Online Appendix to

“History Dependence in the Housing Market”

Philippe Bracke Silvana Tenreyro

April 2020

A Additional Figures and Tables

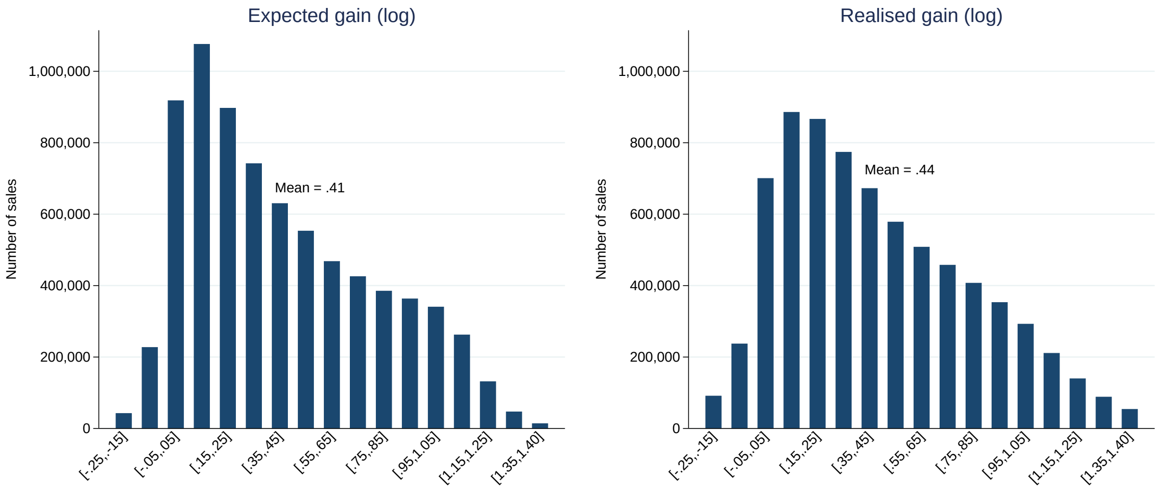


Figure A1: Distribution of gains, 1995-2014

Notes: The left chart shows the distribution of expected gains, $\widehat{GAIN}_{jst}$ in *Sample 1*. Expected gains are computed as the change in the postcode-district house price index between the year of the current sale (t) and the year in which the property was previously purchased (s). The right chart shows the distribution of actual gains, $GAIN_{jst}$, where actual gains are computed as the log house price difference between two pairs of repeat sales.

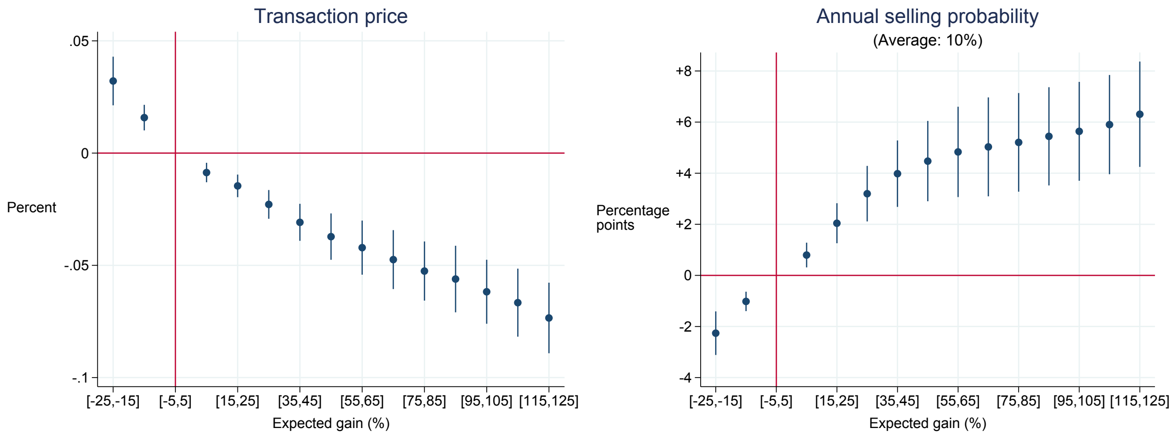


Figure A2: Effects of gains and losses, non-bootstrapped standard errors

Notes: The charts show the coefficients and corresponding 95-percent confidence bands for the k dummy variables associated with different expected gains/losses ($\widehat{GAIN}_{jkst}$'s) in the regression $y_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN}_{jkst} + \eta_t \otimes \eta_s + \lambda_t \otimes \lambda_s + w_{ijt}$, where y_{ijt} is the transaction price (p_{ijt}) in the upper chart and an binary indicator of sale (q_{ijt}) in the bottom chart. The precise values of the coefficients are reported in Table A1 and A2 in the Appendix. All regressions control for individual-property fixed effects, time-varying characteristics (whether the property was purchased new or second-hand; property it was sold as leasehold or freehold) as well as all combinations of purchase and sale year. Regressions have year-by-postcode district fixed effects (δ_{jt} in the regression formula).

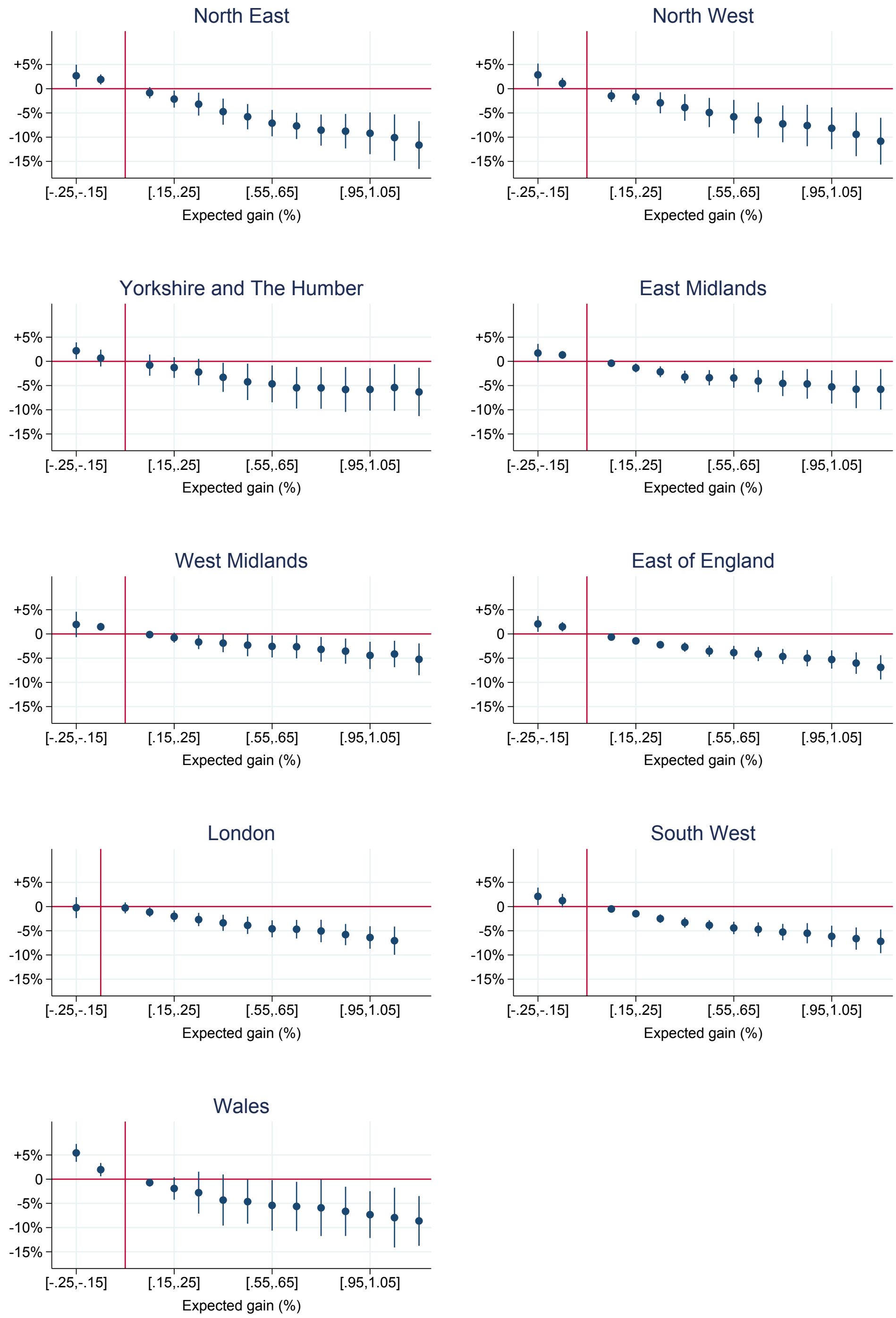


Figure A3: Effects of expected gains and losses on transaction prices, by region

Notes: The charts replicate the analysis of the left chart of Figure 2 for each region in England and Wales. The charts show the coefficients and associated confidence bands for the k dummy variables associated with different expected gains/losses ($\widehat{GAIN}_{kjst}$'s) in the regression $p_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN}_{kjst} + \eta_t \otimes \eta_s + \lambda_s \otimes \lambda_t + w_{ijt}$, run separately for each region. Regressions have year-by-postcode district fixed effects and standard errors are double-clustered by year and postcode district.

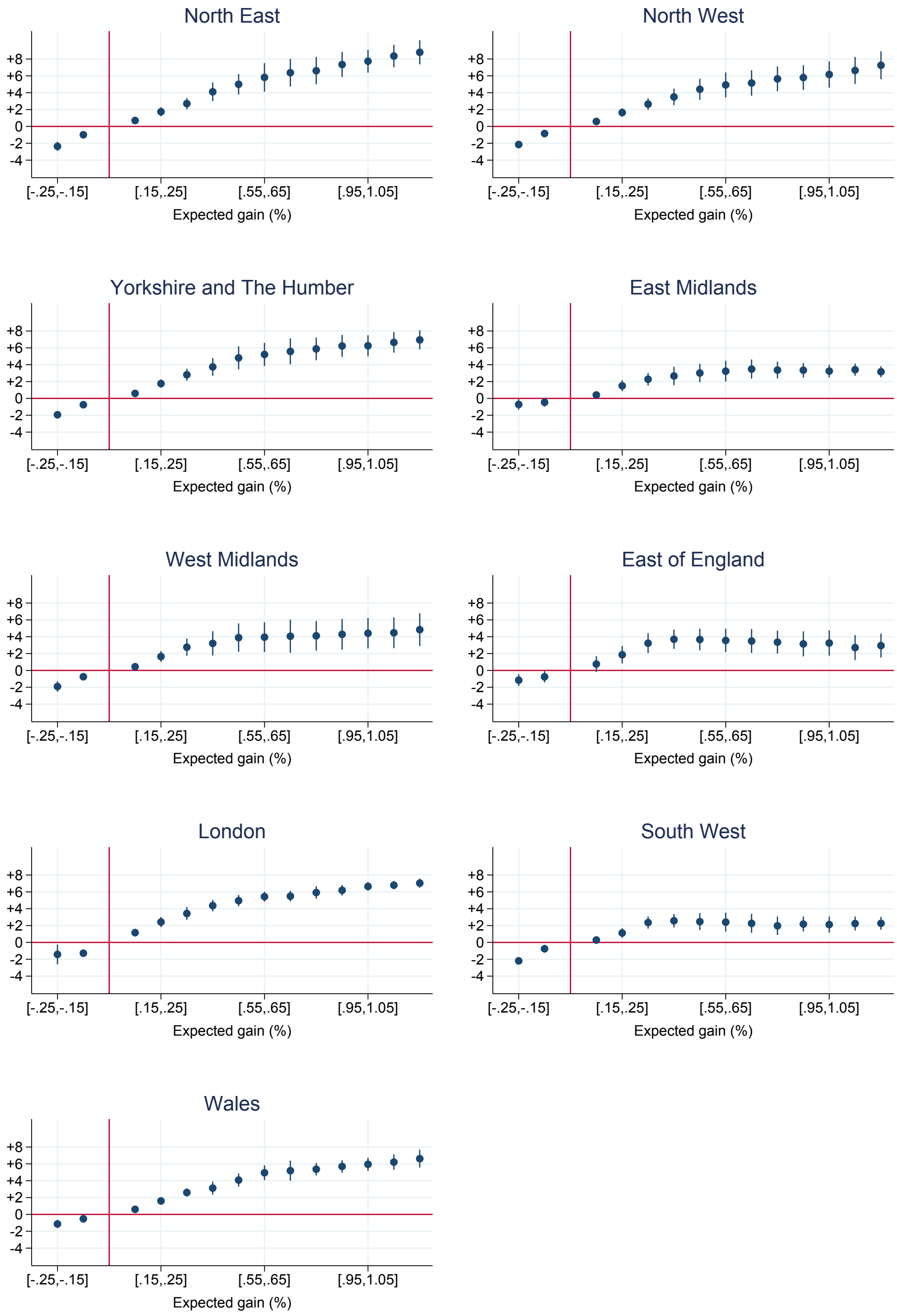


Figure A4: Effects of expected gains and losses on selling propensities, by region

Notes: The charts replicate the analysis of the right hand chart of Figure 2 for each region in England and Wales. The charts show the coefficients and associated confidence bands for the k dummy variables associated with different expected gains/losses ($\widehat{GAIN}_{kjst}$'s) in the regression $q_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN}_{kjst} + \eta_t \otimes \eta_s + \lambda_s \otimes \lambda_t + w_{ijt}$, run separately for each region. Regressions have year-by-postcode district fixed effects and standard errors are double-clustered by year and postcode district.

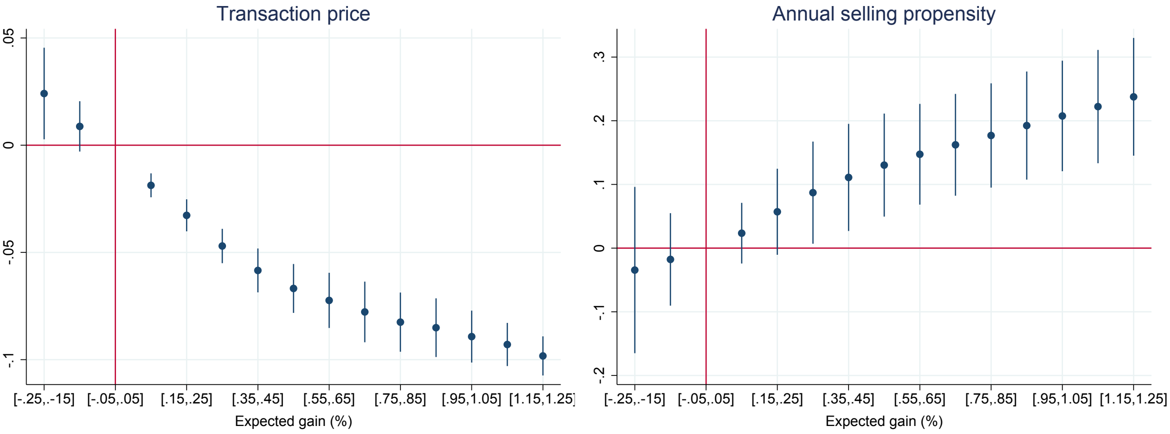


Figure A5: Robustness: No control for holding period

Notes: The two charts display the results of an alternative version of Figure 2, where the regression specification excludes the control for holding period. In other words, we do not include the purchase- and sale-year combinations $\lambda_s \otimes \lambda_t$ from equation (4).

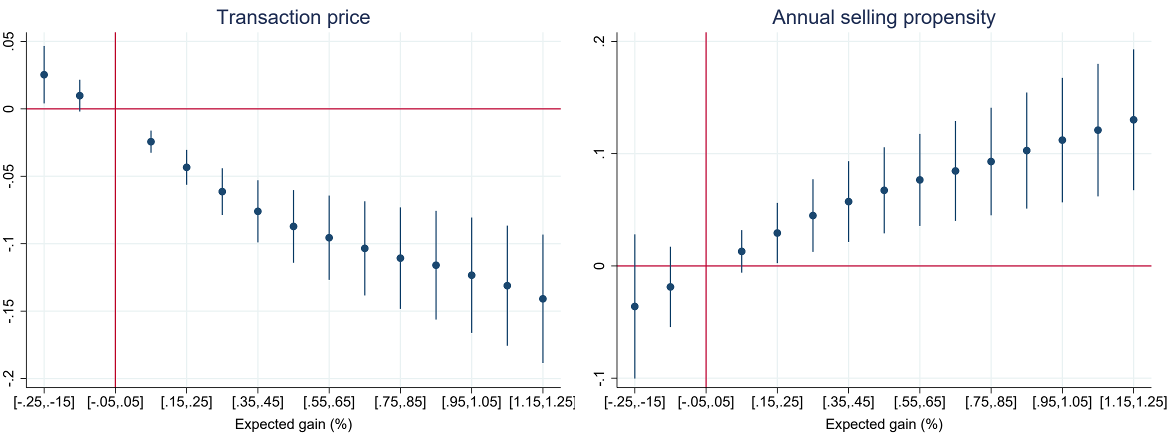


Figure A6: Robustness: Duration dummies for holding period

Notes: The two charts display the results of an alternative version of Figure 2, where the regression specification replaces the purchase- and sale-year combinations $\lambda_s \otimes \lambda_t$ from equation (4) with a series of dummies corresponding to the number of years a property has stayed with the same owner.

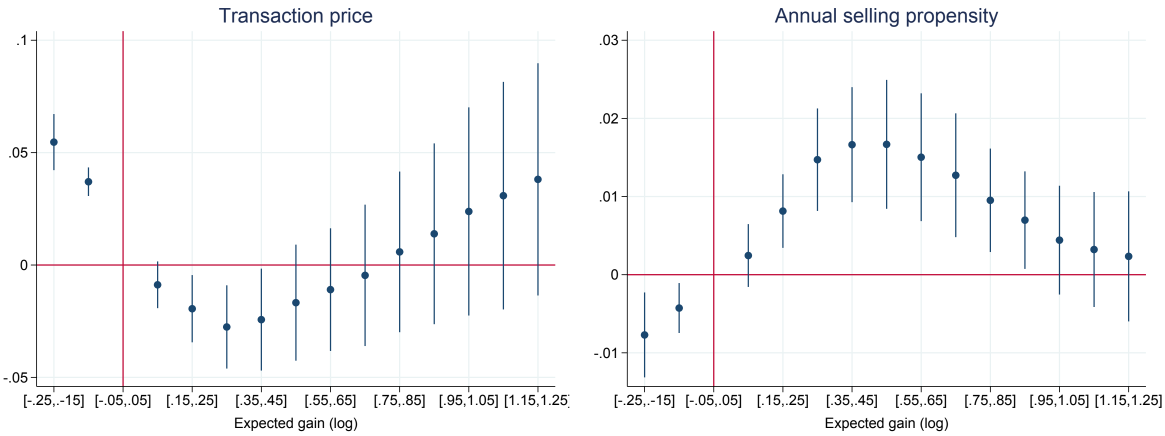


Figure A7: Robustness: No individual-property fixed effects

Notes: The two charts display the results of an alternative version of Figure 2, where the regression in equation (4) excludes individual-property fixed effects (v_{ij}).

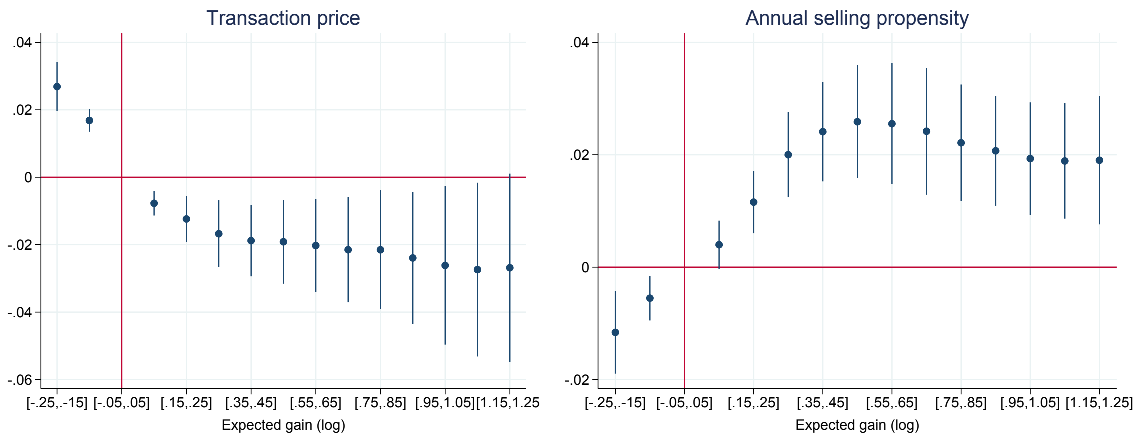


Figure A8: Robustness: full-postcode fixed effects

Notes: The two charts display the results of an alternative version of Figure 2, where the regression in equation (4) has full-postcode fixed effects instead of individual-property fixed effects (v_{ij}).

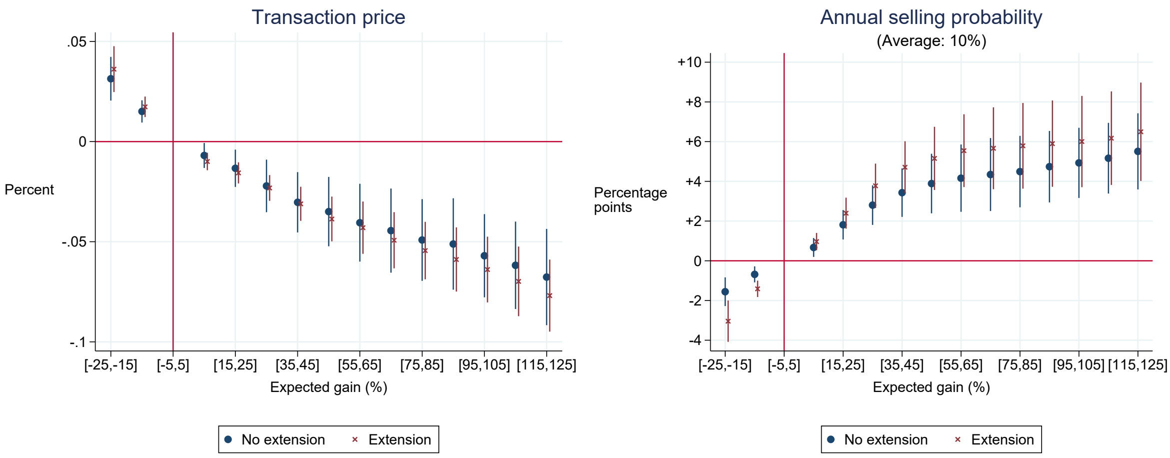


Figure A9: Robustness: No houses with extensions

Notes: The two charts display the results of an alternative version of Figure 2, where the sample only contains properties that are not labeled as "with extension" in the UK Energy Performance Certificate dataset.

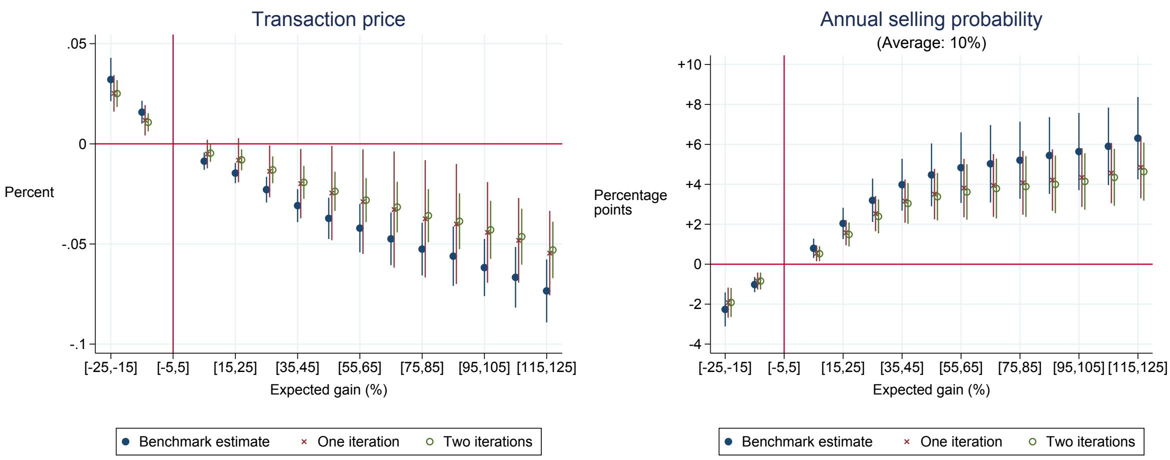


Figure A10: Robustness: Iterating the expected gain measure

Notes: The solid dots replicate the results of Figure 2 in the paper; they show the coefficients and corresponding 95-percent confidence bands for the k dummy variables associated with different expected gains/losses ($\widehat{GAIN}_{kjst}$'s) in the regression $y_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN}_{jkst} + \lambda_t \otimes \lambda_s + w_{ijt}$. The crosses show the coefficients of a similar regression, $y_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN'}_{jkst} + \lambda_t \otimes \lambda_s + w_{ijt}$, where $\widehat{GAIN'}_{jkst}$ is constructed from the local authority-by-year effects δ_{jt} estimated in the previous iteration. The hollow dots show the results from a second iteration of $\widehat{GAIN'}_{jkst}$.

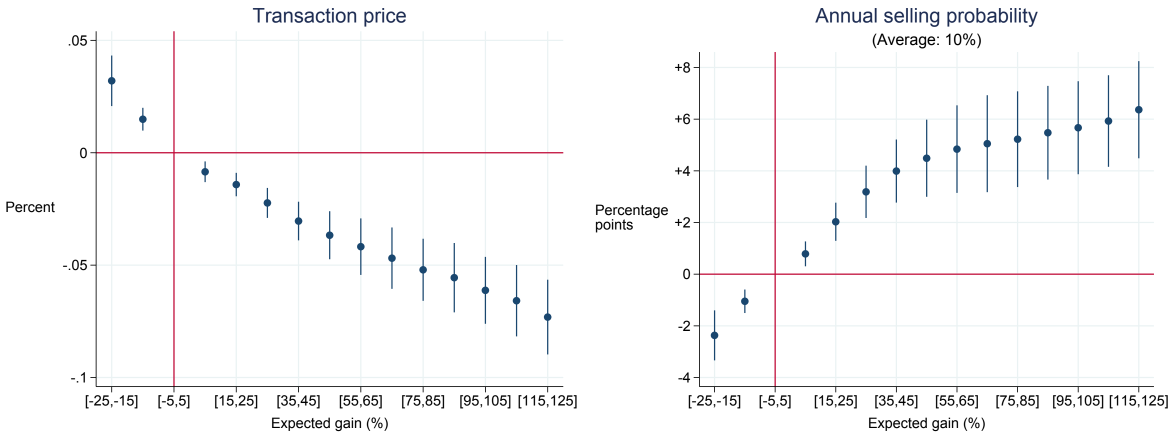
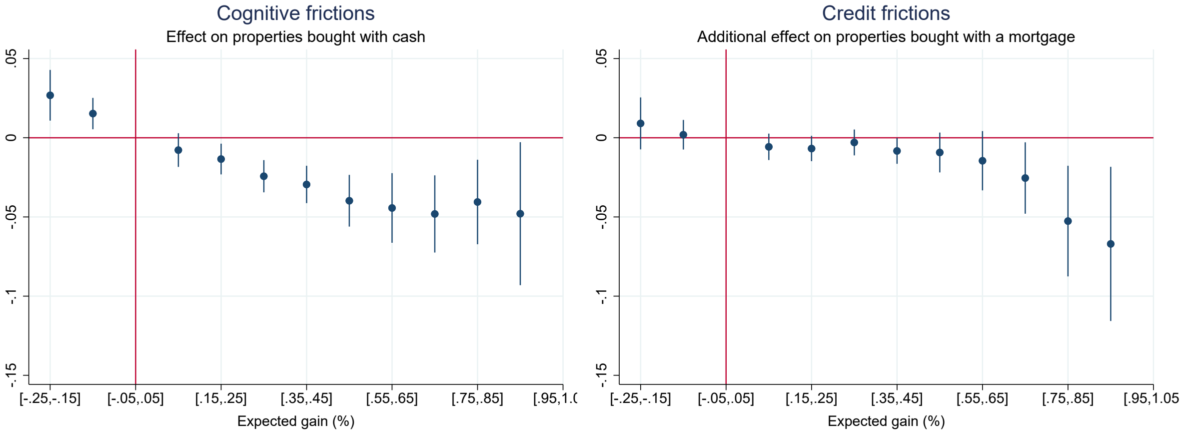


Figure A11: Robustness: No postcode district-year with fewer than 100 sales

Notes: The two charts display the results of an alternative version of Figure 2, where the sample does not include postcode district-year with fewer than 100 sales.

Transaction prices



Selling propensities

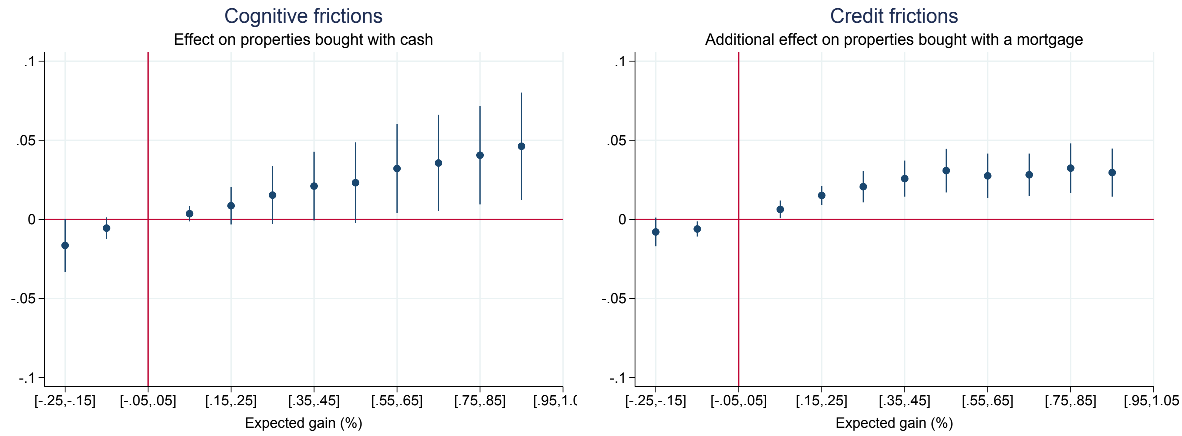


Figure A12: Credit vs cognitive frictions (2002–2014), non-bootstrapped standard errors

Notes: The charts replicate the analysis of Figure A2 but focuses on the results for properties purchased after 2001, for which information is available on whether the transaction was financed with cash or with a mortgage. The regression is $y_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_{0k}\left(\widehat{GAIN}_{jkst} \times \mathit{upto2001}\right) + \sum_k \gamma_{1k}\left(\widehat{GAIN}_{jkst} \times \mathit{post2001}\right) + \sum_k \gamma_{2k}\left(\widehat{GAIN}_{jkst} \times \mathit{mortgage}\right) + \lambda_t \otimes \lambda_s + w_{ijt}$, and the charts report the $\hat{\gamma}_{1k}$ and $\hat{\gamma}_{2k}$'s. The indicator variable *post2001* singles out properties for which a purchase is available after 2001; the indicator variable *mortgage* tags properties bought with a mortgage. The regression coefficients are reported in Table A1 and A2 in the Appendix.

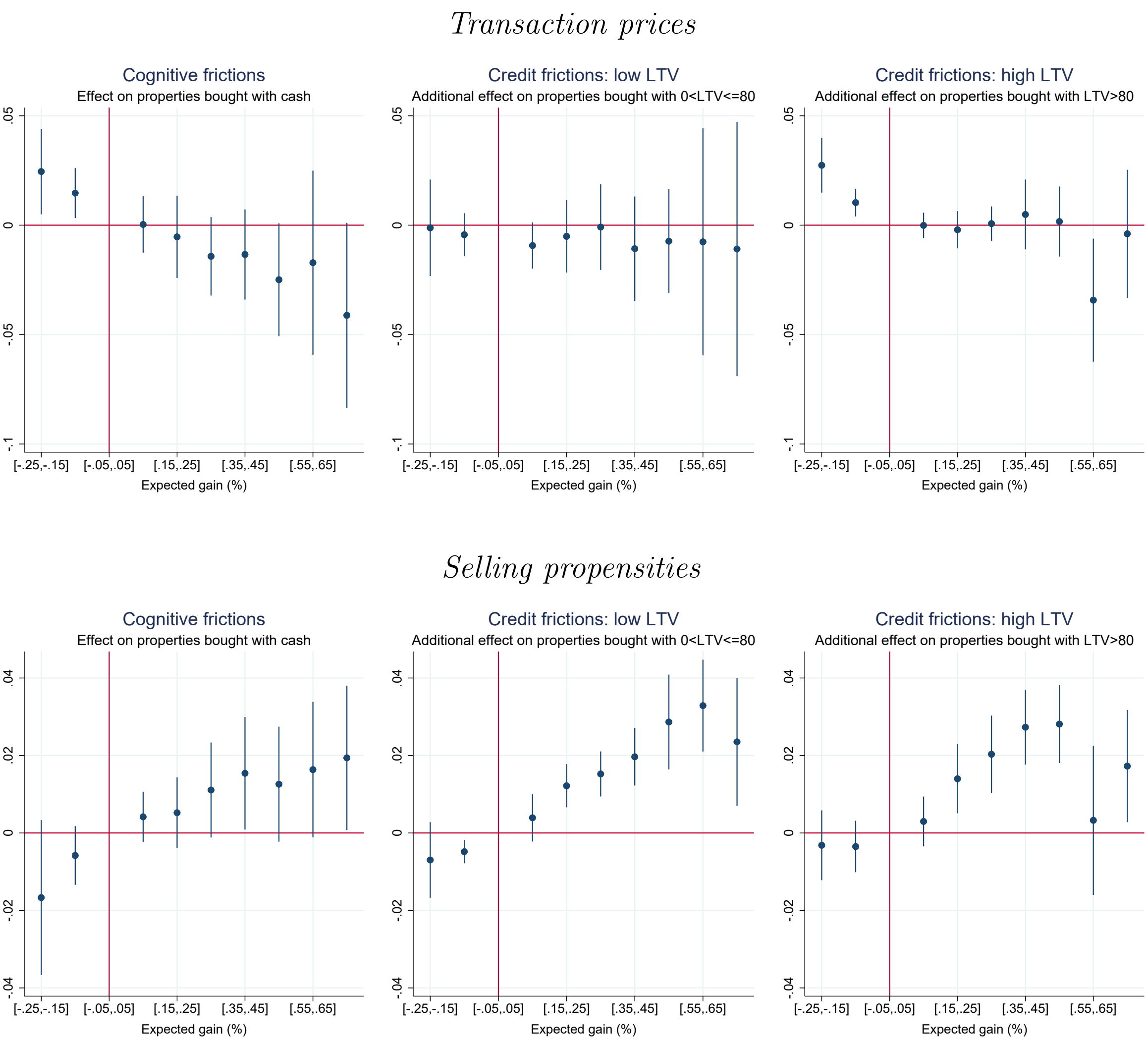


Figure A13: Effects of credit and cognitive frictions after 2005q1, non-bootstrapped standard errors

Notes: The charts replicate the analysis of Figure A2 but focuses on the results for properties purchased after March 2005, for which information is available on the characteristics of the mortgage used to finance the transaction. The regression is $y_{ijt} = X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_{0k}\left(\widehat{GAIN}_{jkst} \times \mathit{upto2001}\right) + \sum_k \gamma_{1k}\left(\widehat{GAIN}_{jkst} \times \mathit{post2005q1}\right) + \sum_k \gamma_{2k}\left(\widehat{GAIN}_{jkst} \times \mathit{mortgage}\right) + \sum_k \gamma_{3k}\left(\widehat{GAIN}_{jkst} \times \mathit{ltv}80\right) + \eta_t \otimes \eta_s + \lambda_t \otimes \lambda_s + w_{ijt}$, and the charts report the $\hat{\gamma}_{1k}$, $\hat{\gamma}_{2k}$ and $\hat{\gamma}_{3k}$'s. The indicator variable *post2005q1* singles out properties for which a purchase is available after March 2005; the indicator variable *mortgage* tags properties bought with a mortgage; $ltv80$ indicates properties that were bought with a loan-to-value (LTV) ratio greater than 80. Information on the characteristics of mortgages is available from the Product Sales Data (PSD) since March 2005. The match between Land Registry (LR) and PSD, described in Appendix B.2, generates four subsets of *post2015q1* transactions: matched properties bought with a high LTV, matched properties bought with a low LTV, properties that were bought with cash according to the LR, and properties that were bought with a mortgage according to the LR but do not match with the PSD. All these groups are included in the regression; the latter group is controlled for through a group-specific dummy. The precise values of the coefficients are reported in Table A1 and A2 in the Appendix.

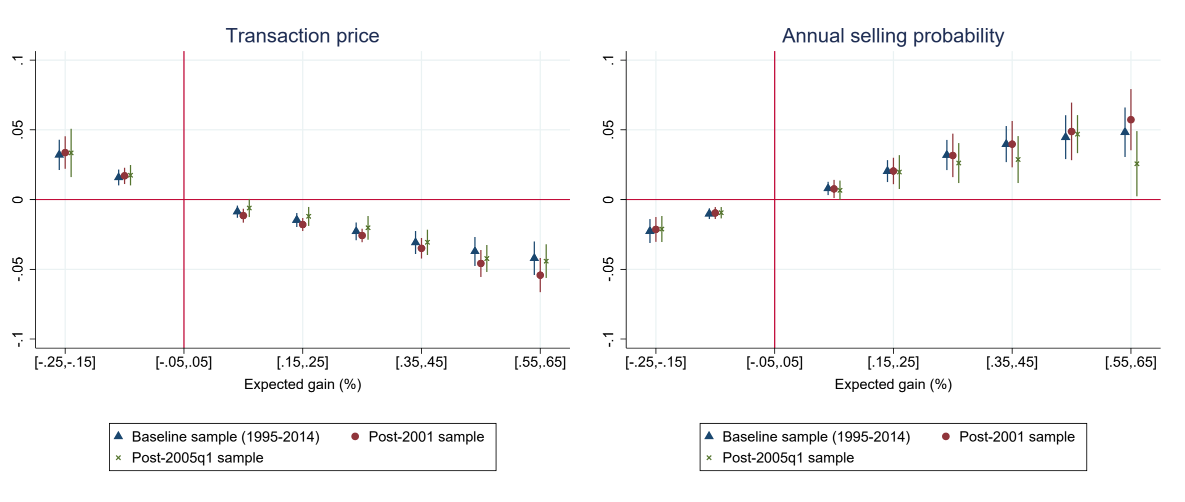


Figure A14: Comparing history dependence across samples

Notes: The charts include three sets of coefficients to show that history dependence persists across different subsamples used throughout the paper. The first set replicates Figure 2. The second set focuses on the post-2001 subsample and the third set is restricted to the post-2005q1 subsample. The coefficients on subsamples are produces by a regression of the form: s $y_{ijt} = X_{ijt}\beta+\delta_{jt}+\sum_k \gamma_{0k}\left(\widehat{GAIN}_{jkst}\times \textit{pre-sample}\right)+\sum_k \gamma_{1k}\widehat{GAIN}_{jkst}+\eta_t\otimes\eta_s+\lambda_t\otimes\lambda_s+w_{ijt}$, where *pre-sample* is an indicator variable covering the sales taking place before the relevant subsample. (These sales are kept in the regression to preserve the same set of property fixed effects across regressions.)

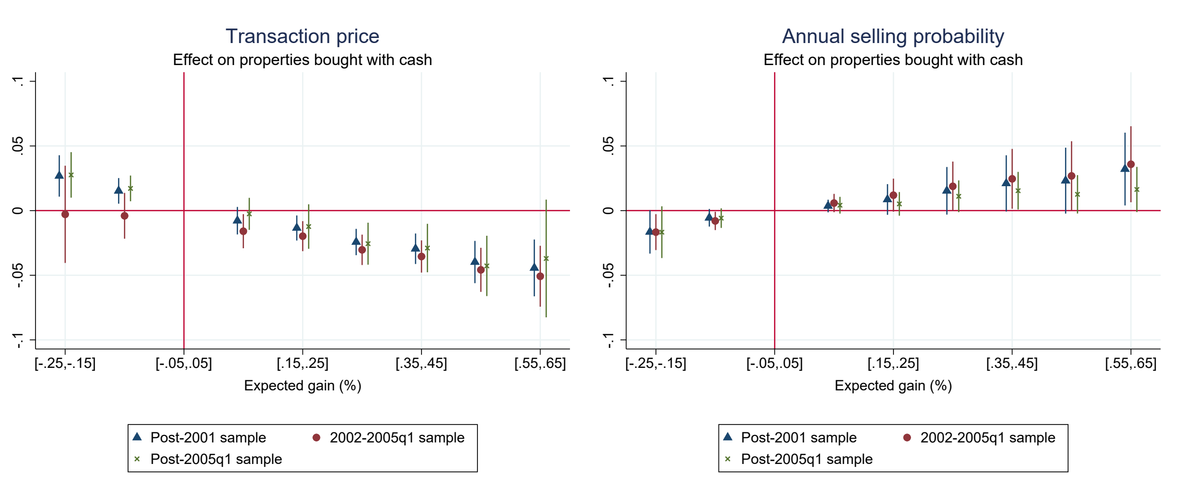
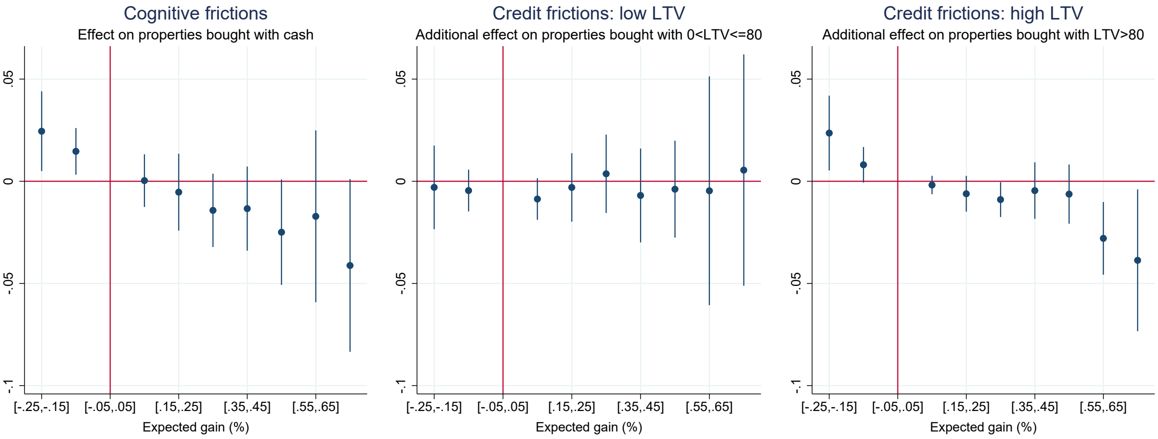


Figure A15: Comparing cognitive frictions across samples

Notes: The charts put together the history-dependence coefficients that refer to properties bought with cash and shown in Figure 3 (post-2001 sample) and 4 (post-2005q1 sample). Because the two sample differ in their coverage of 2001-2005q1 properties, this group is analyzed in a new separate regression and given its own coefficients, which are shown in the charts.

Transaction prices



Selling propensities

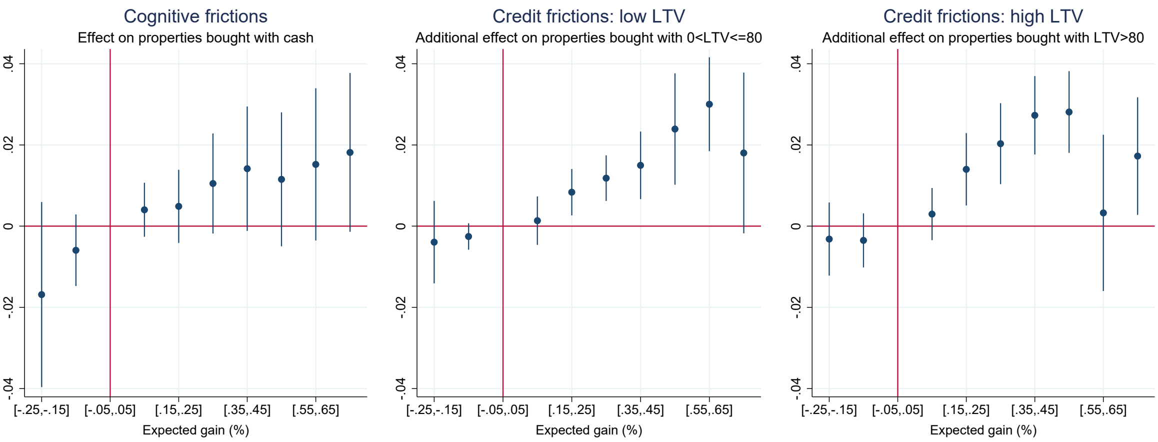


Figure A16: Effects of credit and cognitive frictions after 2005q1, historic rather than contemporaneous LTV

Notes: The charts replicate the analysis of Figure A13 but uses as threshold between constrained and un-constrained home sellers LTV at origination rather than current LTV.

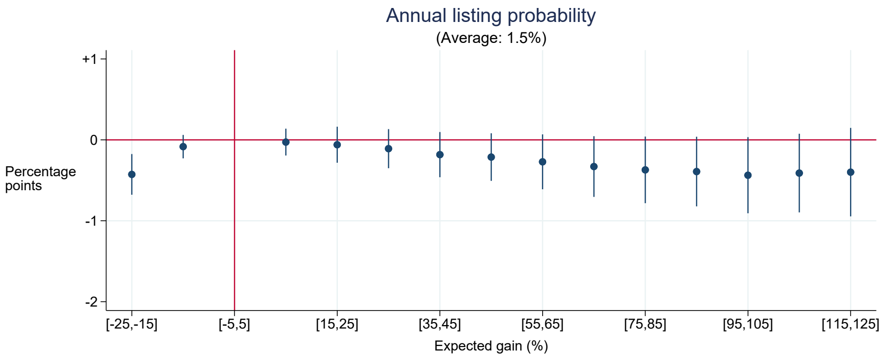


Figure A17: Annual probability analysis of being listed on Zoopla

Notes: The figure shows an analysis similar to the one on the right hand side chart of Figure 2, but the outcome variable for houses is being listed on Zoopla, rather than selling.

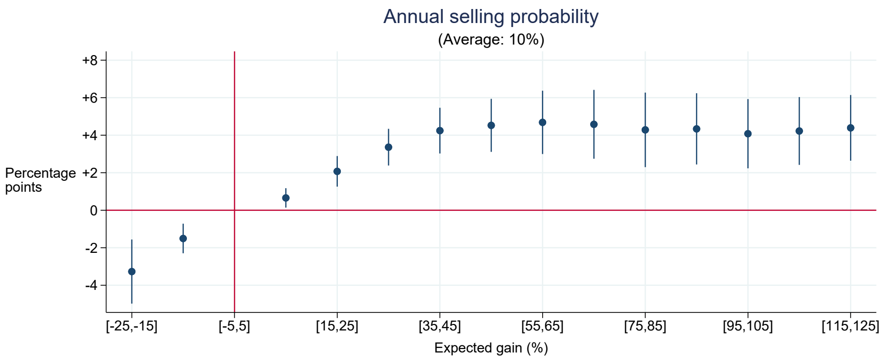


Figure A18: Annual selling propensity analysis on sample which listed on Zoopla

Notes: The figure shows an analysis of selling propensities similar to the one on the right hand side chart of Figure 2. The sample is limited to properties that appear in the WhenFresh/Zoopla dataset of listings.

Table A1: Effects of expected gains and losses on transaction prices

Notes: The first column of the table contains the coefficients and (non-bootstrapped) standard errors for the k dummy variables associated with different gains/losses ($\widehat{GAIN}_{kjst}$'s) in the regression $p_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN}_{kjst} + \eta_t \otimes \eta_s + \lambda_s \otimes \lambda_t + w_{ijt}$, where p_{ijt} is the (log) transaction price. The coefficients are displayed graphically with their bootstrapped confidence bands in the left hand side part of Figure 2 and with their non-bootstrapped confidence bands (derived from double-clustered standard errors by postcode district and year, in parentheses in this table) in Appendix Figure A2.

Column 2 shows the coefficient on the interaction $\widehat{GAIN}_{jst} \times post2001$, where $post2001$ indicates sales whose previous purchase took place after 2001, in the regression $p_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_{1k}(\widehat{GAIN}_{kjst} \times post2001) + \sum_k \gamma_{2k}(\widehat{GAIN}_{kjst} \times Mortgage) + \eta_t \otimes \eta_s + \lambda_s \otimes \lambda_t + w_{ijt}$. Column 3 shows the coefficient on $\widehat{GAIN}_{kjst} \times Mortgage$ on this same regression. Information on whether the buyer used a mortgage to finance the transaction is available from the Land Registry since 2002.

Column 4 shows the coefficient on the interaction $\widehat{GAIN}_{jst} \times post2005q1$, where $post2005q1$ indicates sales whose previous purchase took place after March 2005, in the regression $p_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_{1k}(\widehat{GAIN}_{kjst} \times post2005q1) + \sum_k \gamma_{2k}(\widehat{GAIN}_{kjst} \times Mortgage) + \sum_k \gamma_{3k}(\widehat{GAIN}_{kjst} \times HighLTV) + \eta_t \otimes \eta_s + \lambda_s \otimes \lambda_t + w_{ijt}$, where $HighLTV$ denotes properties bought with a mortgage with a loan-to-value ratio (LTV) greater than 80 percent. Both $Mortgage$ and $HighLTV$ are defined only within $Sample3$, which derives from the match between LR and PSD which is described in Appendix B.2.

Dependent variable:	Transaction price (p_t)					
	(1995-2014)	(2002-2014)		(2005-2014)		
	All	Cash	Mortgage	Cash	Low-LTV	High-LTV
	(1)	(2)	(3)	(4)	(5)	(6)
Gain [-.25,-.15]	0.032	0.027	0.009	0.028	0.000	0.031
	(0.005)	(0.008)	(0.008)	(0.008)	(0.010)	(0.004)
Gain [-.15,-.05]	0.016	0.015	0.002	0.017	-0.003	0.008
	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)
Gain [.05,.15]	-0.009	-0.008	-0.006	-0.003	-0.009	-0.005
	(0.002)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)
Gain [.15,.25]	-0.015	-0.013	-0.007	-0.012	-0.003	-0.014
	(0.002)	(0.005)	(0.004)	(0.008)	(0.008)	(0.002)
Gain [.25,.35]	-0.023	-0.024	-0.003	-0.026	0.003	-0.024
	(0.003)	(0.005)	(0.004)	(0.008)	(0.009)	(0.004)
Gain [.35,.45]	-0.031	-0.029	-0.008	-0.029	-0.006	-0.031
	(0.004)	(0.006)	(0.004)	(0.009)	(0.011)	(0.007)
Gain [.45,.55]	-0.037	-0.040	-0.009	-0.043	-0.006	-0.014
	(0.005)	(0.008)	(0.006)	(0.011)	(0.011)	(0.014)
Gain [.55,.65]	-0.042	-0.044	-0.015	-0.037	-0.014	-0.043
	(0.006)	(0.011)	(0.009)	(0.022)	(0.025)	(0.023)
Gain [.65,.75]	-0.047	-0.048	-0.025	-0.063	-0.011	-0.023
	(0.006)	(0.012)	(0.011)	(0.021)	(0.028)	(0.018)
Gain [.75,.85]	-0.053	-0.041	-0.053	-0.105	0.043	-0.026
	(0.006)	(0.013)	(0.017)	(0.035)	(0.035)	(0.045)
Gain [.85,.95]	-0.056	-0.048	-0.067	-0.081	-0.109	
	(0.007)	(0.022)	(0.023)	(0.050)	(0.069)	
Gain [.95,1.05]	-0.062	-0.042	-0.110	-0.138	0.216	
	(0.007)	(0.028)	(0.033)	(0.063)	(0.086)	
Gain [1.05,1.15]	-0.067	-0.051	-0.132	-0.333	0.090	
	(0.007)	(0.029)	(0.041)	(0.155)	(0.171)	
Gain [1.15,1.25]	-0.073	-0.073	-0.142	-0.176	0.000	
	(0.008)	(0.054)	(0.077)	(232.589)	(1661.778)	
Gain [1.25,1.35]	-0.083	-0.079	-0.209	-0.120	-0.350	
	(0.008)	(0.085)	(0.094)	(0.065)	(0.035)	
Gain [1.35,1.45]	-0.089	0.075	-0.170	0.712	0.000	
	(0.010)	(0.138)	(0.179)	(0.058)	(0.000)	
N	4,280,790	4,280,790	4,280,790	4,280,790	4,280,790	4,280,790

Table A2: Effects of expected gains and losses on selling propensities

Notes: The table is analogous to Table A1 but refers to the regressions of the type $q_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN}_{kt} + \eta_t \otimes \eta_s + \lambda_s \otimes \lambda_t + w_{ijt}$, where q_{ijt} is a binary indicator of sale. The coefficients are displayed graphically with their 95 percent confidence bands in the lower half of Figure 2 (column 1, 2, and 5) , 3 (column 3 and 4), and 3 (column 6 and 7). Regressions have year-by-postcode district (PCD) fixed effects (δ_{jt} in the regression formula) and standard errors are double-clustered by year and postcode district.

Dependent variable:	Selling probability (q_t)					
	(1995-2014)	(2002-2014)		(2005-2014)		
	All	Cash	Mortgage	Cash	Low-LTV	High-LTV
	(1)	(2)	(3)	(4)	(5)	(6)
Gain [-.25,-.15]	-0.023	-0.016	-0.008	-0.017	-0.007	-0.003
	(0.004)	(0.008)	(0.004)	(0.010)	(0.005)	(0.004)
Gain [-.15,-.05]	-0.010	-0.006	-0.006	-0.006	-0.005	-0.003
	(0.002)	(0.003)	(0.002)	(0.004)	(0.001)	(0.003)
Gain [.05,.15]	0.008	0.004	0.006	0.004	0.004	0.003
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Gain [.15,.25]	0.020	0.009	0.015	0.005	0.012	0.014
	(0.004)	(0.006)	(0.003)	(0.004)	(0.003)	(0.004)
Gain [.25,.35]	0.032	0.015	0.021	0.011	0.015	0.020
	(0.005)	(0.009)	(0.005)	(0.006)	(0.003)	(0.005)
Gain [.35,.45]	0.040	0.021	0.026	0.015	0.020	0.027
	(0.006)	(0.010)	(0.005)	(0.007)	(0.004)	(0.004)
Gain [.45,.55]	0.045	0.023	0.031	0.013	0.029	0.028
	(0.007)	(0.012)	(0.007)	(0.007)	(0.006)	(0.005)
Gain [.55,.65]	0.048	0.032	0.027	0.016	0.033	0.003
	(0.008)	(0.013)	(0.007)	(0.008)	(0.006)	(0.009)
Gain [.65,.75]	0.050	0.036	0.028	0.019	0.024	0.017
	(0.009)	(0.015)	(0.006)	(0.009)	(0.008)	(0.007)
Gain [.75,.85]	0.052	0.041	0.032	0.013	0.040	-0.007
	(0.009)	(0.015)	(0.007)	(0.012)	(0.010)	(0.012)
Gain [.85,.95]	0.054	0.046	0.030	0.005	0.053	0.056
	(0.009)	(0.016)	(0.007)	(0.013)	(0.016)	(0.045)
Gain [.95,1.05]	0.056	0.054	0.029	0.051	0.023	-0.047
	(0.009)	(0.019)	(0.008)	(0.019)	(0.057)	(0.058)
Gain [1.05,1.15]	0.059	0.056	0.030	0.035	0.035	-0.065
	(0.009)	(0.020)	(0.008)	(0.012)	(0.025)	(0.035)
Gain [1.15,1.25]	0.063	0.068	0.022	-0.022	1.053	
	(0.010)	(0.023)	(0.011)	(0.046)	(0.081)	
Gain [1.25,1.35]	0.067	0.049	0.055	0.041	0.009	-0.079
	(0.011)	(0.028)	(0.027)	(0.018)	(0.020)	(0.078)
Gain [1.35,1.45]	0.071	0.059	0.051	-0.165	0.217	-0.004
	(0.012)	(0.031)	(0.029)	(0.106)	(0.104)	(0.016)
N	13,704,178	13,704,178	13,704,178	13,704,178	13,704,178	13,704,178

Table A3: Effects of expected gains and losses on list prices

Notes: The regressions are similar to those in Table A1 and A2 but with different dependent variables, samples and controls for average local conditions.
In terms of dependent variables, columns 1 and 2 use WhenFresh/Zoopla list prices (l_t); columns 3 and 4 use a 0/1 variable indicating whether the property was sold in each month after it was advertised for sale on Zoopla; column 5 uses LR transaction prices and column 6 uses the log difference between Zoopla list prices and their final transaction price (for those properties that were sold).
Columns 1 and 3 are based on the sample of all Zoopla listings for which a previous purchase can be found on the LR. The other columns restrict this sample to those listings for which a subsequent sale can be found in the LR.
Because of the more limited size of the sample, we use price indices and fixed effects at the local authority level (δ_{jt} in equation (4)) and full-postcode fixed effects as the granular control for time-invariant property characteristics (v_{ij} in regression (4)).
Standard errors in parentheses are double-clustered at the year and local-authority level.

Dependent variable:	Listing price (l_t)		Sell prob (h_t)		Price (p_t)	Discount ($l_t - p_t$)
	All	Sold	All	Sold	Sold	Sold
	(1)	(2)	(3)	(4)	(5)	(6)
Gain [-.25,-.15]	0.002	0.008	-0.008	0.001	0.010	-0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)
Gain [-.15,-.05]	0.003	0.005	-0.007	-0.000	0.007	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Gain [.05,.15]	-0.005	-0.007	0.007	0.004	-0.008	0.001
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Gain [.15,.25]	-0.006	-0.011	0.012	0.009	-0.012	0.002
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)
Gain [.25,.35]	-0.006	-0.012	0.015	0.011	-0.015	0.003
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Gain [.35,.45]	-0.006	-0.013	0.015	0.008	-0.016	0.003
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Gain [.45,.55]	-0.008	-0.016	0.015	0.010	-0.020	0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Gain [.55,.65]	-0.009	-0.016	0.014	0.008	-0.021	0.005
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)
Gain [.65,.75]	-0.011	-0.018	0.012	0.011	-0.023	0.006
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)
Gain [.75,.85]	-0.012	-0.020	0.011	0.011	-0.026	0.006
	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.001)
Gain [.85,.95]	-0.012	-0.021	0.009	0.008	-0.026	0.005
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)
Gain [.95,1.05]	-0.012	-0.024	0.009	0.009	-0.029	0.005
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)
Gain [1.05,1.15]	-0.014	-0.022	0.011	0.013	-0.028	0.006
	(0.003)	(0.003)	(0.003)	(0.006)	(0.003)	(0.002)
Gain [1.15,1.25]	-0.015	-0.023	0.013	0.011	-0.029	0.005
	(0.003)	(0.004)	(0.004)	(0.006)	(0.004)	(0.002)
Gain [1.25,1.35]	-0.014	-0.024	0.012	0.011	-0.029	0.005
	(0.004)	(0.005)	(0.004)	(0.006)	(0.004)	(0.003)
Gain [1.35,1.45]	-0.013	-0.019	0.009	0.009	-0.026	0.008
	(0.005)	(0.004)	(0.006)	(0.006)	(0.005)	(0.002)
Gain [1.45,1.55]	-0.014	-0.027	0.007	0.015	-0.029	0.002
	(0.004)	(0.008)	(0.005)	(0.011)	(0.007)	(0.004)
Gain [1.55,1.65]	-0.008	-0.022	0.001	0.009	-0.021	-0.001
	(0.007)	(0.008)	(0.006)	(0.013)	(0.010)	(0.005)
Gain [1.65,1.75]	-0.002	-0.014	0.001	-0.009	-0.015	0.001
	(0.011)	(0.003)	(0.004)	(0.010)	(0.006)	(0.006)
N	2,597,866	1,126,859	13,778,554	5,256,126	1,126,859	1,126,859

B Matched-in data sources

B.1 Mortgage v cash additional LR variable

Information on funding of housing transactions can be purchased from the LR. The LR provides a file with complete address, price paid and Deed date, (but no transaction ID) which we watch to the publicly available LR dataset.

Figure B1 shows that the total number of cash purchases in England and Wales is less cyclical than the number of mortgages.

Table B1 shows some descriptive statistics for *Sample 2* grouping properties by funding source (mortgage or cash). Properties bought with cash are usually less expensive, except at the top of the price distribution (above the 99th percentile).

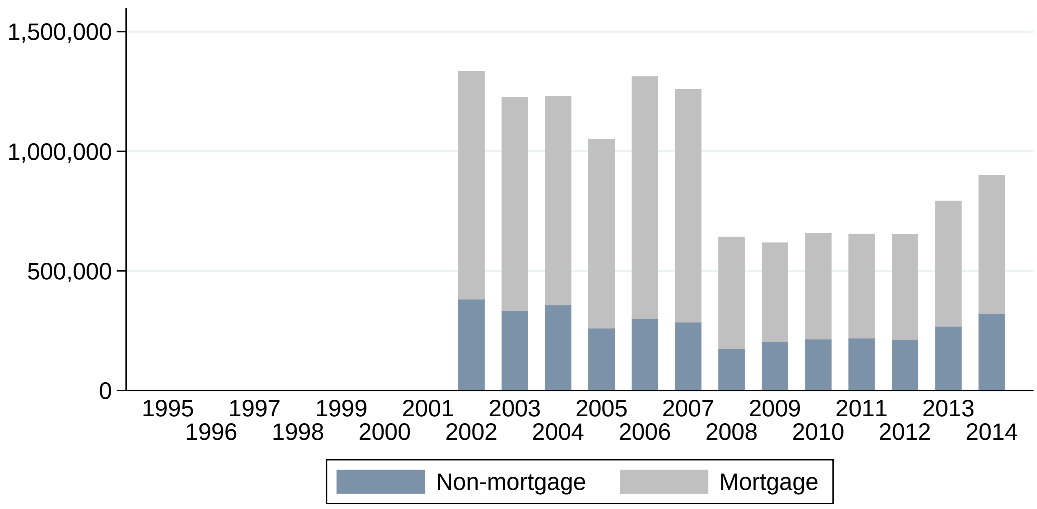


Figure B1: Mortgage vs non-mortgage purchases, 2002-2014

Notes: The bars represent the number of sales in the England and Wales Land Registry (LR) since information on the funding of housing transaction has been available (2002). This information is collected in a variable denoted 'charge', which indicates whether an additional ownership claim (on top of the owner's) is present on the property in question.

Table B1: Summary statistics: bought with a mortgage vs bought with cash

Notes: This table repeats the analysis of Table 1, focusing on *Sample 2* and contrasting properties that were bought with a mortgage with properties that were bought with cash.

	Previous purchase in 2002-2014	
	Bought with a mortgage	Bought with cash
Sales	2,299,688	899,701
Properties	1,941,359	811,728
Current sale price (p_{ijt})		
Mean	214,981	204,092
p1	49,500	27,000
p25	121,000	110,000
p50	168,950	159,950
p75	245,000	235,000
p99	925,000	940,000
Property type (proportion)		
Flat	0.22	0.25
Terraced	0.34	0.31
Semi	0.26	0.23
Detached	0.19	0.22
Lease	0.27	0.30
New	0.00	0.00
Expected Log log capital gains ($\widehat{GAIN}_{jst}$)		
Mean	0.18	0.16
Median	0.14	0.10
p01	-0.16	-0.16
p10	-0.04	-0.03
p90	0.47	0.46
p99	0.75	0.75
Years btw previous purchase and current sale (DUR_t)		
Mean	3.74	3.13
p01	0	0
p10	1	0
p50	3	2
p90	8	8
p99	11	11

B.2 Mortgage information from the Product Sale Data

To match in information on mortgages from the PSD to the LR we perform a record linkage exercise between the two datasets.

Data preparation As a preliminary step, we restrict the PSD to initial mortgages and exclude remortgages; we limit the sample to England and Wales and exclude Scotland and Northern Ireland. These exclusions leave us with a dataset of 6.2m observations between 31 March 2005 (the start day of the PSD data collection) and 31 December 2014 (the end of the sample analyzed in this paper). We call this dataset *Relevant PSD*. In the same period, the LR contains 8.3m observations. Since we can identify which LR sales were funded with a mortgage, we restrict our attention to those, leading to a reduction of the relevant LR observations to 6.3m, a number similar to the size of the *Relevant PSD*.

The LR contains information on:

- sale price
- address
- sale date (completion)
- type of property

The PSD variables that could be related to LR information are:

- sale price or property value
- postcode
- date of mortgage account opening
- type of property.

In the *Relevant PSD* The sale price variable is missing for 2.3m sales, but the property value variable is missing for only 554 observations. Comparing sale price with property value for records were both of these are non-missing reveals that the two numbers coincide most of the times; hence we create a new price variable which equates the purchase price when it is available, and the property value otherwise. In theory, the price variable should match with the corresponding sale price in the LR. In practice, in a preliminary analysis we tabulated all the specific values of price found in the PSD, compared them with all the individual sale prices found in the LR, and found that around 30% of price values found in the PSD are not found in the LR.[1]

The postcode variable is never missing in the PSD. As a preliminary step in the analysis, we found that around 90% of postcodes found in the PSD are found in the LR—a better result than the one on prices.[2]

The date in which a bank transfer the mortgage amount to the buyer is the completion date or a few days before. Figure B2 shows that, on a monthly scale, there is a 1:1 relation between observations in the LR and the PSD.

Finally, data on property type are missing for 40 percent of the observations in the PSD, hence we do not use them for the matching.

Data matching We assign an ID to every combination of postcode, date, and price in the LR and the PSD.[3] We proceed in steps, from the best matches to less precise ones:

1. We first select observations that match on all three variables (postcode, date, and price)—there are 1.5m of them. We create a variable indicating matching quality and assign these observations the maximum value (4). We then remove their IDs from the list of LR and PSD observations to be matched.

2. We select observations that match on postcode and price, which sometimes results in multiple matches (the same combination of postcode and price can be associated with different dates). For each LR ID, we select the observation where the distance

[1]Manual inspection of those prices revealed no noteworthy pattern. Their distribution was similar to the price distribution in the LR.

[2]Again, manual inspection of non-matching postcodes revealed no noteworthy pattern.

[3]There are around 60,000 duplicates in postcode, date, and price in both the LR and the PSD, corresponding to 1 percent of observations. We eliminate duplicates before proceeding.

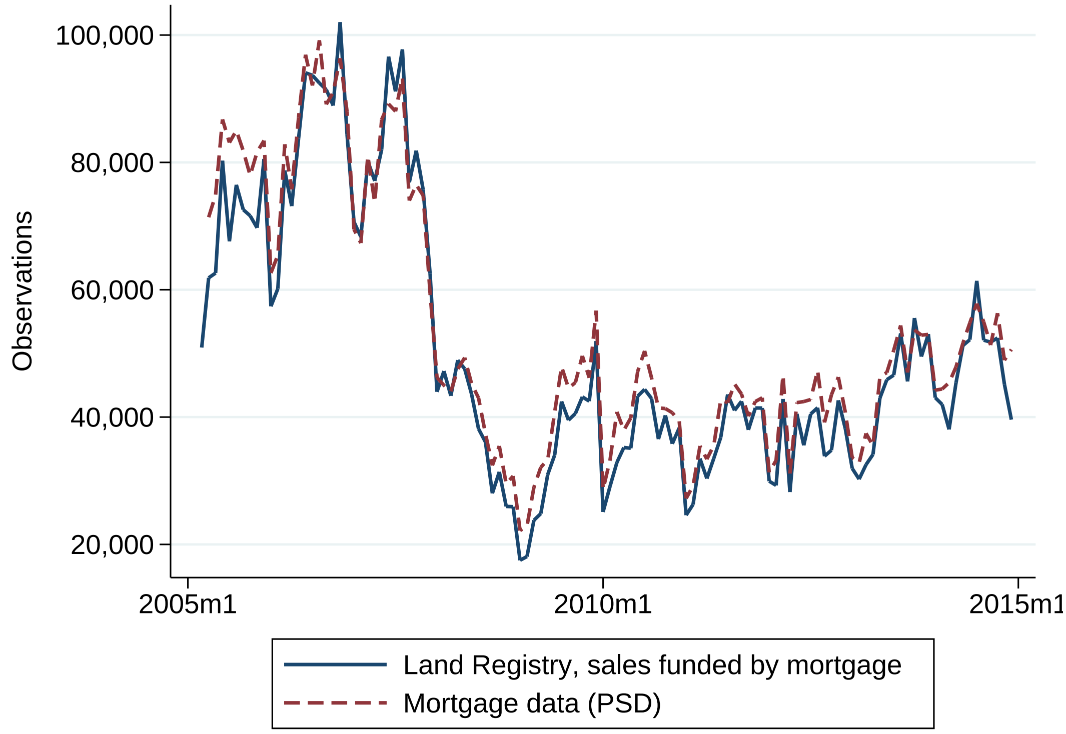


Figure B2: Number of observations by month in the Land Registry and Product Sale Data before matching

Notes: The Land Registry (LR) sample is made of all England and Wales registered sales between March 2005 and the end of 2014. The PSD sample is made of all mortgages for house purchase (excluding remortgages) in England and Wales for the same period. (The PSD started to collect data on mortgages on April 1st, 2005. We keep March 2005 sales in the LR because we allow for a maximum difference of 30 days, in both directions, between the sale date in the LR and the mortgage starting date in the PSD.)

between the LR and PSD date is the lowest, limiting the selection to instances where this distance does not exceed 30 days. We do the same for each PSD ID. Once we have a group of uniquely matched IDs (in this case, 2.5m sales), we assign them match quality 3 and remove them from the list of IDs that still need to be matched.

3. We select observations that match on postcode and date. We eliminate duplicate IDs similarly to the previous step, by selecting for each ID the observation where the percentage difference between the LR and PSD price is the lowest, limiting

the selection to differences of plus or minus 10 percent. This step of the process produces 150,000 additional matches with match quality 2.

4. Finally, we create all the combinations of the remaining observations that match on postcode only. Within duplicates observations of the same ID, we select the observation with the lowest date difference. If there are ties, we select the observation with the lowest price difference. All the observations where the differences between variables exceed the thresholds (30 days for dates, 10 percent for prices) are eliminated. This step produces 270,000 additional matches with quality 1.[4]

There are in total 4,540,412 matched sales, which correspond to 73 percent of all PSD mortgages. In the paper, we show results based on matches with qualities from 4 to 1. Running the analysis only on matches with quality 4 to 3 yields almost identical results (this group corresponds to 90 percent of matched properties).

Descriptive statistics of matching results Table B2 shows the characteristics of properties in *Sample 3* (transaction price analysis). The aggregate statistics for this sample are showed in the third column of the upper half of Table 1; this table splits the sample into four groups: properties that match with the PSD and were purchased with an initial LTV greater than 80 percent, properties that match with the PSD and were purchased with an initial LTV lower or equal to 80 percent, properties that the LR indicates as having been purchased with a mortgage but that do not match with the PSD, and properties that according to the LR were bought with cash. In general, properties purchased with a higher LTV are cheaper and have longer holding periods.

Figure B3 shows the distribution of mortgage LTVs in the relevant PSD dataset, the subset of observations that match with the Land Registry, and the observations belonging

[4]This matching algorithm is implicitly assuming that postcodes exactly match. In other words, we have not made any attempt to allow for errors in postcodes. To check whether these errors are likely to be relevant, we joined the two datasets on price and date and then compared the postcodes in the LR and PSD. If errors in postcodes were a relevant issue, we would expect to see several instances among the combined observations where postcodes in the two datasets were similar but not identical. A visual inspection of these observations revealed no such instances in the first 100 rows of the dataset.

to *Sample 3* used in the transaction price analysis. Spikes are apparent next to important LTV values such as 75, 80, 85, 90 and 95 percent. This bunching is due to the way in which UK mortgages are priced (see Best et al, 2020).

Additional references

Best, M. C. J. Cloyne, E. Ilzetzki, and H. Kleven (2019). "Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches," *The Review of Economic Studies*, forthcoming.

Table B2: Summary statistics: *Sample 3* subgroups generated by Land Registry-Product Sales Data match

Notes: This table repeats the analysis of the upper half of Table 1, focusing on *Sample 3* and distinguishing between the four subgroups of sales which derived from the Land Registry (LR)-Product Sales Data (PSD) match. The first two groups refer to repeat sales where the previous purchase matches with a PSD mortgage: properties that were bought with a high LTV (>80%) and properties that were bought with a low LTV. The third and the fourth group refer to repeat sales where the previous purchase does not match with a PSD mortgage: either properties that according to the LR were purchased with a mortgage (third column) or properties that according to the LR were bought with cash (fourth column).

	Sample 3 (previously purchased in 2005-2014)			
	Matched		Not matched	
	Bought with LTV>80%	Bought with LTV≤80%	Bought with Mortgage	Bought with Cash
Sales	377,241	366,426	237,134	404,852
Properties	362,682	354,297	230,259	381,419
Current sale price (p_{ijt})				
Mean	204,169	269,705	232,902	222,231
p1	60,000	68,000	43,000	41,000
p25	123,500	150,000	117,000	120,000
p50	166,000	208,000	167,500	168,950
p75	239,960	300,000	250,000	248,000
p99	765,000	1,250,000	1,300,000	1,100,000
Property type (proportion)				
Flat	0.24	0.17	0.30	0.26
Terraced	0.39	0.29	0.34	0.28
Semi	0.26	0.29	0.22	0.24
Detached	0.10	0.26	0.15	0.22
Lease	0.28	0.20	0.35	0.31
New	0.00	0.00	0.00	0.00
Expected Llog capital gains ($\widehat{GAIN}_{jst}$)				
Mean	0.05	0.04	0.04	0.03
p1	-0.19	-0.19	-0.19	-0.18
p25	-0.03	-0.04	-0.02	-0.02
p50	0.04	0.03	0.03	0.01
p75	0.11	0.10	0.10	0.07
p99	0.44	0.43	0.47	0.38
Years btw previous purchase and current sale (DUR_t)				
Mean	3.82	3.60	2.81	2.51
p1	0	0	0	0
p25	2	2	1	0
p50	4	3	2	2
p75	6	5	5	4
p99	8	8	8	8

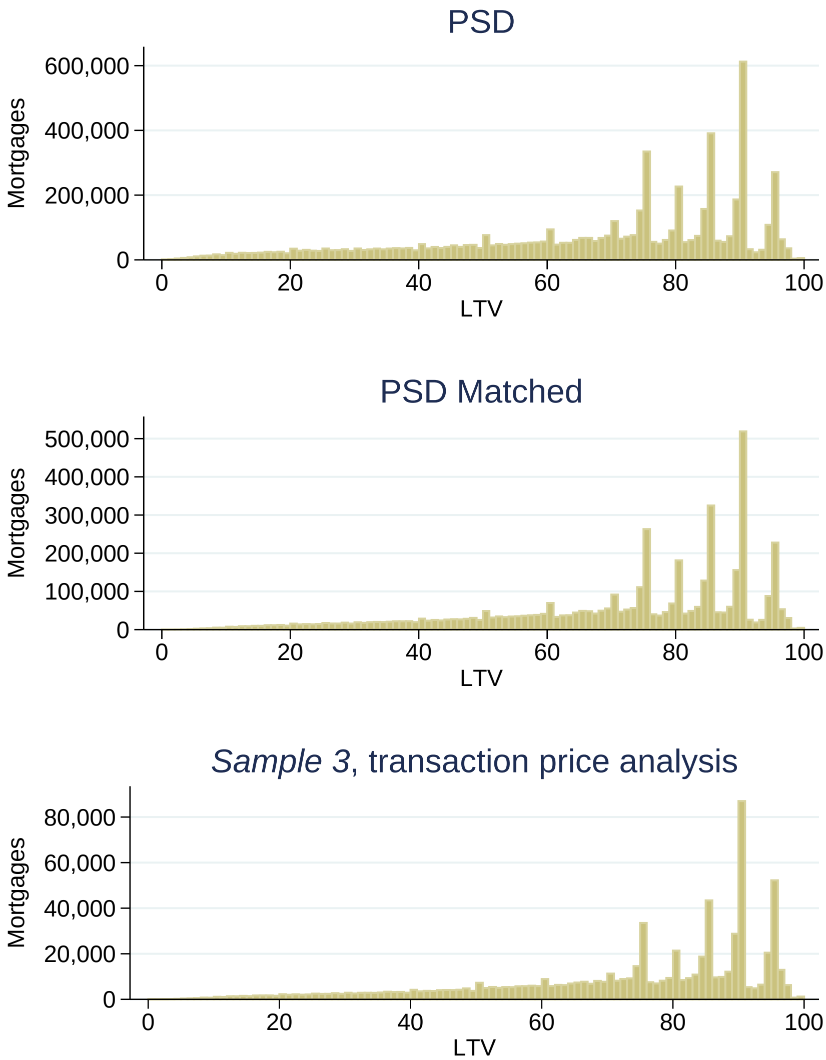


Figure B3: LTV distributions in the Product Sales Data and the matched observations

Notes: The top chart reports the distribution of loan-to-value (LTV) ratios of mortgages for house purchases in the Product Sales Data (PSD), which covers the universe of homeowner mortgages since April 2005. The middle chart refers to the mortgages that match a sale in the Land Registry (LR) according to the matching algorithm described in Appendix B.2. The bottom chart reports the distribution of LTVs for purchases of properties that belong to *Sample 3* in the analysis of LR transaction prices in this paper.

B.3 Whenfresh/Zoopla data

The raw data is provided by data company WhenFresh and corresponds to all listings appeared on property portal Zoopla. For each listing we would like to know:

1. whether the previous purchase of the property is on the LR, and
2. whether the listing attempt successfully resulted in a subsequent sale recorded in the LR.

We perform two matches, which we call *match 1* and *match 2*, corresponding to the two objectives above. (An alternative and equivalent approach would be to perform just one of the Zoopla-LR matches and then retrieve the other matches by exploiting repeat sales in the LR).

Data cleaning We initially restrict the dataset to sale listings in England and Wales with a complete address which appeared on the website in 2009-2014[5]—this corresponds to 6,861,663 observations. Excluding listings where the creation date is after the deletion date or where the initial price or the number of bedrooms are missing brings the number of observations to 6,770,311. In order to avoid duplicates, we eliminate listings on the same address happening before 180 days of the first one—ending with 4,405,445 listings. Furthermore, to avoid outliers we eliminate listings corresponding to the first and 99th percentile of the list price distribution. We have now 4,317,919 listings to be matched with the LR.

Data matching Property addresses in the WhenFresh/Zoopla do not have the same format as addresses in the LR. Moreover addresses are provided to Zoopla by estate agents and may occasionally contain errors.

After trying different matching approaches, we obtained the best performance by requiring an exact match on (1) the two postcodes (the one in the LR and the one in

[5]Zoopla was launched in November 2008 but given that most of our specifications are based on local authority × year fixed effects, 2008 observations are too sparse to be used.

the WhenFresh/Zoopla dataset) and (2) the first part of the address, which corresponds to the street number for a house and the apartment number for a flat. The combination of these two variables is likely to identify a unique property,[6] allowing us to sidestep the problem of complete addresses being written in different formats.

The combination of property address and listing date identifies a listing in the WhenFresh/Zoopla dataset. After having joined the two dataset through postcode and the first part of the address, duplicates in listings and LR sales still exist. In the context of *match 1*, we eliminate all combinations where the listing date occurs before the LR date, and then we choose the match where the two dates are closest—we end up with 2,610,073. For *match 2*, we only keep combinations where the listing date occurs before the LR sale date and keep the observations where the distance in days between the two days is shortest. Furthermore, we eliminate all instances where the sale occurred more than one year after the first listing, because it becomes less clear whether these two events should be grouped together as the same sale attempt.

[6] A complete UK postcode identifies around 10-15 units. In theory, for postcodes encompassing more than one street, the combination postcode-street number would not be sufficient to identify a unit; a similar issue would occur for two apartment small buildings being located in the same postcode and using the same apartment numbering convention. In practice, visual inspection of the matching results demonstrated that these instances are extremely rare, at least within the group of observations and the time frame which are relevant for us.

C Comparison with Genesove and Mayer (2001, GM) empirical methodology

In this section of the Appendix we show results using the GM specification for expected gains, $\widehat{GAIN}^{GM}_{ijst} = \hat{\delta}_{jt} - \hat{\delta}_{js} - w_{ijs}$. In the GM setting, the gain measure is specific to the property i because it includes the error estimated in the first stage, w_{ijs}.

The original GM paper does not include property-specific fixed effects. In that setting, the role of unobserved property characteristics that affect both the first and second transaction is likely to be prominent. This makes the inclusion of w_{ijs} of primary importance.

In a setting with property-level fixed effects, the case for the inclusion of such term is less straightforward. Consider the main estimating equation of the paper with the GM measure:

$$p_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + f(\widehat{GAIN}^{GM}_{ijst}) + \eta_t \otimes \eta_s + \lambda_t \otimes \lambda_s + w_{ijt}. \tag{1}$$

This is a dynamic panel model which is prone to bias as in Nickell (1981). A high price in the second sale translates into a high fixed effect v_{ij}, a high error w_{ijt} and a low $\hat{w}_{ijs}$. Hence $\widehat{GAIN}^{GM}_{ijst}$ becomes correlated with the error. When we run this regression we get the effects labeled as Model I in Figure C1. The effects in the price regression are large and in the opposite direction with respect to those in the paper (larger gains are associated with higher prices). The effects in the selling propensity regression replicate the pattern in the paper; quantitatively, the coefficients are a bit larger than in our specification.

While not specifically referred to a fixed-effect estimation as here, the discussion in GM (p.1239-1241) contemplates the possibility of a bias due to the presence of $\hat{w}_{ijs}$ in $\widehat{GAIN}^{GM}_{ijst}$. For this reason, GM suggest an alternative model (Model II) in which $\hat{w}_{ijs}$ is

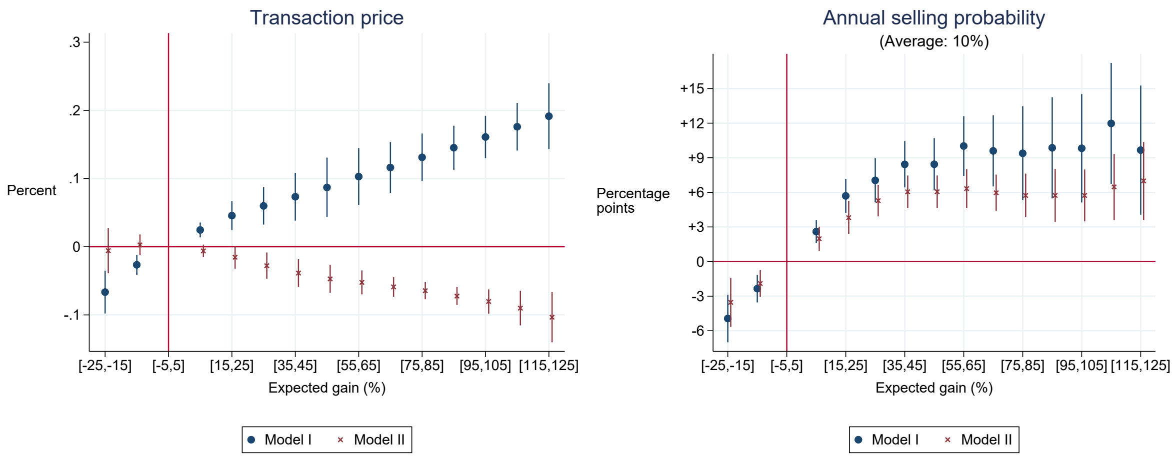


Figure C1: Results adding GM residuals with property fixed effects

explicitly controlled for in the regression:

$$p_{ijt} = v_{ij} + X_{ijt}\beta + \delta_{jt} + f(\widehat{GAIN}^{GM}_{ijst}) + \gamma \hat{w}_{ijs} + \eta_t \otimes \eta_s + \lambda_t \otimes \lambda_s + w_{ijt}. \tag{2}$$

If we run such an estimation we get the coefficients labeled as Model II in Figure C1, which are more similar to the ones reported in the main estimation of the paper, except perhaps the lack of a significant positive effect on prices for losses. This result shows that the bias from the presence of the error term affects the price regression more than the selling probability regression.

If our intuition regarding the additional bias introduced by the residual when the regression has property fixed effects is correct, results should be more in line with those of the paper if we were to re-run the estimation without property fixed effects. Figure C2 shows the results of running Model I without fixed effects in the first-stage estimation of the price indices and in the second-stage estimation of the main effects. In accordance with our story, the pattern of the coefficients in the price regression is now consistent with the results in the paper (larger gains are associated with lower prices) both in the price and selling propensity regression, but it is still the case that estimated effects for Model I are significantly larger than with our specification.

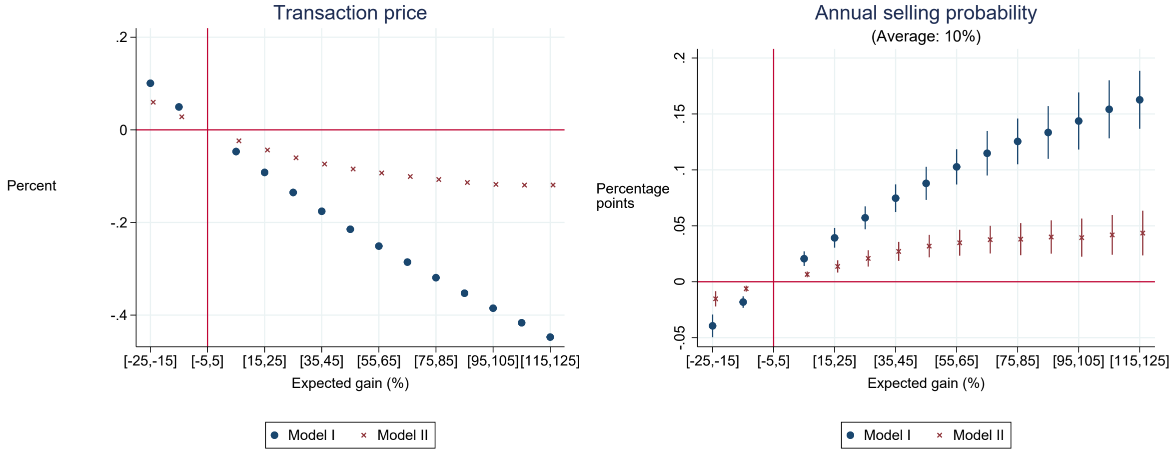


Figure C2: Results adding GM residuals in a setting without property fixed effects

Additional references

Nickell, S. (1981). "Biases in dynamic models with fixed effects," *Econometrica*, 49(6): 1417-1426.