium ℓ such that $\beta(\ell) = -\widehat{G}$. Conditional on a sale, the realized gain for this range of $\widehat{G}$ values is $G = 0$. For sellers with potential gains below $\widehat{G}_1$, the expected costs associated with transactions failure probabilities become unacceptably high, relative to the expected benefits of avoiding losses. It therefore becomes sub-optimal to aim for realized gains of zero in this range. For levels of $\widehat{G} < \widehat{G}_1$, sellers have no choice but to accept losses, but still set marginally higher listing premia.

We also note that very similar intuition applies along the potential home equity dimension; the impact of potential home equity is modulated by the different penalty function shown in (**??**).

Bunching around Realized Gains of Zero

Household preference parameters also have implications for transactions quantities. Different parameter values result in shifts in mass in the distribution of completed transactions along the G dimension. In our estimation approach, we use this insight to harness an additional set of moments which help to pin down the preference parameters of the model. The right-hand side diagram in Panel A of Figure **??** illustrates how the distribution of realized transactions varies with preference parameters in the simple version of the model described just above. Detailed solutions associated with this figure are provided in online appendix section B.2.

When $\eta = 0$, sellers choose a constant listing premium ℓ^*, which results in a constant realized premium $\beta(\ell^*)$ of actual gains G over potential gains $\widehat{G}$. In the linear reference dependence model ($\eta > 0, \lambda = 1$), sellers with $\widehat{G} < 0$ choose relatively higher ℓ^*. This lowers the likelihood that willing buyers are found, meaning that the likelihood of observing transactions in this domain of $\widehat{G}$ is lower. However, if these transactions *do* go through, the associated G is then higher, shifting mass in the final sales distribution towards $G > 0$ (in the right-hand plot, the dashed line becomes the dotted line).

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

In Panel B of Figure **??**, we consider the effects of introducing imprecise targeting of the final price. As mentioned earlier, in this case we assume that for a group of sellers $\beta(\ell)$ is known with

certainty, whereas a fraction $1-\pi$ imprecisely target the final price, i.e., they do not know $\beta(\ell)$ with certainty at the time of listing. This inclusion of noise in $\beta(\ell)$ has a number of effects. First, because of the mixture of precise and imprecise targeters, sharp kinks in the listing premium profile are smoothed out. Second, this force in the model generates risk aversion, because sellers with positive but low potential gains also worry about realizations of $\beta(\ell)$ that push the final price below the reference point; this pushes some sellers to list more aggressively. Finally, and importantly, sellers that face uncertainty realize final prices that are close to, but often above the reference point. With the introduction of imprecise targeting λ plays a dual role, generating strict bunching for those precise targeting sellers who can anticipate and steer negotiation outcomes towards their desired price, and diffuse bunching for those sellers who can only imprecisely steer final outcomes.

These predictions are important for the structure of uncertainty in negotiation outcomes in our model. We need to assume that targeting ability is a fixed feature of each seller type, as opposed to simply reflecting the heterogeneity of final outcomes. This is because, unlike in the public finance literature where manipulation amounts are not observable (see, e.g., **??**), we observe seller decisions made at the point of listing as well as final sale outcomes, and need to reconcile the two with our model. On the one hand, we see evidence in the data that suggests extensive negotiations between buyers and some sellers, with plausibly uncertain final outcomes. This is consistent with evidence on negotiations from real estate (**??**) as well as other markets (**?**). On the other hand, as we show below, we also observe strict bunching precisely at the reference point in the data; this evidence points to least a fraction of sellers being able to precisely anticipate market conditions and steer transactions towards their desired final outcome.

Concave Demand

The demand functions $\alpha(\ell)$ and $\beta(\ell)$ are an important determinant of listing behavior, and their functional form can affect the shape of the ℓ^* schedule in this model. In the simple case of linear demand functions discussed thus far, when the probability of sale $\alpha(\ell)$ is less responsive to ℓ, the marginal cost of choosing a larger listing premium is lower, and therefore the level of ℓ^* is higher. However, the literature has shown that $\alpha(\ell)$ is nonlinear. In particular, it is "concave".

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

More generally, the model predicts a tight link between the shape of $\alpha(\ell)$ and the slope of ℓ^*. A steep negative slope of $\alpha(\ell)$ for ℓ above $\underline{\ell}$ leads to a gradual slope of ℓ^* in the loss domain, since the marginal cost of increasing the listing premium is higher in this case, and vice versa. In the online appendix section B.5 we illustrate this mechanism, positing a concave shape for $\alpha(\ell)$ and showing the effect of varying $\alpha(\ell)$ around $\underline{\ell} = 0$, i.e., the point at which $L = \widehat{P}$.

Concave demand also has an effect on transaction volumes in the model. This shape of $\alpha(\ell)$ predicts shifts of mass in final sales towards positive values of realized gains, depending on the level of $\underline{\ell}$, the point at which concavity "kicks in." However, it is not associated with sharp bunching of the type associated with loss aversion, as demand concavity is assumed (and seen in the data to be) smooth.[9]

Extensive Margin

In the model, any force inducing a wedge between the expected utility from a successful listing and the outside option $\underline{u}$ affects both intensive and extensive margins. In particular, the model predicts that sellers with lower $\widehat{G}$ are less likely to list at all. This prediction points to the relationship between the propensity to list and $\widehat{G}$ as an additional moment to inform structural estimates of preference parameters.

[9] A more subtle point here is that any change in the precise specification of the reference point R in the presence of loss aversion will also change the location at which bunching is observed, and heterogeneity in reference points will make it hard to observe the precise location of bunching. To complicate matters further, variations in the level of $\underline{\ell}$ are a confound, potentially rendering it difficult to distinguish models with heterogeneous reference points from models with spatial or temporal variation in $\underline{\ell}$. We avoid this complexity in our setup by simply taking the stance that R is the nominal purchase price of the property and evaluating the extent to which we see bunching given this assumption. As we will later see, this turns out to be a reasonable assumption—we observe significant evidence in the data of bunching using this assumption about R, confirming its relevance to sellers.

Modelling the extensive margin decision also helps to account for selection effects that can drive patterns of observed intensive margin listing premia in the data, an issue that prior literature (e.g., **????**) has been unable to control for as a result of data limitations. For example, if sellers that decide not to list are more conservative (i.e., they set lower listing premia), and those who decide to list are more aggressive (i.e., setting higher listing premia) the resulting selection effect would lead to a higher observed non-linearity in listing premia around reference points, which would bias parameter estimates and inferences conducted only using the intensive margin.[10]

There are more subtle implications of the model linking the extensive and the intensive margins. High realizations of θ affect the listing decision, and push the seller towards setting higher listing premia. However, this force can move ℓ into regions of concave demand in which the response of buyers is more (or less) pronounced, because of non-linearities in $\alpha(\ell)$. This in turn means that variation in θ can affect the observed magnitude of the seller's responses to $\widehat{G}$, another force which smooths and blurs kinks in the model-implied ℓ^* profile. Online appendix section B.1 illustrates this analytically with a specific example, showing how a smooth "hockey stick" average listing premium profile can result from averaging the three-piece-linear listing premium profile (for $\lambda > 1$) across the distribution of sellers with different θ.

Overall, the model provides clear guidance as to how the first three structural parameters η, λ and μ can be separately identified in the data: (i) $\eta > 0$ leads to a negative slope of the listing premium profile along the entire range of potential gains, (ii) $\lambda > 1$ leads to excess bunching of transactions for realized prices at and above the nominal reference point, and an additional contribution to the slope of the listing premium profile in the loss domain, (iii) $\mu > 0$ leads to a negative slope of the listing premium profile with respect to potential home equity.

The remaining four parameters $\theta_m, \theta_\sigma, \pi$, and ϕ can be thought of as "fitting parameters," which allow us to pin down important quantities that are not the main focus of our model. A simple characterization is that (i) θ_m mainly determines the average level of the listing premium, (ii) θ_σ drives the degree to which the kinks and non-linearities that our assumptions on preferences and constraints imply are smoothed out in the data, (iii) π pins down the relative magnitudes of strict and diffuse bunching of realized prices around the reference point, and (iv) ϕ is a key determinant of the extensive margin probability that a given property is listed for sale.

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

Beyond these simplified characterizations, there are complex interactions between different parameters which determine ultimate outcomes. We next describe the data and key moments visible in the data as a precursor to a more rigorous structural estimation of the model's parameters using these moments.

3.2. Data

We obtain high-quality administrative data on Danish housing listings, transactions, and the housing stock, as well as demographic and financial information about house sellers. We briefly describe these data below, and online appendix C contains detailed descriptions of data sources, data construction and filtering, and the process of matching involved in assembling the final dataset.

3.2.1. Property Transactions and the Housing Stock

We acquire comprehensive administrative data on the ownership and hedonic characteristics of the housing stock of all registered properties in Denmark between 1992 and 2016, as well as all transactions of these properties from the Danish Tax and Customs Administration (SKAT) register and the Danish housing register (Bygnings-og Boligregisteret, BBR). We also obtain the assessment values of each property. This information is provided by SKAT, which assesses property values every second year.[11]

3.2.2. Property Listings Data

Property listings data from 2008 to 2016 are provided to us by RealView. We link these transactions to the cleaned/filtered sale transactions in the official property registers; 79.6% of all sale transactions have associated listing data. We describe these data more fully in the online appendix, noting here that unmatched transactions generally occur off-market as direct private transfers.

[11] Tax-assessed property values are used for determining tax payments in Denmark. Online appendix D.1 describes the property taxation regime in Denmark in greater detail including inheritance taxation; we simply note here that there is the usual "principal private residence" exemption on capital gains on real estate, and that property taxation does not have important effects on our inferences.

3.2.3. Mortgage Data

We obtain data on any mortgages attached to each property from the Danish central bank (Danmarks Nationalbank), which collects these data from mortgage banks. The data are available annually for each owner from 2009 to 2016, cover all mortgage banks and all mortgages in Denmark, and contain information on the mortgage principal, outstanding mortgage balance each year, the LTV ratio, and the mortgage interest rate. If several mortgages are outstanding for the same property, we simply sum them, and calculate a weighted average interest rate and loan-to-value ratio for the property and mortgage in question.[12]

Danish households can take up a mortgage covering up to 80% of the property price. Banks require a cash down payment of 5% and households can bridge the remaining 80%-95% using bank loans or "Pantebreve" (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. A third (typically unobservable) possibility is that households can bring additional funds to the table by liquidating other assets, or by borrowing from friends and family.

In **?**, M represents the outstanding mortgage debt net of any liquid assets that the household can put towards the down payment. The granular data that we employ allow us to measure the net financial assets that households can use to supplement realized home equity. In online appendix section K we verify using these data that our inferences are sensible when taking these additional funds into account.

3.2.4. Owner/Seller Demographics

Demographic data on individuals and households come from the official Danish Civil Registration System (CPR Registeret). Individual income and wealth data from the official records at SKAT, which hold detailed information by CPR numbers for the entire Danish population. In addition to each individual's personal identification number (CPR), the records also contain

[12]Online appendix D.2 provides a description of several features of the Danish mortgage market including the conditions under which mortgages are assumable, as well as the effects of the Danish refinancing system (studied in greater detail in **?**) on sale and purchase incentives.

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

3.2.5. Final Merged Data

Our sample is restricted to transactions for which we can measure both nominal losses and home equity. Transactions data are available from 1992 to the present, meaning that we can only measure the purchase price (i.e., reference price) for properties that were bought during or after 1992. The mortgage data run from 2009 to 2016, which further restricts the sample to this time period. We also restrict our analysis to properties for which we know both the ID of the owner, as well as that of the owner's household, in order to be able to match housing transactions to wealth and income data. We exclude data from foreclosure transactions, properties with a registered size of 0, and properties that are sold at prices which are unusually high or low (below 100,000 DKK and above 20MM DKK in 2015, or marked as having an extreme price by Statistics Denmark). For listings that end in a final sale, we also drop within-family transactions, transactions that Statistics Denmark flag as anomalous or unusual, and transactions where the buyer is the government, a company, or an organization. We also restrict our analysis to residential households, dropping summerhouses and listings from households that own more than three properties in total, as they are more likely to be property investors than owner-occupiers.[13]

In online appendix C.5, we describe the data construction filters and their effects on our final sample in more detail. Once all filters are applied, the sample comprises 214,103 listings of Danish owner-occupied housing between 2009 and 2016, for both sold (70.4%) and retracted (29.6%) properties, matched to mortgages and other household financial information. Of these, 172,225 listings have an attached mortgage, and 41,878 listings have no associated mortgage (i.e., are owned entirely by the seller).

The listings correspond to a total of 191, 507 unique households, and 178,933 unique properties.

[13]We apply these filters because transactions involving a registered corporation, a government entity or a foreclosed property are often conducted at non-market prices. This is, for example, for tax efficiency or avoidance purposes in the case of corporations, and for eminent domain reasons in the case of government purchases. Moreover, the market for such properties is extremely thin, meaning that predicting the price using a hedonic or other model is particularly prone to error. As concerns investment homes, we note that **?** separately estimate loss aversion across groups of owner-occupiers and speculators, but we choose to focus our parameter estimation in this paper on owner-occupiers.

Most households sell one property during the sample period, but roughly 9% of households sell two, and roughly 1.5% of households sell three or more properties. As mentioned, we also use the entire housing stock, filtered in the same manner as the listing data. These data comprise 5,538,052 observations of 807,345 unique properties. This enables us to capture the determinants of the propensity to sell, i.e., the extensive margin decision in the model.

3.2.6. Hedonic Pricing Model

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

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

$$\ln(P_{it}) = \xi_{tm} + \beta_{ft}\mathbb{1}_{i=f} + \mathbf{fi_x}\mathbf{X_{it}} + \mathbf{fi_{fx}}\mathbb{1}_{i=f}\mathbf{X_{it}} + \Phi(v_{it}) + \mathbb{1}_{i=f}\Phi(v_{it}) + \varepsilon_{it}, \tag{3.6}$$

where $\mathbf{X_{it}}$ is a vector of time-varying property characteristics (these characteristics are recorded and updated each year), namely ln(lot size), ln(interior size), number of rooms, bathrooms, and showers, a dummy variable for whether the property was unoccupied at the time of sale or retraction, ln(age of the building), dummy variables for whether the property is located in a rural area, or has been marked as historic, and ln(distance to the nearest major city). ξ_{tm} are year cross municipality fixed effects (there are 98 municipalities in Denmark), and $\mathbb{1}_{i=f}$ is an indicator variable for whether the property is an apartment (denoted by f for flat) rather than a house.

Finally, $\Phi(v_{it})$ is a third-order polynomial of the previous-year tax assessor valuation of the property, which also includes a square-root term.

Online appendix E.1 discusses the hedonic model in greater detail. We later on also check robustness to a range of different cases, variation in the hedonic model.

The R^2 of the model equals 0.88 in the full sample, a high degree of accuracy which we verify in various ways in online appendix E.[14]

[14]Briefly, when we estimate the model in levels rather than logs, we obtain an R^2 of 0.85. The R^2 when

An important confound when structurally estimating model parameters is noise or bias in the estimation of $\widehat{P}$, arising from factors such as time-varying unobserved house quality (see e.g., **????**). We describe our approach to dealing with these issues later in the paper.

3.3. Reduced-Form Facts

As a precursor to full-blown structural estimation, we document patterns in the data, primarily focusing on listing premia and sales transactions volumes in relation to measured G and $\widehat{G}$. We informally discuss how these patterns relate to the predictions of the model, with a specific focus on our main parameters of interest, η and λ. We also explore how the patterns in the data change when we account for (i) sellers' down-payment constraints (i.e., $\widehat{H}$), (ii) concave demand, and (iii) robustness to changes in measurement. We also discuss how these changes might influence parameter estimation.

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

We compute listing premia as $\ell = \ln L - \widehat{\ln P}$, where L is the reported initial listing price observed in the data, and $\widehat{\ln P}$ is estimated using the hedonic model. We focus on listing premia, as listing prices move virtually one-to-one with hedonic values, akin to **?** (see online appendix Table A.3). Mean ℓ is 12.1%, and ℓ is greater than zero for 74% of the sample.

We compute potential gains as $\widehat{G} = \widehat{\ln P} - \ln R$, where R is set to the nominal purchase price of the property. Mean $\widehat{G}$ estimated in this way is 38%, and 23% of property-years exhibit negative gains. Online appendix Figure A.2 plots the distributions of these variables.

Figure **??** plots the average listing premium on the y-axis associated with each level of potential gains on the x-axis. The figure shows that prospective sellers of properties that have appreciated since the initial purchase choose lower listing premia, while the reverse is true for those facing potential losses. The downward slope of the listing premium along $\widehat{G}$ is visible throughout the domain, including when $\widehat{G} > 0$. This downward slope is consistent with the predictions

we eliminate the tax assessor valuation from the hedonic characteristics is 0.77. We achieve a similarly high fit when estimating the model on random 50% samples of the data and fitting the remainder out-of-sample to address overfitting concerns.

of a model with reference dependence $\eta > 0$. Moving from the gain to the loss domain, the slope becomes much more pronounced, i.e., listing premia react more aggressively to every unit decrease in potential returns when $\widehat{G} < 0$, and a kink is visible at $\widehat{G}$=0. This "hockey stick" pattern is seemingly consistent with the predictions of a model with loss aversion $\lambda > 1$. In the piecewise linear formulation of preferences, however, loss aversion also predicts a flattening out of the listing premium profile deeper into the loss domain, which is not visible in the plot.

While these patterns seem to be consistent with reference-dependent and loss-averse preferences, the model tells us that various other parameters could also be at work. We continue our investigations below.

To see the effects of home equity constraints, for all observations in the data, we calculate $\widehat{H} = \widehat{\ln P} - \ln M$, where $\widehat{\ln P}$ is estimated using our hedonic model as before, and M is the outstanding mortgage balance reported by a given household's mortgage bank each year.

The average level of potential home equity is $\widehat{H} = 41\%$, with a median level of $\widehat{H} = 37\%$. Across the full sample of property × year observations, 35% of sellers potentially face a binding financial constraint (i.e., $\widehat{H} < 20\%$). The modal $\widehat{H}$ conditional on having a mortgage is around 18%, which is consistent with the Danish constraint on the issuance of mortgages—the Danish Mortgage Act specifies that LTV at issuance by mortgage banks is restricted to be 80% or lower.[15]

$\widehat{G}$ and $\widehat{H}$ are jointly dependent on $\widehat{\ln P}$, but there are multiple other factors that can influence the correlation of these variables, including the LTV ratio at origination (i.e., variation in initial down payments), and households' post-initial-issuance remortgaging decisions. In online appendix Figure A.4, we plot the joint distribution of $\widehat{G}$ and $\widehat{H}$, and show that there is substantial variation in the four regions defined by $\widehat{G} \lessgtr 0$ and $\widehat{H} \lessgtr 20$, which permits identification of their independent impacts on listing decisions. In online appendix Figures A.4 and A.5, we also show that this variation is not confined to one particular part of the sample period—there is substantial mass in all four quadrants of the joint distribution defined by thresholds in $\widehat{G} \lessgtr 0$ and $\widehat{H} \lessgtr 0$, and the relative mass remains fairly stable over the sample period. This alleviates

[15]This constraint does not change over our sample period. The online appendix table A.3 documents the changes in the Danish Mortgage Act over the 2009 to 2016 sample period. While the constraint does not move during this period, there are a few changes in the wording of the act, a change in the maximum maturity of certain categories of loans in February 2012 from 35 to 40 years, and the revision of certain stipulations on the issuance of bonds backed by mortgage loans. None of these changes materially affect our inferences.

concerns that identification simply comes from different time periods in the data; identification is likely to arise mainly from the large cross section rather than the relatively more limited time series. We also confirm that the inclusion of cohort and cohort-cross-municipality fixed effects in the hedonic model does not materially affect our inferences.

Figure **??** shows a 3-D representation of ℓ against both $\widehat{G}$ and $\widehat{H}$ in the data, averaged in bins of 3 percentage points. ℓ declines in both $\widehat{G}$ and $\widehat{H}$, and there is evidently both *independent* and *interactive* variation along both dimensions. Quantitatively, the conditional variance of $\widehat{G}$ given $\widehat{H}$ is 0.255, and of $\widehat{H}$ given $\widehat{G}$ is 0.130. To better see the independent variation of ℓ along both dimensions, the dotted lines on the 3-D surface indicate two cross-sections in the data corresponding to $G = 0\%$ and $H = 20\%$. These "marginals" reveal that the "hockey stick" profile of ℓ along the $\widehat{G}$ dimension survives, controlling for $\widehat{H}$, and there is also a pronounced downward slope in ℓ along the $\widehat{H}$ dimension, controlling for $\widehat{G}$. In terms of the interactive variation, Figure **??** (left plot) shows how the "marginals" of the listing premium along the $\widehat{G}$ dimension vary as we vary the level of $\widehat{H}$; we discuss this and the other marginal again at the end of the paper, where we evaluate the extent to which we can match these interactive relationships using the model.

3.3.2. Fact 2: "Hockey Stick" and Optimization Frictions

The left-hand plot in Figure **??** shows the probability of a house sale within six months, which we map to $\alpha(\ell)$ in the model on the y-axis, as a function of ℓ on the x-axis. These probabilities are derived from underlying data on the time-on-the-market (TOM) that elapses between sale and listing dates for the properties in the sample.[16] To smooth the average point estimate at each level of ℓ, we use a simple nonlinear function which is well-suited to capturing the shape of $\alpha(\ell)$, namely, the generalized logistic function or GLF (**???**).[17] The solid line corresponds to this set of smoothed point estimates.

[16]Mean (median) TOM in the data is 35 weeks (24 weeks). We pick six months in the computation of $\alpha(\ell)$ to match the median TOM observed in the sample, but the shape is robust to using other windows such as 3, 9, or 12 months. Online appendix Figure A.2 shows the distribution of TOM, which is winsorized at 200 weeks, meaning that we view properties that spend roughly 4 years on the market as essentially retracted.

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

The right-hand plot in Figure **??** shows how $\ln P(\ell) - \widehat{\ln P}$, i.e., the "realized premium" of the final sales price over the hedonic value varies with ℓ. This "conversion" of listing prices to sale prices is the function $\beta(\ell) = \beta_0 + \beta_1(\ell)$ in the model. The plot shows that the average $\beta(\ell)$ rises close to one-for-one with ℓ, corresponding to the estimated coefficients $\beta_0 = -0.083$ and $\beta_1 = 0.865$. The dotted lines in the plot show percentiles of the distribution of price realizations $B(\ell)$ around the average $\beta(\ell)$ in the data, i.e., the distribution $B(\ell)$. As discussed in the model section, we use this distribution to capture the range of possible price outcomes for each level of the listing premium.

The two plots together reveal that in Denmark low list prices appear to reduce seller revenue with little corresponding decline in time-on-the-market, a virtually identical pattern to that in **?**, who studies three U.S. markets.[18] This "concavity" of demand is a confound for estimating λ, because the model predicts two distinct possible drivers for the differential slopes of ℓ^* across gains and losses. When $\lambda > 1$, there are kinks in ℓ^* around $\widehat{G} = 0$, which can be smoothed into a "hockey stick" pattern by variation in θ. The other possibility is buyer sensitivity to ℓ, captured by the shape of $\alpha(\ell)$.

How plausible is the second channel? To check, we exploit regional variation across sub-markets of the Danish housing market. We separately estimate the slope of ℓ in the domain $\widehat{G} < 0$, as well as separate $\alpha(\ell)$ functions (in particular, the slope of $\alpha(\ell)$ when $\ell \geq 0$) in different municipalities of Denmark.[19] The left-hand plot of Panel B of Figure **??** shows how $\alpha(\ell)$ varies with ℓ for municipalities ranked on the magnitude of the slope of $\alpha(\ell)$ when $\ell \geq 0$. This slope has a value between -1.4 and -1.2 for the 5% of municipalities with strong demand concavity, between -1.2 and -0.3 for the middle group, and between -0.3 and 0.1 for the 5% of municipalities with weak demand concavity. The corresponding right-hand plot of Panel B shows the relationship between ℓ and $\widehat{G}$ for municipalities grouped in this fashion. Consistent with the predictions of the model, in areas with weak demand concavity, a relatively steeper slope is visible in the listing premium schedule, and vice versa.

To better pin down the "pass through" of demand concavity to listing premia, we later calculate

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

[19] Municipalities are a natural local market unit—there are 98 in Denmark, with 60,000 inhabitants and roughly 1,800 listings on average. We also re-do this exercise using shires, which are a smaller geographical delineation covering 80 listings on average as a cross-check. For computational efficiency, we group municipalities in three categories, when we incorporate local variation in the structural estimation.

moments in the model that reflect variation of demand concavity across these three municipality groups. We assume that seller preferences are consistent across locations, meaning that the variation of listing behavior that is unexplained by differences in demand concavity helps to pin down the degree of loss aversion. We also note here that the degree of demand concavity should vary with the ease of value estimation and hence price comparison in a given market. If comparable properties are readily available in a local area, buyers in that area should be quick to penalize unusually high premia with sharp declines in the probability of a quick sale. Conversely, if the market is less homogeneous, buyers face a more difficult inference problem to discern whether high listing premia are indeed warranted for specific properties. In support of this logic, we show in Figure **??** that geographic variation in demand concavity is strongly positively related to the homogeneity of the local housing stock, as measured by the share of apartments and "cookie cutter" row houses listed in a given sub-market. An IV strategy confirms that instrumented demand concavity predicts variation in the listing premium slope across different regions of Denmark.[20]

3.3.3. Fact 3: Bunching in Realized Sales

The left-hand panel in Figure **??** plots the frequency distribution of property sales across the dimension of realized gains ($\ln P - \ln R$). Each dot shows the empirical frequency of sales (y-axis) occurring in each 1 percentage point bin of realized gains (x-axis). Observing a counterfactual for this distribution is difficult, as in other settings which attempt to estimate loss aversion using bunching estimators. Our approach is to overlay on this plot (as a dotted line) the empirical frequency of realized sales (i.e., the same y-axis) occurring in each 1 percentage point of *potential* gains $\widehat{\ln P} - \ln R$ (i.e., a different x-axis), as a counterfactual distribution. The counterfactual shows the shape of the realized sales distribution that would result if households were to sell their properties at their model-implied hedonic values.

Using this counterfactual, we see clear evidence of both reference dependence and loss aversion in Figure **??**. First, as in the theory-implied right-hand panel of Figure **??**, the distribution of sales around the reference point is shifted to the right when compared to the counterfactual

[20]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 (we show pictures of typical row houses in Denmark in online appendix H). We discuss the IV strategy in detail in online appendix H.

distribution. This is consistent with $\eta > 0$ in the model. Second, the precise position of the pronounced jump in the distribution at $G = 0\%$, and the distribution of mass to the left and right of this point relative to the counterfactual are also informative about λ. In the theory-implied right panel of Figure **??**, when $\lambda > 1$, the model predicts a jump in the final distribution of house sales precisely at $G = 0$, additional mass in this distribution just to the right of this point, and relatively lower mass in the loss domain, to the left of $G = 0$.

The pronounced bunching that we observe precisely at the point $G = 0$ offers empirical support (which is non-parametric, since it does not require reliance on a hedonic or other model) for the choice of R as the nominal purchase price (see **?**, for a discussion of bunching at reference points). The plot also clearly reveals two important additional features of the data: sizeable "diffuse bunching" excess mass just to the right of zero, which we discuss in greater detail below, and a small but visible "hole" just to the left of $G = 0$, which we also later attempt to rationalize by adding a "notch" to seller preferences.[21]

An appealing feature of this counterfactual is that the hedonic model (or a very similar variant) is widely used in this market setting for valuation, and can be verified using realized prices. The standard polynomial counterfactual approach would not allow us to gauge the extent of shifts in mass which are predicted in the presence of reference-dependent loss-aversion, as that approach by definition fits the observed empirical distribution up to the range over which the counterfactual is extrapolated.[22] In this sense, our approach is more similar to **?**, who extracts evidence of loss aversion from U.S. tax returns data, and employs a model to gauge expected tax avoidance costs and benefits—which would otherwise be difficult to measure.

In the right-hand panel of Figure **??** we calculate a measure of excess mass relative to the counterfactual in each bin of realized gains. This allows greater precision on the exact magnitude of bunching. For example, for realized gains of zero, we find that the observed frequency of realized gains is 69.6% higher than under the counterfactual, with a bootstrap standard error

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

[22] We also implement this approach in the online appendix Figure A.7, following **?** and **?**. When doing so, we exclude bins near the threshold, and extrapolate the fitted distribution to the threshold, excluding one bin on each side of the zero gain bin, i.e. $j \in \{-1\%, 1\%\}$, with a polynomial order of 7. The results are robust to other polynomial orders and to variations of the excluded range, and generate similar results on the excess bunching mass at zero.

of 7.5%. In the one-percentage point bin immediately to the left of zero, the observed frequency is 24.3% lower than under the counterfactual, with a bootstrap standard error of 3.8%.

Online appendix Figure A.6 reports sale transaction frequencies which show the degree of bunching in a similar 3-D fashion. We confirm that regardless of the level of $\widehat{H}$, there is a visible shift of mass from the $\widehat{G} < 0$ domain to the $\widehat{G} > 0$ domain.

3.3.4. Fact 4: Extensive Margin: Probability of Listing

To understand the decision to list, we turn to data on the total housing stock in Denmark, using $5,538,052$ property-years in the data. We compute that the unconditional average annual listing propensity is 3.91% of the housing stock. This corresponds to between 3% and 4.5% of the housing stock listed across sample years.

A model-consistent explanation for the average propensity to list is beyond the scope of the paper, because it would require us to take a strong stance on the factors that drive the moving decision, which we currently summarize using our estimates of θ. Instead, we focus on the variation of listing behavior around the reference point. Figure **??** plots the listing propensity at each level of $\widehat{G}$, which comes from estimating $\widehat{\ln P}$ for all properties in Denmark for which we have data on the nominal purchase price R.[23] The figure shows a mild increase in the probability of listing as $\widehat{G}$ increases, which is consistent with $\eta > 0$, and potentially, $\lambda > 1$.

3.3.5. Robustness

Unobserved Quality and Measurement Error

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

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

We make a concerted effort to address these measurement error concerns, outlining in online appendix I how different sources of measurement error conceptually affect our inferences about underlying structural parameters, and proposing specific fixes for each particular case. Following **?**, we differentiate between two types of measurement error, namely, (potentially time-varying) unobserved quality (ν_{it}), and idiosyncratic over- or under-payment by the seller at the point of purchase (ω_{it}). We outline how these sources of error can bias estimated moments and yield misleading inferences about true underlying relationships. We outline assumptions under which alternative empirical models of $\widehat{P}$ can unwind these biases in inferences, permitting recovery of clean estimates of true underlying relationships. As we discuss in greater detail below and in online appendix I, we also simulate the effects of varying degrees of unobserved quality and recover parameter estimates from this exercise, which provides further theoretical confidence in the robustness of our structural estimation approach.[24]

We apply these concepts by comparing how key data moments vary when we i) employ a standard hedonic model; ii) employ repeat sales models to difference out time-invariant unobservable components; iii) use as regressors time-varying hedonic characteristics and novel tax exemption data on home renovation expenses to narrow in on time-varying changes in unobservable quality, and iv) combine all of these features in a single hedonic model. The important patterns in the data, including the hockey stick in listing premia over potential gains, and the relationship between sale probabilities and listing premia (i.e., concave demand) are robustly visible across these different models of $\widehat{P}$, and quantitatively similar.

Perhaps more importantly, we later show that the internal consistency provided by structural estimation leads our identification of parameters to be robust to estimation error in $\widehat{P}$.[25] Unobserved quality continues to bias the slope of the "hockey stick" upwards in our setup, and when viewed through a reduced-form lens, a steeper slope can mistakenly be interpreted as evidence for a high degree of reference dependence. In the model, however, listing premia are not directly informative about underlying preferences; the identification of preference parameters also de-

[24] We thank an anonymous referee for suggesting that we pursue this.

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

pends on the extent to which "fishing" for a higher price pays off in the market, captured by the demand function $\beta(\ell)$. Interestingly, it turns out that unobserved quality leads to offsetting bias in estimated listing premia and the estimated $\beta(\ell)$ function. This is because while unobserved quality can affect sellers' listing premia, it countervailingly affects the prices that buyers will be willing to pay for properties. This provides insight into why structurally recovered underlying preference parameters from the model remain robust to unobserved quality.

Bunching: Round Numbers, Holding Periods and Reference Price

We show in online appendix J.1 that the bunching patterns documented in Figure **??** are robust to a number of issues previously identified in this literature (e.g., **??**). The spike in sales volumes at $G = 0$ and the patterns of excess mass relative to the counterfactual distribution do not appear to be driven by bunching at round numbers, remaining striking and visible when we exclude up to 20% of all observations. We also show that these bunching patterns are robust when we split the sample into five groups based on the time between sale and purchase, i.e., the holding period of the property. Except for the sub-sample with the longest holding period (> 12 years, 20% of the data), we find strong evidence of bunching at $G = 0$. Finally, we find strong evidence of bunching in all cases when we split the sample into quintiles based on the level of R, with quintile cutoffs ranging from around $658,000$ DKK to 1.9MM DKK. Together, these checks assuage concerns that bunching could result purely from these differences in underlying properties.

We now turn to describing our structural estimation approach.

3.4. Structural Estimation

3.4.1. Moments

To transparently map the patterns in the data back to underlying parameters, we use the binscatter "dots", i.e. binned averages visible in Figures 2 to 7 as our moments. This choice avoids the need for additional parametric assumptions about the data that might not be directly consistent with the model. We enumerate these moments below.

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

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

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

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

Finally, for each 1-percentage point bin of potential gains in the full sample of listings, we calculate the average listing premium $\ell_{k\in\{High,Mid,Low\}}(\widehat{G})$ in three municipality groups, distinguishing between the top 5% locations where the slope of demand decreases most sharply ("high" demand concavity), the 5% of locations where this slope is flatter ("low" demand concavity), and the locations where the slope of demand lies between the 5%-95% percentiles (79×3 moments, Figure **??** Panel B, "Cross-sectional variation" in Table 1).

We collect all $79 \times 7 = 474$ empirical moments obtained in this way in the vector $\mathbf{M_d}$:

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

3.4.2. Moments in the Model

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

For computational reasons, we need to place restrictions on the distribution of the moving shock θ. Only a small fraction (around 4%) of the property stock is listed for sale every year. Solving the model for the full range of $\theta's$ in the population is computationally burdensome, so we re-interpret the distribution of $\theta's$ as a conditional version of the full population. To facilitate this, we make the identifying assumption that the marginal value θ_{min} which corresponds to the 1^{st} percentile of the conditional distribution of θ comes from an expected utility of 0, experienced by a financially unconstrained seller with potential gains equal to $\widehat{G}_+ = 40\%$. Moreover, since a model of the unconditional listing probability is outside the scope of this work, we normalize the listing probability for this potential gains level of $\widehat{G}_+$ to equal its observed value in the data.

We compute the solution to the seller's problem numerically on a three-dimensional grid, for each 1-percentage point bin of potential gains and potential home equity, and for each realization of the moving shock θ. We solve the model for both precise and imprecise targeters, using the formulation of expected utility described in equation (**??**). The aggregate model-implied moments $\mathbf{M_m}(\mathbf{x})$ are calculated according to the mixture distribution of the seller population, where a fraction π of sellers are precise targeters, and the remainder are imprecise targeters.

3.4.3. Estimation

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

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

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

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

conditional on a choice for the weighting matrix $\mathbf{W}$. Online appendix section B.8 describes our numerical optimization procedure in greater detail.

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

We first employ an equal-weighting scheme in which $\mathbf{W}$ is the identity matrix. To ensure that the scale of the moments does not influence their relative weights, we normalize each binned value for each moment by its average value across all bins, i.e., we evaluate the model fit in terms of relative prediction errors. In model versions in which we use cross-sectional moments, we assign a weight of 1/3 to each municipality group. This relative over-weighting of the municipality groups with extreme levels of demand concavity is an attempt to force the model to explain variation in the listing premium after appropriately accounting for the "pass-through" of local market conditions. We later check the robustness of our inferences to inverse-variance-weighting, i.e., we use the bootstrap draws to construct a diagonal matrix of moment variances $\mathbf{V}$, and set $\mathbf{W} = \mathbf{V}^{-1}$.

3.4.4. Results

The model is complex, and there are many different patterns in the data. To understand the forces in the model and the sources of parameter identification, we therefore approach full structural estimation gradually. We build up from more simple model variants with fewer parameters to more complicated models, and adding subsets of moments as we go along. This culminates in estimating the entire set of model parameters using the full set of empirical moments. Table **??** shows the 8 model variants that we estimate, listed in rows. The columns labelled "Param-

eters" show the specific subsets of parameters estimated in each model along with associated bootstrap standard errors, the columns labelled "Optimization frictions" indicate whether or not we allow for concave demand and imprecise targeting, and the remaining columns indicate the subset of moments used to estimate the parameters in each row.

Model variant 1 only uses the "hockey stick" moments in Figure **??**. We match these moments using a stripped-down version of the model involving a representative seller with no financial constraints, in a market with linear demand and all sellers able to precisely target outcomes (i.e., $\pi = 1$).[26] In this special case, we solve analytically for the optimal ℓ (see online appendix section B.1), and find that η drives the listing premium slope when $\widehat{G} > 0$; λ generates a difference in slope in the domain $\widehat{G} < 0$; and the level of ℓ is driven by both η and θ_m, the level of the moving "shock" in this simple case. The first row of Table **??** shows that this model delivers $\eta = 0.340$, $\lambda = 2.256$ and $\theta_m = 0.525$, all of which are highly statistically significant using the bootstrap.

We next check how correctly modelling demand using the concave α function estimated in the data and visible in Figure **??** affects parameter estimates. When we incorporate this feature, Model 2 in Table **??** shows that $\eta = 0.848$ is estimated higher, loss aversion $\lambda = 1.468$ is lower, and the fitting parameter θ_m adjusts upwards to a level of 0.774.[27] As surmised, accounting for concave demand substantially decreases estimated loss aversion (which nonetheless remains statistically significantly greater than 1), but it also increases measured reference dependence. This is because the relative "flattening out" of ℓ when $\widehat{G} > 0$ (and thus, the asymmetry between the slopes in the gain and loss domains) can partly be explained by optimal responses to demand concavity in addition to loss aversion. This in turn increases the degree of reference dependence needed to explain the negative slope of ℓ when $\widehat{G} < 0$.

Listing premia are ex-ante choices by sellers. An alternative ex-post measure is offered by bunching in the distribution of realized prices in actual transactions. Model 3 estimates reference dependence and loss aversion parameters using bunching and the hockey stick moments together. Using this additional moment increases the precision of the λ estimate, but the point estimate $\lambda = 1.142$ is now *lower*, tightly controlled by the bunching seen precisely at zero

[26] The slope of $\beta(\ell)$ that we assume in this case is plotted as a straight dotted line in Figure **??**.

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

gains. This tight control comes from the assumption of precise targeting, i.e. $\pi = 1$. The lower estimated λ decreases the fit to the slope of ℓ when $\widehat{G} < 0$. This means that a higher $\eta = 0.998$ is now required to match the hockey stick (i.e., the curve rotates, giving up fit in the domain $\widehat{G} > 0$ to better match the domain $\widehat{G} < 0$).

Model 4 adds imprecise targeting as an additional optimization friction, meaning that a fraction $1 - \pi$ of sellers are unable to either anticipate or precisely control the outcome of price negotiations in the market. This allows the diffuse mass observed to the right of $G = 0$ to be used to pin down λ. However, as illustrated in Panel B of Figure **??**, imprecise targeting generates an identification problem. More specifically, the parameter π is not uniquely identified by the hockey stick pattern when $\pi < 1$. To understand this, consider the fact that a purely downward-sloping straight line can be obtained in two cases. The first is when $\pi = 0$ with linear reference dependence, and the second is when $\pi > 0$ and $\lambda > 1$. We therefore allow the fraction π to be estimated jointly by the listing premium profile ("Hockey stick") and the pattern of mass shifts in the distribution of realized gains ("Bunching"). With $\pi > 0$, some sellers who would otherwise bunch at the reference point now do not have an ability to precisely target this point, and therefore the relatively low magnitude of strict bunching at the reference places less of a restriction on the magnitude of loss aversion in the model. With this assumption, we find a higher value of $\lambda = 2.169$, consistent with a significantly lower value of $\pi = 0.106$, and a lower degree of overall reference dependence $\eta = 0.512$.

Model 5 in Table **??** adds down-payment constraints into the model, and simultaneously adds in the variation of the listing premium along the home equity dimension as additional moments. This change slightly reduces the estimated η to 0.403, since responses to down-payment constraints can explain some of the variation in listing premia previously attributed solely to potential gains. Moreover, estimated $\lambda = 2.596$ rises in this case, as the model attempts to match the observed slope of listing premia in the loss domain, but with a lower level of reference dependence overall. Model 5 shows that the estimated penalty parameter for the home equity constraint is $\mu = 2.764$, which is empirically realistic.[28]

An unrealistic feature of all model variants considered thus far is the assumption of a representative seller, with a single value of the moving "shock" $\theta = \theta_m$. This is very restrictive, because

[28]This coefficient provides an important validation opportunity for the model. In online appendix section B.9, we show that the penalty implied by the estimated value of μ is consistent with its counterpart in the data.

it does not allow for an extensive margin decision of whether to list or not, and any resulting selection effects. If we were to assume just $\phi > 0$, without heterogeneity in θ, either all owners would list (for a low enough value of ϕ), or no owner would list (for a high enough value of ϕ). The extensive margin allows us to jointly identify the two parameters θ_σ and ϕ, through the level and the slope of the listing probability profile. Model 6 adds in the extensive margin moments seen in Figure **??** to the set, to help estimate these parameters. Estimated $\theta_m = 1.345$ in this model variant, and heterogeneity $\theta_\sigma = 0.711$. To interpret magnitudes, note that this is normalized relative to the hedonic value of the house, meaning that the lifetime contribution to utility discounted back to the present is 1.345 of the hedonic value of the house, or equivalently 7.5% per year, assuming a discount rate equal to 3.80%, i.e., the average return to long-term mortgage bonds between 2009 and 2016, and a holding period of 30 years. Additionally, we find search costs equal to $\phi = 0.050$, i.e., 5% of the property price, which is once again empirically realistic.[29] This is accompanied by a slightly lower $\lambda = 2.319$, while both the estimated degree of reference dependence and the down-payment constraint increase. To understand this, note that the effect of higher dispersion in θ is to smooth the observed listing premium slopes, which means that an overall higher level of reference dependence is necessary to match the data. The higher η now also increases the magnitude of strict bunching generated by the model, leading the loss aversion parameter λ to decrease.

Figure **??** also exhibits a small but clearly evident volume of missing mass for realized gains in the one-percentage point bin just below zero. In the bottom two rows of the table, we extend the model with an additional feature, a "notch" in seller preferences around zero. This is modelled as a discontinuous jump of magnitude ζ for realized gains just below zero, i.e.:

$$u(P(\ell), R) = \begin{cases} P(\ell) + \zeta + \lambda\eta G(\ell), \text{ if } G(\ell) < 0 \\ P(\ell) + \eta G(\ell), \text{ if } G(\ell) \geq 0 \end{cases}. \tag{3.7}$$

Using this augmented model, we find that the parameter $\zeta = -0.006$ is quite imprecisely estimated, despite the very clearly visible effects in the distribution of realized gains in the data. This is consistent with the observation of **?** that small magnitudes of notches can be observed to have large effects on the observed shifts of mass. To show how the model fits the

[29] We note here that the magnitude of the "fitting parameters" θ and ϕ also embeds the effect of dynamic features such as time discounting and the sequential bargaining process between buyers and sellers, as well as other soft factors like the hassle of allowing visitors in the house, the value of regret, and any other drivers of the option value of the listing.

data, Figure **??** shows the moments and the model predictions together.

Finally, Model 8 additionally considers the set of cross-sectional (i.e., regional) moments, i.e., it uses "local" demand conditions across the three groups of municipalities. Here, estimated $\lambda = 2.473$ increases slightly. In the data, listing premia in municipalities in which demand is very close to linear do not show significant variation in slopes across gain and loss domains; and in municipalities in which demand is highly concave, the listing premium is non-linear and precisely consistent with the non-linearity of demand in those locations. Figure **??** shows that the model is well able to capture this pattern. Moreover, as before, the strong rejection of $\lambda = 1$ arises from the sharp bunching seen in Figure **??**. With the inclusion of cross-municipality moments the model is better able to separate the role of concave demand from the impact of heterogeneity in moving shocks. This leads to lower estimated dispersion of θ, and a slightly higher estimate for the search cost, with all other parameters only being marginally affected.

Overall, the model is well able to capture the broad contours of the data, with an average prediction error of 6.95%.[30] In the next section, we dig deeper into validating the model, and identify where it is unable to match the data. We also consider the usefulness of model extensions, and discuss what we learn from any gaps that remain between the model and the rich patterns observed in the data.

3.5. Validating the Model

3.5.1. Measurement Error and Robustness

We earlier discussed measurement error and unobserved quality. The top rows of Table **??** alter the estimation of $\widehat{P}$ to account for different sources of measurement error, recompute all affected moments using the re-estimated $\widehat{P}$, and then re-estimate structural parameters on the new moments. Model 'Renovations' augments the baseline hedonic model in equation **??** with lagged information on tax exemptions sought by owners for renovation expenses to capture time-varying unobserved quality changes occasioned by renovations. Model 'Repeat Sales I'

[30] In online appendix Table A.4, we report prediction errors for a wider set of moments. In online appendix section F we show that an alternative concave formulation for the financial penalty function allows us to match listing premia along the home equity dimension better, but it also entails a more computationally burdensome parameterization, without any material impact on the identification of the main structural parameters.

is a pairwise repeat sales model which includes both time-varying hedonic characteristics and lagged renovation expenses. The model includes the lagged pricing residual from the same house's previously traded price as a right-hand side variable, a strategy similar to **?**. Model 'Repeat Sales II' generalizes the repeat sales approach, and additionally includes the average of *all* past pricing residuals available since 1992 for each house where available, to use information from all past repeat sales observed. A more comprehensive discussion of the rationalization for and implementation of these models is in the online appendix section I.3.

Through simulation, we also check robustness of the estimated model parameters to information about property quality known by the seller, and embedded in their choice of list price, but hidden from the econometrician, and therefore not reflected in $\widehat{P}$. We simulate the extent of this bias, re-construct the empirical moments based on de-biased values of $\widehat{P}$, and report estimated structural parameters using this set of adjusted moments in rows 4-6 of Table **??**. Across different plausible levels of such unobserved quality, we find that the structural parameters we recover are roughly of similar magnitudes to our baseline estimates, except for a somewhat less pronounced down-payment constraint penalty, and a higher level of the "fitting parameter" θ_m, which adjusts to capture a downward shift of the average listing premium.

This exercise also provides deeper insights into how unobserved quality affects parameter identification. Sellers that are asymmetrically informed about property quality will exhibit steep listing premia that contain the wedge between true quality and measured quality, but the insight is that the same factor strengthens the correlation between listing prices and final outcomes. When we de-bias the listing premium to account for unobserved quality, its average level decreases, and the "hockey stick" flattens out, but the very same bias correction also leads to a flatter slope of $\beta(\ell)$. A lower $\beta(\ell)$ reduces the "fishing" incentive of sellers, meaning that even without any adjustment of preference parameters, sellers optimally choose lower listing premia.[31] This line of reasoning explains why we do not find substantial changes in preference parameters in this simulation, and suggests that the internal consistency provided by structural estimation can help to deal with the contaminating effects of unobserved quality. Online appendix section I.7 discusses this issue in more detail.

[31] If unobserved quality is at play, sellers will have an informational advantage relative to the econometrician, but this information is revealed to the buyer during property inspection, and will be reflected in prices negotiated between buyers and sellers. This means that final transaction prices are more tightly linked to listing prices. In contrast, $\alpha(\ell)$ is determined both by buyer arrival rates as well as by post-inspection buyer offers, meaning that this function is relatively less affected by unobserved property quality.

Finally, the bottom row of Table **??** shows values of estimated structural parameters under an alternative inverse-variance weighting matrix for empirical moments. Reassuringly, across all robustness checks, relative to Model 8 in Table **??**, the point estimates of most parameters remain similar, with no material changes to the qualitative interpretations from that model.[32]

3.5.2. Interactions Between Preferences and Constraints

The three-dimensional patterns of listing premia observed in Figure **??** suggest that there is considerable variation in the slope of the relationship between ℓ and $\widehat{G}$ as $\widehat{H}$ varies. It appears as if the effects of losses and constraints *interact* with one another, and that the factors affecting household behavior are neither one nor the other variable viewed in isolation.

The left-hand plot in Panel A of Figure **??** compares the model predicted and observed listing premium profiles along the $\widehat{G}$ dimension, as the potential home equity position of the seller varies. The important new fact is that less constrained sellers are likely to respond more strongly in their listing decision to their gain/loss position, while unconstrained households exhibit seemingly greater levels of reference dependence.[33]

The corresponding plot in Panel B of the figure reports more formal estimation results from a regression of listing premia on potential gains, conditioning the coefficient on the level of potential home equity. This new fact on the interaction of preferences and constraints appears to require a more intricate model of preferences and/or constraints than the literature has thus far proposed. Our model cannot rationalize this feature of the data, even though it includes many of the forces proposed in the literature. With a view towards motivating theoretical work on a broader class of preference and constraint specifications, we conjecture that the luxury of being unconstrained appears to allow more psychological motivations such as loss aversion to come to the fore. We discuss this further in the conclusion to the paper.

[32] One visible change concerns the estimated magnitude of down-payment constraints. Compared to Model 8 in Table **??**, the estimated values of the parameter μ in the case of hedonic models "Repeat Sales I" and "Repeat Sales II" appear slightly lower. We can attribute this to the fact that the repeat sales sample is restricted (it contains only properties for which multiple past sales are observed), and happens to have a greater number of younger, urban households with high net financial assets—consistent with lower down-payment constraints for this sample.

[33] The only real way to justify this interaction in the model is that constrained households already choose higher listing premia, thus leaving less room to react in the face of demand concavity. However, the quantitative magnitude that this line of logic can generate is very modest, meaning that virtually all of the interaction effect cannot be explained by our canonical model.

In contrast, the two right-hand plots of Figure ?? compare the model predicted and observed listing premium profiles along the $\widehat{G}$ dimension, as the potential home equity position of the seller varies. The reaction of sellers to changes in their potential home equity position seems to be independent of the realization utility associated with their accumulated nominal gains, an observation that the model captures well.

3.5.3. Price-Volume Correlation

Housing market volume is significantly lower in falling than in rising housing markets, a finding which originally motivated the investigations in ? and ?. Reference dependence explains this finding by households listing their houses at levels that depend on the reference point. This can make housing markets sluggish when house values decline—making list prices unrealistically high—and more active when markets rise. Down-payment constraint-based explanations posit that when households are faced with down-payment constraint-imposed adjustments when housing markets fall, they may set higher listing prices and be prepared to tolerate longer times-on-the-market, or decide not to list at all, generating the price-volume correlation.

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

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

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

$$N_i = s(\widehat{G}_i, \widehat{H}_i) \times \alpha(\ell(\widehat{G}_i, \widehat{H}_i)) \times N_{stock,i}. \tag{3.8}$$

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

In Denmark, aggregate prices and volumes declined prior to the start of our sample window in 2009, remained relatively flat until 2014, and then rose until the end of our sample window in 2016 (see appendix Figure A.8). To check whether our model can capture this price-volume correlation in the data, we first estimate the coefficient ρ from the following regression:

$$\Delta \ln N_{m,t} = \mu + \rho \Delta \ln \widehat{P}_{m,t} + \epsilon_{m,t}, \tag{3.9}$$

where $N_{m,t}$ are the numbers of transaction in municipality m in year t, and $\widehat{P}_{m,t}$ is the corresponding average hedonic price level in municipality m in year t. The empirical correlation between prices and volumes measured this way is reported in the left-most column of Panel A in Table **??**, and equals $\rho = 0.515$.

We then use the observed change in valuations $\Delta\widehat{P}_{m,t}$ to calculate model-implied changes in transaction volumes implied by equation (**??**), and re-estimate the model-implied correlation coefficient in equation (**??**).[34] We use the set of structural parameters estimated in row 8 of Table **??**, with $\eta = 0.629$ and $\lambda = 2.473$. This delivers a price-volume correlation equal to $\rho = 0.498$, which is a modest under-prediction of the value of ρ identified in the data.

In the bottom rows of Table **??**, we then decompose this model-implied correlation into the different model ingredients. First, we confirm that in a frictionless version of the model with no financial constraints ($\mu = 0$) and no reference dependence ($\eta = 0$), transaction volumes

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

are completely independent of price movements in the model. We then augment the model, and find that down-payment constraints account for 58.5% of the co-movement, with reference dependence and loss aversion accounting for 18.2% and 23.2%, respectively.

The important role of down-payment constraints for the aggregate dynamics of prices and volumes over the sample period that we consider is not surprising, given that constrained sellers are very prevalent in the data. They account for a large share (48%) of all sellers with an outstanding mortgage, which is 35% of the entire sample. For comparison, a much lower share of 23% of sellers in the data face the possibility of realizing a loss. Perhaps more importantly, we also find that in the data constrained sellers respond more strongly to an improvement in their home equity position than loss-averse sellers do to the possibility of realizing a loss. On average, the listing premium decreases by 0.37% in response to a 1% increase in potential home equity, and by only 0.27% for a 1% increase in potential gains.

3.6. Conclusion

We structurally estimate a new model of house selling decisions on comprehensive Danish housing market data, and acquire new estimates of key behavioral parameters and household constraints from this high-stakes household decision. Our parameter estimates are consistent with a high degree of reference dependence in house sellers: they appear to care greatly about nominal gains and losses relative to the original purchase price. We also find strong evidence of loss aversion in this important field setting, complementing previous field evidence in the literature (**??????**), with a point estimate for the disutility contribution of losses equal to 2.473 times the utility contribution of gains.

We use these estimated parameters to quantify the relative contributions of down-payment constraints and reference dependence to housing market turnover, helping to pin down the sources of the asymmetric effects of house price changes on house selling decisions. This provides more precise answers for why property owners appear "locked in" to their houses during market downturns (**??**), thus aiding a better understanding of labor mobility, and informing mortgage market design and policy (**??**).

Perhaps most intriguingly, our model cannot completely match new facts that we bring to light

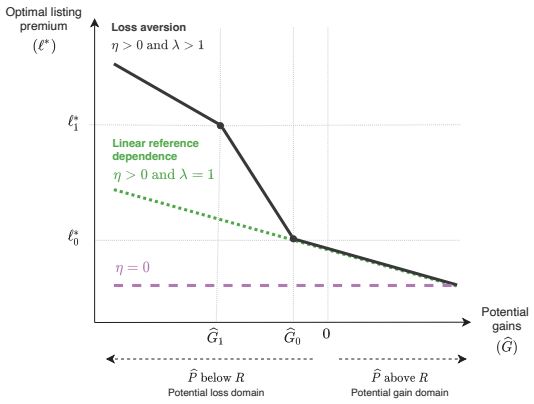
in the high-quality administrative data that we employ. We view these facts as a new target for behavioral economics theory and structural behavioral economics (**?**). We find that nominal losses and down-payment constraints interact with one another, in the sense that reference dependence and loss aversion are less evident when households face more severe constraints, but quite pronounced when households are relatively unconstrained. In micro terms, this interaction between reference dependence and constraints could have implications for the way we model behavior. We tend to assume that agents optimize their (potentially behavioral) preferences subject to constraints, and in numerous models, agents may also wish to impose constraints on themselves to "meta-optimize" (**?????**). However, if constraints affect the incidence of behavioral biases, or if being in a zone that is more prone to bias affects the response to constraints, our models may need to become richer to predict such behavior.

FIGURE 3.1: REFERENCE DEPENDENCE AND LOSS AVERSION

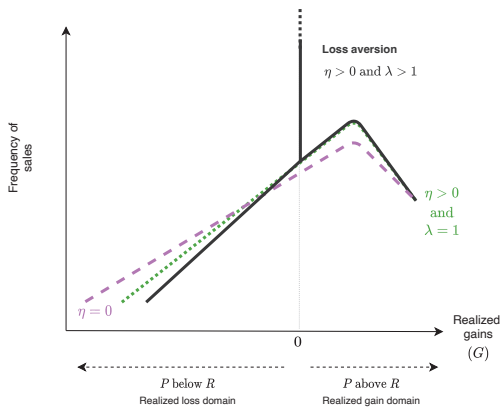
Panel A

Precise targeting of the final price

Optimal Listing Premium by Potential Gains



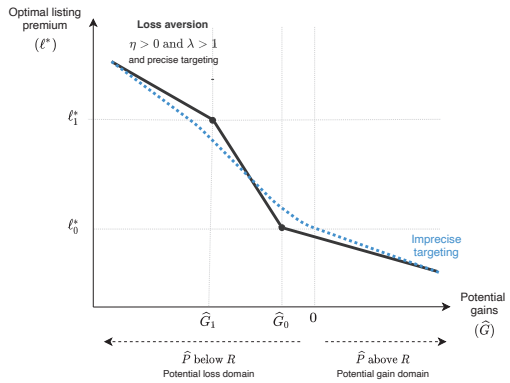
Realized Gains Distribution



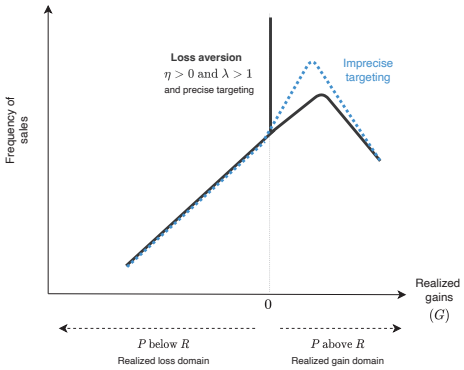
Panel B

Imprecise targeting of the final price

Optimal Listing Premium by Potential Gains



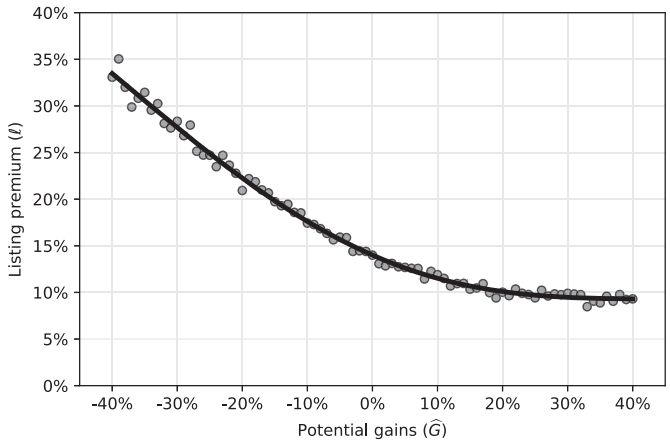
Realized Gains Distribution



The figure illustrates how each specification of the utility function is reflected in the seller's optimal choice of listing premia (left-hand side panel) and distribution of realized gains (right-hand side panel). The interval delimited by $\widehat{G}_0$ and $\widehat{G}_1$ illustrates the incentive of loss averse sellers to bunch at realized gains of exactly zero. Sellers with potential gains $\widehat{G} \in [\widehat{G}_1, \widehat{G}_0]$ choose listing premia ℓ such that $\beta(\ell) = -\widehat{G}$ and thus, conditional on a sale, their realized gain is equal to $G = 0$. We plot a stylized version of listing premium profiles, for the case in which demand functions $\alpha(\ell)$ and $\beta(\ell)$ are linear, the household is not facing financing constraints, and property prices are increasing on average. This generates greater mass in the nominal gain relative to the nominal loss domain, even in the case with no reference dependence ($\eta = 0$). In the online appendix, we describe and solve an analytical version of this model.

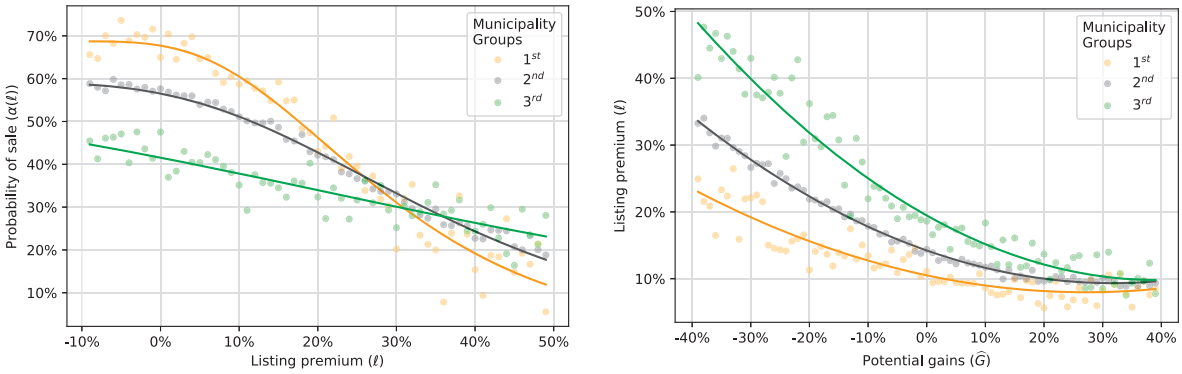
FIGURE 3.2: LISTING PREMIA AND POTENTIAL GAINS

Panel A



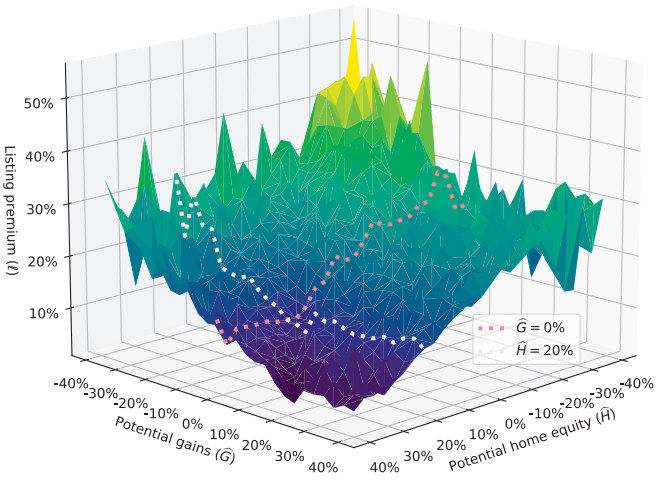
Panel B

Cross-Sectional Variation



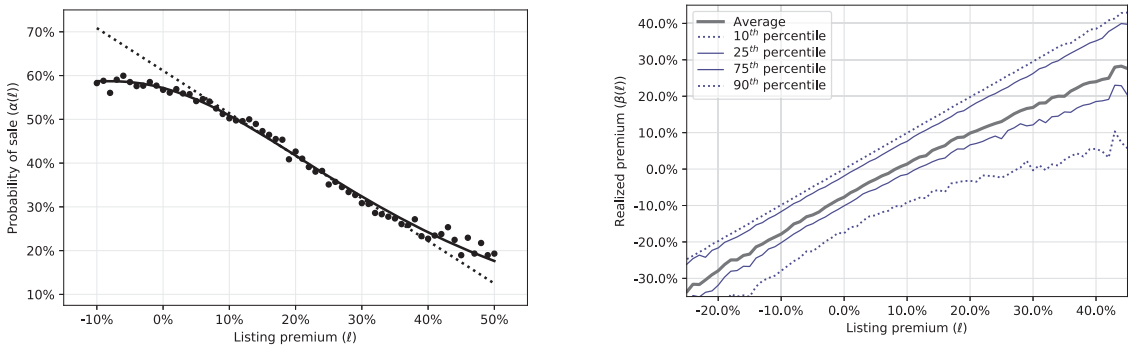
Panel A reports binned average values (in 1 percentage point steps) for the listing premium (ℓ) for different levels of potential gains ($\widehat{G}$). The solid line corresponds to a polynomial fit of order three. Panel B shows demand concavity (left-hand side panel), i.e. the probability of sale within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

Figure 3.3: Listing Premia Across Gains and Home Equity



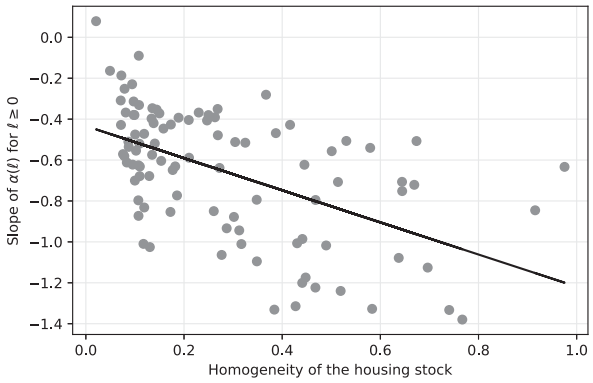
The figure reports binned average values (in steps of 3 percentage points) for the listing premium (ℓ) along both levels of potential gains and home equity.

Figure 3.4: Optimization Frictions in the Data



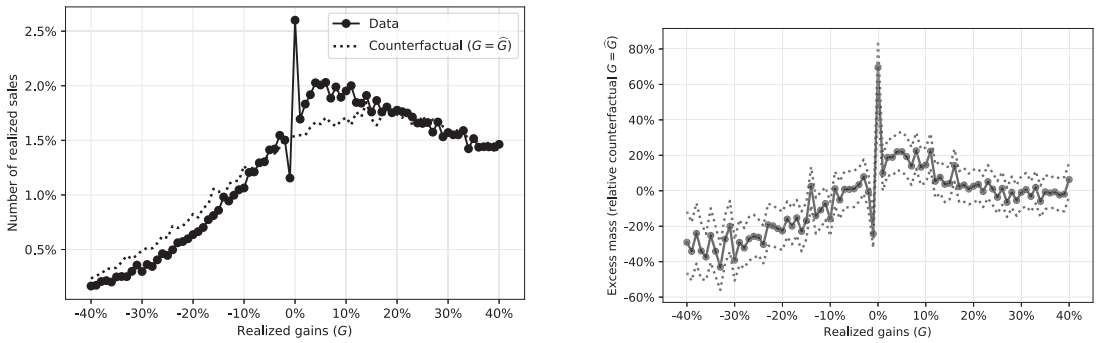
The left-hand side figure reports the average probability of sale within six months $\alpha(\ell)$ across 1 percentage point bins of the listing premium in the sample. The right-hand side of the figure shows the average realized premium $\beta(\ell)$ across 1 percentage point bins of the listing premium.

Figure 3.5: Geographic Variation in Concave Demand and Housing Stock Homogeneity



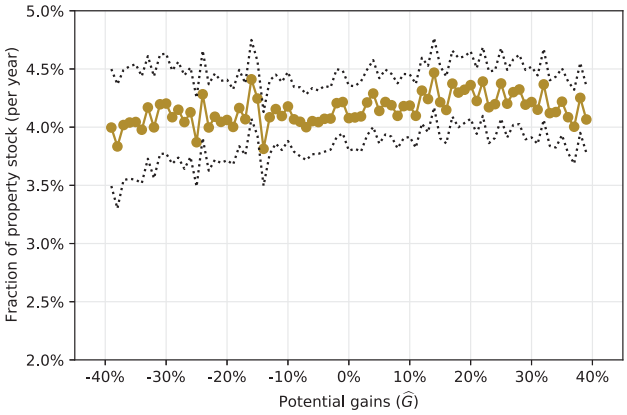
The figure plots the estimated degree of demand concavity, measured as the slope coefficient of the effect of an increase in the listing premium on the probability of sale within six months, for positive listing premia ($\ell \in [0, 40]$) on the y-axis, against the degree of homogeneity of the housing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

Figure 3.6: Bunching Around Realized Gains of Zero



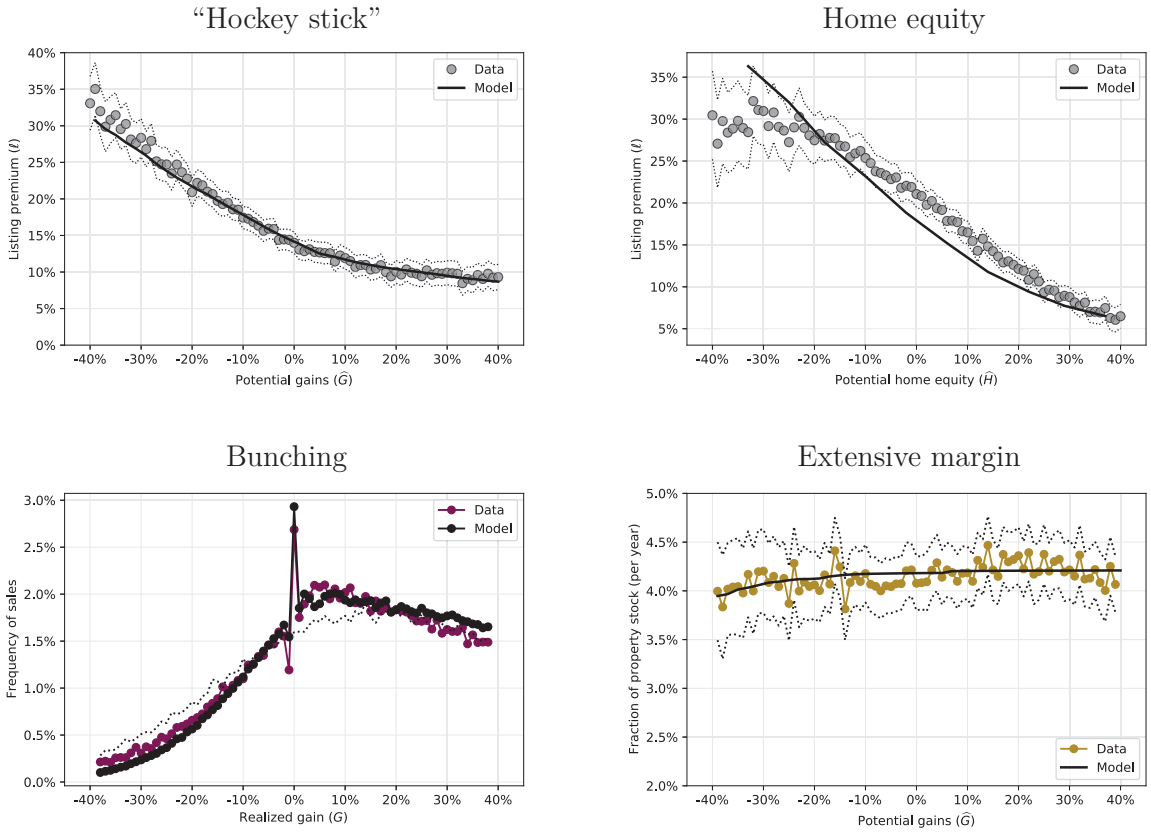
The left-hand panel reports binned frequencies of observations (in 1 percentage point steps) for different levels of realized gains (G). The dotted line shows the counterfactual corresponding to the distribution of potential gains ($\widehat{G}$) in the sample of realized sales. The right-hand panel reports frequencies of observations of realized gains relative to the level of the counterfactual. Dotted lines indicate 95% confidence intervals based on bootstrap standard errors.

FIGURE 3.7: EXTENSIVE MARGIN



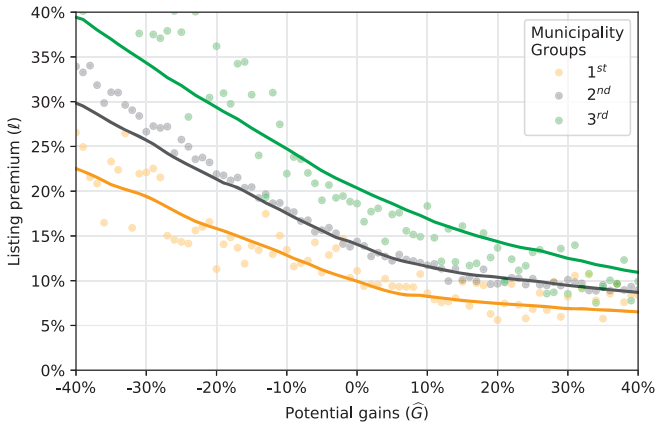
The figure reports the average annual probability of listing a property for sale. We first calculate the potential gain level for each unit in the stock of properties in Denmark, for each year covered by our sample of listings, using the same hedonic model used to calculate potential gains in the sample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

FIGURE 3.8: OVERVIEW OF MODEL FIT



The figure reports our set of moments in the data and in the model, evaluated at the set of parameters which correspond to the complete version of the model and the complete set of empirical moments, as indicated in line 7 of Table **??**. Dotted lines show 95% confidence intervals based on bootstrap standard errors.

FIGURE 3.9: "HOCKEY STICK" ACROSS MUNICIPALITIES

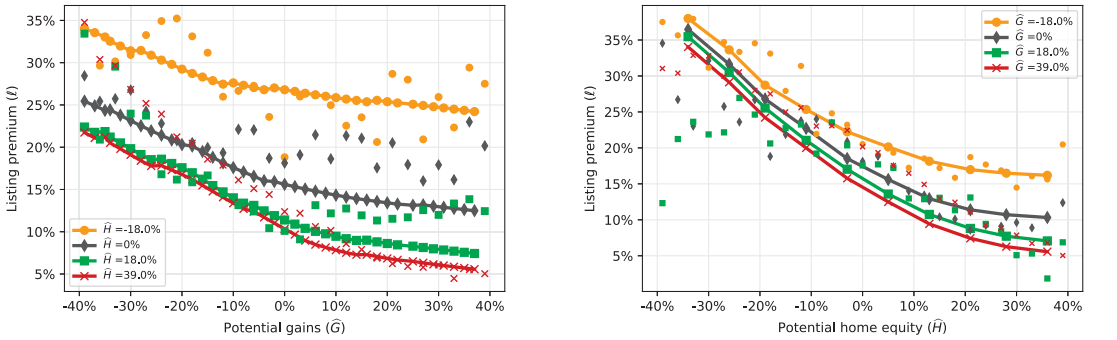


The figure reports binned average listing premia by potential gains, distinguishing between three different municipality groups, according to their concavity of demand: bottom 5%, middle 90%, top 5%. The degree of demand concavity is measured as the slope coefficient of the effect of an increase in the listing premium on the probability of sale within six months, for positive listing premia ($\ell \in [0, 40]$), for each municipality. Dotted lines indicate average listing premium levels in the data, calculated for 1-percentage point bins, and the solid line their counterparts implied by the model. We adjust the average level of the model prediction to correspond to the data, since the level of the listing premium depends on other parameters, most notably the average size θ_m of the moving shock, which may be heterogeneous across locations. We evaluate the model fit at the set of parameters which correspond to the complete version of the model, and the set of empirical moments indicated in the bottom row of Table **??**.

FIGURE 3.10: OUT-OF-SAMPLE FIT: INDEPENDENT AND INTERACTIVE VARIATION

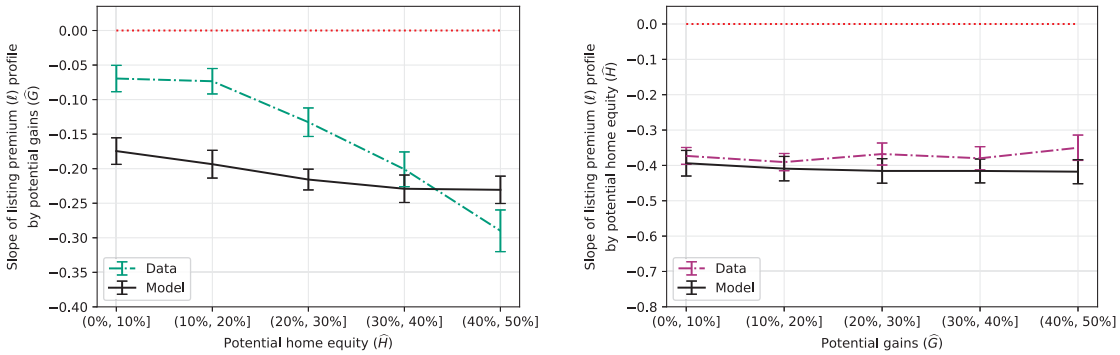
Panel A

Listing premia by potential gains and home equity



Panel B

Conditioning on home equity and potential gains



The figure reports the model fit for conditional listing premia profiles, conditioning on different levels of home equity and potential gains. The model is evaluated at the set of parameters estimated for the complete set of empirical moments, as indicated in row 8 of Table **??**. In Panel A, individual dots indicate average levels of listing premia in the data, and solid lines with corresponding markers are their model-implied counterparts. Interactive effects between home equity and the degree of reference dependence are visible in the left-hand plot, where we observe that, in the data, the slope of the listing premium along the potential gains dimension varies greatly with the level of home equity. This is a feature that is not matched by the corresponding model-implied moments. In Panel B, we report the estimated slope of the listing premium along the potential gains and home equity dimension, respectively, measured in sub-samples that correspond to the interval of the conditioning variable shown on the horizontal axis. Error bars indicate 99% confidence intervals for each point estimate.

Table 3.1: Variables and parameters in the model

State variables:	R	Nominal reference price (observed)
	M	Outstanding value of the mortgage (observed)
	$\widehat{P}$	Property value (estimated)
	$\widehat{G}$	Potential gain (estimated, $\widehat{P} - R$)
	$\widehat{H}$	Potential home equity (estimated, $\widehat{P} - M$)
	θ	Magnitude of the moving shock (unobserved)
Exogenous market conditions:	$\alpha(\ell)$	Concave demand (estimated)
	$\beta(\ell)$	Realized premium (estimated)
Endogenous model variables:	ℓ	Listing premium
	s	Listing decision (extensive margin)
	P	Realized price (observed)
	G	Realized gain (observed, $P - R$)
	H	Realized home equity (observed, $P - M$)
Calibrated model parameters:	γ	Down-payment constraint (set at 20%)
Structural model parameters:	η	Reference dependence
	λ	Loss aversion
	π	Fraction of precise targeters
	μ	Financial constraint
	θ_m	Average value of the moving shock
	θ_σ	Standard deviation of the moving shock
	ϕ	Magnitude of the search and listing cost

TABLE 3.2: STRUCTURAL PARAMETERS

Panel A

Estimation results

Model variant	Parameters							
	Reference dependence η	Loss aversion λ	Average moving shock θ_m	Fraction of targeters π	Financial constraints μ	S.d. of moving shock θ_σ	Search cost ϕ	Preference notch ζ
1	0.340	2.256	0.525					
	(0.025)	(0.137)	(0.025)					
2	0.848	1.468	0.774					
	(0.065)	(0.125)	(0.036)					
3	0.998	1.142	0.760					
	(0.148)	(0.051)	(0.076)					
4	0.512	2.169	0.698	0.106				
	(0.076)	(0.219)	(0.045)	(0.031)				
5	0.404	2.596	0.816	0.127	2.764			
	(0.044)	(0.166)	(0.038)	(0.031)	(0.135)			
6	0.668	2.319	1.345	0.159	4.925	0.711	0.050	
	(0.032)	(0.056)	(0.096)	(0.031)	(0.106)	(0.079)	(0.014)	
7	0.640	2.317	1.353	0.154	4.938	0.730	0.042	-0.005
	(0.028)	(0.049)	(0.093)	(0.022)	(0.077)	(0.080)	(0.013)	(0.007)
8	0.629	2.473	1.335	0.138	4.972	0.672	0.064	-0.006
	(0.043)	(0.128)	(0.104)	(0.029)	(0.117)	(0.099)	(0.018)	(0.006)

Panel B

Frictions and moments used to estimate model variants

Model variant	Demand		Moments used in estimation				
	Concave demand	Imprecise targeting	Hockey stick	Bunching	Home equity	Extensive margin	Cross-sect. variation
1	×	×	✓	×	×	×	×
2	✓	×	✓	×	×	×	×
3	✓	×	✓	✓	×	×	×
4	✓	✓	✓	✓	×	×	×
5	✓	✓	✓	✓	✓	×	×
6	✓	✓	✓	✓	✓	✓	×
7	✓	✓	✓	✓	✓	✓	×
8	✓	✓	✓	✓	✓	✓	✓

Panel A of the table reports structural parameter estimates obtained through classical minimum distance estimation. We recover concave demand $\alpha(\ell)$ and $\beta(\ell)$ from the data and set the down-payment constraint $\gamma = 20\%$. Each row corresponds to a different model variant (1-8), as described in Panel B. The "hockey stick" and "home equity" moments refer to the pattern of listing premia by potential gains and potential home equity, respectively. "Bunching" refers to the distribution of realized gains. The "extensive margin" refers to the variation of the probability of listing in the housing stock by potential gains. In the specifications in which we seek to exploit cross-municipality variation, we distinguish between three different municipality groups, according to their concavity of demand: bottom 5%, middle 90%, top 5%. In parentheses, we report standard errors based on re-estimating the model across bootstrap draws, clustered at the shire level.

Table 3.3: Alternative Model Specifications

Model variant	**Parameters**							
	Reference dependence η	Loss aversion λ	Average moving shock θ_m	Fraction of targeters π	Financial constraints μ	S.d. of moving shock θ_σ	Search cost ϕ	Preference notch ζ
Alternative valuation models								
Renovations	0.633	2.286	1.323	0.138	4.766	0.661	0.063	-0.009
	(0.038)	(0.104)	(0.087)	(0.024)	(0.207)	(0.083)	(0.011)	(0.004)
Repeat sales I	0.605	2.323	1.245	0.150	3.525	0.560	0.050	-0.005
	(0.057)	(0.098)	(0.139)	(0.030)	(0.316)	(0.103)	(0.016)	(0.008)
Repeat sales II	0.586	2.265	1.245	0.147	3.675	0.605	0.048	-0.007
	(0.054)	(0.101)	(0.119)	(0.028)	(0.203)	(0.062)	(0.013)	(0.005)
Controlling for unobserved quality								
Conservative	0.619	2.339	1.426	0.140	4.590	0.774	0.067	-0.007
	(0.033)	(0.099)	(0.108)	(0.023)	(0.221)	(0.100)	(0.014)	(0.002)
Moderate	0.660	2.357	1.695	0.140	4.066	0.880	0.085	-0.003
	(0.046)	(0.093)	(0.096)	(0.022)	(0.174)	(0.084)	(0.009)	(0.006)
Aggressive	0.719	2.333	2.150	0.134	3.039	0.995	0.093	0.006
	(0.058)	(0.107)	(0.065)	(0.020)	(0.141)	(0.112)	(0.021)	(0.007)
Inverse-variance weighting	0.534	2.211	1.244	0.185	5.197	0.808	0.042	-0.007
	(0.045)	(0.083)	(0.085)	(0.037)	(0.203)	(0.062)	(0.013)	(0.004)

The table reports structural parameter estimates obtained through classical minimum distance estimation. We recover concave demand $\alpha(\ell)$ and $\beta(\ell)$ from the data and set the down-payment constraint $\gamma = 20\%$. In parentheses, we report standard errors based on re-estimating the model across bootstrap draws, clustered at the shire level. We re-estimate the model version in row 8 of Table **??**, using alternative ways to recover empirical moments, or to weight them in estimation. We consider three alternative specifications for the hedonic model of $\widehat{P}$. First, we include information on renovation expenses to proxy for potentially time-varying unobserved heterogeneity. Second, we estimate a pairwise repeat sales model with time-varying hedonic characteristics and lagged renovation expenses, by including the lagged past pricing residual, similar to Genesove and Mayer (2001). Third, we generalize the repeat sales approach and include the average of all past pricing residuals to use information from all past repeat sales observed, plus lagged renovation expenses.

Table 3.4: Price-Volume Correlation in the Data and the Model

	Data	Model	
Price-volume correlation (ρ)	0.515^{***}	0.498	
	(0.183)		
Decomposition of effect in the model		Cumulative	Marginal (%)
Frictionless version		0.000	–
+ Reference dependence ($\eta > 0$)		0.091	18.2%
+ Loss aversion ($\lambda > 1$)		0.207	23.2%
+ Financial constraints ($\mu > 0$)		0.498	58.5%

The table reports the estimated coefficient ρ from the regression specification $\Delta \ln N_{m,t} = \mu + \rho \Delta \ln \widehat{P}_{m,t} + \epsilon_{m,t}$, where $N_{m,t}$ are transaction volumes and $\widehat{P}_{m,t}$ is the average hedonic price level in municipality m in year t. *, **, *** indicate statistical significance at the 10%, 5% and 1% confidence levels, respectively. We repeat the calculation of the model-implied value of ρ, starting with a frictionless version of the model ($\eta = \mu = 0$ and $\lambda = 1$), and sequentially adding structural ingredients. The column labeled "Cumulative" reports the magnitude of the price-volume correlation for the respective model version. In the column labeled "Marginal", we report the share of the model-implied price-volume correlation that can be attributed to the respective structural ingredient.

References

Abaluck, J., and J. Gruber (2011): "Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program," *American Economic Review*, 101(4), 1180–1210.

Acharya, V., L. Pedersen, T. Philippon, and M. Richardson (2009): "Regulating systemic risk," *Restoring financial stability: How to repair a failed system*, pp. 283–304.

Acharya, V. V., D. Gale, and T. Yorulmazer (2011): "Rollover Risk and Market Freezes," *The Journal of Finance*, 66(4), 1177–1209.

Agarwal, S., G. Amromin, I. Ben-David, S. Chomsisengphet, T. Piskorski, and A. Seru (2017a): "Policy intervention in debt renegotiation: Evidence from the home affordable modification program," *Journal of Political Economy*, 125(3), 654–712.

Agarwal, S., I. Ben-David, and V. Yao (2017b): "Systematic mistakes in the mortgage market and lack of financial sophistication," *Journal of Financial Economics*, 123(1), 42–58.

Agarwal, S., S. Chomsisengphet, N. Mahoney, and J. Stroebel (2014): "Regulating consumer financial products: Evidence from credit cards," *The Quarterly Journal of Economics*, 130(1), 111–164.

Agarwal, S., J. C. Driscoll, X. Gabaix, and D. Laibson (2009): "The age of reason: Financial decisions over the life cycle and implications for regulation," *Brookings Papers on Economic Activity*, 2009(2), 51–117.

Agarwal, S., J. C. Driscoll, and D. I. Laibson (2013): "Optimal Mortgage Refinancing: A Closed-Form Solution," *Journal of Money, Credit and Banking*, 45(4), 591–622.

Agarwal, S., J. Grigsby, A. Hortaçu, G. Matvos, A. Seru, and V. Yao (2017c): "Search and Screening in Credit Markets," *Unpublished working paper*.

Akerlof, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84(3), 488–500.

Alexandrov, A., and S. Koulayev (2018): "No shopping in the US mortgage market: Direct and strategic effects of providing information," *CFPB Working Paper*.

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

Allen, F., E. Carletti, I. Goldstein, and A. Leonello (2015): "Moral hazard and government guarantees in the banking industry," *Journal of Financial Regulation*, 1(1), 30–50.

Allen, J., R. Clark, and J.-F. Houde (2014): "Price dispersion in mortgage markets," *The Journal of Industrial Economics*, 62(3), 377–416.

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

Anagol, S., A. Davids, B. B. Lockwood, and T. Ramadorai (2022): "Diffuse Bunching with Optimisation Frictions," .

Anagol, S., and H. H. Kim (2012): "The impact of shrouded fees: Evidence from a natural experiment in the Indian mutual funds market," *American Economic Review*, 102(1), 576–593.

Andersen, S., J. Y. Campbell, K. M. Nielsen, and T. Ramadorai (2020a): "Sources of inaction in household finance: Evidence from the Danish mortgage market," *American Economic Review*, 110(10), 3184–3230.

——— (2020b): "Sources of Inaction in Household Finance: Evidence from the Danish Mortgage Market," *American Economic Review*, 110(10).

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

Argyle, B., T. Nadauld, and C. Palmer (2017): "Real effects of search frictions in consumer credit markets," *Unpublished working paper*.

Aron, J., and J. Muellbauer (2016): "Modelling and forecasting mortgage delinquency and foreclosure in the UK," *Journal of Urban Economics*, 94, 32–53.

Ashraf, N., D. Karlan, and W. 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, M., T. Blake, B. Larsen, and S. Tadelis (2020): "Sequential bargaining in the field: Evidence from millions of online bargaining interactions," *Quarterly Journal of Economics*, 135(3), 1319–1361.

Badarinza, C., J. Y. Campbell, and T. Ramadorai (2016): "International comparative household finance," *Annual Review of Economics*, 8, 111–144.

——— (2018): "What calls to ARMs? International evidence on interest rates and the choice of adjustable-rate mortgages," *Management Science*, 64(5), 2275–2288.

Bailey, M., E. Dávila, T. Kuchler, and J. Stroebel (2019): "House price beliefs and mortgage leverage choice," *The Review of Economic Studies*, 86(6), 2403–2452.

Barber, B. M., T. Odean, and L. Zheng (2005): "Out of sight, out of mind: The effects of expenses on mutual fund flows," *The Journal of Business*, 78(6), 2095–2120.

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

Bartik, T. (1991): *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute.

Baye, M. R., J. Morgan, and P. Scholten (2006): "Information, search, and price dispersion," *In: Hendershott T. (Ed.), Handbook on Economics and Information Systems*, 1, 323–77.

Becker, B., M. Campello, V. Thell, and D. Yan (2021): "Credit Risk and the Life Cycle of Callable Bonds: Implications for Corporate Financing and Investing," .

Belgibayeva, A., T. Bono, P. Bracke, J. Cocco, T. Majer, et al. (2020): "When Discounted Rates End: The Cost of Taking Action in the Mortgage Market," *FCA Occasional Papers in Financial Regulation*.

Benetton, M. (2021): "Leverage regulation and market structure: An empirical model of the UK mortgage market," *Journal of Finance (forthcoming)*.

Beraja, M., A. Fuster, E. Hurst, and J. Vavra (2019): "Regional Heterogeneity and the Refinancing Channel of Monetary Policy," *The Quarterly Journal of Economics*, 134(1), 109–183.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, pp. 242–262.

Best, M. C., J. S. Cloyne, E. Ilzetzki, and H. J. Kleven (2020): "Estimating the elasticity of intertemporal substitution using mortgage notches," *The Review of Economic Studies*, 87(2), 656–690.

Bhutta, N., A. Hizmo, and A. Fuster (2019): "Paying Too Much? Price Dispersion in the US Mortgage Market," *Unpublished working paper*.

Bhutta, N., and B. J. Keys (2022): "Moral hazard during the housing boom: Evidence from private mortgage insurance," *The Review of Financial Studies*, 35(2), 771–813.

Blundell, R. (2014): "Income dynamics and life-cycle inequality: mechanisms and controversies," *The Economic Journal*, 124(576), 289–318.

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

Bordalo, P., N. Gennaioli, and A. Shleifer (2013): "Salience and consumer choice," *Journal of Political Economy*, 121(5), 803–843.

——— (2019): "Memory, attention, and choice," *Unpublished working paper*.

Borusyak, K., P. Hull, and X. Jaravel (2018): "Quasi-experimental shift-share research designs," *Unpublished working paper*.

Bracke, P., and S. Tenreyro (2020): "History dependence in the housing market," *Bank of England Working Paper*.

Brown, J. R., and A. Finkelstein (2008): "The Interaction of Public and Private Insurance: Medicaid and the Long-Term Care Insurance Market," *American Economic Review*, 98(3), 1083–1102.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018): "Fintech, regulatory arbitrage, and the rise of shadow banks," *Journal of Financial Economics*, 130(3), 453–483.

Bucks, B., and K. Pence (2008): "Do borrowers know their mortgage terms?," *Journal of Urban Economics*, 64(2), 218–233.

Button, R., S. Pezzini, and N. Rossiter (2010a): "Understanding the price of new lending to households," *Bank of England Quarterly Bulletin*, p. Q3.

——— (2010b): "Understanding the price of new lending to households," *Bank of England Quarterly Bulletin 2010 Q3*.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997): "Labor supply of New York City cabdrivers: One day at a time," *Quarterly Journal of Economics*, 112(2), 407–441.

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

——— (2013): "Mortgage market design," *Review of Finance*, 17(1), 1–33.

Campbell, J. Y., N. Clara, and J. F. Cocco (2021): "Structuring mortgages for macroeconomic stability," *The Journal of Finance*, 76(5), 2525–2576.

Campbell, J. Y., and J. F. Cocco (2003): "Household risk management and optimal mortgage choice," *The Quarterly Journal of Economics*, 118(4), 1449–1494.

Campbell, J. Y., and J. F. Cocco (2015): "A Model of Mortgage Default: A Model of Mortgage Default," *The Journal of Finance*, 70(4), 1495–1554.

Campbell, J. Y., and R. J. Shiller (1991): "Yield spreads and interest rate movements: A bird's eye view," *The Review of Economic Studies*, 58(3), 495–514.

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

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

Carlin, B. I. (2009): "Strategic price complexity in retail financial markets," *Journal of Financial Economics*, 91(3), 278–287.

Carroll, C., J. Slacalek, K. Tokuoka, and M. N. White (2017): "The distribution of wealth and the marginal propensity to consume," *Quantitative Economics*, 8(3), 977–1020.

Carroll, C. D. (1997): "Buffer-stock saving and the life cycle/permanent income hypothesis," *The Quarterly journal of economics*, 112(1), 1–55.

Célérier, C., and B. Vallée (2017): "Catering to investors through security design: Headline rate and complexity," *The Quarterly Journal of Economics*, 132(3), 1469–1508.

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

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011): "Adjustment costs,

firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records," *Quarterly Journal of Economics*, 126(2), 749–804.

Chetty, R., A. Looney, and K. Kroft (2009): "Salience and taxation: Theory and evidence," *American Economic Review*, 99(4), 1145–77.

Chioveanu, I., and J. Zhou (2013): "Price competition with consumer confusion," *Management Science*, 59(11), 2450–2469.

Choi, J., D. Hackbarth, and J. Zechner (2018): "Corporate Debt Maturity Profiles," *Journal of Financial Economics*, 130(3), 484–502.

Choi, J. J., D. Laibson, and B. C. Madrian (2009): "Why does the law of one price fail? An experiment on index mutual funds," *The Review of Financial Studies*, 23(4), 1405–1432.

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

Cloyne, J., K. Huber, E. Ilzetzki, and H. Kleven (2019): "The Effect of House Prices on Household Borrowing: A New Approach," *American Economic Review*, 109(6), 2104–2136.

Cocco, J. F. (2013): "Evidence on the benefits of alternative mortgage products," *The Journal of Finance*, 68(4), 1663–1690.

Cocco, J. F., F. J. Gomes, and P. J. Maenhout (2005): "Consumption and portfolio choice over the life cycle," *The Review of Financial Studies*, 18(2), 491–533.

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

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

DellaVigna, S., A. Lindner, B. Reizer, and J. F. Schmieder (2017): "Reference-dependent job search: Evidence from Hungary," *Quarterly Journal of Economics*, 132(4), 1969–2018.

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

Diamond, D. W. (1991): "Debt Maturity Structure and Liquidity Risk," *The Quarterly Journal of Economics*, 106(3), 709–737.

Drechsler, I., A. Savov, and P. Schnabl (2017): "The deposits channel of monetary policy," *The Quarterly Journal of Economics*, 132(4), 1819–1876.

Duarte, F., and J. S. Hastings (2012): "Fettered consumers and sophisticated firms: evidence from Mexico's privatized social security market," *National Bureau of Economic Research Working Paper*.

Dunn, K. B., and C. S. Spatt (1985): "An Analysis of Mortgage Contracting: Prepayment Penalties and the Due-on-Sale Clause," *The Journal of Finance*, 40(1), 293–308.

——— (1988): "Private information and incentives: Implications for mortgage contract terms and pricing," *The Journal of Real Estate Finance and Economics*, 1(1), 47–60.

Eberly, J., and A. Krishnamurthy (2014): "Efficient Credit Policies in a Housing Debt Crisis," *Brookings Papers on Economic Activity*, 2014(2), 73–136.

Egan, M. (2018): "Brokers vs. retail investors: Conflicting interests and dominated products," *Journal of Finance, Forthcoming*.

Egan, M., G. Matvos, and A. Seru (2019): "The market for financial adviser misconduct," *Journal of Political Economy*, 127(1).

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

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

Ellison, G. (2005): "A model of add-on pricing," *The Quarterly Journal of Economics*, 120(2), 585–637.

Ellison, G., and S. F. Ellison (2009): "Search, obfuscation, and price elasticities on the internet," *Econometrica*, 77(2), 427–452.

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

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

Fabozzi, F. J., R. J. Shiller, and R. S. Tunaru (2020): "A 30-Year Perspective on Prop-

erty Derivatives: What Can Be Done to Tame Property Price Risk?," *Journal of Economic Perspectives*, 34(4), 121–45.

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

FCA (2018): "Mortgage Market Study Interim Report," .

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

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

Fisher, J., A. Gavazza, L. Liu, T. Ramadorai, and J. Tripathy (2021): "Refinancing Cross-Subsidies in the UK Mortgage Market," *Unpublished working paper*.

Flannery, M. J. (1986): "Asymmetric information and risky debt maturity choice," *The Journal of Finance*, 41(1), 19–37.

Foà, G., L. Gambacorta, L. Guiso, and P. E. Mistrulli (2015): "The supply side of household finance," *BIS Working Paper*.

Fons, J. S. (1994): "Using default rates to model the term structure of credit risk," *Financial Analysts Journal*, 50(5), 25–32.

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

Fuster, A., and P. S. Willen (2017): "Payment Size, Negative Equity, and Mortgage Default," *American Economic Journal: Economic Policy*, 9(4), 167–191.

Gabaix, X., and D. Laibson (2006): "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *The Quarterly Journal of Economics*, 121(2), 505–540.

Galenianos, M., and A. Gavazza (2017): "A structural model of the retail market for illicit drugs," *American Economic Review*, 107(3), 858–96.

Ganong, P., and P. Noel (2020): "Liquidity versus wealth in household debt obligations: Evidence from housing policy in the Great Recession," *American Economic Review*, 110(10), 3100–3138.

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

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

Gerardi, K. (2017): "GSE Mortgage Insurance Pricing," *Slides*.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2017): "Bartik Instruments: What, When, Why and How," *National Bureau of Economic Research Working Paper*.

Gomes, F. (2020): "Portfolio Choice Over the Life Cycle: A Survey," *Annual Review of Financial Economics*, 12, 277–304.

Gomes, F., M. Haliassos, and T. Ramadorai (2020): "Household finance," *Journal of Economic Literature, forthcoming*.

Green, R. K., and S. M. Wachter (2005): "The American Mortgage in Historical and International Context," *Journal of Economic Perspectives*, 19(4), 93–114.

Greenwald, D. L., T. Landvoigt, and S. Van Nieuwerburgh (2021): "Financial Fragility with SAM?," *The Journal of Finance*, 76(2), 651–706.

Grubb, M. D. (2015): "Failing to choose the best price: Theory, evidence, and policy," *Review of Industrial Organization*, 47(3), 303–340.

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

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

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

Guren, A. M., A. Krishnamurthy, and T. J. McQuade (2021): "Mortgage design in an equilibrium model of the housing market," *The Journal of Finance*, 76(1), 113–168.

Gurun, U. G., G. Matvos, and A. Seru (2016): "Advertising expensive mortgages," *The Journal of Finance*, 71(5), 2371–2416.

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

——— (2016): "What is the role of the asking price for a house?," *Journal of Urban Economics*, 93, 115–130.

Handel, B., I. Hendel, and M. D. Whinston (2015): "Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk," *Econometrica*, 83(4), 1261–1313.

Handel, B., I. Hendel, and M. D. Whinston (2017): "The Welfare Effects of Long-Term Health Insurance Contracts," *Unpublished working paper*, p. 56.

Harimohan, R., M. McLeay, and G. Young (2016): "Pass-through of bank funding costs to lending and deposit rates: lessons from the financial crisis," *Bank of England Staff Working Paper*.

Harris, M., and B. Holmstrom (1982): "A Theory of Wage Dynamics," *The Review of Economic Studies*, 49(3), 315.

——— (1987): "On the Duration of Agreements," *International Economic Review*, 28(2), 389.

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

Hastings, J. S., A. Hortaçsu, and C. Syverson (2017): "Advertising and competition in privatized social security: The case of Mexico," *Econometrica*, 85(6), 1723–1761.

He, Z., and W. Xiong (2012): "Rollover risk and credit risk," *The Journal of Finance*, 67(2), 391–430.

Heidhues, P., B. Kőszegi, and T. Murooka (2017): "Inferior Products and Profitable Deception," *The Review of Economic Studies*, 84(1), 323–356.

Helwege, J., and C. M. Turner (1999): "The slope of the credit yield curve for speculative-grade issuers," *The Journal of Finance*, 54(5), 1869–1884.

Hendel, I. (2017): "Dynamic Selection and Reclassification Risk: Theory and Empirics," *Unpublished working paper*.

Hendel, I., and A. Lizzeri (2003): "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," *The Quarterly Journal of Economics*, pp. 299–327.

Henderson, B. J., and N. D. Pearson (2011): "The dark side of financial innovation: A case study of the pricing of a retail financial product," *Journal of Financial Economics*, 100(2), 227–247.

Hertzberg, A., A. Liberman, and D. Paravisini (2018): "Screening on Loan Terms: Evidence from Maturity Choice in Consumer Credit," *The Review of Financial Studies*, 31(9), 3532–3567.

Hortaçsu, A., and C. Syverson (2004): "Product differentiation, search costs, and com-

petition in the mutual fund industry: A case study of S&P 500 index funds," *The Quarterly Journal of Economics*, 119(2), 403–456.

Hosseini, R. (2015): "Adverse selection in the annuity market and the role for social security," *Journal of Political Economy*, 123(4), 941–984.

Hurst, E., B. J. Keys, A. Seru, and J. Vavra (2016): "Regional redistribution through the US mortgage market," *American Economic Review*, 106(10), 2982–3028.

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

Iscenko, Z. (2018): "Choices of dominated mortgage products by UK consumers," *FCA Occasional Paper*.

Iscenko, Z., and J. Nieboer (2018): "Effects of the advice requirement and intermediation in the UK mortgage market," *FCA Occasional Paper*.

Jensen, T. L., and N. Johannesen (2017): "The Consumption Effects of the 2007–2008 Financial Crisis: Evidence from Households in Denmark," *American Economic Review*, 107(11), 3386–3414.

Jordà, Ò., M. Schularick, and A. M. Taylor (2017): "Macrofinancial history and the new business cycle facts," *NBER Macroeconomics Annual*, 31(1), 213–263.

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

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

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig (2010): "Did securitization lead to lax screening? Evidence from subprime loans," *The Quarterly Journal of Economics*, 125(1), 307–362.

Keys, B. J., D. G. Pope, and J. C. Pope (2016): "Failure to refinance," *Journal of Financial Economics*, 122(3), 482–499.

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

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

Koijen, R. S., O. Van Hemert, and S. Van Nieuwerburgh (2009): "Mortgage timing," *Journal of Financial Economics*, 93(2), 292–324.

Kopecky, K. A., and R. M. Suen (2010): "Finite state Markov-chain approximations to highly persistent processes," *Review of Economic Dynamics*, 13(3), 701–714.

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

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

Laibson, D., P. Maxted, and B. Moll (2021): "Present bias amplifies the household balance-sheet channels of macroeconomic policy," Discussion paper, National Bureau of Economic Research.

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

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

Liu, L. (2019): "Non-salient fees in the mortgage market," *Bank of England Staff Working Paper No. 819*.

Mayer, C., T. Piskorski, and A. Tchistyi (2013): "The Inefficiency of Refinancing: Why Prepayment Penalties Are Good for Risky Borrowers," *Journal of Financial Economics*, 107(3), 694–714.

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

Merton, R. C. (1974): "On the pricing of corporate debt: The risk structure of interest rates," *The Journal of Finance*, 29(2), 449–470.

Mian, A., and A. Sufi (2015): *House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again*. University of Chicago Press.

Miles, D. (2004): "The UK mortgage market: Taking a longer-term view," *Final Report and Recommendations*.

Nelson, S. (2018): "Private information and price regulation in the us credit card market," *Unpublished working paper*.

Nielsen, H. S., T. Sørensen, and C. Taber (2010): "Estimating the effect of student aid

on college enrollment: Evidence from a government grant policy reform," *American Economic Journal: Economic Policy*, 2(2), 185–215.

Palmer, C. (2015): "Why Did So Many Subprime Borrowers Default During the Crisis: Loose Credit or Plummeting Prices?," *SSRN Electronic Journal*.

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

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

Piccione, M., and R. Spiegler (2012): "Price competition under limited comparability," *The Quarterly Journal of Economics*, 127(1), 97–135.

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

Piskorski, T., and A. Tchistyi (2010): "Optimal mortgage design," *The Review of Financial Studies*, 23(8), 3098–3140.

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

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

Robles-Garcia, C. (2020): "Competition and incentives in mortgage markets: The role of brokers," *Unpublished working paper*.

Rouwenhorst, K. G. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Frontiers of business cycle research*, pp. 294–330. Princeton University Press.

Ru, H., and A. Schoar (2016): "Do credit card companies screen for behavioral biases?," *National Bureau of Economic Research Working Paper*.

Salop, S., and J. Stiglitz (1977): "Bargains and ripoffs: A model of monopolistically competitive price dispersion," *The Review of Economic Studies*, pp. 493–510.

Sarig, O., and A. Warga (1989): "Some empirical estimates of the risk structure of interest rates," *The Journal of Finance*, 44(5), 1351–1360.

Scharfstein, D., and A. Sunderam (2015): "Market power in mortgage lending and the transmission of monetary policy," *Unpublished working paper*.

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

Shiller, R. J. (2014): "Why Is Housing Finance Still Stuck in Such a Primitive Stage?," *American Economic Review*, 104(5), 73–76.

Shiller, R. J., and A. N. Weiss (1999): "Home equity insurance," *The Journal of Real Estate Finance and Economics*, 19(1), 21–47.

Spiegler, R. (2006): "Competition over agents with boundedly rational expectations," *Theoretical Economics*, 1(2), 207–231.

Stahl, D. O. (1989): "Oligopolistic pricing with sequential consumer search," *The American Economic Review*, pp. 700–712.

Stango, V., and J. Zinman (2015): "Borrowing high versus borrowing higher: price dispersion and shopping behavior in the US credit card market," *The Review of Financial Studies*, 29(4), 979–1006.

Stanton, R., and N. Wallace (1998): "Mortgage choice: What's the point?," *Real Estate Economics*, 26(2), 173–205.

Stein, J. C. (1995): "Prices and trading volume in the housing market: A model with down-payment effects," *The Quarterly Journal of Economics*, 110(2), 379–406.

Tauchen, G. (1986): "Finite state markov-chain approximations to univariate and vector autoregressions," *Economics Letters*, 20(2), 177–181.

Tauchen, G., and R. Hussey (1991): "Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models," *Econometrica*, pp. 371–396.

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

Varian, H. R. (1980): "A model of sales," *The American Economic Review*, 70(4), 651–659.

Wong, A. (2019): "Refinancing and the transmission of monetary policy to consumption," *Unpublished working paper*.

Woodward, S. E., and R. E. Hall (2012): "Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence," *American Economic Review*, 102(7), 3249–3276.

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

Appendix 1

Appendix for "The Demand for Long-Term Mortgage Contracts and the Role of Collateral"

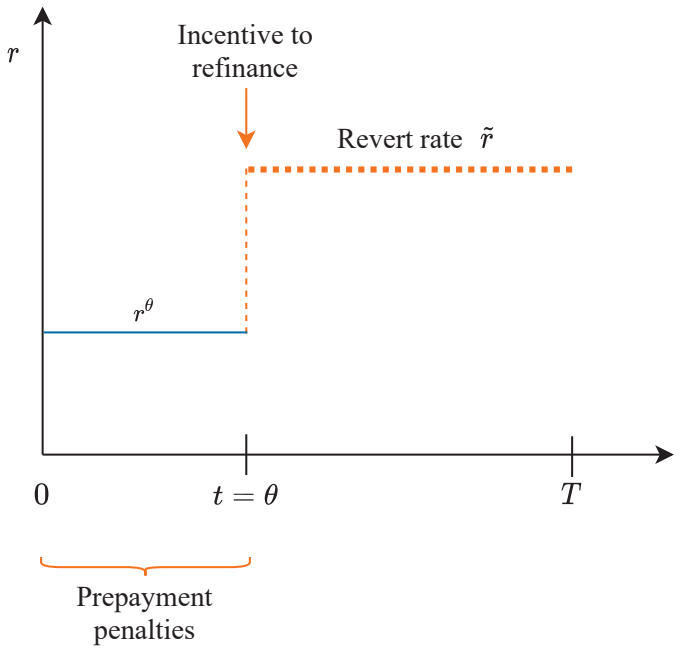
A. Further Background on Mortgage Market Contract Structure, Pricing and Choice

This section provides further background on the contract structure, pricing and contract choice in the UK mortgage market.

1. Mortgage Contract Structure and Payment Profile

Figure **??** illustrates the payment structure of a typical UK fixed-rate mortgage contract. The initial fixed-rate r^θ is fixed over the initial fixation length θ. Throughout this initial fixation period, prepayment penalties apply if the mortgage is prepaid. Once the initial fixation period ends, the interest rate automatically switches to a revert rate $\tilde{r}$ until the end of the loan repayment window T, which is a floating rate that is priced at a premium over the base rate (Bank Rate). Rather than paying this rate at reset, the borrower can choose to refinance, at which point a new contract is priced.

Figure A.1: Illustration of Payment Profile for UK Mortgage Contracts with Initial Fixation Period



2. Mortgage Pricing and the Role of LTV

Mortgage pricing in the UK is predominantly done based on the loan-to-value (LTV) ratio. A higher LTV is correlated with a higher probability of default, as well as a greater loss given default, as it is a direct (inverse) measure of collateralization of the loan, and reflects the size of the loan relative to the value of the collateral. Computing 1 minus the LTV gives the equity buffer that the lender has in case the borrower defaults and the lender needs to recover the loan by repossessing and selling the house. The role of LTV for mortgage pricing can be directly verified in the data. Figure **??** shows the adjusted R^2 from a regression of observed rates on a range of fixed effects. "Base" refers to the regression including year-month, lender, buyer type (first-time buyer, second time buyer or refinance) and fixation length fixed effects, and all interaction effects. "+Income" includes income decile fixed effects and interactions with year-month to the base specification, and "+Age" and "+LTV" do this analogously for borrower age decile, and LTV band fixed effects. The results illustrate that borrower characteristics such as age and income have negligible effects on the overall variation explained over and above base fixed effects, while the inclusion of LTV raises the adjusted R^2 by around 30 percentage points. The inclusion of household-specific creditworthiness via a FICO score yields similar results to age and income (shown by **?**), meaning that mortgage rates do not vary strongly by FICO. Lenders have "full recourse" in the UK, meaning they can recover losses from defaulted borrowers though their assets and incomes for up to seven years, until the debt is paid (**?**), which may help explain why measures of household-specific creditworthiness including the FICO score are only accounted for via a minimum threshold at loan application, but result in limited price variation conditional on LTV.

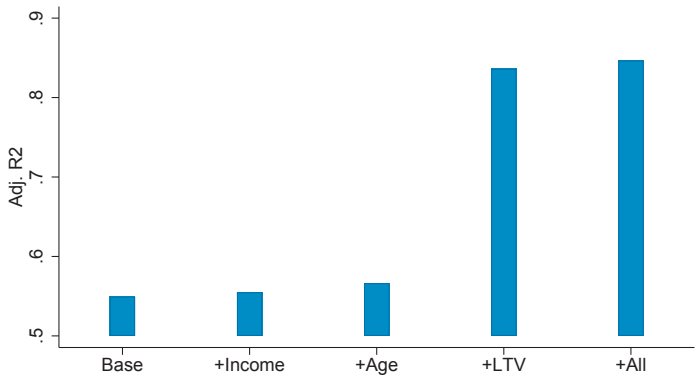
3. Mortgage Contract Choice And The Role of LTV

The raw distribution of mortgage fixation lengths across LTV is shown in Figure **??**.

In addition, Figure **??** illustrates which variables drive variation in 5-year fixed-rate contract choice. It shows the adjusted R^2 of a regression with an indicator that takes the value 1 if the household chooses a 5-year fixed-rate contract, and 0 if the household chooses a 2-year fixed-rate contract, across a range of covariates. The first dot shows the adjusted R^2 with

FIGURE A.2: MORTGAGE PRICING REGRESSION R^2

This figure shows the adjusted R^2 from a regression of observed rates on a range of fixed effects. "Base" refers to the regression including year-month, lender, buyer type (first-time buyer, second time buyer or refinance) and fixation length (less than 1, 2, 3 to 4, 5, more than 5) fixed effects, and all interaction effects. "+Income" includes income decile fixed effects and interactions with year-month to the base specification, and "+Age" and "+LTV" do this analogously for borrower age decile, and LTV band (≤70%, 70-75%, 75-80%, 80-85%, 85-90%, and 90-95%) fixed effects.



time (year-month), lender, location and borrower-type fixed-effects only. The following dots show the adjusted R^2 for any single variable that is added to the set of fixed effects, including the LTV band, log income, log loan size, loan-to-income (LTI) ratio, borrower age, loan term, origination fees, and local house price growth, measured as the local authority house price growth two years prior to the time of choice. The inclusion of LTV band results in the largest increase in the adjusted R^2, to around 0.11.

4. SUMMARY STATISTICS BY CONTRACT CHOICE

Figure A.3: Contract Choice by LTV band

This figure shows the share of newly originated fixed rate contracts by fixation length and LTV band, for loans originated between 2013 to 2017.

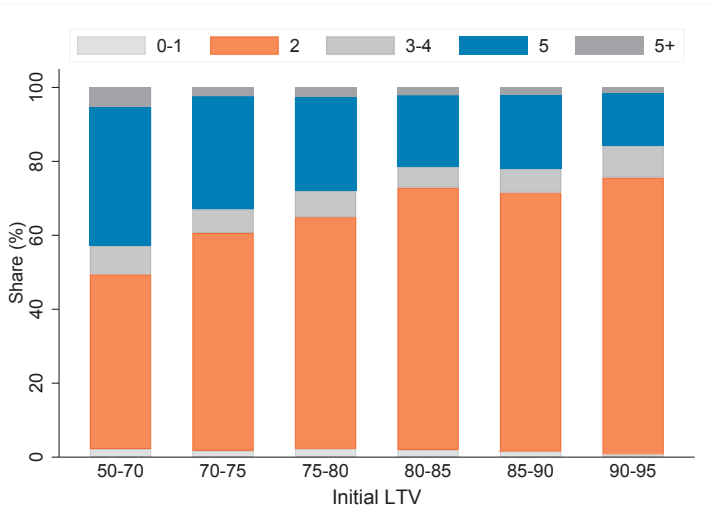


Figure A.4: Explaining Fixation Length Choice - LTV and Other Channels

This figure shows the adjusted R^2 of a regression with an indicator that takes the value 1 if the household chooses a 5-year fixed-rate contract, and 0 if the household chooses a 2-year fixed-rate contract, across a range of covariates. The first dot shows the adjusted R^2 with time (year-month), lender, location and borrower-type fixed-effects only. The following dots show the adjusted R^2 for any single variable that is added to the set of fixed effects, including the LTV band, log income, log loan size, loan-to-income (LTI) ratio, borrower age, loan term, origination fees, and local house price growth, measured as the local authority house price growth two years prior to the time of choice.

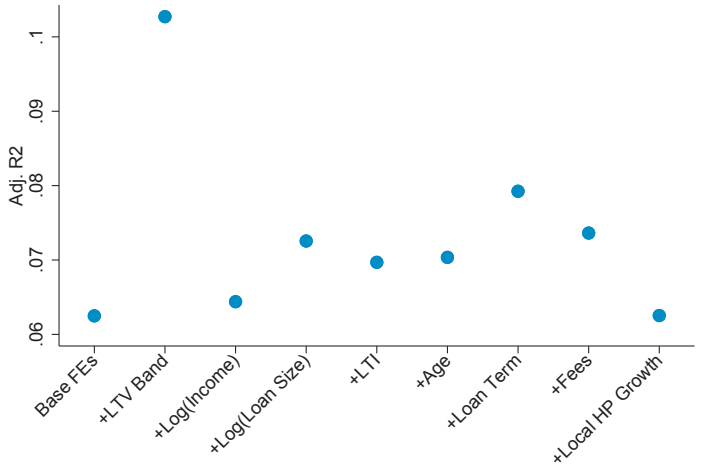


Table A.1: Summary statistics for 2-year and 5-year fixed-rate borrowers

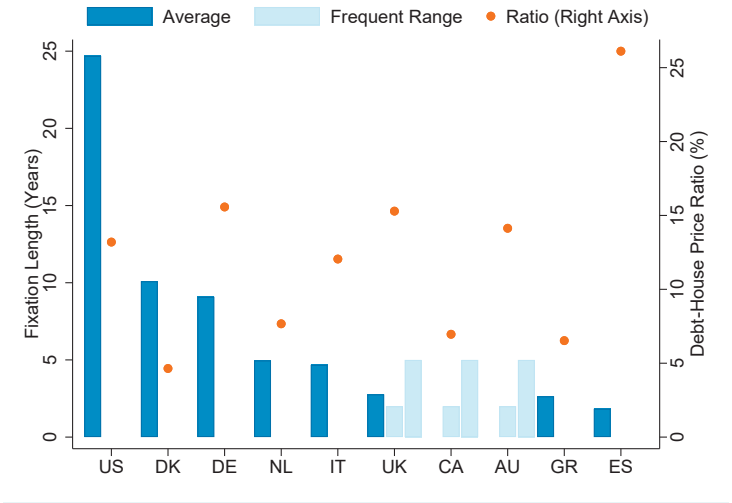
This table compares the average and standard deviation (in parenthesis) of selected observable characteristics between borrowers with a 2-year, and borrowers with a 5-year fixed-rate mortgage, in the main origination data (2013-2017).

	2-Year Fixed Rate	5-Year Fixed Rate
Age	36.61	38.79
	(9.03)	(9.75)
Joint Income	0.59	0.59
	(0.49)	(0.49)
Income	56,316	55,980
	(36,441)	(37,201)
Interest Rate	2.49	2.84
	(0.93)	(0.76)
Property Value	254,979	272,840
	(168,880)	(186,342)
Loan Size	172,614	157,629
	(109,938)	(106,686)
Loan Term	25.23	22.64
	(7.50)	(7.96)
LTV	71.10	61.76
	(19.95)	(21.74)
LTI	3.19	2.95
	(1.03)	(1.09)
N	1,898,660	967,001

B. Mortgage Fixation Period Across Countries

Figure B.1: Mortgage Fixation Period Across Countries

This figure shows average initial mortgage fixed-rate lengths across advanced economies, based on administrative data obtained from **?**, as at December 2013 (with the exceptions of Greece and Denmark, where data is from 2010). For Canada and Australia, this data is not available so the most frequent range is plotted, which akin to the UK, is 2 to 5 years. The household-debt-to-house-price ratio is calculated as average household debt (in USD) divided by the average house price (in USD) in 2013. Average household debt is calculated using household debt per capita (as a percentage of net household disposable income) and household disposable income from the OECD database, with the exception of Italy, where data for 2013 is not available and is replaced using total debt do household and population data from the Macrohistory Database (**?**). Average house price levels are obtained from national statistics bureaus, and where price level data is not available, from industry data service providers.



C. Dataset Construction

This section details the construction of the borrower panel and data cleaning steps.

1. Merge (PSD007 and PSD001) for Borrower Panel

To analyze borrower refinancing behavior and repricing of rates at the point of refinance, I merge two administrative datasets on the universe of UK mortgage borrowers. First, I observe new mortgage originations in the Product Sales Data (PSD) that is accessed through the Bank of England via a data sharing agreement with the Financial Conduct Authority. The PSD001 dataset collects detailed borrower characteristics such as income, age, address, loan amount, property value, and detailed loan characteristics such as the loan maturity, interest rate, fixation length, and which lender originated the mortgage. This allows me to identify first-time buyer cohorts who newly originate their mortgage between 2013H2 to 2015H1. The origination data is available from 2005Q2 and updated at quarterly frequency up to today.

I then use a more recent additional dataset, PSD007, that is part of the PSD which tracks the entire stock of UK mortgages outstanding, available from 2015H1 and updated at half-yearly frequency up to today. The stock data contains information on the current interest rate type, current interest rate paid, current loan amount, current lender, and whether the loan is in arrears. I merge the stock data with the origination data to track refinancing behavior and outcomes (in particular, the interest rate paid and an indicator whether the borrower is in arrears) between the first half-year snapshot, obtained in June 2015 (henceforth, 2015H1) and December 2017 (henceforth, 2017H2) for the first-time borrower cohorts identified in the origination data.

The final data has a panel format which comprises detailed borrower and loan characteristics at origination, and half-yearly updates on outcomes such as interest rates, loan amount remaining and default status. In addition to the characteristics at origination, each refinance that reflects a switch to a different lender is recorded as a new origination, so I observe updated information on income and other borrower characteristics if the borrower does a so-called "external" refinance. This is in contrast to an "internal" refinance where the lender refinances into a different contract and interest rate, but stays with the current lender. I identify internal and external refinancers

as follows: the data records if the borrower is on the revert rate, so a refinance requires the borrower to either move from a revert rate to a fixed-rate contract, or move from an existing fixed-rate contract into a new fixed-rate contract. External refinances are recorded in the origination data, so if a borrower is recorded as a refinancer in the origination data, the lender changes, and the interest rate changes in that period, I classify the borrower as an external refinancer. If only the interest rate changes but there is no entry in the orgination data and the lender remains the same, I classify this as an internal refinance.[1]

For the 2013H2 first-time borrowers who took out a 2-year fixed rate contract, about 55% refinance internally, and around 30% refinance externally by 2016H1, i.e. within six months of the end of the fixed rate period in 2015H2. Each half-year origination cohort comprises around 150,000 first-time borrowers, leaving me with 721,060 unique borrowers, and around 5.5 million borrower-half-year observations between 2015H1 to 2017H2.

The two datasets do not have unique borrower identifiers, but a borrower can be identified up to an (anonymized) date of birth and six-digit postcode which is approximately the building block in which a UK household resides. Table **??** illustrates the quality of the merge. The mortgage stock data starts in 2015H1 but in order to observe a longer time-series of outcomes, I start with a borrower cohort in 2013H2. That means that borrowers are not matched in 2015H1 if there is a pure merging error (e.g. because the borrower identification is not unique), or if borrowers prepay or default and leave the sample prior to 2015H1. In addition, the data is less complete in 2015H1 compared to 2015H2 onwards. I hence compare the observations that are not matched to the 2015H2 stock data across first-time borrower cohorts in Table **??**. From the share of "not matched" observations, it can be seen that 1.8% of first-time borrowers in 2015H2 cannot be matched to the stock data in 2015H2, which provides an estimate of unmatched observations driven by pure merging error. Going from the origination cohort in 2015H1 back to 2013H2, around 1-2% more observations are unmatched from half-year to half-year, providing an estimate of the share of borrowers that leaves the stock data due to prepayment or porting the mortgage (for instance if the borrower moves to another house) or default, at half-yearly frequency. Note that because mortgages are portable in the UK, mortgages could show up in the data with the same borrower but a new house address, which I cannot track due to the borrower identification procedure described above. Table **??** compares

[1]Reassuringly, the resulting numbers are very similar to those provided by **?** who obtain explicit data on internal refinancing through a survey of the 20 largest UK lenders for 2-year fixed rate borrowers in 2013.

Table C.1: Matching to 2015H2 mortgage stock

This table shows the share of borrowers in each first-time borrower cohort between 2013H2 and 2015H2 in the origination data that can be matched to the stock of all mortgages outstanding in 2015H2.

FTB cohort	2015H2 data					
	Not matched		Matched		Total	
	No.	%	No.	%	No.	%
2013H2	8,641	6.0%	135,156	94.0%	143,797	100.0%
2014H1	5,761	4.1%	134,714	95.9%	140,475	100.0%
2014H2	4,832	3.2%	147,732	96.8%	152,564	100.0%
2015H1	4,208	3.3%	121,642	96.7%	125,850	100.0%
2015H2	2,841	1.8%	155,533	98.2%	158,374	100.0%
Total	26,283	3.6%	694,777	96.4%	721,060	100.0%

the average characteristics (with standard deviations) of borrowers and loans for matched and unmatched observations. Unmatched borrowers are slightly older, have larger incomes, loan sizes and property values, and lower loan-to-income and loan-to-value ratios, suggesting that the unmatched borrowers seem to reflect movers rather than more risky borrowers who have defaulted.

Table C.2: Balance of matched observations

This table compares the average and standard deviation (in parenthesis) between the matched and unmatched sample shown in Table **??** across a range of observable characteristics.

	Not matched	Matched
Age	32.43	31.10
	(8.15)	(7.04)
Joint Income	0.62	0.52
	(0.49)	(0.50)
Income	55,941	44,731
	(37,997)	(26,664)
Interest Rate	3.44	3.39
	(0.92)	(0.98)
Property Value	231,813	196,274
	(169,121)	(128,351)
Loan Size	163,467	146,385
	(104,402)	(87,692)
Loan Term	26.71	28.10
	(6.67)	(6.14)
LTV	74.10	77.11
	(17.90)	(16.67)
LTI	3.14	3.38
	(1.10)	(0.94)
Origination Year	2014	2014
	(0.77)	(0.74)
N	26,283	694,777

Lastly, the origination data prior to 2015H1 does not require to report the fixation length,

so I do not observe the fixed-rate period for about 40% of first-time borrowers. The sample for which fixation lengths are observed appears to be a highly balanced sample compared to where it is not observed, as noted by **?** and demonstrated in Table **??**. I hence proceed with the remaining sample of 414,643 first-time borrowers for which this key variable is observed, resulting in a panel of around 2.8 million borrower-half-year observations.

Table C.3: Sample selection balance (fixation length observed)

This table compares the average and standard deviation (in parenthesis) between the subset of observations for which the fixation length is observed and that for which it is not, in the matched sample across a range of observable characteristics.

	Observed	Not observed
Age	31.17	31.12
	(7.08)	(7.10)
Joint Income	0.53	0.52
	(0.50)	(0.50)
Income	45,520	44,563
	(27,052)	(27,402)
Interest Rate	3.35	3.50
	(1.00)	(0.93)
Property Value	202,989	190,188
	(133,626)	(125,063)
Loan Size	149,795	143,213
	(91,227)	(84,286)
Loan Term	28.08	28.01
	(6.09)	(6.26)
LTV	76.60	77.54
	(17.52)	(15.56)
LTI	3.38	3.36
	(0.95)	(0.94)
Origination Year	2015	2014
	(0.65)	(0.50)
N	414,643	306,417

2. Data Cleaning in PSD007

For approximately 50,000 of observations in 2016 to 2017, the loan balance and interest rate are erroneously reported twice and summed, meaning they are reported as twice the actual level. I check for loan growth of more than 50% in a given half-year and replace the reported loan balance with 0.5 times the reported loan balance, which lines up with typical loan amortization trends. I use the same procedure to correct interest rates for these observations. Similarly, for approximately 15,000 observations, mortgage rates reported as 7.98% are corrected to 3.99% as mortgage rates rarely exceed 5% over the sample period and there are no mortgage offers that match a level of 7.98%. A mortgage rate of 3.99% is in line with mortgage rates by other

borrowers with similar levels of LTV. The estimate of the current house price is obtained by scaling the house price at origination with the local-authority-level house price index. The current LTV is computed as the reported loan balance, divided by the current house price. To account for outliers, mortgage rates are winsorized at the 1st and 99th percentile, and current LTV levels are winsorized between 0 and 200%.

D. Evidence On Alternative Choice Mechanisms

1. Within-Household Contract Choice

Table D.1: 5-Year Fixation Length Contract Choice Regressions - Panel

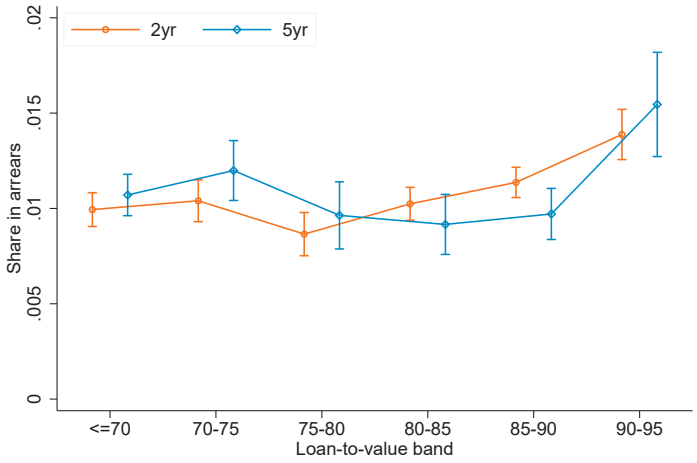
This table shows three regressions of the probability of taking out a 5-year contract, using the borrower panel data between 2015H1 to 2017H2. The dependent variable is an indicator that takes the value 1 if the contract fixation length is 5 years, and 0 if the fixation length is 2 years. The current LTV ratio is computed as the reported current loan balance, divided by the current house price, obtained by scaling the initial house price with the local-authority-specific house price index. Column (1) estimates the regression using all observations in the panel data, column (2) imposes borrower-fixed effects and only contains pairs of observation for which a repeat choice is observed, and column (3) conditions on the subset of fixation choice pairs where the household is observed to initially originate a 2-year fixed-rate contract.

	(1) I[5yr]	(2) I[5yr]	(3) I[5yr]
I[LTV≥80%]=0 × LTV	-0.024***	-0.014	-0.017
	(0.006)	(0.010)	(0.009)
I[LTV≥80%]=1 × LTV	-0.092***	-0.028***	-0.021**
	(0.015)	(0.007)	(0.008)
Time FE	✓	✓	✓
Local-Authority FE	✓	✓	✓
Loan-Decile FE	✓	✓	✓
Lender FE	✓	✓	✓
Borrower FE		✓	✓
Observations	727,094	360,496	298,504
R^2	0.07	0.65	0.55

2. Ex Post Default by Fixation Choice and LTV

Figure D.1: Ex Post Default by Fixation Choice and LTV

These figures plot the ex post share of loans in arrears, across LTV bands and initial fixation length (2 and 5 years), for first-time borrower cohorts between 2013H2 to 2015H2. The share refers to a loan being in arrears at any point in the sample window (2013H2 to 2017H2).



3. Household Refinancing Behavior with Prepayment Penalties

I use the borrower panel to study refinancing behavior over time. Figure **??** plots the cumulative share of borrowers who have refinanced at least once over time, by contract fixation period.[2] For borrowers with an initial 5-year contract, the share is very low throughout the sample window between 2013H2 and 2017H2, with only around 5% of borrowers refinancing out of their initial contract by the end of 2017, i.e. after four years since origination. For borrowers with an initial 2-year contract, there is a slight increase in refinancers in 2015H1, and a large jump in refinancers in 2015H2 as expected, as the initial fixation period ends and borrowers are moved onto the revert rate unless they refinance at this point.[3] At 2015H2, in the half-year reporting window that tracks outcomes two years after the initial contract origination, the share of refinancers jumps to around 75%. If one takes a 6-month window around the scheduled refinance date, i.e. including 2016H1, that share rises to around 85%. The share rises further to around 90 to 95% when looking at the full four-year reporting window.[4] In sum, almost all first-time borrowers remain in the contract until the end of the initial fixation window, and only around 5% of borrowers exit the contract early and pay a prepayment penalty. Inspecting this subset of borrowers further, these households have larger incomes and smaller loan balances, which could also be consistent with prepayment in order to move. Mortgages are portable in the UK and so some of the households that exit early could be porting their mortgage to another property without paying a prepayment penalty. I cannot verify the share of porters as these transactions would show up as a new loan in the data with a different location, and the data does not allow to track households across locations.

I can further illustrate the binding nature of prepayment penalties during a period of strong house price growth. Figure **??** shows ex post interest rate realizations for the cohort of 2013H2 first-time borrowers, split by initial fixation window. Consistent with contract features, average rates remain stable over the sample window from 2013H2 to 2017H2 for borrowers with a 5-year contract who lock in the initial rate at 2013H2. For borrowers with a 2-year fixation window,

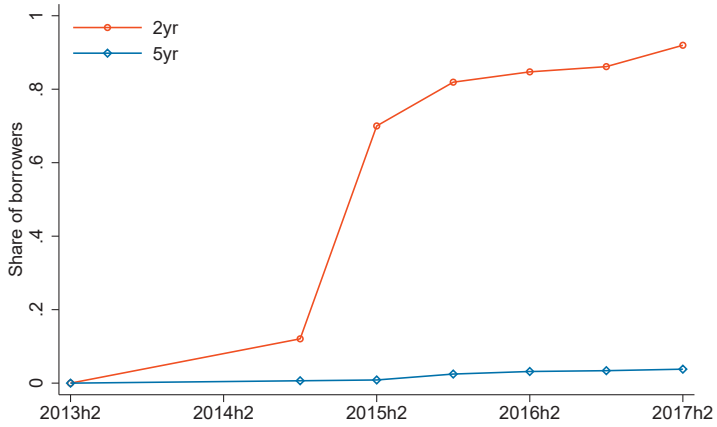
[2] I focus on the first-time borrower cohort of 2013H2 in order to maximize the sample window over which outcomes can be observed (four years until 2017H2), and confirm that the results are robust when using other cohorts or when pooling all cohorts.

[3] The slight increase in 2015H1 is partly driven by some 2-year windows ending in that half-yearly reporting period, but that were originated in 2013H2, so comprises many "on schedule" refinances.

[4] This gradual increase over time comprises both refinancers who exhibit inertia, i.e. refinance late but could have refinanced and potentially saved cost relative to the revert rate, and borrowers who were not able to refinance at that time, for instance if their LTV exceeded 100%, but were able to do so at a later point (**???**).

FIGURE D.2: REFINANCING BEHAVIOR

This figure shows the cumulative share of borrowers who refinance (either with their existing lender or with a different lender), for borrowers who chose a contract with an initial 2-year, or 5-year fixation length, respectively, based on the 2013H2 cohort of first-time borrowers.

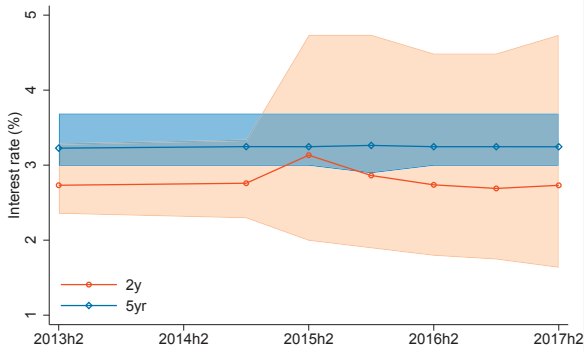


the interdecile range widens visibly in 2015H2. Panel B shows that for borrowers with a high initial LTV (85-90% LTV band), borrowers with a 2-year fixed rate window experience a sharp decrease in average rates paid in 2015H2, while borrowers with a 5-year contract continue to pay the fixed rate, despite the incentive to switch into a lower 2-year fixed rate, absent prepayment penalties.

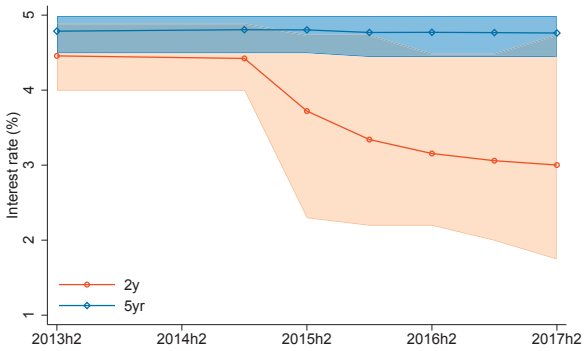
Figure D.3: Ex Post Repricing

This figure tracks the 2013H2 cohort of first-time borrowers and the distribution of mortgage rates paid based on the interdecile range (shaded area), and average rate (connected dots), over time, for 2- and 5-year fixed-rate borrowers who stay with their initial lender, respectively. Panel A shows rates paid for borrowers with an initial LTV between 70-75%, and Panel B shows the equivalent for borrowers with an initial LTV between 85-90%.

(a) Low LTV



(b) High LTV



E. Decomposing the Relative Cost of Longer-Term Mortgages

1. Decomposition of Mortgage Yield Difference

Mortgage rates can be written as having a funding cost component, approximated by a long-term government bond, and a credit cost component, in this case simplified to be captured by the LTV ratio. Denote the per-period mortgage rate $r_t^{m,\theta}$ where superscript m refers to the mortgage rate, and $\theta \in \{\theta^{ST}, \theta^{LT}\}$ is the length (in years) over which the rate stays fixed. $r_t^{m,\theta}$ depends on r_t, the base (i.e. aggregate) interest rate r_τ, and LTV_t at the time of pricing t. We can write the expected yield difference between the longer-term θ^{LT}-period mortgage rate, and the average rate when rolling over a sequence of shorter-term θ^{ST}-period contracts, given an initial LTV_t, as:[5]

$$E_t\left[r_t^{m,\theta^{LT}}(r_t, LTV_t) - \frac{1}{n}\sum_{i=0}^{n-1} r_{t+\theta^{ST}\times i}^{m,\theta^{ST}}(r_{t+\theta^{ST}\times i}, LTV_{t+\theta^{ST}\times i})\right], \quad n = \theta^{LT}/\theta^{ST}. \tag{1.1}$$

Denote this expression $\Delta^{\theta^{LT},\theta^{ST}}(LTV_t)$.

For notational simplicity, let $\theta^{LT} = n \cdot \theta^{ST}$ where n is an integer. The longer-term mortgage rate is $r_t^{m,\theta^{LT}}$, and the shorter-term mortgage rate is $r_t^{m,\theta^{ST}}$. For the longer-term mortgage rate, the mortgage is priced in the initial period t. For the shorter-term rate sequence, the rate gets repriced with each new contract, i.e. every θ^{ST} years. By adding and subtracting the average of the sequence of shorter-term rates at current LTV levels, the expression can be rewritten as:

$$\begin{aligned}\Delta^{\theta^{LT},\theta^{ST}}(LTV_t) = E_t\left[r_t^{m,\theta^{LT}}(r_t, LTV_t) - \frac{1}{n}\sum_{i=0}^{n-1} r_{t+\theta^{ST}\times i}^{m,\theta^{ST}}(r_{t+\theta^{ST}\times i}, LTV_t)\right] \\ + E_t\left[\frac{1}{n}\sum_{i=0}^{n-1}\left(r_{t+\theta^{ST}\times i}^{m,\theta^{ST}}(r_{t+\theta^{ST}\times i}, LTV_t) - r_{t+\theta^{ST}\times i}^{m,\theta^{ST}}(r_{t+\theta^{ST}\times i}, LTV_{t+\theta^{ST}\times i})\right)\right].\end{aligned} \tag{1.2}$$

We can further rewrite the first term in equation **??**. Recall that empirically, mortgage rates become essentially risk-free and insensitive to the LTV below a threshold $\underline{x}$, typically being 60 to 70% LTV. We can then split the mortgage rate into the LTV-sensitive (for which $LTV > \underline{x}$)

[5] This builds on the framework by **?** for risk-free bonds.

and in-sensitive part (for which $LTV \leq \underline{x}$). Define $r_t^{m,\theta}(r_t) \equiv r_t^{m,\theta}(r_t, LTV_t \mid LTV_t \leq \underline{x})$, i.e. under the assumption that mortgages below an LTV of $\underline{x}$ are collateral risk-free, LTV can be omitted in the notation for mortgage rates with an LTV below $\underline{x}$. Then denote $\rho_t^{\theta,LTV}$ the LTV credit spread, i.e. the rate difference between a mortgage with fixation length θ with some LTV, and the LTV-insensitive mortgage rate with an LTV below $\underline{x}$:

$$\rho_t^{\theta,LTV} = r_t^{m,\theta}(r_t, LTV_t) - r_t^{m,\theta}(r_t). \tag{1.3}$$

Combining equations **??** and **??**, we obtain:

$$\begin{aligned}\Delta^{\theta^{LT},\theta^{ST}}(LTV_t) = & \underbrace{E_t\left[r_t^{m,\theta^{LT}}(r_t) - \frac{1}{n}\sum_{i=0}^{n-1} r_{t+\theta^{ST}\times i}^{m,\theta^{ST}}(r_{t+\theta^{ST}\times i})\right]}_{\text{(I)}} \\ & + \underbrace{E_t\left[\rho_t^{\theta^{LT},LTV_t} - \frac{1}{n}\sum_{i=0}^{n-1}\rho_{t+\theta^{ST}\times i}^{\theta^{ST},LTV_t}\right]}_{\text{(II)}} \\ & + \underbrace{E_t\left[\frac{1}{n}\sum_{i=0}^{n-1}\left(r_{t+\theta^{ST}\times i}^{m,\theta^{ST}}(r_{t+\theta^{ST}\times i}, LTV_t) - r_{t+\theta^{ST}\times i}^{m,\theta^{ST}}(r_{t+\theta^{ST}\times i}, LTV_{t+\theta^{ST}\times i})\right)\right]}_{\text{(III)}}.\end{aligned} \tag{1.4}$$

Equation **??** decomposes the expected yield difference between the longer-term mortgage contract and sequence of shorter-term contracts into three components (I-III). The first component (I) is a familiar-looking expression based on the expectations theory of the term structure of interest rates (**?**), i.e. a bond term premium that is independent of the level of LTV. In the data, the time series variation in the bond term premium tracks variation in the funding cost differential between interest rate swap rates of differing maturities.

The second component (II) reflects the average difference in credit spreads for a given current LTV (i.e. initial LTV at origination time t) across longer-term and shorter term mortgage contracts, e.g. the average credit spread for a 2-year 90% LTV mortgage, compared to a 5-year 90% LTV mortgage. And the third component (III) reflects the yield difference between the shorter-term mortgage sequence without and with LTV repricing. This component is positive if there are decreases in LTV and hence credit risk over the long-term contracting horizon, i.e. when there is a credit repricing benefit associated with the shorter-term contract sequence.

2. Mapping Decomposition of Mortgage Yield Difference to the Data

Next, I map the decomposition of the mortgage yield difference to the data. To provide a concrete example of equation **??**, we can apply it to the yield difference between a 5-year fixed-rate mortgage and a sequence of 2-year fixed-rate contracts, i.e. setting $\theta^{LT} = 2$ and $\theta^{LT} = 5$. For t=0 and an initial LTV (LTV_0) of 90%, we obtain:[6]

$$\begin{aligned}\Delta^{5,2}(LTV_0 = 90) = E_0\left[r_0^{m,5}(r_0) - \frac{1}{2.5}\left(r_0^{m,2}(r_0) + r_2^{m,2}(r_2) + \frac{1}{2}r_4^{m,2}(r_4)\right)\right] \\ + E_0\left[\rho_0^{5,90} - \frac{1}{2.5}\left(\rho_0^{2,90} + \rho_2^{2,90} + \frac{1}{2}\rho_4^{2,90}\right)\right] \\ + E_0\left[\frac{1}{2.5}\left(r_0^{m,2}(r_0, 90) + r_2^{m,2}(r_2, 90) + \frac{1}{2}r_4^{m,2}(r_4, 90)\right)\right] \\ - E_0\left[\frac{1}{2.5}\left(r_0^{m,2}(r_0, 90) + r_2^{m,2}(r_2, LTV_2) + \frac{1}{2}r_4^{m,2}(r_4, LTV_4)\right)\right]\end{aligned} \tag{1.5}$$

Bond Term Premium (I). The first line in equation **??** compares the mortgage yield of a 5-year fixed-rate contract at a collateral-insensitive level of LTV ($\leq$70%) to the equivalent expected yield of a 2-year fixed-rate sequence. Using monthly average data from 2013 to 2017, I find that the 5-year rate mortgage rate at 70% LTV lies above the 2-year 70% LTV rate throughout, consistent with a positive bond term premium included in the 5-year mortgage rate relative to the 2-year rate. Since there is no forward market for 2-year fixed-rate mortgage contracts, I use spot rates to compute an expected yield difference. The difference between the 5-year and 2-year mortgage rate at 70% LTV is around 50 basis points over this period. In addition, I find that both the level of the bond term premium, and variation in the bond term premium over time is strongly correlated with the difference in 5-year and 2-year swap rates, reflecting lenders' relative funding cost. Lenders in the UK typically enter a swap contract which matches the initial fixation period of the mortgage contract to hedge interest rate exposure, by paying a floating rate plus premium and receiving a fixed rate for funding. The 5-year fixed-rate mortgage contract requires a 5-year swap contract, while the 2-year fixed rate requires a 2-year swap contract.

Pricing Difference in Credit Spreads (II). The second line in equation **??** compares the

[6]Note that because in this case θ^{LT}/θ^{ST} is not an integer, the last 2-year contract is divided by two to reflect the same contract horizon as $\theta^{LT} = 5$.

difference in credit spreads, holding constant the level of LTV. I compute $\rho_t^{\theta,LTV}$ by extracting the credit spread paid across LTV bands for 2 and 5-year contracts relative to the lowest LTV band ($\leq 70\%$) from the data, using the following regression:

$$r_{ijt}^m = \alpha + \kappa \mathbb{I}[5yr] + \rho^{2yr} \sum_{c=1}^{C} \mathbb{I}[LTV = c] + \rho^{5yr} \sum_{c=1}^{C} \mathbb{I}[LTV = c] \times \mathbb{I}[5yr = 1] \tag{1.6}$$
$$+ \gamma_t + \gamma_j + \gamma_1 \sum_{i=1}^{N} \mathbb{I}[i = type] + \gamma_2 \sum_{q=1}^{Q} \mathbb{I}[fees = q] \times \mathbb{I}[fees = 1] + \epsilon_{ijt},$$

where mortgage rates r^m by household i, lender j, originated at year-month t are regressed on an indicator for a 5-year contract ($\mathbb{I}[5yr]$), LTV band-fixed effects in steps of five percentage points, starting from the LTV pricing threshold $\underline{x} = 70\%$, i.e. $c \in \{[0-70], (70-75], (75-80], (80-85], (85-90], (90-95]\%)\}$, controlling for year-month, lender, buyer-type (first-time buyer, second-time buyer or refinance) and fee-quintile fixed effects, using data from 2013 to 2017. As a robustness check, the magnitudes remain very similar when estimating the regression on the full origination data between 2005Q2 and 2017Q4, and including additional loan-specific fixed effects and household controls.[7] The bond term premium κ can also be directly extracted from this regression, as the average difference between 2 and 5-year fixed rate contracts for the $\leq$70% LTV band, yielding 53 basis points, which is very similar to aggregate data of 52 basis points using the Bank of England Database.

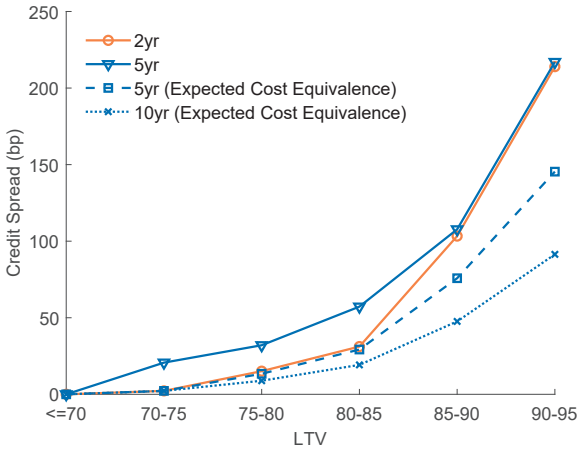
The resulting credit spreads are plotted in Figure **??**. The pricing difference in credit spreads across 5-year and 2-year contracts across LTV bands (i.e., the difference between the credit pricing curves) is positive at an LTV up to 85%, and around zero for high LTV levels greater than 85%. A pure credit pricing differential could be consistent with some degree of adverse selection into 5-year fixed-rate contracts at low levels of LTV, but the effect seems less relevant at high levels of LTV.

Credit Repricing Effect (III). Lastly, the third and fourth line in equation **??** require a comparison between the average yield for a 2-year contract sequence, holding the LTV constant,

[7]Note that the credit spreads are estimated jointly for 2-year and 5-year fixed rate contracts in the same regression, with 2-year fixed rate contracts as the base category, and 5-year fixed rate contracts with an additional interaction term.

FIGURE E.1: OBSERVED AND COUNTERFACTUAL CREDIT SPREAD PRICING

This figure plots credit spreads across loan-to-value (LTV) bands (≤70%, (70-75]%, (75-80]%, (80-85]%, (85-90]%, and (90-95]%) and fixation length, by extracting LTV-band fixed effects from a regression of mortgage rates on LTV bands and fixation period length (2 or 5 years), controlling for year-month, lender, buyer-type and fee-quintile fixed effects, using data from 2013 to 2017 (see equation **??**). It further shows counterfactual credit spreads for 5-year and 10-year fixed-rate contracts that would equalize the expected average cost of rolling over a matching sequence of 2-year fixed rate contracts given 2-year credit spreads over five (dashed line) and ten years (dotted line), respectively.

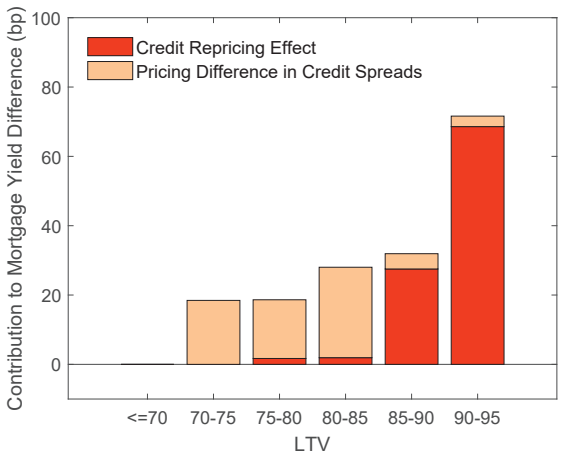


compared to the average yield for a 2-year contract sequence, while repricing the LTV over time. This requires an expected future path of LTV, which is obtained using calibrated UK house price growth with $\mu_h = 0.0258$ (data from 1987 to 2017, nominal house prices deflated using RPI), and computing a fully-amortizing loan repayment path (assuming repayment over 30 years). Figure **??** decomposes the total mortgage yield difference into the pricing difference in credit spreads, and credit repricing effect, across LTV bands. The credit repricing effect is relatively small for an LTV below 85%, and rises to 69 basis points for an LTV of 95%.

Mortgage Yield Difference Across LTV. The total mortgage yield difference across LTV is computed as the sum of the credit spread pricing difference and credit repricing effect, and is also plotted in Figure **??**. The total mortgage yield difference rises from 18 basis points at 75% LTV, to 72 basis points at 95% LTV. The overall mortgage yield difference is hence increasing in LTV, implying that the cost of insurance via longer-term contracts is increasing in borrower riskiness.

FIGURE E.2: DECOMPOSITION OF MORTGAGE YIELD DIFFERENCE

This shows the decomposition of the total mortgage yield difference into the credit spread pricing differential across 2-year and 5-year contracts, and the credit repricing effect, across LTV bands.



Expected Cost Equivalence. An alternative way to interpret the credit repricing effect is as the magnitude by which the 5-year contract should be cheaper than the 2-year contract - i.e. for the expected cost to be equivalent, 5-year contracts would have to price in the declining expected rate path of the 2-year contract sequence, and the 5-year LTV pricing curve would be weakly lower than the 2-year curve at each LTV band. This effect would be even more pronounced when comparing the short-term rate path with LTV repricing over a 10-year fixation window. This counterfactual pricing scheme is shown for 5- and 10-year fixed-rate contracts in Figure **??**.

Effect on Household Contract Choice. The decomposition shows that the credit repricing effect, due to the expected decline in credit spreads over time, raises the relative cost of insurance for high-LTV borrowers. The decomposition provides a stylized way of illustrating the relative cost by comparing the longer-term contract with a matching shorter-term contract sequence. When households make an optimal contract choice, they take into account the relative cost, but also need to account for the fact that they can choose again between a shorter-term and longer-term contract after every contract, i.e. they are not locked into a shorter-term contract sequence and gain a differential option value from a 2-year compared to a 5-year contract. They also need to trade off the insurance benefit that they obtain, together with the joint distribution of risks over time, all of which is more fully considered in the model section.

F. Model Appendix

1. Value Function for $\theta^{LT} = 5$

The value function captures the temporal dependencies across repricing states. Using an example of a 5-year fixed-rate contract, the value function in the fifth period from choosing the contract is

$$\begin{aligned} &V_{t+4}\left(X_{t+4}, LTV_{t+4}, r_{t+4}, \mathcal{S}_{t+4} | \mathcal{S}_{t+4} = 1\right) \\ &= \underset{C_{t+4}, R_{t+4}}{max}\ U(C_{t+4}) + \beta E_{t+4}\left[V^*_{t+5}\left(X_{t+5}, LTV_{t+5}, r_{t+5}\right)\right], \end{aligned}$$

with repricing at the end of the period, indicated by $\mathcal{S}_t = 1$. Over the fixation length of the contract, the household is protected from repricing, indicated by $\mathcal{S}_t \in \{2, ...5\}$. The continuation value, where $V^*_t = \max\left\{V^{ST}_t, V^{LT}_t\right\}$, takes into account that the household can optimally choose a short- or long-term contract after this period, and does not depend on the repricing state.

In the fourth period, the value function is

$$\begin{aligned} &V_{t+3}\left(X_{t+3}, LTV_{t+3}, r_{t+3}, \mathcal{S}_{t+3} | \mathcal{S}_{t+3} = 2\right) \\ &= \underset{C_{t+3}, R_{t+3}}{max}\ U(C_{t+3}) + \beta E_{t+3}\left[V^{LT}_{t+4}\left(X_{t+4}, LTV_{t+4}, r_{t+4}, \mathcal{S}_{t+4} \mid \mathcal{S}_{t+4} = 1\right)\right]. \end{aligned}$$

In the third period, the value function is

$$\begin{aligned} &V_{t+2}\left(X_{t+2}, LTV_{t+2}, r_{t+2}, \mathcal{S}_{t+2} | \mathcal{S}_{t+2} = 3\right) \\ &= \underset{C_{t+2}, R_{t+2}}{max}\ U(C_{t+2}) + \beta E_{t+2}\left[V^{LT}_{t+3}\left(X_{t+3}, LTV_{t+3}, r_{t+3}, \mathcal{S}_{t+3} \mid \mathcal{S}_{t+3} = 2\right)\right]. \end{aligned}$$

In the second period, the value function is

$$\begin{aligned} &V_{t+1}\left(X_{t+1}, LTV_{t+1}, r_{t+1}, \mathcal{S}_{t+1} | \mathcal{S}_{t+1} = 4\right) \\ &= \underset{C_{t+1}, R_{t+1}}{max}\ U(C_{t+1}) + \beta E_{t+1}\left[V^{LT}_{t+2}\left(X_{t+2}, LTV_{t+2}, r_{t+2}, \mathcal{S}_{t+2} \mid \mathcal{S}_{t+2} = 3\right)\right]. \end{aligned}$$

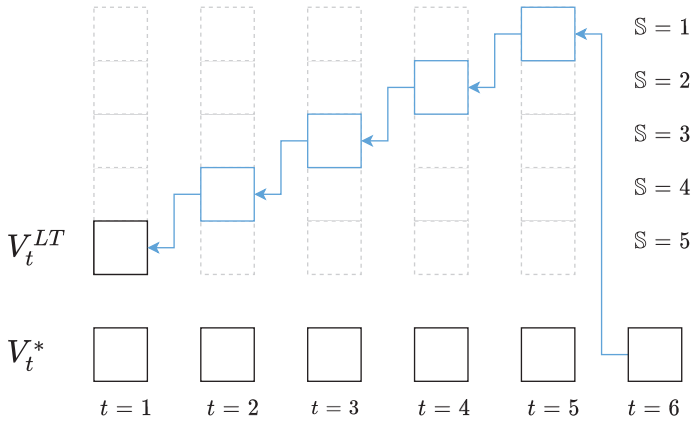
In the current period, the value function for choosing the 5-year fixed-rate contract is

$$V_t\left(X_t, LTV_t, r_t, \mathcal{S}_t \mid \mathcal{S}_t = 5\right)$$
$$= \max_{C_t, R_t} U(C_t) + \beta E_t \left[V_{t+1}^{LT}\left(X_{t+1}, LTV_{t+1}, r_{t+1}, \mathcal{S}_{t+1} \mid \mathcal{S}_{t+1} = 4\right)\right].$$

The dependencies across time and repricing states are further illustrated for V^{LT} with $\theta^{LT} = 5$ in Figure **??** (starting with $t = 1$), with arrows to indicate the dependencies of the value function and continuation values as described above.

FIGURE F.1: ILLUSTRATION OF VALUE FUNCTION AND REPRICING STATES

This figure illustrates the time×repricing state dependencies of the value function, for $\theta^{LT} = 5$. Horizontally, each box represents a value of the value function in the time dimension. Vertically, each box represents a value of the value function in the repricing state dimension, for V_t^{LT}. Since $V_t^* = \max\{V_t^{LT}, V_t^{ST}\}$, there is only one value across all repricing states for V_t^*. V_t^{LT} tracks the value function if the household chooses a long-term contract at each point in time t, using its value across a linked chain of repricing states from $\mathcal{S} = 1$ to $\mathcal{S} = 5$.



2. NUMERICAL SOLUTION

Discretization. The six state variables are discretized as follows. Net financial wealth (X) is normalized by permanent income at age 35. Grid points are equally spaced on a grid between 0 to 22.5 in steps of 0.025, yielding 901 grid points. Time is measured in years between 30 to 80 (working age from 30 to 60, and retirement from 60 to 80), yielding 51 grid points. The LTV grid takes values between 30 to 150, in steps of 1 percentage points, yielding 121 grid points. The interest rate process is discretized using five states using the method by **?**, which has been shown by **?** to yield better results when approximating very persistent AR(1) pro-

cesses compared to **??**. The repricing state variables $\mathcal{S}^{\theta^{LT}}$ and $\mathcal{S}^{\theta^{ST}}$ take 5 (or 10) and 2 states, respectively. Transitory income shocks and house price shocks are discretized on an equal-spaced grid between -4 and 4 standard deviations.[8] Consumption is placed on the same grid as net financial wealth, while mortgage contract choice is discrete with two outcomes (short- or longer-term contract).

Model Solution. Optimal consumption and mortgage fixation choice policy functions are found as the maximum for each combination of discretized states in the state space, i.e. $901 \times 51 \times 121 \times 5 \times (5 + 2)$ (for the 5-year contract) or $901 \times 51 \times 121 \times 5 \times (10 + 2)$ (for the 10-year contract), yielding around 195 to 334 million combinations. The model is solved separately under different specifications, and under different pricing assumptions, which is computationally intensive and parallelized on the Imperial HPC cluster.

Simulation. I use the optimal policy functions for consumption and mortgage fixation choice to simulate the model. Households are initialized at $t = 1$ with zero net financial wealth and a random distribution of transitory income shocks, and no house price or interest rate shocks, in order to allow households to start at the same initial LTV band and interest rate. The initial base interest rate places households in the second lowest out of five states, in order to reflect an economic environment with a greater emphasis on the risk of rising rates. The simulation is done separately for households starting at different LTV levels, but uses the same shocks. Each simulation is done for 10,000 households.

[8] Tail probabilities exceeding the LTV grid are added to the lowest and highest grid point, respectively.

G. Additional Model Results - Robustness

Table G.1: Long-Term Contract Shares (Robustness)

This table shows the simulated long-term contract shares given optimal household choice under different scenarios. The columns show results taking pricing as given for low- (70%) and high-(90%) LTV borrowers, and under pricing that equalizes the expected cost of the long-term contract and short-term sequence (shown in Figure **??**), respectively. The rows show additional scenarios. The long-term contract shares are computed as the share of household-year observations that are under a long-term, compared to a short-term contract, over the first ten years since loan origination (Panel A) and over the entire maturity of the loan, 30 years (Panel B). The simulation tracks household optimal behavior given realization of shocks, based on 10,000 households for each scenario and LTV band.

	Baseline		Expected Cost Equivalence	
	Low LTV	High LTV	Low LTV	High LTV
Panel A: Share on Long-Term Contract Over Initial 10 Years				
Higher risk aversion ($\gamma = 10$)	0.95	0.35	0.96	0.94
Higher time discount rate ($\beta = 0.9$)	0.95	0.35	0.96	0.93
Higher revert rate ($\rho^{REV} = 538bp$)	0.95	0.40	0.96	0.94
Panel B: Share on Long-Term Contract Over 30 Years				
Higher risk aversion ($\gamma = 10$)	0.84	0.63	0.84	0.83
Higher time discount rate ($\beta = 0.9$)	0.83	0.62	0.84	0.83
Higher revert rate ($\rho^{REV} = 538bp$)	0.83	0.63	0.84	0.83

Table G.2: Value of Long-Term Contract (Robustness)

This table shows the consumption-equivalent of introducing a long-term contract to an existing short-term contract under different scenarios. The columns show results taking pricing as given for low- (70%) and high-(90%) LTV borrowers, and under pricing that equalizes the expected cost of the long-term contract and short-term sequence (shown in Figure **??**), respectively. The rows show additional scenarios. The consumption certainty equivalent is computed as the percentage increase in consumption across states that a household would require to reach the same life-time expected utility when a longer-term contract is available, in addition to the shorter-term contract. Life-time expected utility is simulated by tracking household optimal behavior given realization of shocks, based on 10,000 households for each scenario and LTV band.

	Baseline		Expected Cost Equivalence	
	Low LTV	High LTV	Low LTV	High LTV
Panel A: Value of 5-year Contract				
Higher risk aversion ($\gamma = 10$)	1.08	0.31	1.08	0.88
Higher time discount rate ($\beta = 0.9$)	0.91	0.22	0.91	1.04
Higher revert rate ($\rho^{REV} = 538bp$)	0.72	0.21	0.72	0.75

Appendix 2

Appendix for "Non-Salient Fees in the Mortgage Market"

A. Further Data Description

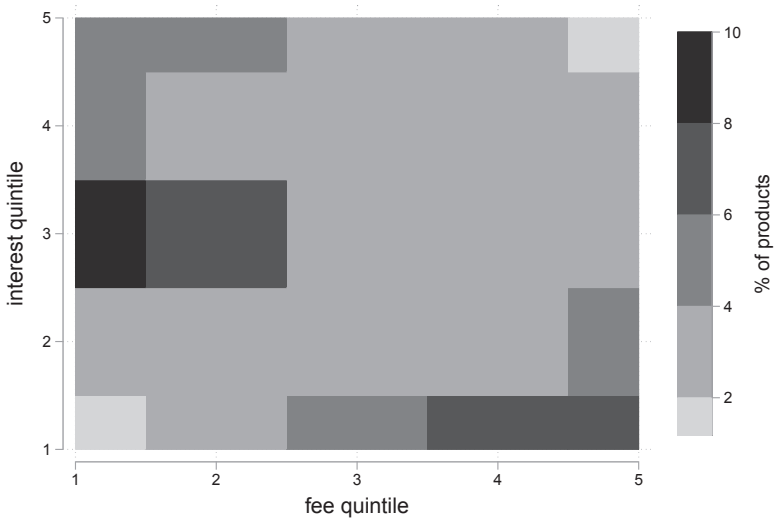
MoneyFacts product-level data. The full Moneyfacts dataset from 2009 to 2016 contains 364,750 observations. Duplicates and mortgages with non-standard eligibility criteria such as shared ownership or buy to let mortgages are dropped, in order to come close to the homogeneous product benchmark as possible and to avoid additional product characteristics that affect a very small share of products. I further restrict my sample to fixed-rate mortgages (approximatey 70% of the sample), available to first time buyers, with a 2-year fixation period (70% and 40% of the remaining sample, respectively). For the lender-level analysis, I only keep mortgage offers by the 27 largest lenders which make up about half of the observations. The resulting main Moneyfacts sample contains 26,451 unique mortgage offers, with approximately 280 observations on average each month. Some product characteristics, in particular fees and prepayment penalties, are extracted via a keyword search of raw text variables in the Moneyfacts data, as shown in Table **??**, with the extracted values marked in blue. "Arrangement Fee Notes" is a text variable that records different arrangement fee components and all fee components are added up, where composite fees serve as the main fee variable. There are four different arrangement fee components in the data, "Arrangement", "Booking", "Completion"

and "Reservation". However, they mostly reflect differences in naming convention across lenders where no lender has more than two different fee components, and they are all considered as part of the arrangement fees required at loan issuance. "Incentive Notes" captures additional incentives and rebates. An example for an additional incentive is a cash rebate of e.g. £100, but the incentive does not seem to affect the interest rate-fee trade-off, i.e. does not seem to be *priced* in terms of differential fees or interest rates and so should not affect the analysis within a given lender, which I confirm more formally in a regression setup. One potential explanation is that these products were offered as part of a temporary marketing campaign. "Prepayment penalty" specifies the terms of the early repayment penalty. I compute a £ cost for the product-level regressions for an average loan size and repayment structure. Prepayment penalties rarely vary within a given lender and are identical for most lenders across products. They do not seem to significantly affect the interest rate-fee trade-off in the regression analysis.

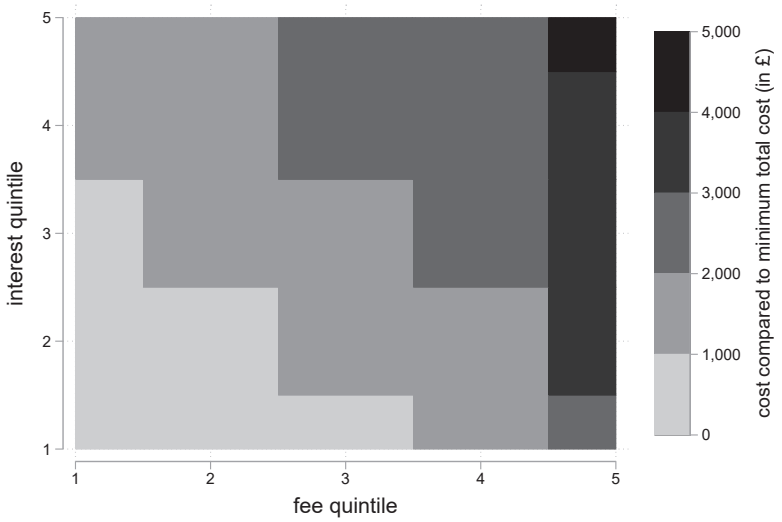
B. Additional Figures and Tables

Figure B.1: Products By Interest and Fee Quintile

(a) Share of Products By Interest and Fee Quintile



(b) Cost Differential For Average Product In Interest And Fee Quintile



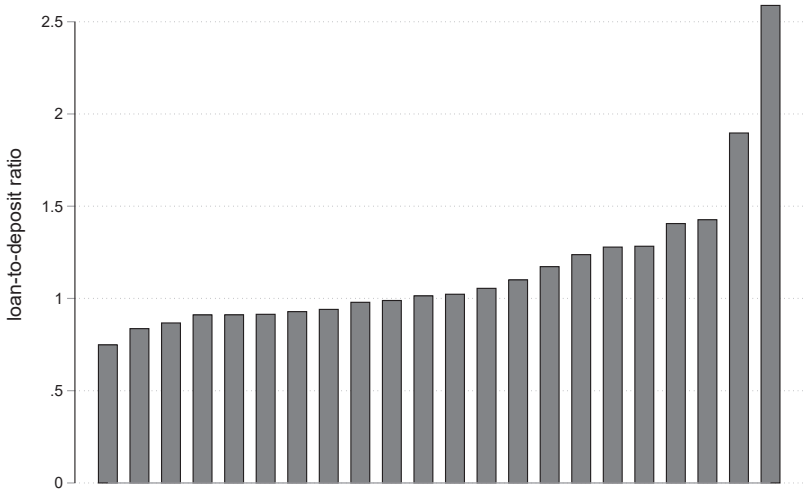
Notes: Figure **??** shows the frequency of products by its position in the rate and fee distribution in a given month, based on 75% LTV 2-year fixed rate products from January 2009 to December 2016. For instance, the lower left corner represents products in the lowest interest and fee quintile and make up around 2% of total products. Figure **??** shows the average (across products in a given interest and fee quintile) cost differential in £, measured as the difference between total cost and the minimum total cost product in a given month, based on a 75% LTV 2-year fixed rate product over two years, for a loan size of £150,000 amortized over 25 years.

Figure B.2: Total Cost Dispersion by Fee Category



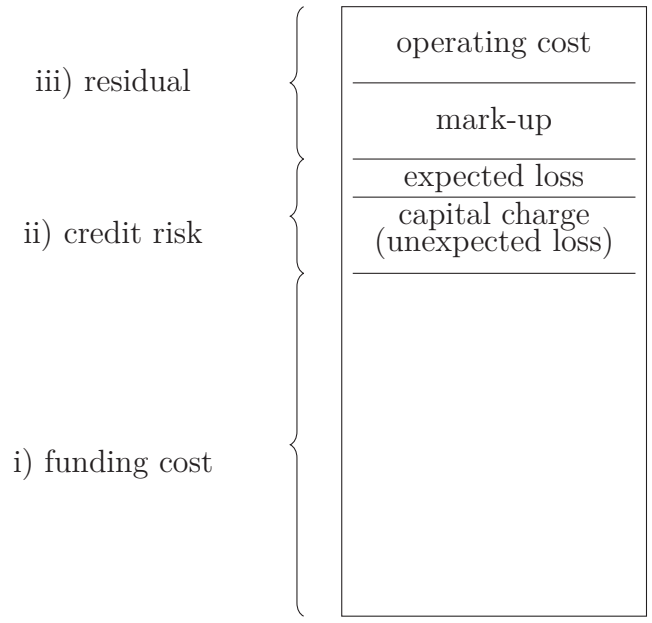
Notes: This figure shows the total cost dispersion relative to the average product, by loan value and fee category, for the set of 2yr, 70-75% LTV products on offer in February 2014 (corresponding to Figure **??** and **??**).

FIGURE B.3: LOAN-TO-DEPOSIT RATIO (2008)



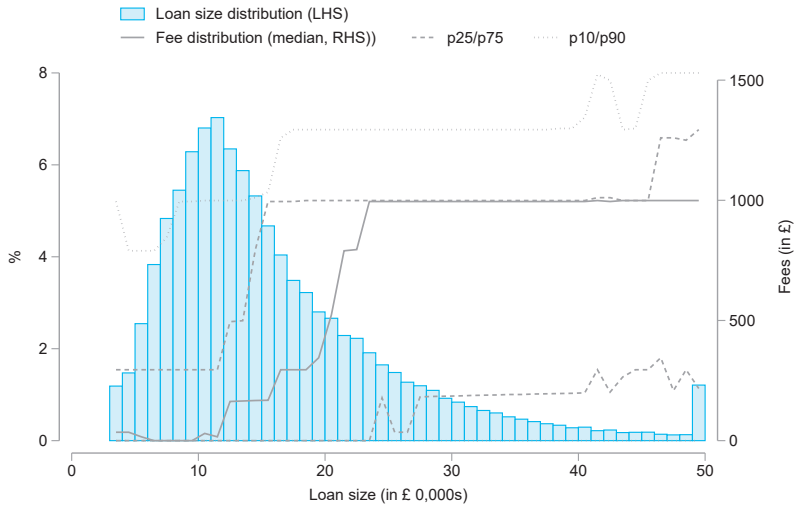
Notes: This figure shows loan-to-deposit ratios in 2008 across lenders.

FIGURE B.4: ILLUSTRATION OF MORTGAGE PRICE COMPONENTS



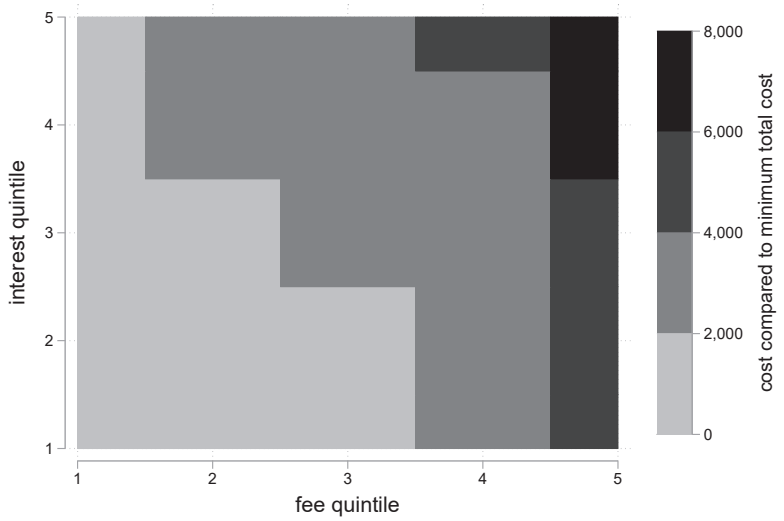
Notes: Adapted from **?**. The proportions are stylized, but similar to what they estimate for the period between 2010 and 2012.

Figure B.5: Loan Value Histogram and Fee Distribution



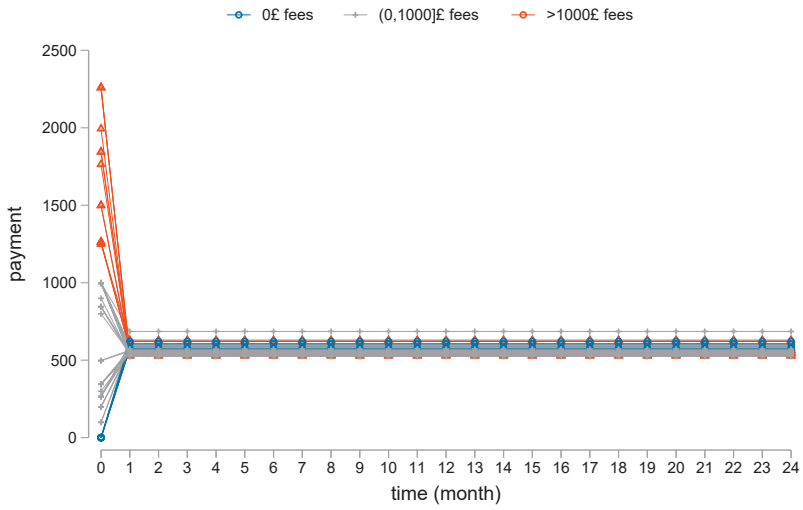
Notes: This figure shows the realised loan size and fee distribution (based on mortgages originated between 2015-2016).

Figure B.6: Cost Differential For Average Product In Interest And Fee Quintile (High Loan Value)



Notes: This figure shows the average (across products in a given interest and fee quintile) cost differential in £, measured as the difference between total cost and the minimum total cost product in a given month, based on a 75% LTV 2-year fixed rate product over two years, for a loan size of £250,000 amortized over 25 years.

FIGURE B.7: MORTGAGE PAYMENT PROFILE OVER TIME



Notes: This figure shows the payment profile for a borrower who takes out a 2yr fixed rate mortgage at 75% LTV for an average loan size, for the range of 2yr 75% LTV products in February 2014 (corresponding to Figure **??** and **??**).

Table B.1: Example of Moneyfacts Data Structure

Variable	75% LTV, low fee	75%, high fee
Collected date	30/04/2013	30/04/2013
Lender	Halifax	Halifax
Initial Period (in months)	27	27
Max LTV % FTB	75	75
Initial Rate	3.39	2.69
Arrangement Fee Notes	Completion GBP295	Arrangement GBP995, Completion GBP295
Incentive	GBP150 rebate. No Higher Lending Charge (HLC).	GBP150 rebate. No Higher Lending Charge (HLC).
Prepayment penalty	to 31/8/15: 3/2% Mortgage Advance	to 31/8/15: 3/2% Mortgage Advance
	90% LTV, low fee	90% LTV, high fee
Collected date	30/04/2013	30/04/2013
Lender	Halifax	Halifax
Initial Period (in months)	28	28
Max LTV % FTB	90	90
Initial Rate	5.59	4.99
Arrangement Fee Notes	Completion GBP295	Arrangement GBP995, Completion GBP295
Incentive	No Higher Lending Charge (HLC).	No Higher Lending Charge (HLC).
Prepayment penalty	to 31/8/15: 3/2% Mortgage Advance	to 31/8/15: 3/2% Mortgage Advance

Notes: This table shows four illustrative mortgage offers by the same lender at the same point in time. It shows the data structure and provides examples of the text variables "Initial Text", "Arrangement Fee Notes", "Incentives", and "Redemption Penalty". The main variables extracted via keyword search are interest rates, fee components and total fee amount (in blue).

Table B.2: Proportion of Cost-Dominated Products By Fee Category (in %) (High Loan Value)

Excess cost	Fees					Total
	zero fees	(0,500]	(500,1000]	(1000,1500]	>1500	
<=500	0.9	2.2	3.6	0.4	0.0	7.3
(500,1000]	2.6	3.5	6.1	0.8	0.2	13.2
(1000,2000]	5.8	6.0	11.9	3.4	2.0	29.0
(2000,4000]	7.2	7.1	10.5	4.9	2.6	32.3
>4000	2.7	5.1	6.4	2.7	1.3	18.2
total	19.1	23.9	38.5	12.3	6.2	

Notes: This table shows the proportion of all products (in %) between January 2009 and December 2016 split by excess cost compared to the lowest cost product available in a given month, and fee categories. The cost are computed for a 2-year fixed rate mortgage over two years, at 75% LTV for first-time buyers, for a loan size of £250,000, which corresponds to about the 90th percentile of the loan size distribution, amortized over 25 years.

Table B.3: Descriptive Statistics For Discrete Choice Analysis

Variable	Observations	Mean	Std. Dev.	Min	Max
Interest	23067	2.19	0.59	1.44	5.29
Fees	23067	691	582	0	2250
Loan value	23067	171689	93186	38000	499999
Income	23067	52779	29154	15184	158420
Age	23067	32.95	5.49	25.00	45.00
LTI	23067	3.34	0.77	1.25	4.50
LTV (adjusted)	23067	72.43	2.96	65.00	75.00
Bunching indicator	23067	0.23	0.42	0.00	1.00
Distance to nearest branch (km)	23067	2.56	2.67	0.05	74.70
Amortization period (term)	23067	26	4	12	30

Notes: This table displays descriptive statistics for the main sample of borrowers underlying the discrete choice analysis, based on 2-year fixed rate borrowers at 75% LTV, with standardised characteristics (age 25 to 45, LTI ≤ 4.5, amortization period ≤ 30 years), between 2015 and 2016. The bunching indicator is 1 for borrowers who are able to bunch at the 75% LTV band using fees. The distance to the nearest branch (of a given lender) is computed using the borrower's location and a 2016 snapshot of all UK bank branch locations.

Table B.4: Balance Tests For Funding Shock ϕ

	Lagged levels			Changes		
	2009	2012	2016	2009	2012	2016
log(assets)	6.03	0.36	0.55	-0.52	-0.03	-0.09
	(6.98)	(1.80)	(0.98)	(0.39)	(0.06)	(0.16)
return on assets	-10.85	-9.51***	1.96	-0.16	-0.01	0.01
	(10.06)	(1.55)	(1.77)	(0.14)	(0.02)	(0.02)
net interest margin	-2.04	1.71	-2.71*	-0.30	-0.01	0.08
	(7.77)	(3.93)	(1.42)	(0.47)	(0.12)	(0.09)
leverage	0.48	0.19	-0.17	-0.01	0.00	-0.00
	(0.50)	(0.19)	(0.10)	(0.01)	(0.00)	(0.00)
big 6 lenders (indicator)	-21.40	11.08	2.33	0.11	0.06	-0.01
	(29.06)	(8.21)	(3.58)	(0.32)	(0.05)	(0.03)
Observations	16	22	27	16	22	22
R^2	0.443	0.726	0.393	0.160	0.236	0.084

Notes: */**/*** denote p<0.1, p<0.05 and p<0.01, respectively. This table reports results from 6 different regressions, based on cross-sectional regressions of funding shock ϕ_j on lender characteristics for the years 2009, 2012 and 2016 separately, in lagged levels and changes. The table is based on Table 3 in **?**. Standard errors are clustered at lender level.

Table B.5: Changes In Price Components And Total Cost (High LTV)

Dep. Var:	$\Delta\bar{r}$	$\Delta\bar{f}$	$\Delta\bar{C}_{1yr}$	$\Delta\bar{C}_{2yr}$	Δr_l	Δf_h	$\Delta C_{1yr}^{r_l,f_h}$	$\Delta C_{2yr}^{r_l,f_h}$	Δr_h	Δf_l	$\Delta C_{1yr}^{r_h,f_l}$	$\Delta C_{2yr}^{r_h,f_l}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\Delta\phi$ (std.)	0.09***	-10.41	84.13***	178.66***	0.10***	11.82	82.17***	166.89***	0.08***	-25.09*	85.86***	202.16***
	(0.02)	(11.54)	(20.73)	(38.33)	(0.02)	(28.07)	(26.74)	(42.98)	(0.02)	(12.81)	(18.66)	(38.85)
Year-Quarter FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lender FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	515	515	515	515	515	515	515	515	515	515	515	515
R^2	0.271	0.083	0.255	0.274	0.246	0.100	0.233	0.258	0.247	0.085	0.236	0.262

Notes: This table reports results from 12 different regressions of changes in average interest rates, fees and total cost (over one and two years) on changes in funding shock ϕ (columns (1)-(4)), and the equivalent for $\{r_l, f_h\}$-products (columns (5)-(8)), and $\{r_h, f_l\}$ products (columns (9)-(12)), based on lender-level panel data between 2009 Q1 and 2016 Q4, at quarterly frequency. Prices are for 2-year fixed rates and 90-95% LTV product offers only. All regressions absorb time (year-quarter) and lender fixed effects. Standard errors are reported in parentheses and clustered at the lender level. */**/*** denote p<0.1, p<0.05 and p<0.01, respectively.

Table B.6: Fee-Rate Trade-Off

Dep. var: interest cost	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fees	-0.2022	-0.4742***	-0.4371***	-0.2419**	-0.4567***	-0.3878***	-0.9848***
	(0.1255)	(0.1115)	(0.1198)	(0.1049)	(0.1015)	(0.1033)	(0.1427)
Early prepayment penalty				-0.0251	0.0612	0.0618	0.0440
				(0.0365)	(0.0890)	(0.2523)	(0.2577)
Cashback indicator				-0.0141	-0.1015	-0.1152	-0.1144
				(0.2867)	(0.1862)	(0.1382)	(0.1087)
Incentive (length)				-1.4142	1.0558	2.5928***	1.5745**
				(1.1843)	(0.6393)	(0.8425)	(0.6952)
Fees^2							0.0003***
							(0.0001)
Year-Month FE	✓	✓	✓	✓	✓	✓	✓
LTV FE	✓	✓	✓	✓	✓	✓	✓
LTV×Year-Month FE	✓	✓	✓	✓	✓	✓	✓
Lender FE		✓	✓		✓	✓	✓
Lender×Year-Month FE			✓			✓	✓
Observations	26450	26450	26450	26450	26450	26450	26450
Adj. R^2	0.838	0.896	0.919	0.840	0.896	0.920	0.925

Notes: This table reports results from 7 different regressions of interest cost, calibrated to the average loan size in a given month, on fees, based on product-level panel data for the full sample, at monthly frequency. The early prepayment penalty is computed for the average loan size. The cashback indicator is 1 if a product provides additional cash incentives (such as a £100 rebate). The incentive length is measured as the text length of the description of additional incentives. Standard errors are clustered at lender level. */**/*** denote p<0.1, p<0.05 and p<0.01, respectively.

Appendix 3

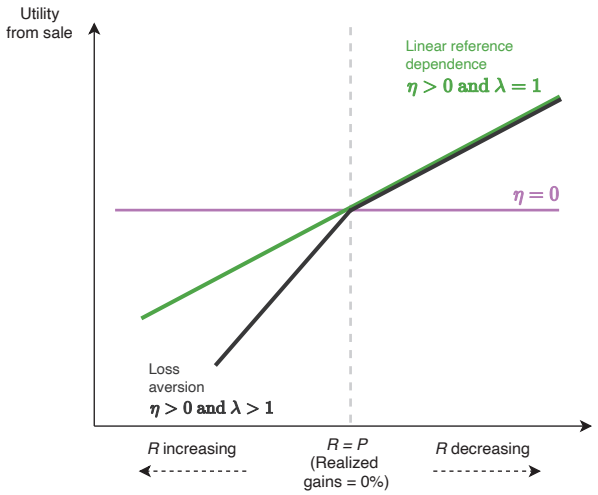
Appendix for "Reference Dependence in the Housing Market"

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

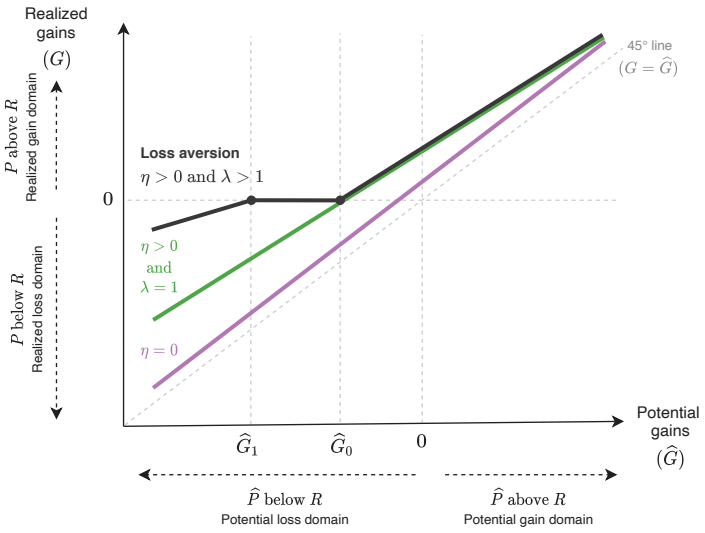
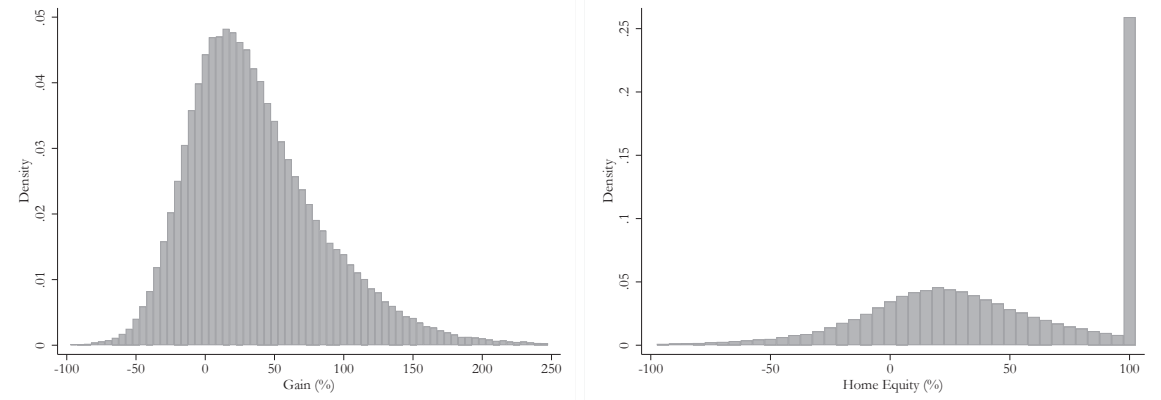


FIGURE A.2: GRAPHICAL SUMMARY STATISTICS:
POTENTIAL GAINS, POTENTIAL HOME EQUITY, LISTING PREMIUM AND TIME-ON-THE-MARKET

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

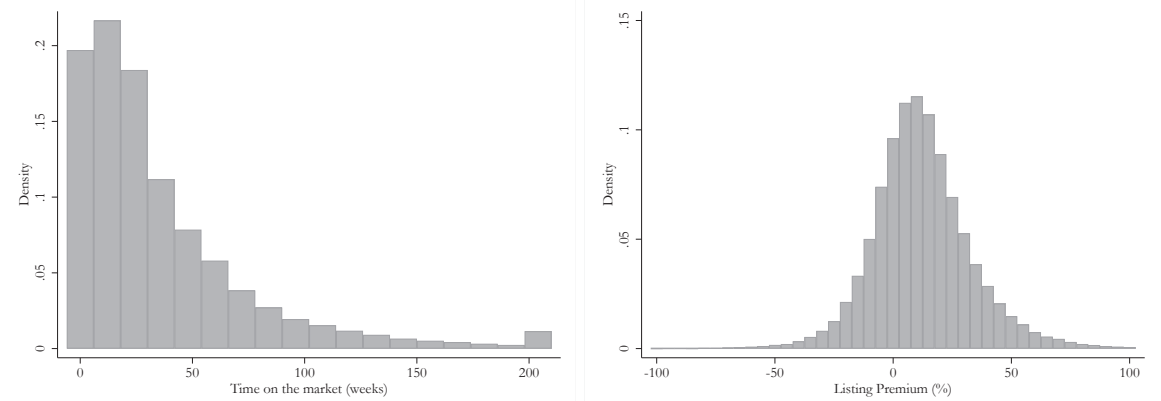


Figure A.3: Bunching of Listing Prices around Reference Point

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.

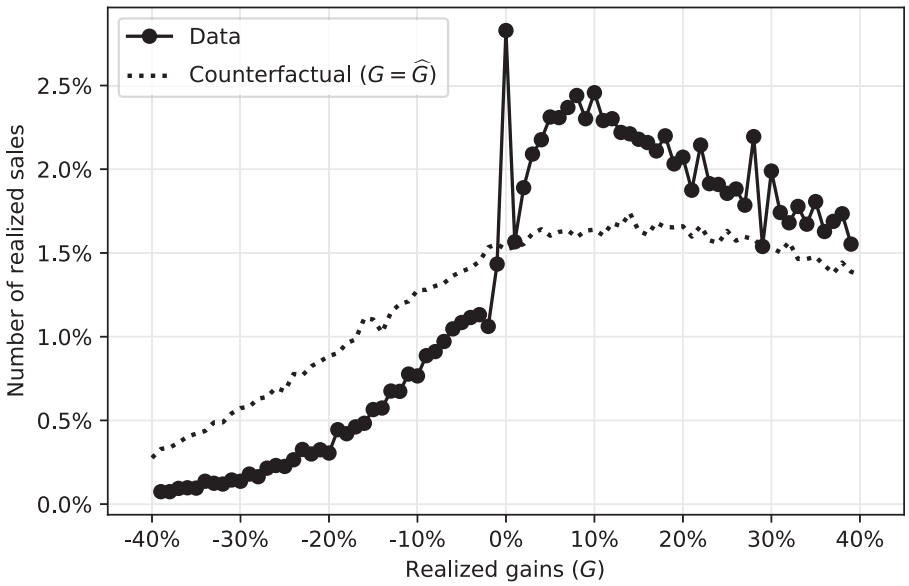


FIGURE A.4: JOINT DISTRIBUTION OF GAINS AND HOME EQUITY AND REGIONS WITH $\widehat{G} \lessgtr 0$ AND $\widehat{H} \lessgtr 20$

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.

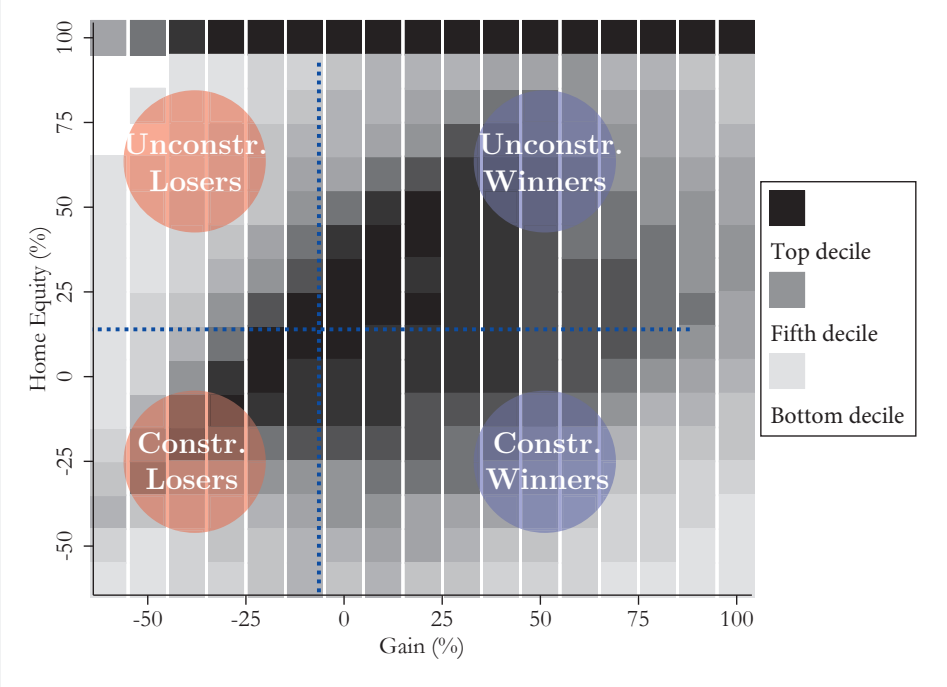


Figure A.5: Seller Groups - Listed (Relative Shares)

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

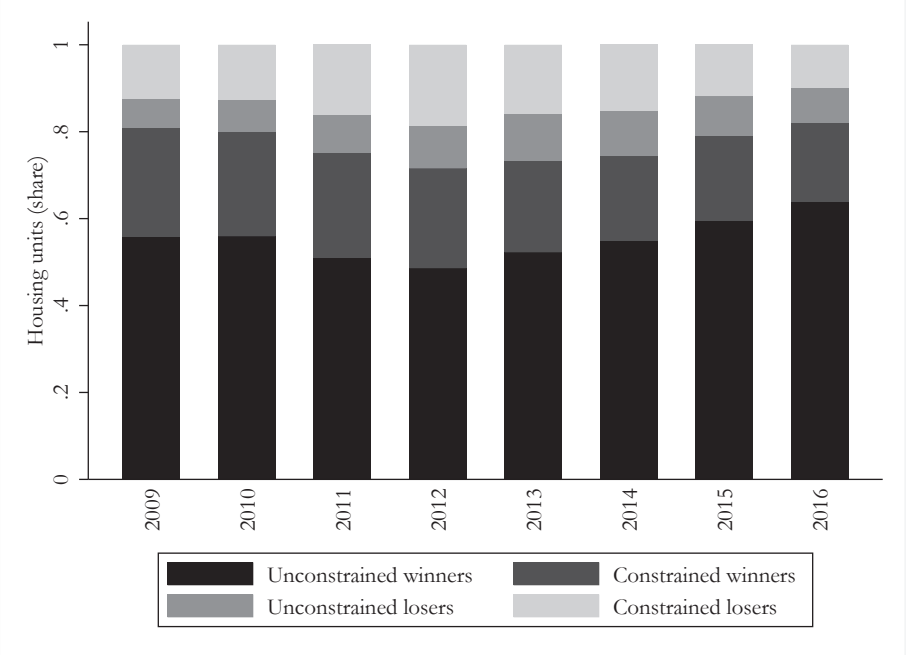


Figure A.6: Realized Gains vs. Realized Home Equity: Bunching

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.

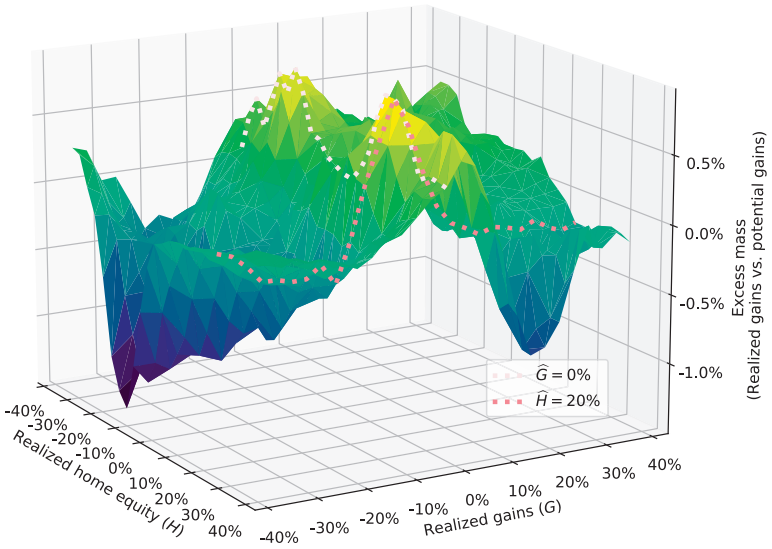


Figure A.7: Bunching around Realized Gains of Zero: Polynomial Counterfactual

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%].

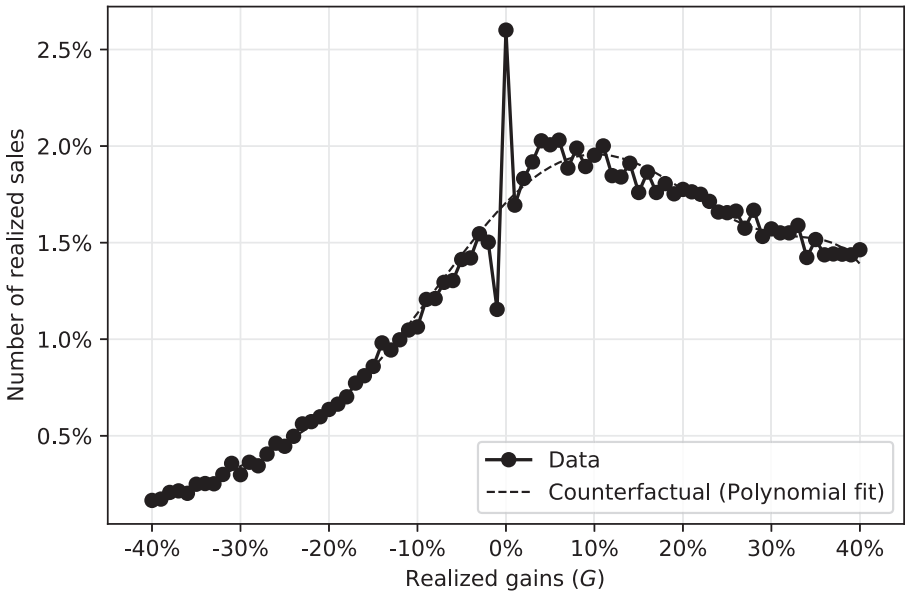


Figure A.8: Price-Volume Correlation

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.

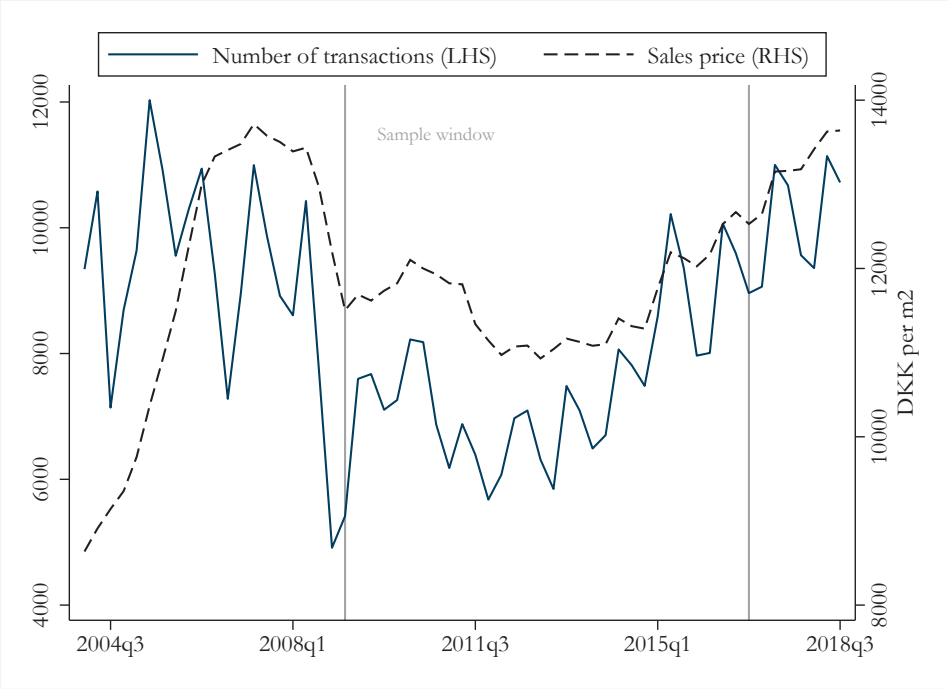


Figure A.9: Extensive Margin - Residualized

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.

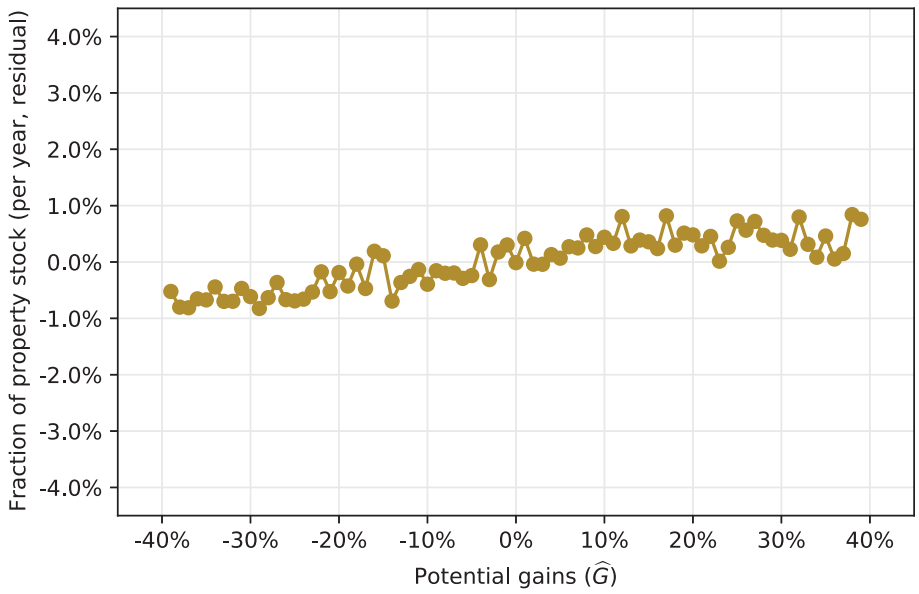


Table A.1: Literature Overview

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 ≥ 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 ≥ 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

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:

$$U^*(\widehat{G}) = \max_{\ell}(\alpha_0 - \alpha_1\ell)\left[\underbrace{\widehat{P} + \beta_0 + \beta_1\ell}_{P(\ell)} + \eta\underbrace{(\widehat{G} + \beta_0 + \beta_1\ell)}_{G(\ell)} + \theta\right] + (1 - \alpha_0 + \alpha_1\ell)\widehat{P}. \tag{3.1}$$

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

$$\alpha_0(1+\eta)\beta_1 - \alpha_1\left[\widehat{P} + (1+\eta)\beta_0 + \eta\widehat{G} + \theta - \widehat{P}\right] - 2(1+\eta)\alpha_1\beta_1\ell^* = 0, \tag{3.2}$$

which implies the optimal solution:

$$\begin{aligned}\ell^*(\widehat{G}) &= \frac{\alpha_0(1+\eta)\beta_1 - \alpha_1\left[(1+\eta)\beta_0 + \eta\widehat{G} + \theta\right]}{2(1+\eta)\alpha_1\beta_1} \\ &= \frac{1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\eta} - \frac{1}{\beta_1}\frac{\eta}{1+\eta}\widehat{G}\right).\end{aligned} \tag{3.3}$$

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

by:

$$U^*(\widehat{G}) = \max_{\ell}(\alpha_0 - \alpha_1\ell)\left[P(\ell) + \eta\underbrace{(\widehat{G} + \beta_0 + \beta_1\ell)}_{G(\ell)}\left(\lambda 1_{\widehat{G}+\beta_0+\beta_1\ell<0} + 1_{\widehat{G}+\beta_0+\beta_1\ell\geq 0}\right) + \theta\right] + (1 - \alpha_0 + \alpha_1\ell)\widehat{P}. \tag{3.4}$$

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 (**??**), "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 (**??**), but with a higher degree of overall reference dependence ($\lambda\eta$):

$$\ell_\lambda^*(\widehat{G}) = \frac{1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\lambda\eta} - \frac{1}{\beta_1}\frac{\lambda\eta}{1+\lambda\eta}\widehat{G}\right). \tag{3.5}$$

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

$$\ell^*(\widehat{G}) = \begin{cases} \frac{1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\eta}\right) - \frac{1}{2\beta_1}\frac{\eta}{1+\eta}\widehat{G}, & \text{if } \widehat{G} \geq \widehat{G}_0 \\ -\frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\widehat{G}, & \text{if } \widehat{G} \in (\widehat{G}_1, \widehat{G}_0) \\ \frac{1}{2}\left(\frac{\alpha_0}{\alpha_1} - \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1}\frac{\theta}{1+\lambda\eta}\right) - \frac{1}{2\beta_1}\frac{\lambda\eta}{1+\lambda\eta}\widehat{G}, & \text{if } \widehat{G} \leq \widehat{G}_1. \end{cases} \tag{3.6}$$

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 (**??**) 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.

2. Mapping Between Potential and Realized Gains

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

$$G(\widehat{G}) = \widehat{G} + \beta_0 + \beta_1 \ell^*(\widehat{G}). \tag{3.7}$$

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:

$$G(\widehat{G}) = \gamma_0 + \gamma_1 \widehat{G} \tag{3.8}$$

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

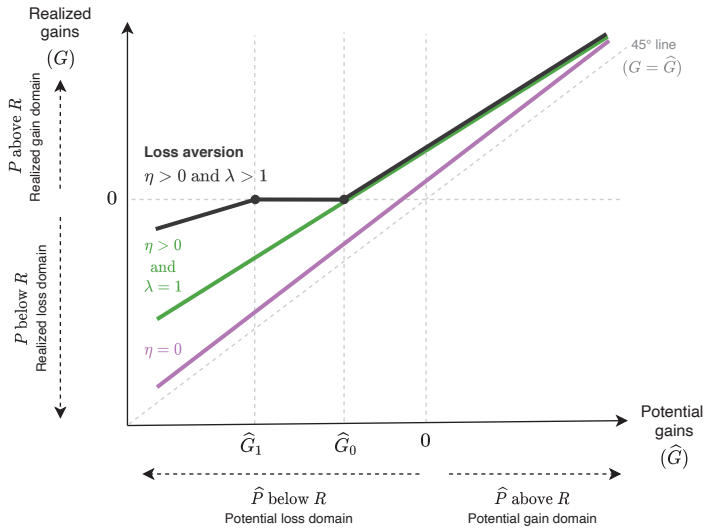
$$G(\widehat{G}) = \begin{cases} \gamma_0 + \gamma_1 \widehat{G} & \text{if } \widehat{G} > \widehat{G}_0, \\ 0 & \text{if } \widehat{G} \in [\widehat{G}_1, \widehat{G}_0], \\ \gamma_{\lambda,0} + \gamma_{\lambda,1} \widehat{G} & \text{if } \widehat{G} < \widehat{G}_1. \end{cases} \tag{3.9}$$

Here, we have:

$$\widehat{G}_0 = -\frac{\gamma_0}{\gamma_1} \text{ and } \widehat{G}_1 = -\frac{\gamma_{\lambda,0}}{\gamma_{\lambda,1}}, \tag{3.10}$$

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:

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



3. Extensive Margin Decision

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

$$U^*(\widehat{G}) = \widehat{P} + \left[\alpha_0 - \alpha_1 \ell^*(\widehat{G})\right]\left[\eta \widehat{G} + (1+\eta)\left(\beta_0 + \beta_1 \ell^*(\widehat{G})\right) + \theta\right] \tag{3.11}$$

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:

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

$$\widetilde{\theta}(\widehat{G}) = -\eta \widehat{G} - (1+\eta)\left(\beta_0 + \beta_1 \ell^*(\widehat{G})\right) \tag{3.12}$$

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

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

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.

5. The Role of Concave Demand

Figure **??** 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 **??** 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 ($\eta = 0$).

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:

$$\max_{L} \alpha(L - \widehat{P}) \left(\widehat{P} + \beta(L - \widehat{P}) + \theta \right) + \left(1 - \alpha(L - \widehat{P}) \right) \widehat{P}, \tag{3.13}$$

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:

$$\alpha_0\beta_1 - \alpha_1\beta_0 + 2\alpha_1\beta_1\widehat{P} - \alpha_1\theta = 2\alpha_1\beta_1 L^*, \tag{3.14}$$

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

which implies that:

$$L^* = \widehat{P} + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\theta}{2\beta_1} \tag{3.15}$$

The main message of equation (**??**) 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:

$$\max_L \alpha(L - \widehat{P}) \Big(\widehat{P} + \beta(L - \widehat{P}) + \eta(\widehat{P} + \beta(L - \widehat{P}) - R) + \theta\Big) + \Big(1 - \alpha(L - \widehat{P})\Big)\widehat{P}, \tag{3.16}$$

and the optimal solution is:

$$L^* = \widehat{P} + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\widehat{P} - R) + \theta}{2(1+\eta)\beta_1}, \tag{3.17}$$

so these conclusions carry through to a setup with reference dependence. Table **??** validates these observations in the data.

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 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:

$$\widehat{P} = R + \delta_t - \delta_s, \tag{3.18}$$

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:

$$L^* = R + \delta_t - \delta_s + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\cancel{R} + \delta_t - \delta_s - \cancel{R}) + \theta}{2(1+\eta)\beta_1}, \tag{3.19}$$

and therefore:

$$L^* - \widehat{P} = \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\delta_t - \delta_s) + \theta}{2(1+\eta)\beta_1}. \tag{3.20}$$

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:

$$L^* = \begin{cases} \widehat{P} + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\widehat{P}-R)+\theta}{2(1+\eta)\beta_1}, & \text{if } \widehat{P} - R > \widehat{G}_0 \\ \widehat{P} - \frac{\beta_0}{\beta_1} + \frac{\widehat{P}-R}{\beta_1}, & \text{if } \widehat{P} - R \in [\widehat{G}_1, \widehat{G}_0] \\ \widehat{P} + \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta\lambda(\widehat{P}-R)+\theta}{2(1+\lambda\eta)\beta_1}, & \text{if } \widehat{P} - R < \widehat{G}_0 \end{cases} \tag{3.21}$$

and with repeat sales:

$$L^* - \widehat{P} = \begin{cases} \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta(\delta_t-\delta_s)+\theta}{2(1+\eta)\beta_1}, & \text{if } \widehat{P} - R > \widehat{G}_0 \\ -\frac{\beta_0}{\beta_1} + \frac{\delta_t-\delta_s}{\beta_1}, & \text{if } \widehat{P} - R \in [\widehat{G}_1, \widehat{G}_0] \\ \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{2\alpha_1\beta_1} - \frac{\eta\lambda(\delta_t-\delta_s)+\theta}{2(1+\lambda\eta)\beta_1}, & \text{if } \widehat{P} - R < \widehat{G}_0 \end{cases} \tag{3.22}$$

All conclusions from above also carry through to this case.

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

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

However, if R enters the seller's property value estimate (denoted by $\widehat{P}^{Seller}$ below) differently from how it enters $\widehat{P}$ we can distinguish between three cases: First, the seller correctly uses R when valuing the property, but we don't. This is possible, but we believe unlikely, given that our results hold strongly and robustly across a large number of alternative models for $\widehat{P}$, including repeat sales. But even if our hedonic model may miss relevant price variation coming from R, this only affects estimated effects in terms of potential gains $\widehat{G}$, and such a model cannot be reconciled with the evidence of excess bunching in *realized* gains G exactly around observed prices $P = R$. Second, sellers misperceive the importance of R, i.e. they

weight it differently: $\widehat{P}^{seller} = (1-\kappa)\widehat{P} + \kappa R$. The optimal listing premium function is then given by $\ell^* = f((\eta+\kappa)(\widehat{P}-R)) = f((\eta+\kappa)\widehat{G})$. In this case, reference dependence and irrational over-weighting of R have observationally equivalent effects on the *average slope* of the listing premium with respect to potential gains, but such a model of misspecified seller beliefs cannot explain the *variation* in slopes ("kinks"), and the bunching of realized prices around the reference point. Third, if both the econometrician and the seller incorrectly use R (and in different ways), we still extract the behaviour of interest, albeit potentially with considerable noise. More importantly, such a version of the model is also unable to explain the observed bunching of prices around the reference point.

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	$\ell^*_{k\in\{High,Med,Low\}}(\widehat{G}, \mathbf{x})$

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: $\widehat{\mathbf{x}}^s = \kappa\left(F(\mathbf{x}^s), \mathbf{x}^s_0\right)$. 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 **??**, 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 $\mathbf{x}^s_{i=1,...,N}$ and evaluating the function $F(\mathbf{x}^s_{i=1,...,N})$ at each draw, choosing as a starting point $\mathbf{x}^s_0$ 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 **??** for each moment separately, and for the full model jointly.

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 **??** shows the average interest rate profile in the data, for different levels of the LTV ratio. Figure **??** 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.

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

[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.

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 **??** we show the price-volume correlation in Denmark over a broader period containing our sample period. The plot looks very similar to the broad patterns observed in the US.

2. Property Listings Data

Property listings are provided to us by RealView (http://realview.dk/en/), who attempt to comprehensively capture all electronic listings of owner-occupied housing in Denmark. RealView data cover the universe of listings in the portal www.boligsiden.dk, in addition to additional data collected directly from brokers. The data include private (i.e., open to only a selected set of prospective buyers) electronic listings, but do not include off-market property transactions, i.e., direct private transfers between households. Of the total number of cleaned/filtered sale transactions in the official property registers (described below), 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.

[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.

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.

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.

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

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

[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 **??** 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.

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 **??** documents the cleaning and sample selection process from the raw listings data to the final matched data.

6. Data Citations

- FinansDanmark (2019a). "Boligmarkedsstatistikken, Ejendomspriser pa boligmarkedet, BM010, 2002-2018". https://rkr.statistikbank.dk/201 (accessed in 2019)
- FinansDanmark (2019b). "Boligmarkedsstatistikken, Bevaegelser pa boligmarkedet, BM020, 2002-2018". https://rkr.statistikbank.dk/201 (accessed in 2019)
- FinansDanmark (2019c). "Boligmarkedsstatistikken, Salgstider pa 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)

[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).

- Statistics Denmark (2021c). "Boligtaellingen (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/ForskningVariabe
ster/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

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)

[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

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.

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 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 (kursskaering) 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

[13] https://www.bolius.dk/omkostninger-ved-at-koebe-bolig-18145
[14] https://www.mybanker.dk/sammenlign/bolig/bidragssatser/

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 **??** for an overview of fees and Figure **??** for an illustration of average costs over LTV.

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

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

between 200 and 500 bp over the mortgage rate. For reference, see categories *DNRNURI* and *DNRNUPI* in the Danmarks Nationalbank's statistical data bank.

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 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.

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 **??** 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 **??** and **??** 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 **??** 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.

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.

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 **??**.

[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.

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 **??** 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 **??**.

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 **??**. 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

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

this in our robustness checks by including these variables in the pricing model.

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 **??** 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 **??** 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 **??** and **??** show that the listing premium over gains and home equity relationships, as well as the pattern for demand concavity are preserved when the hedonic price is predicted out-of-sample.

F. Functional Form of Down-Payment Constraints

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

[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 **??**. 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.

issue, we verify the implications of the model using a concave version of the penalty function:

$$\kappa(P(\ell)) = \begin{cases} \mu(\gamma - H(\ell))^{1/2}, \text{ if } H(\ell) < \gamma \\ 0, \text{ if } H(\ell) \geq \gamma \end{cases}. \tag{3.23}$$

Figure **??** 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 **??** 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.

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 **??** 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 **??** 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 **??** 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 **??** 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 **??**, 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:

$$\alpha(\ell) = A + \frac{K - A}{(C + Qe^{-B\ell})^{1/\nu}}. \tag{3.24}$$

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 **??** 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 **??** 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 **??**, 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.

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:

$$\underbrace{L_{ist} - \widetilde{P}_{it}}_{\widetilde{\ell}_{ist}} = \mu_0 + m(\underbrace{\widetilde{P}_{it} - R_{is}}_{\widetilde{G}_{ist}}) + \varepsilon_{ist}. \tag{3.25}$$

When bringing this model to the data, the problem is that the "true" $\widetilde{P}$ is imperfectly observed.

[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 $R_{is} \equiv P_{is}$.

Let $\widehat{P}$ be the "feasible" valuation model, and ξ_{it} the potentially time-varying estimation error:

$$\widetilde{P}_{it} = \widehat{P}_{it} + \xi_{it}. \tag{3.26}$$

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

$$\begin{aligned} \widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\ &= \widetilde{G}_{ist} - \xi_{it}, \end{aligned} \tag{3.27}$$

$$\begin{aligned} \widehat{\ell}_{ist} &= L_{ist} - \widehat{P}_{it} \\ &= \widetilde{\ell}_{ist} + \xi_{it}. \end{aligned} \tag{3.28}$$

Assuming that the shocks ε_{it} and ξ_{it} are uncorrelated with "true" potential gains $\widetilde{G}_{ist}$, the estimated coefficient $\widehat{m}$ is then given by:

$$\begin{aligned} \widehat{m} &= \frac{Cov(\widehat{G}_{ist}, \widehat{\ell}_{ist})}{Var(\widehat{G}_{ist})} = \frac{Cov(\widetilde{G}_{it} - \xi_{it}, \widetilde{\ell}_{ist} + \xi_{it})}{Var(\widetilde{G}_{ist} - \xi_{it})} = \frac{Cov(\widetilde{G}_{it} - \xi_{ist}, \mu_0 + m\widetilde{G}_{it} + \varepsilon_{it} + \xi_{it})}{Var(\widetilde{G}_{ist} - \xi_{it})} \\ &= m \underbrace{\frac{Var(\widetilde{G}_{ist})}{Var(\widetilde{G}_{ist}) + Var(\xi_{it})}}_{\substack{\text{Classical measurement} \\ \text{error}}} - \underbrace{\frac{Var(\xi_{it})}{Var(\widetilde{G}_{ist}) + Var(\xi_{it})}}_{\substack{\text{Over-estimation bias} \\ (\text{because } m < 0)}}. \end{aligned} \tag{3.29}$$

Equation (**??**) 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.

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 qual-

ity[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:

$$P_{it} = \underbrace{X_i\beta + \delta_t + \nu_{it}}_{\substack{\text{"True" hedonic value} \\ \text{of property } (=\widetilde{P}_{it})}} + \underbrace{\omega_{it}}_{\substack{\text{Idiosyncratic over-} \\ \text{or under-payment}}}, \tag{3.30}$$

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:

$$\widetilde{P}_{it} = X_i\beta + \delta_t + \nu_{it}. \tag{3.31}$$

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:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \nu_{it} = 0, \text{ and } \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \omega_{it} = 0. \tag{3.32}$$

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:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \omega_{it} = 0. \tag{3.33}$$

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}

[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.

over time directly inform which model of prices should be preferred, as discussed further below.

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.

Model Descriptions

1. Standard hedonic regression:

    a. Time-invariant observables

    $$\widehat{P}_{it} = X_i\beta + \delta_t. \tag{3.34}$$

    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

    $$\widehat{P}_{it} = X_{it}\beta + \delta_t. \tag{3.35}$$

    Equation **??** 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:

$$\widehat{P}_{it} = X_i\beta + \delta_t + \bar{\nu}_{it_{\tau<t}} + \bar{\omega}_{it_{\tau<t}}, \tag{3.36}$$

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

$$\widehat{P}_{it} = X_i\beta + \delta_t + \nu_{is} + \omega_{is}, \tag{3.37}$$

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):

$$\widehat{P}_{it}^{level} = R_{is}^{level} \cdot \frac{d_t}{d_s}. \tag{3.38}$$

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

$$\begin{aligned}
\widehat{P}_{it} &= R_{is} + \delta_t - \delta_s. \\
&= X_i\beta + \delta_s + \nu_{is} + \omega_{is} + \delta_t - \delta_s \\
&= X_i\beta + \delta_t + \nu_{is} + \omega_{is},
\end{aligned} \tag{3.39}$$

which is equivalent to equation (??).

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

$$P_{it} = X_{it}\beta + \delta_t + \nu_{it} + \omega_{it}. \tag{3.40}$$

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

$T = 2$:

$$\nu_{is} + \omega_{is} = \underbrace{R_{is} - \widehat{P}_{is}}_{\text{Hedonic model pricing residual at time } s} \tag{3.41}$$

And more generally:

$$\bar{\nu}_{it_{\tau<T}} + \bar{\omega}_{it_{\tau<T}} = \underbrace{\frac{1}{T-1}\sum_{\tau<T} R_{i\tau} - \widehat{P}_{i\tau}}_{\substack{\text{Average of the hedonic model pricing} \\ \text{residuals across repeat sales for which } \tau < T}} \tag{3.42}$$

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

$$\widehat{P}_{it} = X_{it}\beta + \delta_t + \frac{\nu_{is} + \nu_{is'}}{2} + \frac{\omega_{is} + \omega_{is'}}{2}, \tag{3.43}$$

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

$$\begin{aligned}
\widehat{P}_{it} &= \underbrace{X_{it}\beta + \delta_t}_{\text{Hedonic model with time-varying observables}} + \frac{1}{T-1}\sum_{\tau<T} R_{i\tau} - \widehat{P}_{i\tau} \\
&= X_{it}\beta + \delta_t + \frac{R_{is} - \widehat{P}_{is} + R'_{is'} - \widehat{P}_{is'}}{2} \\
&= X_{it}\beta + \delta_t + \frac{X_{is}\beta + \delta_s + \nu_{is} + \omega_{is} - (X_{is}\beta + \delta_s) + X_{is'}\beta + \delta_{s'} + \nu_{is'} + \omega_{is'} - (X_{is'}\beta + \delta_{s'})}{2} \\
&= X_{it}\beta + \delta_t + \frac{\nu_{is} + \nu_{is'}}{2} + \frac{\omega_{is} + \omega_{is'}}{2}.
\end{aligned} \tag{3.44}$$

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 **??**),

$$\widehat{P}_{it} = X_i\beta + \delta_t + \overline{r}_{it}, \tag{3.45}$$

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.

Model Estimation

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

1. The standard hedonic regression:

$$\widehat{P}_{it} = X_i\beta + \delta_t, \tag{3.46}$$

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

$$\begin{aligned}
\widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\
&= X_i\beta + \delta_t - (X_i\beta + \delta_s + \nu_{is} + \omega_{is}) \\
&= \delta_t - \delta_s - \nu_{is} - \omega_{is} \\
&= \widetilde{G}_{ist} - \nu_{it}.
\end{aligned} \tag{3.47}$$

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

$$\begin{aligned}
\widehat{m} &= \frac{Cov(\widehat{G}_{ist}, \widehat{\ell})}{Var(\widehat{G}_{ist})} \\
&= \frac{Cov(\tilde{G}_{ist} - \nu_{it}, \nu_{it} + \tilde{\ell}_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\
&= \frac{Cov(\tilde{G}_{ist} - \nu_{it}, \nu_{it} + \mu_0 + m\widetilde{G}_{ist} + \epsilon_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\
&= m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} - \underbrace{\frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\substack{\text{Over-estimation bias} \\ \text{(because } m < 0\text{)}}}
\end{aligned} \tag{3.48}$$

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

$$\widehat{P}_{it} = X_i\beta + \delta_t + \nu_{is} + \omega_{is}. \tag{3.49}$$

Recall, the true gain is:

$$\begin{aligned}\tilde{G}_{ist} &= \tilde{P}_{it} - R_{is} \\ &= X_i\beta + \delta_t + \nu_{it} - (X_i\beta + \delta_s + \nu_{is} + \omega_{is}) \\ &= \delta_t - \delta_s + \nu_{it} - \nu_{is} - \omega_{is}\end{aligned} \tag{3.50}$$

The potential gain is:

$$\begin{aligned}\widehat{G}_{ist} &= \widehat{P}_{it} - R_{is} \\ &= \delta_t - \delta_s \\ &= \tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}\end{aligned} \tag{3.51}$$

Hence the coefficient estimate is:

$$\begin{aligned}\widehat{m} &= \frac{Cov(\widehat{G}_{ist}, \widehat{\ell})}{Var(\widehat{G}_{ist})} \\ &= \frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \nu_{it} + \tilde{\ell}_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\ &= \frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \nu_{it} + \mu_0 + m\tilde{G}_{ist} + \epsilon_{ist})}{Var(\tilde{G}_{ist} - \nu_{it})} \\ &= m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \nu_{is} + \omega_{is}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} - \underbrace{\frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\nu_{is}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0\text{, bias}}\end{aligned} \tag{3.52}$$

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

$$\widehat{m} = m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \bar{\nu}_{it_{\tau<T}} + \bar{\omega}_{it_{\tau<T}}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} - \underbrace{\frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\bar{\nu}_{it_{\tau<T}}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0, \text{ bias}} \tag{3.53}$$

   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 (**??**) 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:

$$\widehat{P}_{it} = X_i\beta + \delta_t + \overline{r}_{it}, \tag{3.54}$$

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

$$\widehat{G}_{ist} = \widetilde{G} - (\nu_{it} - \overline{r}_{it}), \tag{3.55}$$

and hence the estimated coefficient is

$$\widehat{m} = m \underbrace{\frac{Cov(\tilde{G}_{ist} - \nu_{it} + \bar{r}_{it}, \tilde{G}_{ist})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{Classical measurement error}} - \underbrace{\frac{Var(\nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})} + \frac{Cov(\bar{r}_{it}, \nu_{it})}{Var(\tilde{G}_{ist}) + Var(\nu_{it})}}_{\text{if } \neq 0\text{, bias}} \tag{3.56}$$

We can think of the inclusion of $\bar{r}_{it}$ 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 $\bar{r}_{it}$ for ν_{it}. If the renovations variable perfectly captures time-varying unobserved quality, then this model also generates a bias-free approach.

Discussion

We implement these different models and compare them below.[24] Table **??** provides an overview of all the models that we implement. Figures **??**, **??** and **??** 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 **??** provides a quantitative comparison of the main models, by estimating summary statistics of the moment relationships that we use to estimate our structural model. In particular, we estimate piecewise linear slopes for the listing premium over negative (row 2) and positive (row 3) potential gains, and for the probability of sale within six months for positive listing premia (row 6), using the same support as we use for the individual moments. Our main models are: the baseline model (Ia), the baseline model augmented with lagged 1-year renovation expenses (Ib), a simple repeat sales model based on shire-level house price changes (II), the combined baseline hedonic and repeat sales model with renovation expenses and the lagged pricing residual for one repeat sale (IIIa), and the equivalent for any number of repeat sales and average lagged residuals (IIIb). Table **??**

[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. $T \geq 2$).

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 **??**, 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 **??** 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 **??**, 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.

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 **??** 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.

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:

$$L_{ist} - \widetilde{P}_{it} = \mu_0 + f(\eta, \lambda, \underbrace{(\widetilde{P}_{it} - R_{is})}_{(\widetilde{G})}) + \varepsilon_{it}, \tag{3.57}$$

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:

$$\underbrace{L_{ist} - \widetilde{P}}_{\widetilde{\ell}_{ist}} = \mu_0 + m_0 \underbrace{(\widetilde{P}_{it} - R_{is})^-}_{\widetilde{G}^-_{ist}} + m_1 \underbrace{(\widetilde{P}_{it} - R_{is})^+}_{\widetilde{G}^+_{ist}} + \varepsilon_{it}, \tag{3.58}$$

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:

$$\begin{aligned}
\widehat{m}_0 &= \frac{Cov(\widehat{G}^-_{ist}, \widehat{\ell}_{ist})}{Var(\widehat{G}^-_{ist})} = \frac{Cov((\widetilde{G}_{it} - \xi_{it})^-, \widetilde{\ell}_{ist} + \xi_{it})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)} \qquad (3.59)\\
&= \frac{Cov((\widetilde{G}_{it} - \xi_{ist})^-, \mu_0 + m_0\widetilde{G}^-_{it} + m_1\widetilde{G}^+_{it} + \varepsilon_{it} + \xi_{it})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)}\\
&= m_0 \underbrace{\frac{Cov((\widetilde{G}_{ist} - \xi_{ist})^-, \widetilde{G}^-_{it})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)}}_{\substack{\text{Measurement}\\ \text{error/Smoothing}}} - \underbrace{\frac{Cov((\widetilde{G}_{ist} - \xi_{ist})^-, \xi_{ist})}{Var((\widetilde{G}_{ist} - \xi_{it})^-)}}_{\substack{\text{Over-estimation bias}\\ \text{(because } m < 0\text{)}}} \qquad (3.60)
\end{aligned}$$

Equation (**??**) 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.

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 **??**). We also show that the distribution of the running variable is smooth around the threshold (no bunching in *potential* gains) (Figure **??**).

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

$$\tau = \lim_{v \to \overline{v}_+} \left. \frac{dE[\ell_{it}|V_{it} = v]}{dv} \right|_{V_{it}=v} - \lim_{v \to \overline{v}_-} \left. \frac{dE[\ell_{it}|V_{it} = v]}{dv} \right|_{V_{it}=v}, \tag{3.61}$$

[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):

$$E[\ell_{it}|V_{it} = v] = \kappa_m + \kappa_t + \text{,}\mathbf{X_{it}} + \left[\sum_{p=1}^{\overline{p}} \gamma_p(\nu - \overline{\nu})^p + \nu_p(v - \overline{v})^p \mathbb{1}_{V \geq \overline{v}}\right]. \tag{3.62}$$

$$\text{where } |v - \overline{v}| < b. \tag{3.63}$$

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 **??** reports results across different bandwidths within which we fit a local linear function on each side of the threshold. Figure **??** 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.

[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 **??**.

6. Demand Concavity

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

$$\alpha_{ist}(\widetilde{\ell}) = \mu_1 + n_0(\underbrace{L_{ist} - \widetilde{P}_{it}}_{\widetilde{\ell}^-_{ist}})^- + n_1(\underbrace{L_{ist} - \widetilde{P}_{it}}_{\widetilde{\ell}^+_{ist}})^+ + \epsilon_{ist}, \tag{3.64}$$

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

$$\widehat{\ell}_{ist} = L_{ist} - (\widehat{P}_{it} + \xi_{it}) = \widetilde{\ell}_{ist} + \xi_{it}, \tag{3.65}$$

the feasible regression is

$$\alpha_{ist}(\widehat{\ell}) = \mu_1 + n_0(\widetilde{\ell}_{ist} + \xi_{it})^- + n_1(\widetilde{\ell}_{ist} + \xi_{it})^+ + \epsilon_{ist}, \tag{3.66}$$

with estimated main coefficient of interest

$$\begin{aligned} \widehat{n}_1 &= \frac{\alpha_{ist}(\widehat{\ell}), (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+} = \frac{Cov(\mu_1 + n_0\widetilde{\ell}^-_{ist} + n_1\widetilde{\ell}^+_{ist} + \epsilon_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+} \\ &= n_1 \underbrace{\frac{Cov(\widetilde{\ell}^+_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+}}_{\text{Measurement error}} + \underbrace{\frac{Cov(\epsilon_{ist}, (\widetilde{\ell}_{ist} + \xi_{it})^+)}{Var(\widetilde{\ell}_{ist} + \xi_{it})^+}}_{0 \text{ if } \xi_{it} \perp \epsilon_{it}} \end{aligned} \tag{3.67}$$

Equation **??** 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.

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 $\tilde{\ell}_{ist}$ is equal to:

$$\widetilde{\ell} = \ell - \zeta, \tag{3.68}$$

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:

$$\widetilde{P} = \widehat{P} - \zeta, \widetilde{G} = \widehat{G} - \zeta, \widetilde{H} = \widehat{H} - \zeta. \tag{3.69}$$

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^2_ζ is then given by the variance $\sigma^2_{\Delta\varepsilon}$ of $\varepsilon - \varepsilon_L$. Now, $\sigma^2_{\Delta\varepsilon}$ 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 **??** give an overview of the effects of the de-biasing procedure, adjusting the empirical moments as indicated in equations (**??**) and (**??**) 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

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

Lastly, in Figure **??**, 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 **??**, 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).

2. Bunching of Listing Prices around Nominal Purchase Price

Figure ?? 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 ??, 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

1. Liquid Financial Wealth

Figure ?? 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.

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 **??** 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 **??** and Table **??** 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 **??** 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.

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.

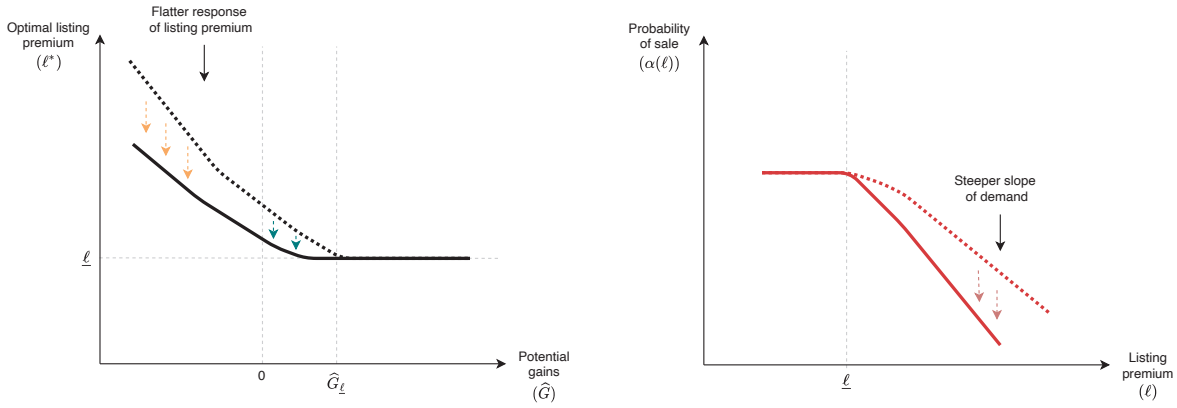


Figure L.2: Cost Of Borrowing

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.

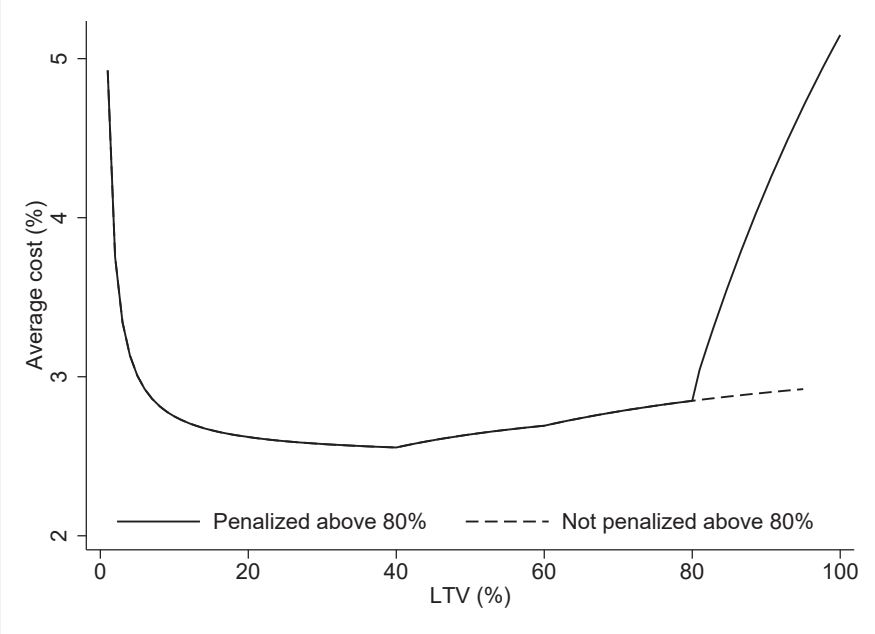


Figure L.3: Listing premia by home equity: Alternative parameterization

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 $\mu = 1.125$.

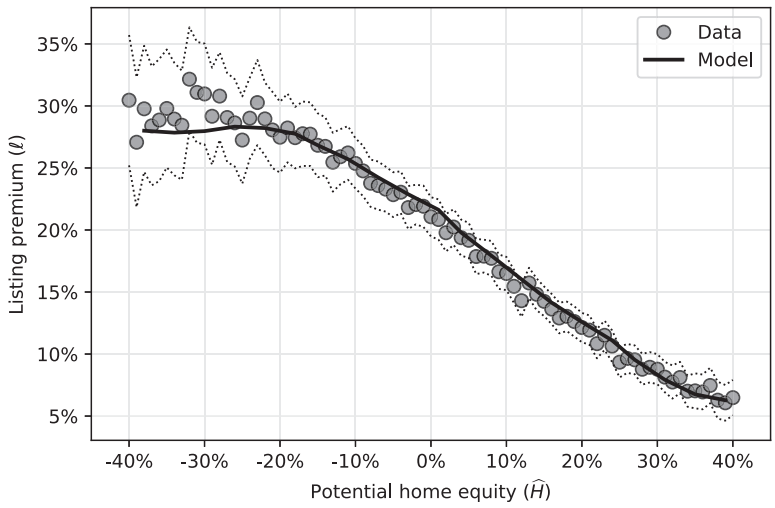


Figure L.4: Actual vs. Predicted Price of Sold Properties

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

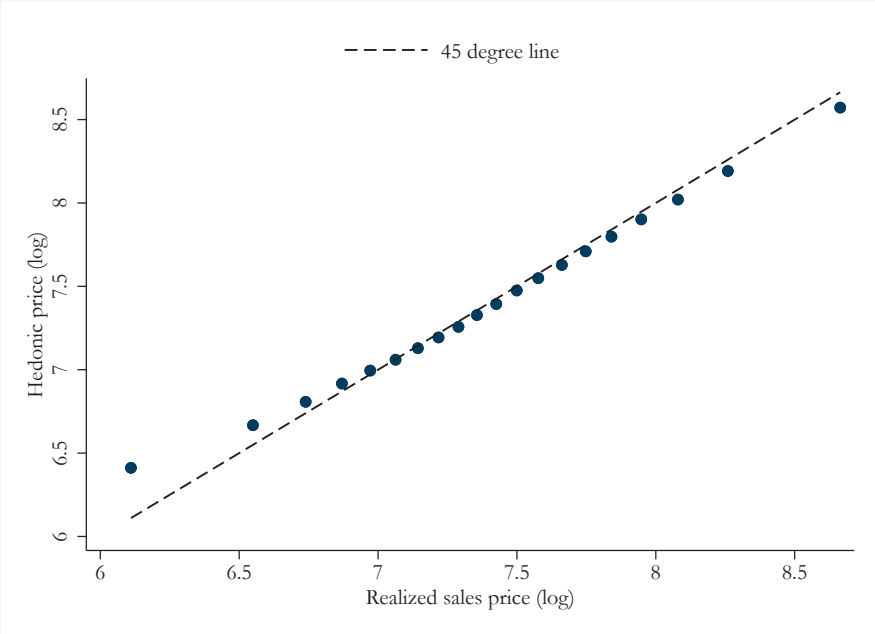
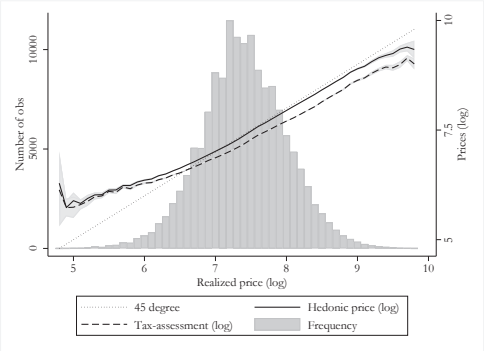


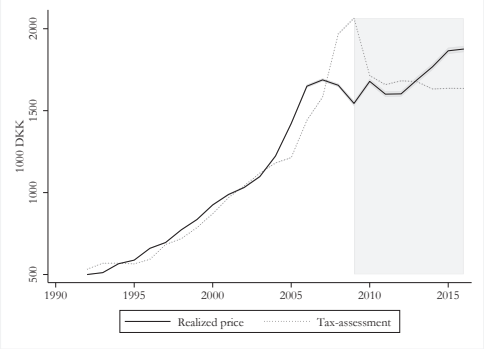
Figure L.5: Accuracy of Tax-Assessed Value

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

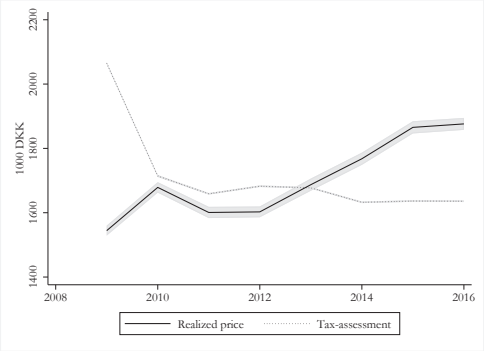
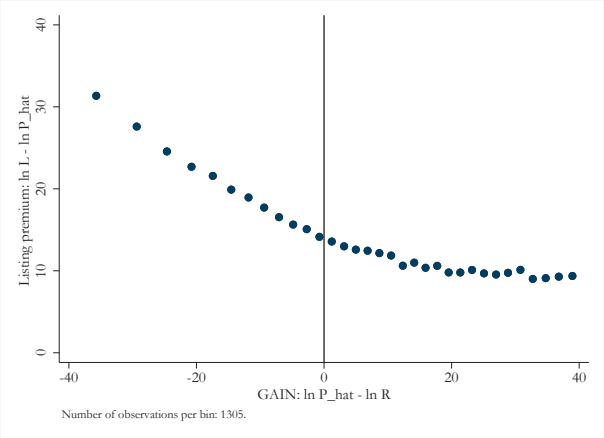


Figure L.6: Listing Premia across Potential Gains and Tax-Assessed Value

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

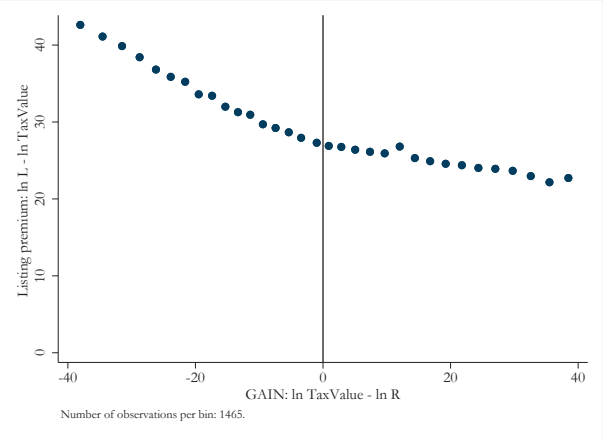
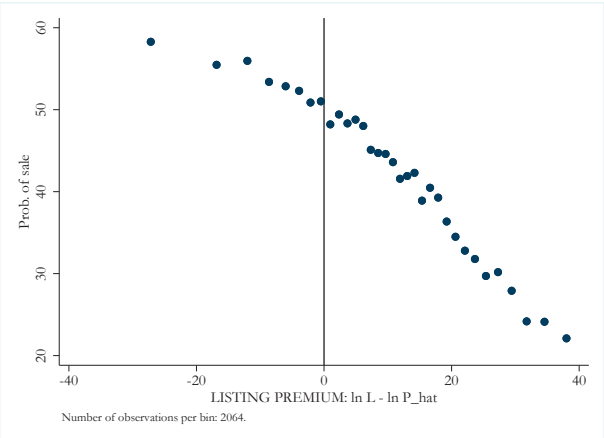


Figure L.7: Probability of Sale by Listing Premia (Concave Demand) and Tax-Assessed Value

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

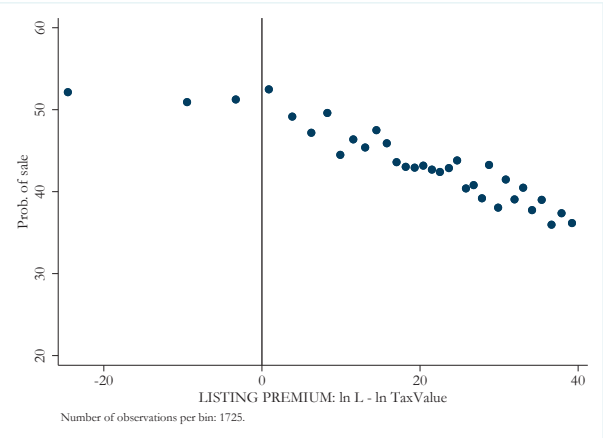
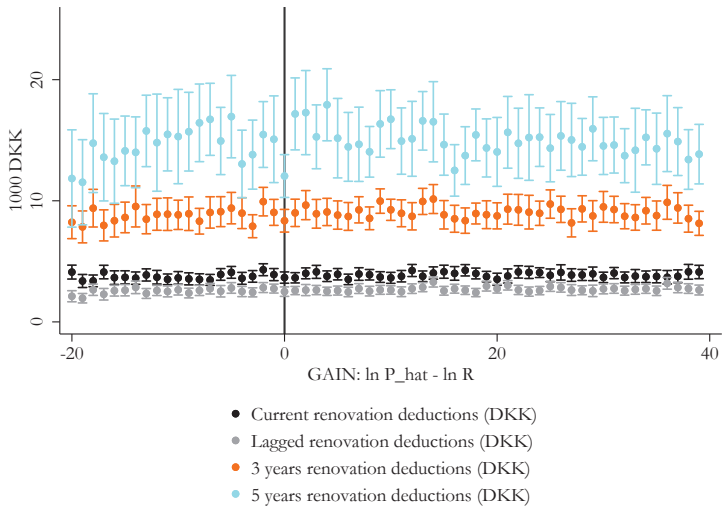


FIGURE L.8: RENOVATION EXPENSES ACROSS POTENTIAL GAINS AND LISTING PREMIA

$\widehat{G}$ 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 $\widehat{\ell}$

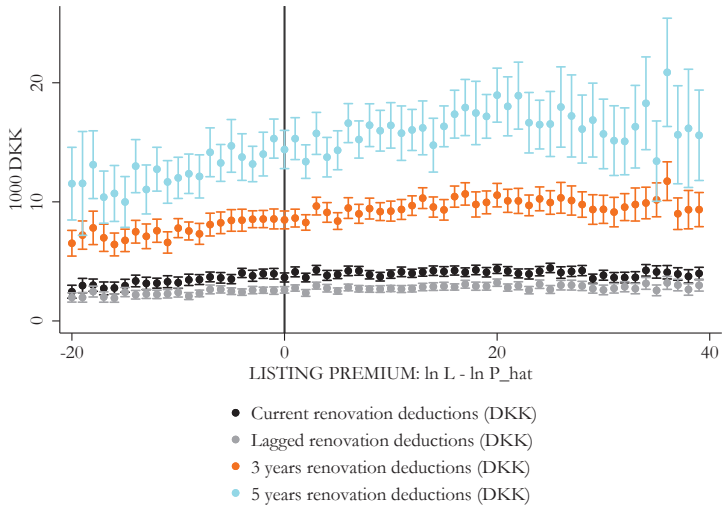
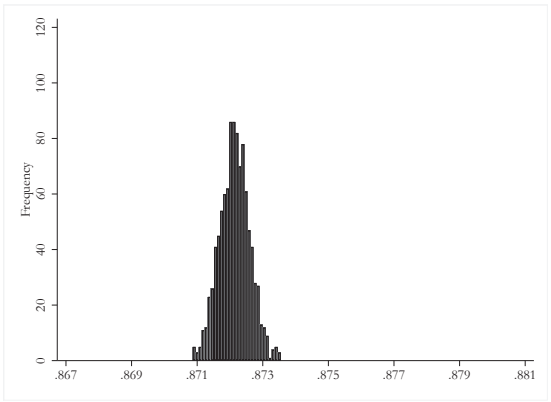


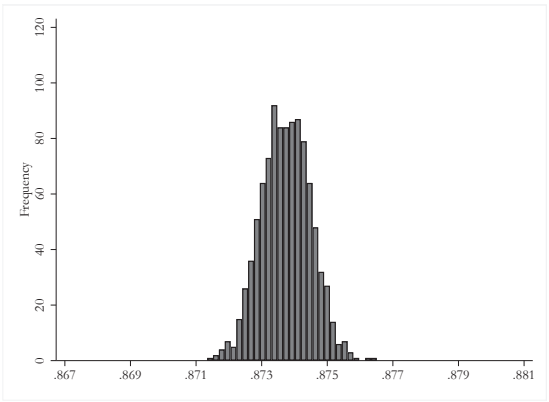
Figure L.9: Distribution of R^2s from Out-of-Sample Estimation of the Hedonic Model

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

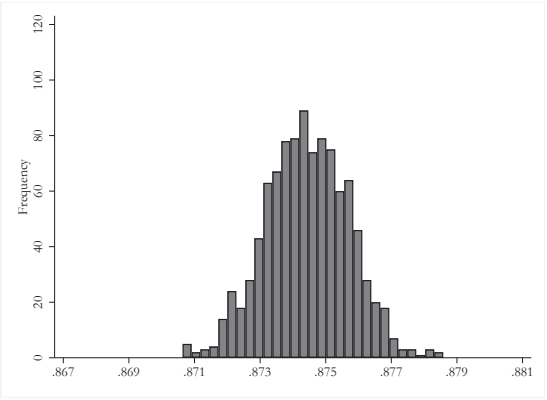
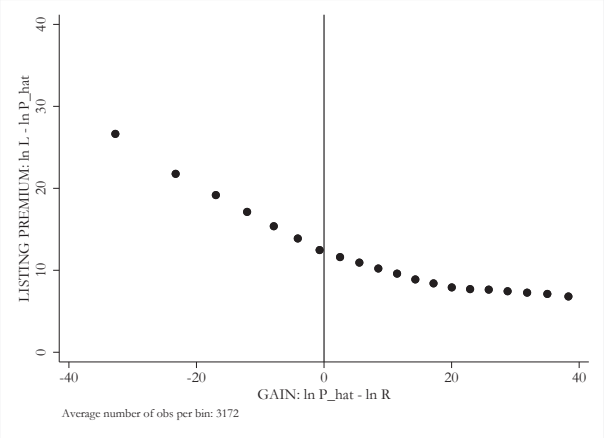


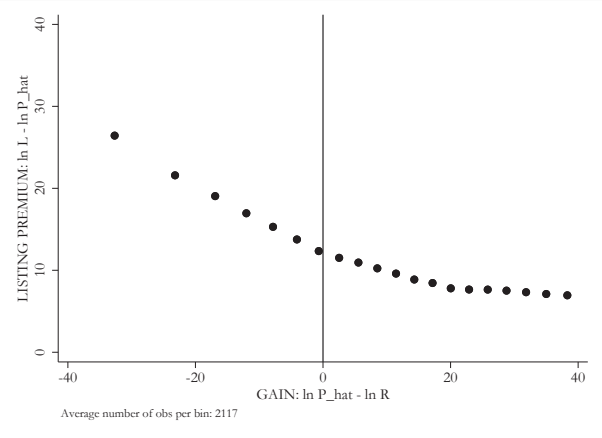
Figure L.10: Listing Premia across Potential Gains - Out-of-Sample Predictions

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

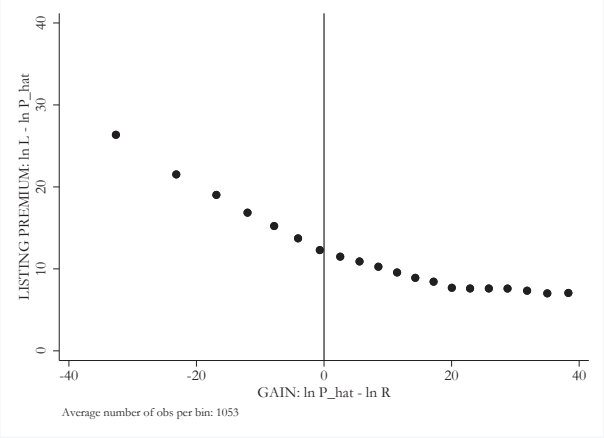
(a) 25 percent out of sample



(b) 50 percent out of sample



(c) 75 percent out of sample



(d) Main data, only sold properties

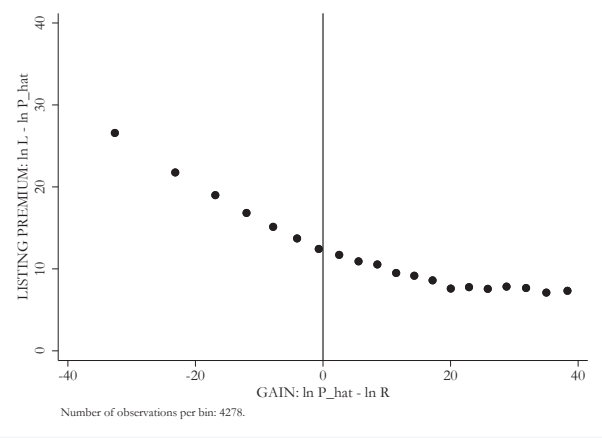
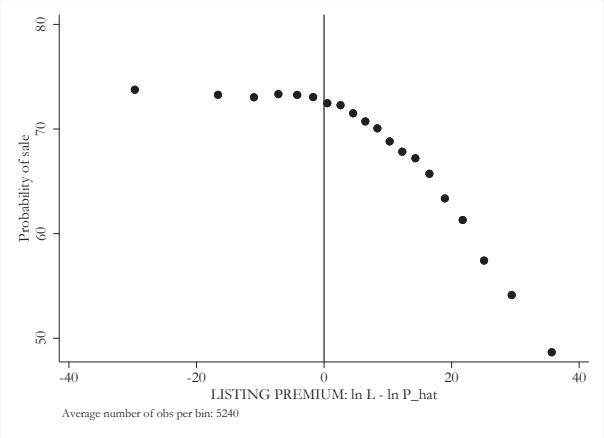


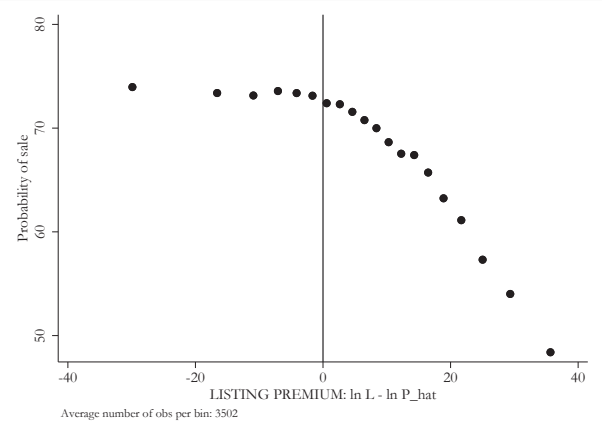
Figure L.11: Probability of Sale by Listing Premium (Concave Demand) - Out-of-Sample Predictions

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.

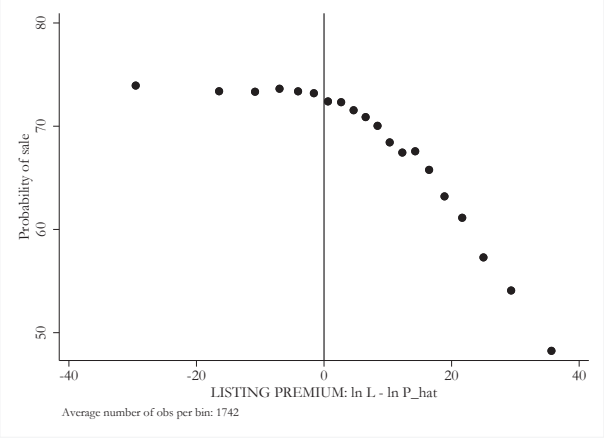
(a) 25 percent out of sample



(b) 50 percent out of sample



(c) 75 percent out of sample



(d) Main data, only sold properties

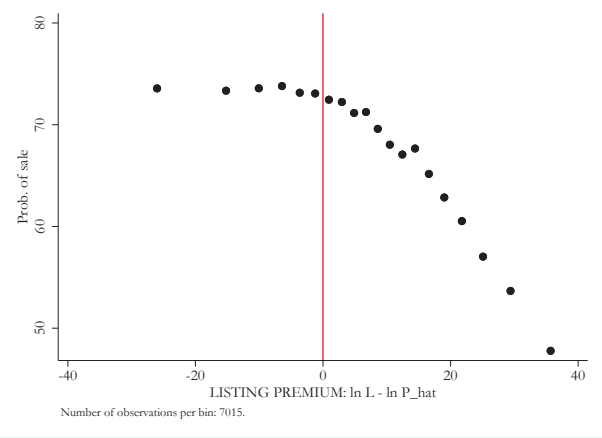
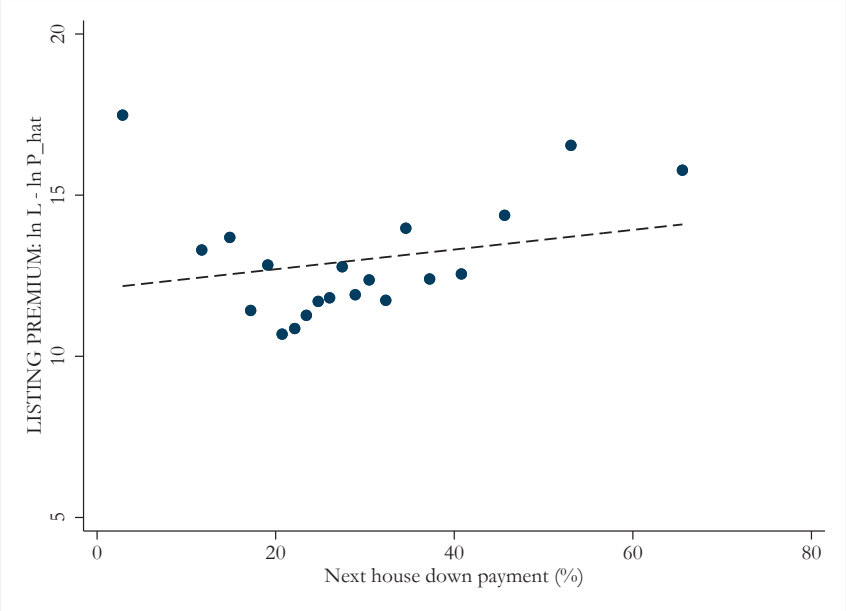


Figure L.12: Listing Premia and Down Payment, and Current and Next House Price

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

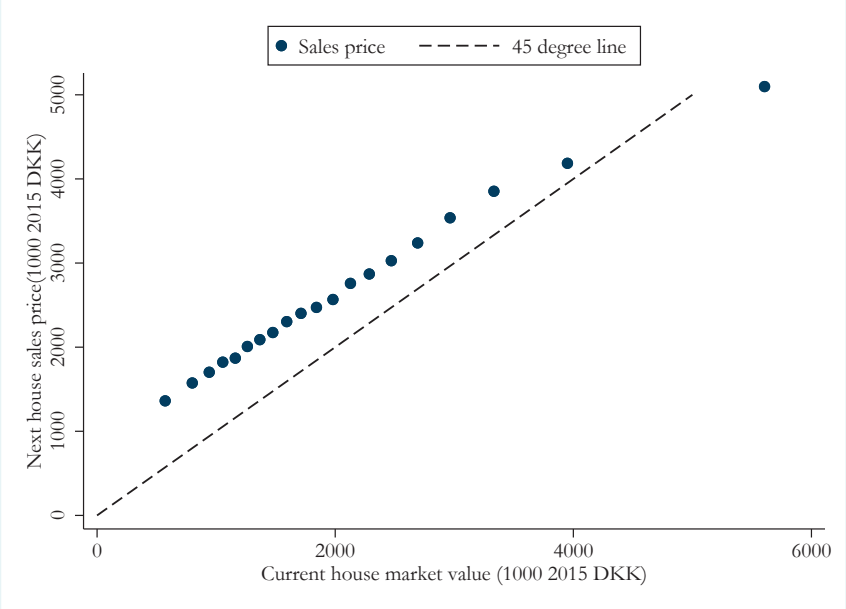


Figure L.13: Time-On-the-Market and Retraction Rate

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

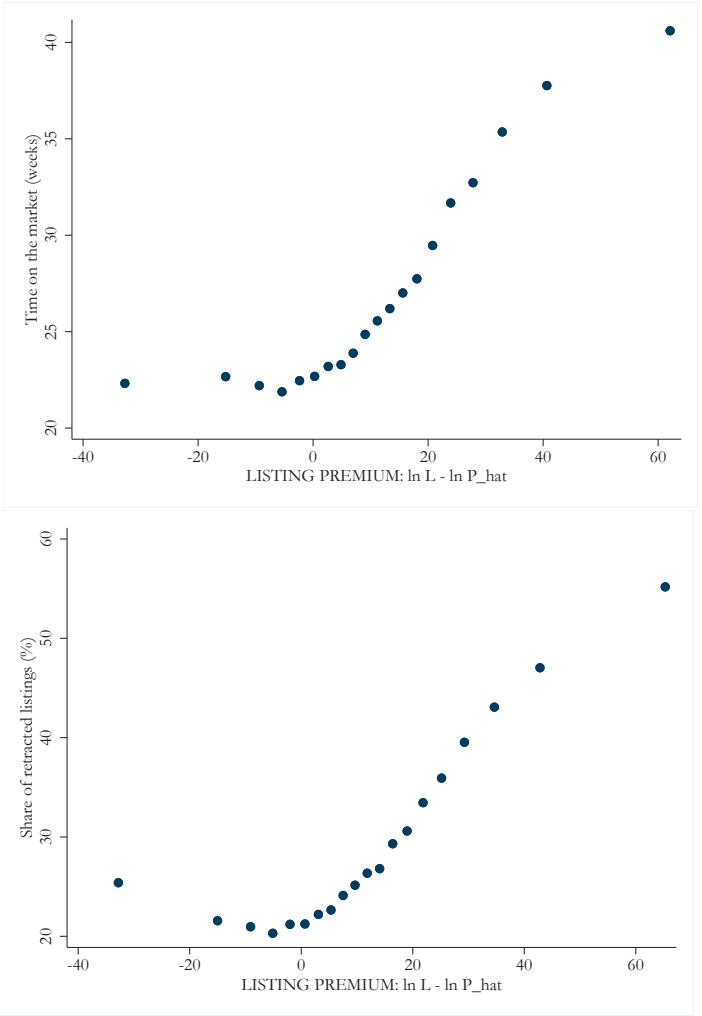


Figure L.14: Illustration of Homogeneity of Housing Stock for IV Estimation

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

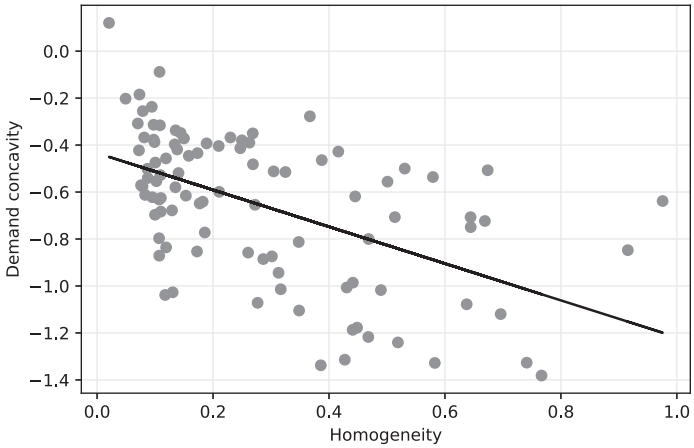


FIGURE L.15: REGIONAL VARIATION IN DEMAND CONCAVITY, LISTING PREMIUM-GAIN SLOPE AND HOUSING STOCK HOMOGENEITY

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

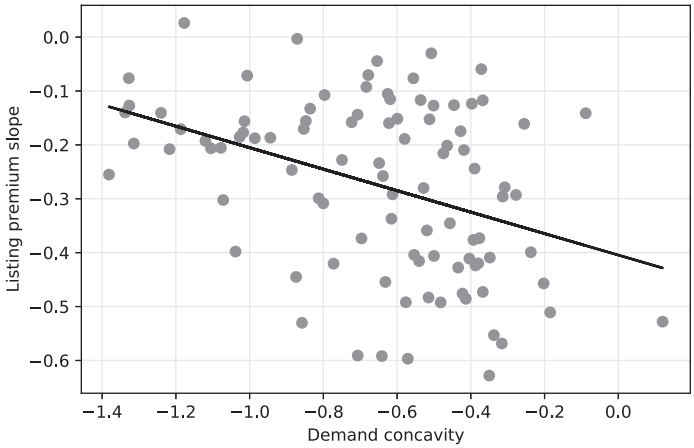
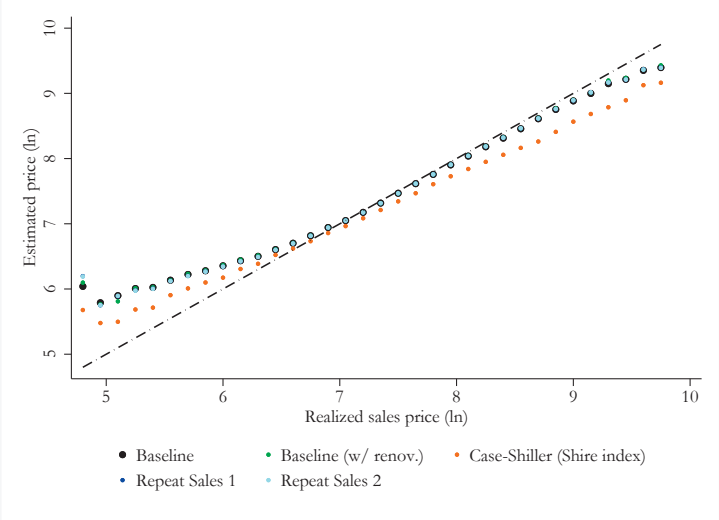


FIGURE L.16: ESTIMATED VS. REALIZED LN(PRICE) ACROSS MAIN MODELS

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

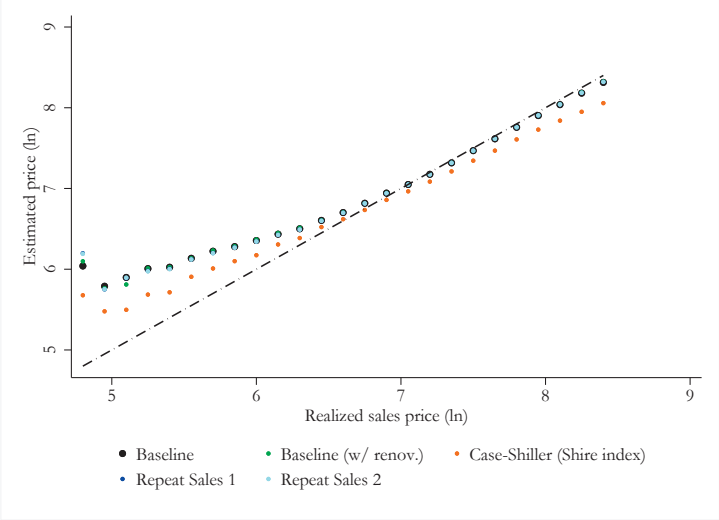
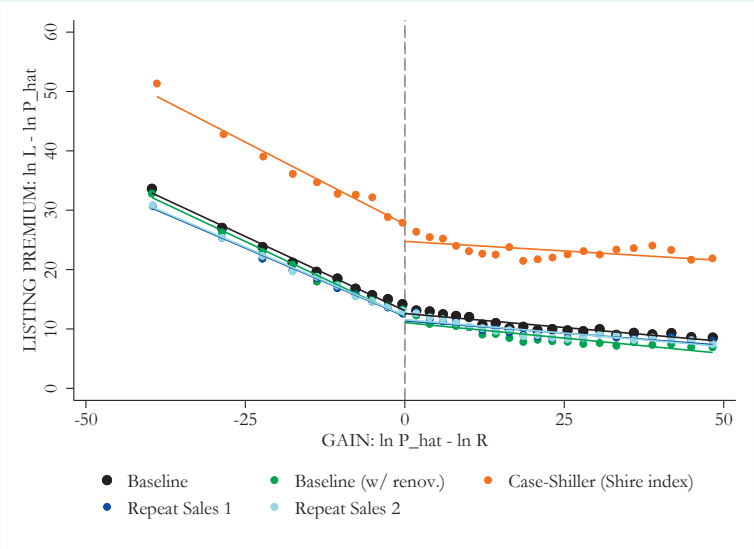


Figure L.17: Hockey Stick and Demand Concavity Across Main Models

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

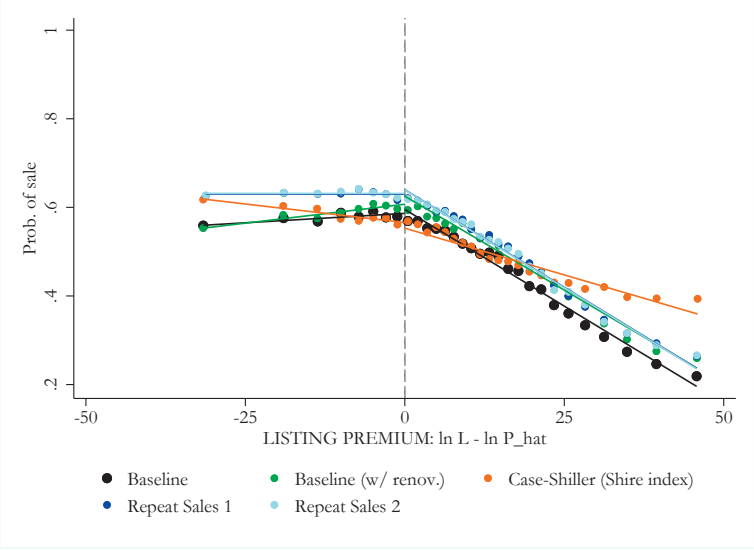
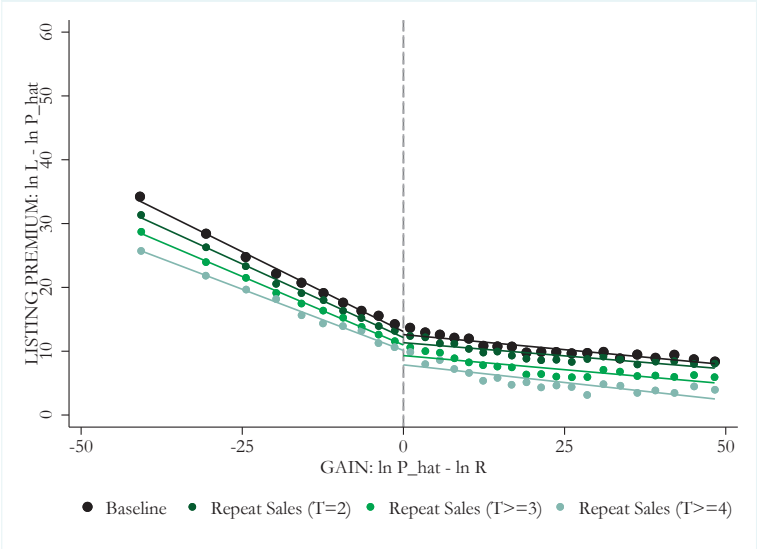


FIGURE L.18: HOCKEY STICK AND DEMAND CONCAVITY ACROSS REPEAT SALES MODELS

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

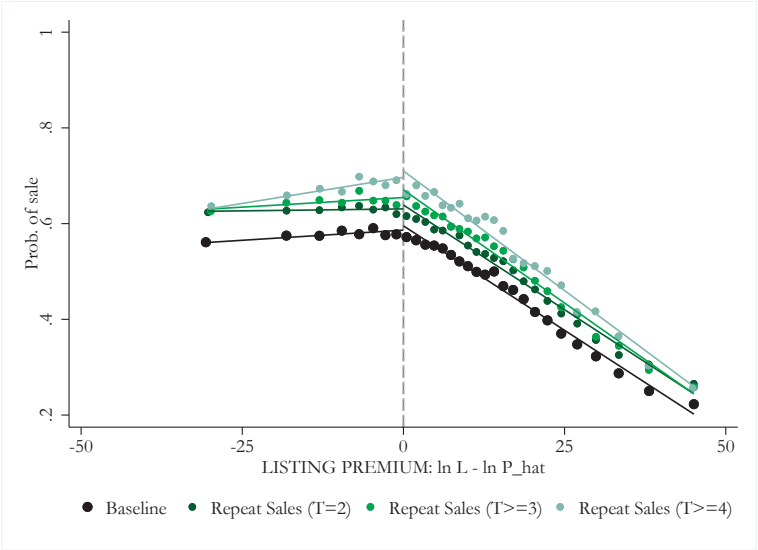


Figure L.19: RKD Validation: Smooth Density of Assignment Variable

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]

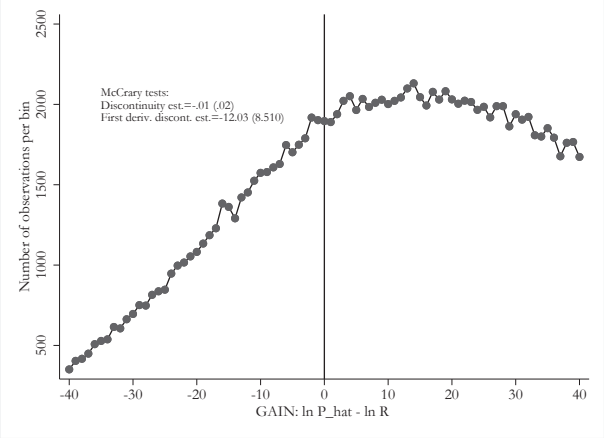


Figure L.20: RKD Validation: Covariates Smooth around Cutoff

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.

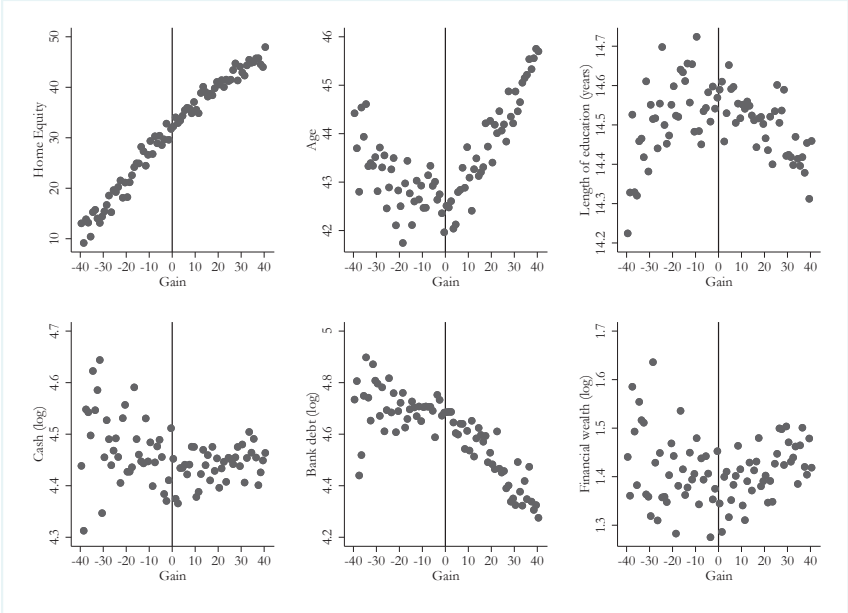


Figure L.21: RKD Robustness: Estimates for Different Bandwidths (Gain)

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).

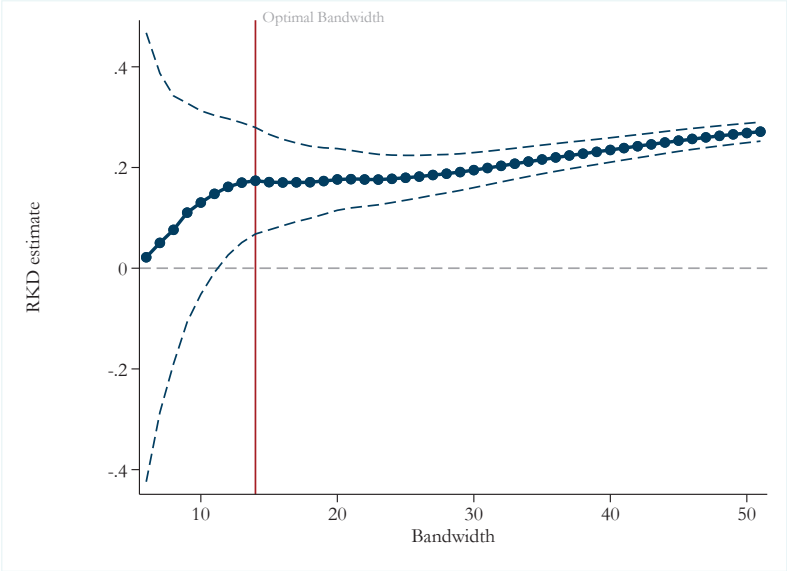


Figure L.22: RKD Estimation: Local Linear vs. Local Quadratic Estimation Results

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).

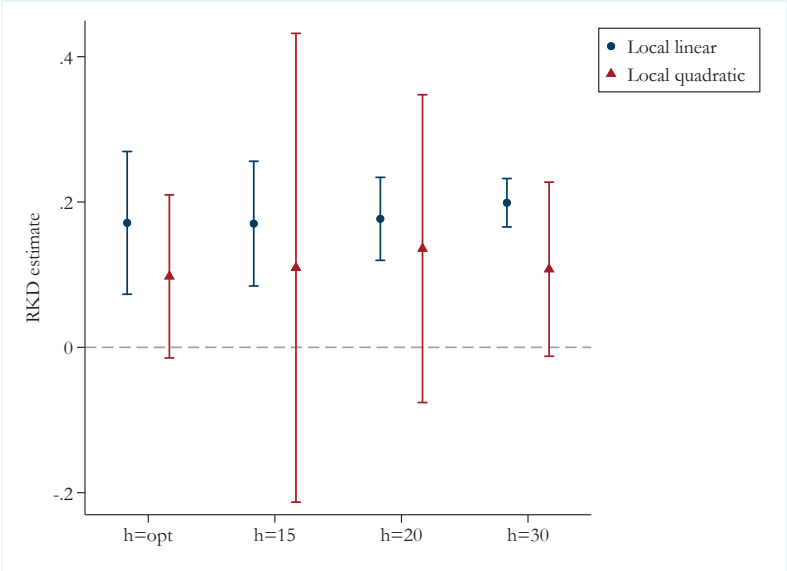


FIGURE L.23: INCIDENCE OF ROUND NUMBERS BY ROUNDING MULTIPLE

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

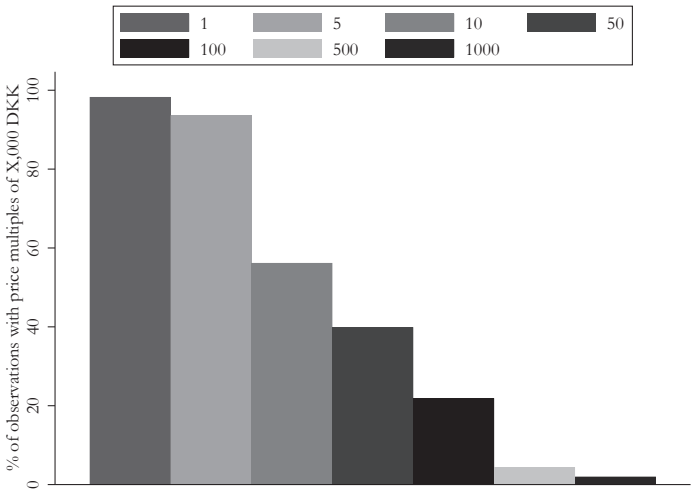
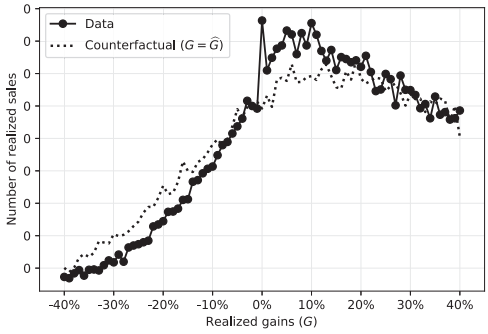


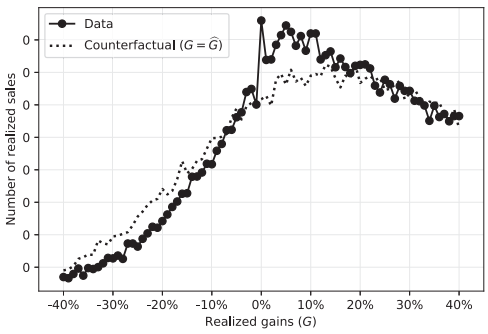
Figure L.24: Bunching Robustness: Excluding Sales at Rounded Prices

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.

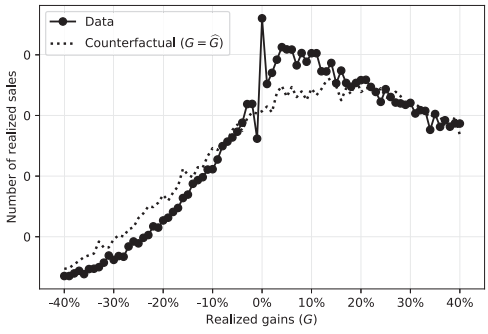
10,000 DKK



50,000 DKK



100,000 DKK



500,000 DKK

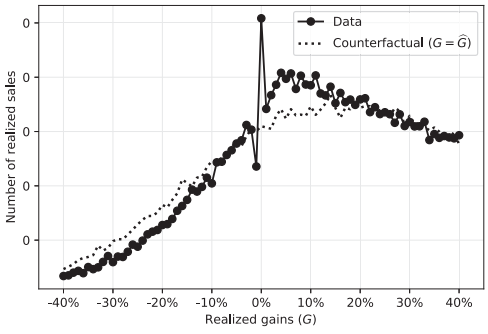
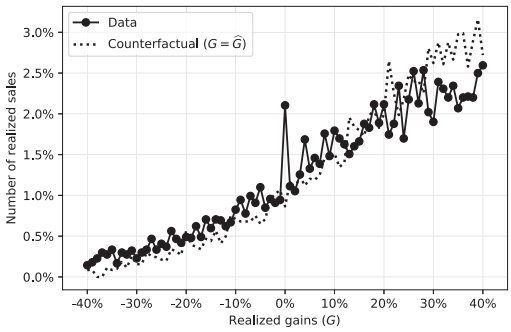


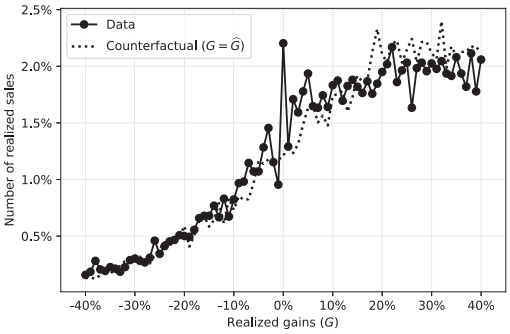
Figure L.25: Bunching Robustness: Across Previous Sales Price

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.

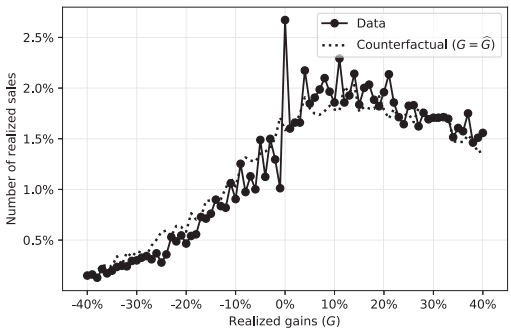
Below DKK 659,000



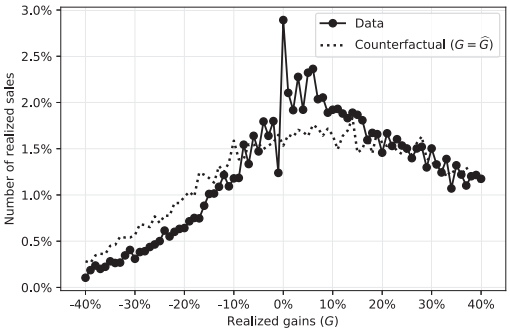
DKK 659,000 – DKK 953,000



DKK 953,000 – DKK 1,313,000



DKK 1,313,000 – DKK 1,901,000



Above DKK 1,901,000

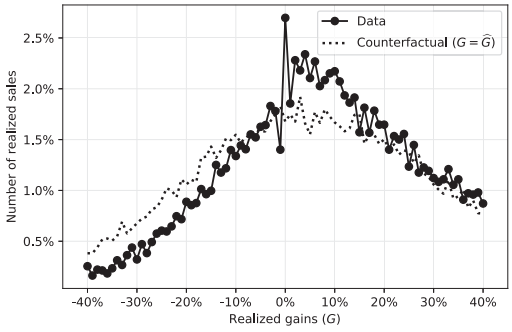
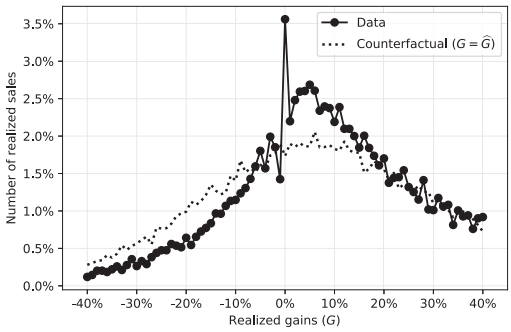


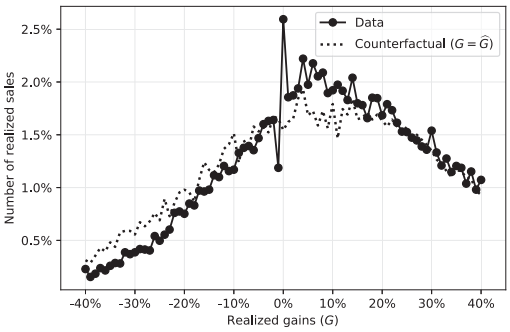
Figure L.26: Bunching Robustness: Across Holding Periods

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.

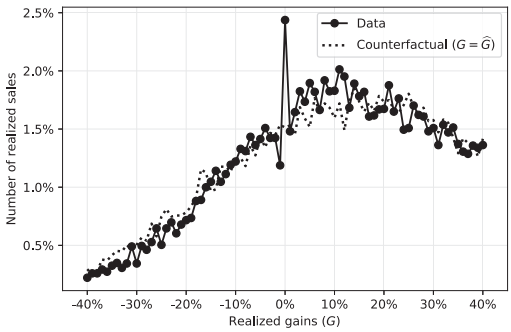
Below 3 years



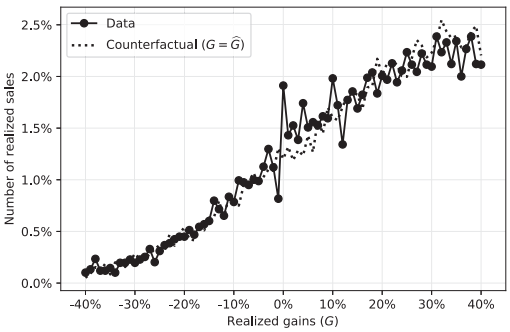
3–6 years



6–9 years



9–12 years



Above 12 years

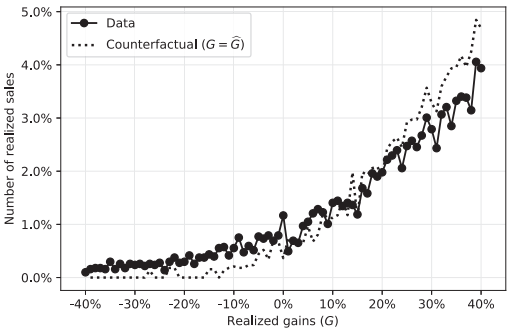


Figure L.27: Bunching Robustness: Excess Mass across Holding Periods

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.

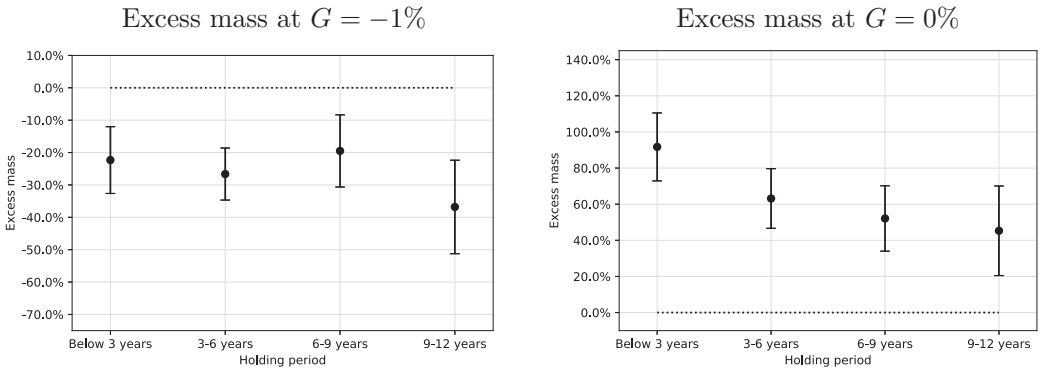


FIGURE L.28: BUNCHING ROBUSTNESS: MODEL WITH COHORT FIXED EFFECTS

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.

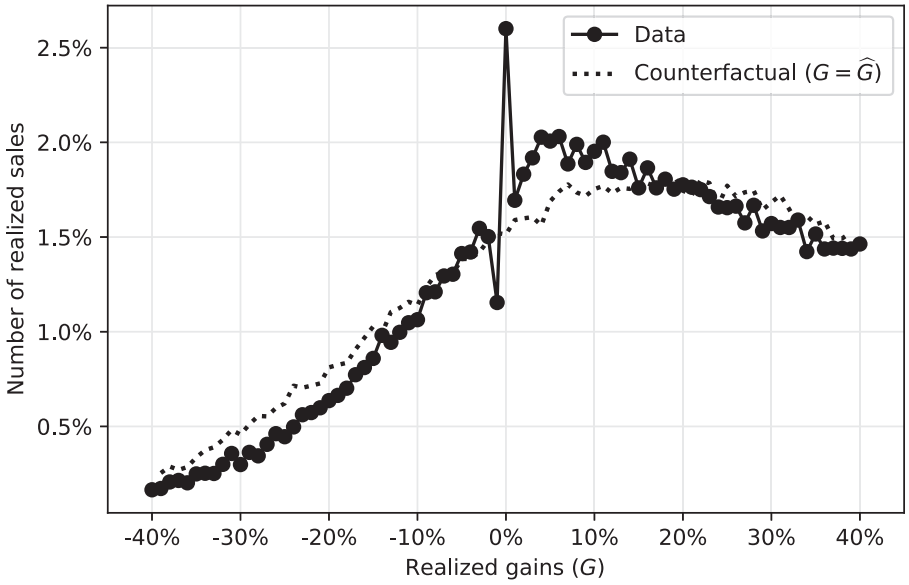
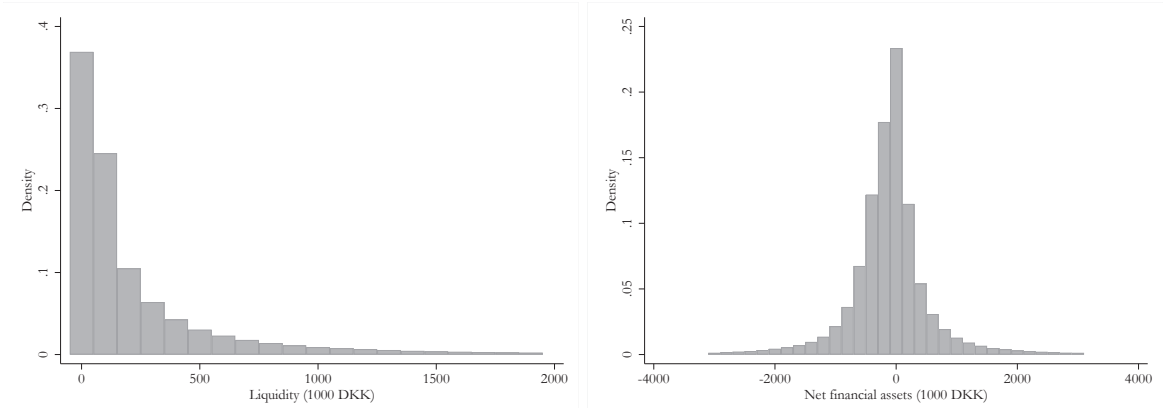


Figure L.29: Summary Statistics: Household Demographics

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

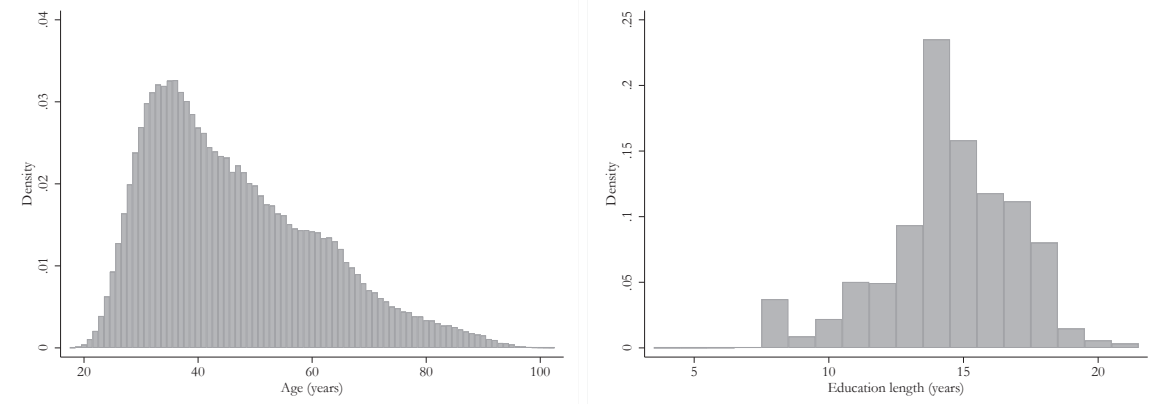


Figure L.30: Residualized Listing Premium and Gains and Home Equity

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.

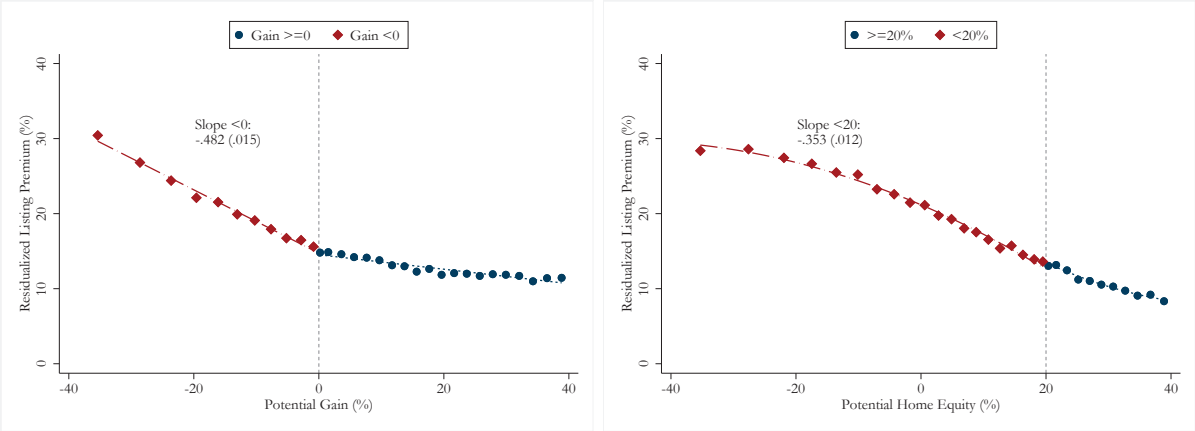


Figure L.31: Listing Premium "Hockey Stick" for Sellers Without Mortgage

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

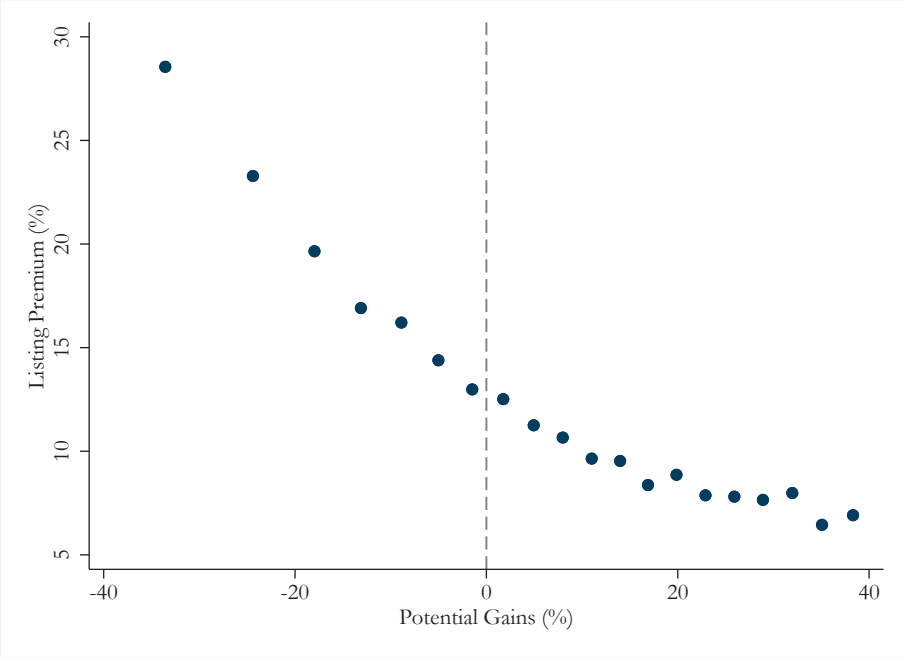


FIGURE L.32: DECOMPOSITION OF MODEL-IMPLIED MOMENTS

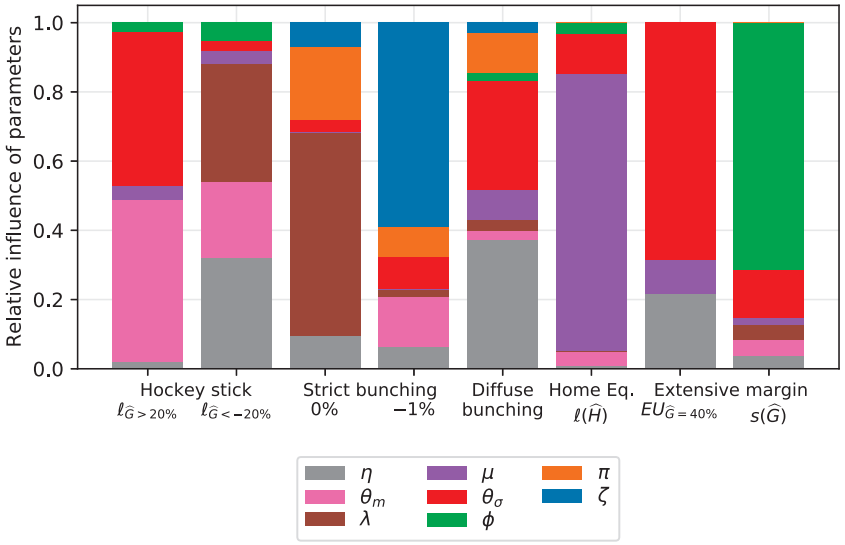


FIGURE L.33: UTILITY PENALTY CALCULATED IN THE DATA AND THE MODEL

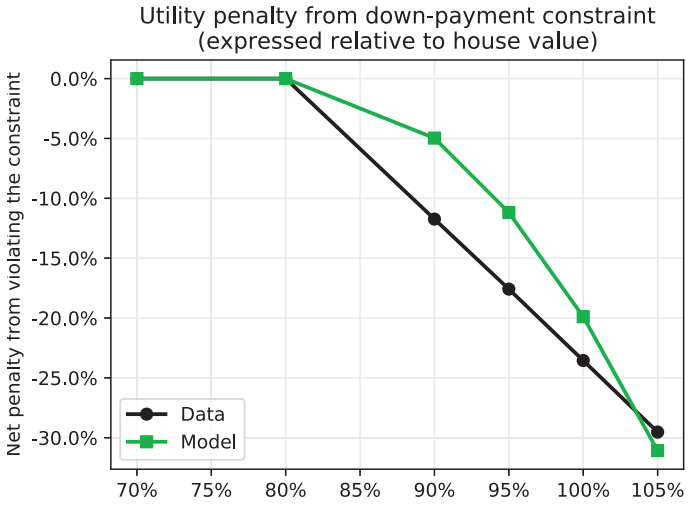


FIGURE L.34: SIMULATION RESULTS: UNOBSERVED QUALITY

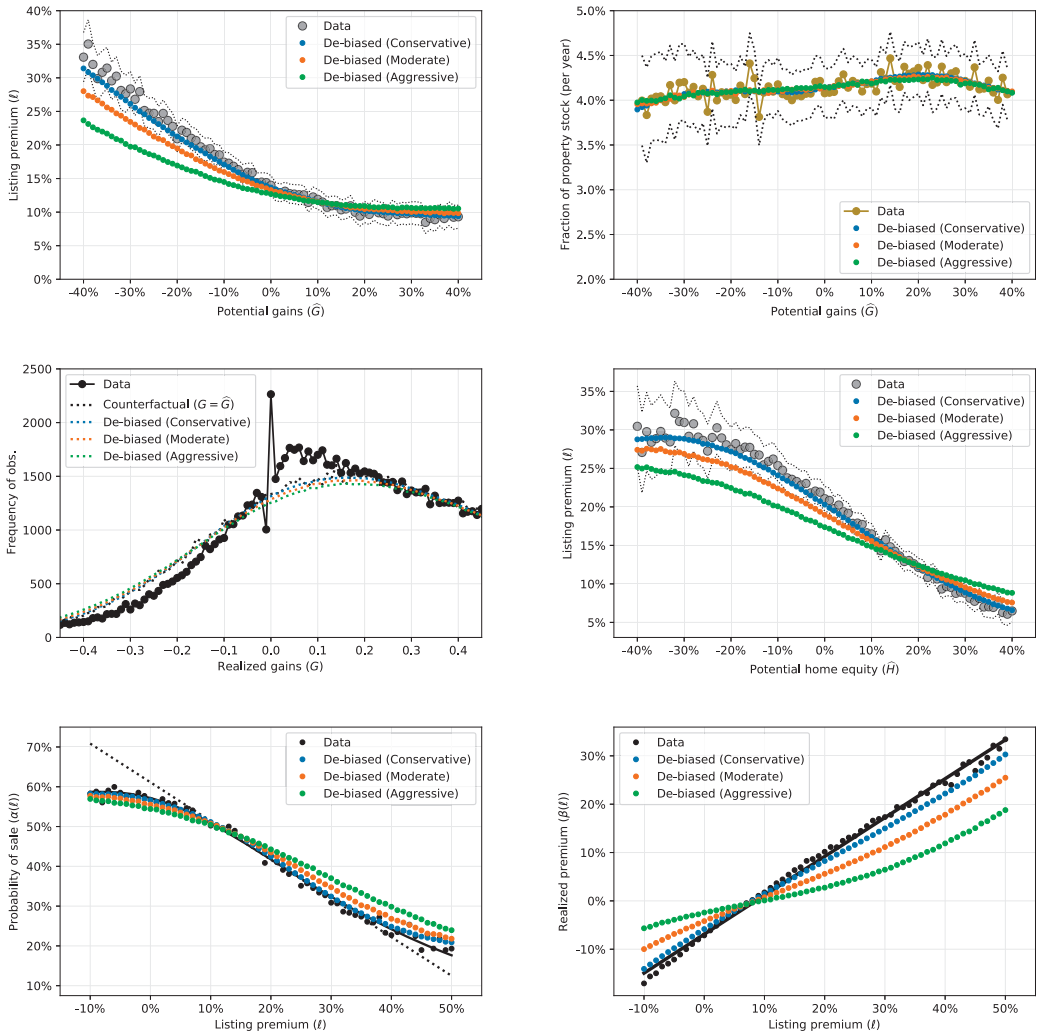


Table L.1: Construction of Main Dataset

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:

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

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

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

Panel A

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

Panel B

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