BANK OF ENGLAND

Staff Working Paper No. 630

History dependence in the housing market

Philippe Bracke and Silvana Tenreyro

July 2018

This is an updated version of the Staff Working Paper originally published on 2 December 2016

BANK OF ENGLAND

Staff Working Paper No. 630

History dependence in the housing market

Philippe Bracke[1] and Silvana Tenreyro[2]

Abstract

Using data on the universe of housing transactions in England and Wales over a 20-year period, we document a robust pattern of history dependence in housing markets. Sale prices and selling propensities are affected by house prices prevailing in the period in which properties were previously bought. We investigate the causes of history dependence complementing our analysis with administrative data on mortgages and online house listings, which we match to actual sales. We find that cognitive and financial frictions explain the history dependence in the data. Both contributed to the collapse and slow recovery of the volume of housing transactions in the post-crisis period.

Key words: Housing market, fluctuations, down-payment effects, reference dependence, anchoring, loss aversion.

JEL classification: R31, E32, D03.

(1) Bank of England. Email: philippe.bracke@bankofengland.co.uk
(2) London School of Economics, Centre for Macroeconomics, and CEPR. Email: s.tenreyro@lse.ac.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. For helpful comments, we would like to thank Felipe Carozzi, Francesco Caselli, Andreas Fuster, Hans Koster, Per Krusell, Roger Loh, Benedikt Vogt and conference and seminar participants at the Bank of England, Bureau for Economic Policy Analysis (The Hague), European Economic Association (Geneva), Ghent Workshop on Empirical Macroeconomics, Lancaster University, LSE, Money Macro and Finance Conference (King's College London), Norges Bank Workshop on Housing and Household Finance, National University of Singapore, Royal Economics Association (Bristol), Society of Economic Dynamics (Edinburgh), Tinbergen Institute, and Urban Economics Association (Vancouver). Tenreyro acknowledges financial support from the ERC Consolidators Grant 2016.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/working-paper/staff-working-papers

Publications and Design Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

This paper documents a pattern of history dependence in house prices and transactions by studying the universe of housing sales in England and Wales over a twenty-year period. Specifically, house prices in the year a house was previously bought influence the individual price at which the house sells next, as well as the owner's propensity to sell. The results are based on twenty million housing sales and are not driven by changes in the composition of the houses transacted. We complement our analysis with matched administrative data on mortgages and on-line house listings. The effects of history dependence on house prices and the probability of sale can be material. Consider two identical houses in the same location in 2014, one previously acquired in 2007, when aggregate prices peaked and the other in 2001.[1] Our results show that, all else equal, the house bought in 2007 will carry a price premium of 5 percent over the one bought in 2001. Moreover, the house bought in 2007 will have, on average, 50 percent less chance of selling. (We control for tenure duration so the results are not driven by shorter durations in the more recent period.)

In aggregate, history dependence has the potential to contribute to the persistence in prices and the pronounced volatility in sales volumes that we observe in housing markets. History dependence is clearly at odds with a frictionless model in which the value of a house and its liquidity depend exclusively on the future stream of dividends (rental value) the property delivers. Two types of frictions can help us explain the presence of history dependence.

The first is credit frictions, among which a leading explanation is the so-called down-payment effect, a mechanism proposed by Stein (1995). For repeat buyers, a large percentage of their down payment comes from the sale of their previous homes, and, importantly, a majority of home sales are to repeat buyers. Hence, owners who bought at high prices will have, all else equal, limited home equity; they will then have higher reservation prices

[1] On average, the owner of the first house will expect a 2% loss in nominal terms, whereas the owner of the second property will be facing a potential capital gain of 89%.

and be less likely to sell than owners of comparable houses bought at lower prices, as they have less money left after their property sale.

The second type is cognitive frictions and include mechanisms such as anchoring and learning. The notion of anchoring or reference dependence goes back to Tversky and Kahneman (1982) and builds on a well-established result from laboratory experiments: in estimating the value of an asset agents tend to show a bias that overweighs possibly irrelevant initial cues. In the context of the housing market, sellers may give excessive weight to the price they paid (vis-à-vis the market evolution of prices) when posting new prices; if they bought at high prices, this will lead to higher advertised prices and more time in the market. A particular kind of reference dependence is loss aversion, whereby losses have greater impact on preferences than gains (Tversky and Kahneman, 1991). With learning, reservation prices are updated slowly following specific rules as in Davis and Quintin (2016). In this framework history dependence arises because the previous purchase price of a property is an important prior in evaluating its current value.

To disentangle the two groups of mechanisms, we study a sample of properties previously bought exclusively with cash, for which the down-payment effect should be muted. We find evidence of history dependence on prices in this cash-only sample. However, for these properties we find limited evidence of history dependence on selling propensities; our results seem to be mostly driven by properties bought with leverage. We measure leverage both along the extensive margin (whether the property was bought with a mortgage) and the intensive margin (the loan-to-value ratio at purchase). This evidence is consistent with a role for a down-payment effect.

Understanding history dependence is a first step to inform the design of policies aimed at preventing or reacting to future crises. In the context of the UK economy, the post-crisis period led to a collapse in the volume of transactions, illustrated in Figure 1. Transactions reached their peak in 2007 and then declined sharply. Prices reached their peak slightly afterwards, subsequently fell, and only after 2009 experienced a recovery. We investigate the quantitative implications of history dependence for the post-crisis

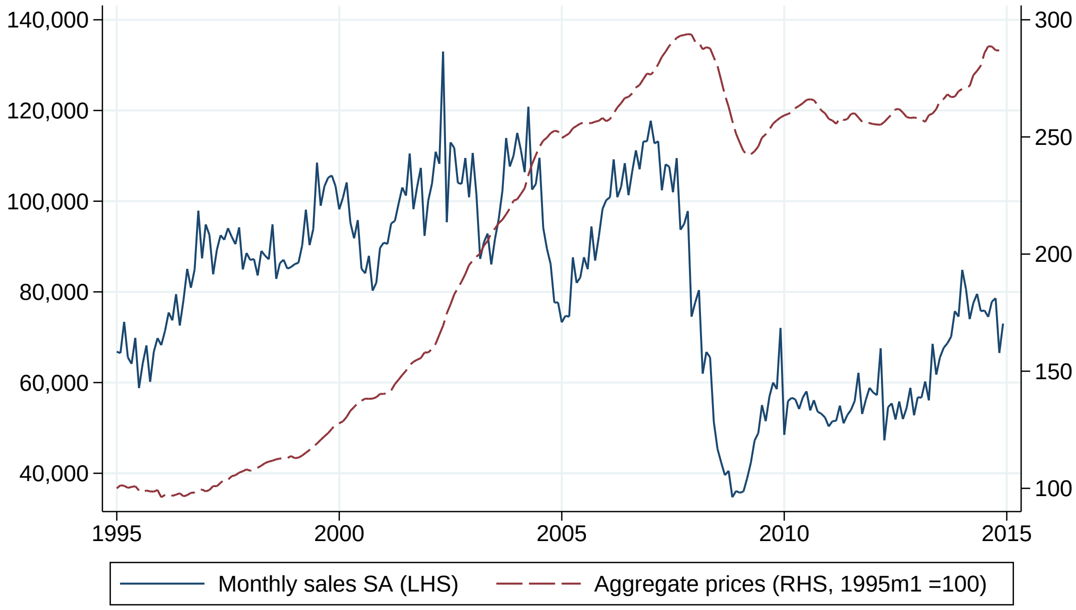


Figure 1: Monthly house prices and sales, England and Wales

Notes: The figure shows the monthly quality-adjusted average price and the monthly total number of transactions in England and Wales over 1995-2014. Data are taken from the England and Wales Land Registry and quality-adjusted through an hedonic regression as described in Section 3.

recovery of the housing market for different regions in England and Wales and measure the relative strengths of the mechanisms at play.

Related literature On conceptual grounds, our paper builds closely on the seminal contributions of Stein (1995) and Tversky and Kahneman (1982), both providing the foundations for the underlying mechanisms behind history dependence that we analyze,[2] and more recently on the literature exploring learning in a housing context (Anenberg, 2015; Davis and Quintin, 2016). On empirical grounds, our paper relates to the seminal work of Genesove and Mayer (2001), who find strong evidence of loss aversion in the context of the Boston condominium market between 1990 and 1997. The authors report significant effects of loss aversion on list prices and time on the market and slightly less

[2] Ortalo-Magne and Rady (2006) also explore the consequences of down-payment constraints in a theoretical model. Head et al. (2018) propose a dynamic model with housing search and defaultable mortgages that produces a positive relation between outstanding debt, asking prices and time-to-sell.

sharp effects on transacted prices. They also find a small role for down-payment effects. Relatedly, Anenberg (2011) analyzes the San Francisco Bay Area housing market and reports significant effects of loss aversion and leverage on transacted prices. We add to the evidence provided by these two studies in a number of important ways. First, we uncover the presence of history dependence for the universe of housing transactions in an important market outside the US. Second, we investigate the quantitative implications of history dependence and its underlying channels on the aggregate volume of transactions, through our estimation of selling propensities. Third, we analyse history dependence on the whole range of gains and losses (rather than focusing on loss aversion only) and show significant and concave effects for expected gains, consistent with prospect theory. Fourth, our novel focus on properties bought with cash allows us to convincingly argue that not all history dependence is due to credit frictions.

In a recent contribution, Guren (2017) examines the relation between local house price appreciation and list price, and use it as an instrument to study the relation between list price and time on the market. In another recent paper, Hong et al. (2016) find some suggestive evidence in the Singaporean condominium market of a kink in the selling propensities at zero gains consistent with realization utility (Barberis and Xiong, 2012). Despite the differences in scope and markets studied, our paper finds strong evidence of cognitive frictions in line with Beggs and Graddy (2009), who study price anchoring in art auctions of Modern, Impressionist, and Contemporary paintings in London and New York (the authors do not study selling propensities). In focusing on the role played by leverage in explaining economic activity, we join a vast literature that has documented the adverse effects of financial frictions during the crisis and post crisis recovery. (See, for example, Mian and Sufi, 2009, and the references therein.)

The gyrations in the housing market of the recent years have stimulated a number of studies on the relation between house prices and mobility, in which the role of financing and cognitive frictions is often critical. Two examples in that line of research are Engelhardt (2003) and Ferreira et al. (2012) for the US economy. Their focus is on household mobility with an eye on its labour market consequences. In this paper, we focus specifically on housing sales, but clearly they would have repercussion for the mobility of households.

The rest of paper is organized as follows. Section 2 describes the methodology. Section 3 presents the data and documents the patterns of history dependence. It next studies the potential channels underlying history dependence and their quantitative relevance. Section 4 contains a similar analysis on house listings from a major UK online property portal matched to the database on actual property sales, where we can examine history dependence in list prices and time on market. Section 5 presents concluding remarks. The Appendix contains additional material to complement the information in the text, as well as a disaggregated analysis of the England and Wales' regional housing markets.

2 Identifying history dependence

The (log) house price is usually modeled as:

$$p_{it} = v_i + X_i\beta + \delta_{jt} + w_{it}, \tag{1}$$

where p_{it} is the transaction price of house i sold at time t in local area j, v_i is a property-specific fixed effect capturing time-invariant features, X_i is a vector of (time-varying) housing characteristics, δ_{jt} is the aggregate house price level at time t in local area j where i sits, and w_{it} is an idiosyncratic error component which contains both unobserved (time-varying) property characteristics and idiosyncratic price effects due to the features of specific transactions.

To study history dependence we augment the standard hedonic regression above with a function of the difference between today's expected sale price, $\hat{p}_{it}$, and the house's previous transaction price p_{is}:

$$p_{it} = v_i + X_i\beta + \delta_{jt} + f(\hat{p}_{it} - p_{is}) + w_{it}, \tag{2}$$

where s denotes the period when the house was previously purchased. The practical implementation of (2) hinges on the definition of $\hat{p}_{it}$; in other words, on how owners assess the expected value of their property. A simple approach is to assume that owners apply to the purchase price they paid at time t the appreciation of the local price index between s and t: $\hat{p}_{it} = p_{is} + (\hat{\delta}_{jt} - \hat{\delta}_{js})$. Equation (2) becomes:

$$p_t = v_i + X_i\beta + \delta_{jt} + f(\widehat{GAIN}_{jst}) + w_{it}, \tag{3}$$

where $\widehat{GAIN}_{jst} = \hat{\delta}_{jt} - \hat{\delta}_{js}$ is the (log) difference in aggregate house prices between time t and when the property was bought (s).[3] Notice that these are expected, rather than realized, gains. To estimate the effect of gains and losses in a non-linear, non-parametric way, we split $\widehat{GAIN}_{jst}$ into equally-sized bins for the different magnitudes of expected gains/losses (ie losses between -0.25 and -0.15 per cent, between -0.15 and -0.05 per cent, and so on).

To measure the effect of history dependence on selling propensities, we start from an equation similar to (3) but with a 0/1 indicator as dependent variable. This indicator takes the value one when the property was sold in a given year, and zero otherwise. Using this approach, a property appears in the dataset each year after its first registered sale (we cannot compute the ongoing $\widehat{GAIN}_{jst}$ before this first sale).

Figure 1 reveals that, for most of the sample period, England and Wales house prices have been trending upwards. Keeping current sale year constant, such a trend leads to a correlation between property tenure and $\widehat{GAIN}_{jst}$. To control for this effect, we could

[3] In Genesove and Mayer (2001) $\widehat{GAIN}_{jst} = \hat{\delta}_{jt} - \hat{\delta}_{js} - w_{is}$, where w_{is} is the error coming from estimating (1) on the previous purchase of the property. Because their specification does not use repeat sales, this term includes time-invariant property characteristics not captured by their hedonic model. In our methodology, w_{is} only includes time-varying characteristics of the property or noise specific to its previous transaction; the two measures of $\widehat{GAIN}_{jst}$ have a correlation of 92%.

insert in the regression the duration of the tenure, measured as the number of years between two sales. Such a variable would capture the time-invariant effect of tenure, but would not address the potential change in tenure effects over the twenty years covered by our sample. This is more easily seen in terms of selling propensities: any change in the mobility rate of households over time would have an impact on the tenure effect. We therefore control for all possible combinations of current year (denoted with dummy variables λ_t) and year of purchase (denoted with λ_s):

$$y_{it} = v_i + X_i\beta + \delta_{jt} + f(\widehat{GAIN}_{jst}) + \lambda_t \otimes \lambda_s + w_{it}, \tag{4}$$

where y denotes either p, the log price, or q, the sale indicator. Our measure of gains and losses, estimated at the local level j, is still identified. The term $\lambda_t \otimes \lambda_s$ has the added advantage of controlling for time-varying unobserved property characteristics that are homogeneous across England and Wales between a given pair of years. This additional control could end up absorbing a substantial amount of variation in $\widehat{GAIN}_{jst}$ but, in practice, we show that results are similar whether we include it in the regression or not.

Our coefficient of interests on $\widehat{GAIN}_{jst}$ could still be biased by other time-varying property characteristics not captured by $\lambda_t \otimes \lambda_s$. For instance, a possible correlation between home improvements and house prices (as in Choi et al., 2014) would affect the analysis, to the extent that the rate at which properties are renovated or extended differs across postcode districts. To address this remaining threat, we run a robustness check on the subsample of properties that have never been extended.[4]

Mechanisms History dependence could be driven by credit or cognitive frictions. To disentangle the two mechanisms, we look for sellers in the data that are likely to be credit constrained in their next house purchase. (The next section explains how we implement this in practice.) We denote this group of borrowers with an indicator variable, *constr*,

[4]England and Wales Energy Perfiormance Certificate data for residential properties record whether an extension has been carried out on a house or apartment.

and run the following regression:

$$y_{it} = v_i + X_{it}\beta + \delta_{jt} + \underbrace{f\left(\widehat{GAIN}_{jst}\right)}_{\text{cognitive frictions}} + \underbrace{f\left(\widehat{GAIN}_{jst}\right) \times constr}_{\text{credit frictions}} + constr + \lambda_t \otimes \lambda_s + w_{it}. \quad (5)$$

The non-interacted term, $f\left(\widehat{GAIN}_{jst}\right)$, captures the effect of history dependence that are common to all properties, independently of whether the owner is credit constrained. We therefore take it as a measure of the part of history dependence that depends on cognitive frictions. The part that depends on credit frictions is captured by the interaction $f\left(\widehat{GAIN}_{jst}\right) \times constr$.

3 History dependence in transaction prices and selling propensities

The first part of this section describes our main data source, the England and Wales Land Registry (LR), which contains twenty years of residential transactions from January 1995 to December 2014. We explain how we compute our measure of local aggregate house prices and how we construct our two estimation datasets—one to analyze transaction prices and one to analyze selling propensities. We then show the results for history dependence and explore its quantitative relevance.

3.1 Data and summary statistics

The LR records all residential property transactions, with few exceptions:[5] The dataset contains close to twenty million sales for twenty years of data, that is, approximately one million sales per year. For each sale, the LR contains the precise postcode, the street name, the street number, and the apartment number if the property belongs to

[5] The exceptions are listed at http://www.landregistry.gov.uk/market-trend-data/public-data/price-paid-data, where a public version of the dataset is available. Most of the excluded transactions refer to sales that were not for full market value, for examples a transfer between parties on divorce.

a multi-unit building. The LR records three attributes of the property: its type (flat, terraced, semi-detached, detached); whether the property is new; and the tenure type of the property (freehold or leasehold).[6] Date of Transfer in LR is the day written on the transfer deed, that is, the date of completion, when keys and funds change hands.

Before analyzing history dependence, we use the LR to compute the aggregate level of house prices needed to create the $\widehat{GAIN}_{jst}$ variable. We do so at the postcode district level, by running regression (1) at an annual frequency for each postcode district in England and Wales. Our procedure has therefore two steps: we first estimate equation (1) to compute expected gains and losses and then estimating (4) to compute the effect of interest.[7] Our dataset includes 2,345 postcode districts; the average postcode districts contains around 10,000 individual addresses. We keep our analysis at the annual frequency because this allows a more straightforward analysis of selling propensities, for which we need to expand our dataset in a property-by-year format. Because the literature has highlighted the seasonality of the housing market (Ngai and Tenreyro, 2014), we add quarterly dummies to our main specification.[8]

Analysis of transaction prices Our empirical analysis relies on the identification of repeat sales. We consider two sales as happening on the same property when they share the same postcode, street name, street number, apartment number (if any), and property type (flat, terraced, semi, detached). Transaction prices from repeat sales allow us to create both a measure of realized gains ($GAIN_t$) and a measure of expected gains for

[6] A leasehold is a tenancy arrangement by which someone buys a property for a limited number of years, usually 99, 125 or 999. It is usually associated with flats and is a time-varying characteristic of the property, because leaseholds can be converted to freeholds (see Giglio et al., 2015; Bracke et al., 2018). As such we use a dummy indicator for leasehold tenancy as part of our control variables.

[7] In our baseline estimates, the first step does not include any measure of gains and losses, as these are not taken into account by market participants when estimating house price indices at the local level. In the Appendix, we show results from an alternative setup which incorporates history dependence in the first step—we run an iterating procedure to make $\hat{\delta}$'s and $\widehat{GAIN}$'s consistent across the two steps. This alternative setup yields similar results to our baseline methodology. In the Appendix we also show that results maintain their statistical significance when standard errors are bootstrapped to take into account that $\widehat{GAIN}_{jst}$ is a generated regressor.

[8] We include both quarter of purchase and quarter of sale, and interact them, resulting in 16 (4 by 4) combinations of dummies.

Table 1: Summary statistics, analysis of transaction prices

Notes: The analysis of transaction prices is based on microdata from the England and Wales Land Registry (LR) for the years 1995-2014. The first column contains information on all the sales included in the LR. The second column describes the sample used in the analysis: it is made of all properties which have at least two sales in the dataset, and excludes for each property the first of such sales. (The first sale provides us with the previous price or the previous aggregate price index to include in the regression that checks for history dependence.) The third column is similar to the second but only refers to properties whose first sale took place after 2001. For this sample we can tell whether the property was purchased with a mortgage and investigate the mechanisms behind history dependence. Finally, the fourth column describes properties whose first sale took place after March 2005 and can potentially be matched to the Product Sales Data (PSD), a dataset of residential mortgages where we can identify the initial LTV with which a house was bought.

	Land Registry All sales 1995-2014	Sales with previous purchase in 1995-2014	Sales with previous purchase in 2002-2014 (funding data)	Sales with previous purchase in 2005-2014 (mortgage data)
Sales (N)	19,628,516	7,527,731	3,199,389	1,385,653
Properties	12,089,086	5,038,658	2,570,092	1,234,381
Current sale price (p_t)				
Mean	161,266	184,100	211,919	231,694
p1	18,500	25,250	40,000	50,000
p25	70,500	93,000	119,000	125,000
p50	124,500	145,000	165,000	176,500
p75	195,000	220,000	243,000	250,000
p99	755,000	825,000	925,000	1,095,000
Property type (proportion)				
Flat	0.18	0.19	0.22	0.24
Terraced	0.31	0.34	0.34	0.32
Semi	0.28	0.27	0.26	0.25
Detached	0.23	0.20	0.19	0.19
Lease	0.23	0.24	0.27	0.28
New	0.10	0.00	0.00	0.00
Expected log capital gains ($\widehat{GAIN}_{jst}$)				
Mean		0.41	0.18	0.04
p1		-0.13	-0.16	-0.19
p25		0.11	0.02	-0.03
p50		0.33	0.13	0.03
p75		0.67	0.29	0.10
p99		1.24	0.75	0.43
Years btw previous purchase and current sale (DUR_t)				
Mean		4.42	3.57	3.21
p1		0	0	0
p25		2	1	1
p50		4	3	3
p75		6	5	5
p99		16	11	8
Matched-in variables (mean)				
Bought with mortgage			0.72	0.71
Bought with LTV>80%				0.25

Table 2: Summary statistics, analysis of selling propensities

Notes: The table shows descriptive statistics of the dataset used to analyze the selling propensity of properties in any given year. The dataset is created by taking the LR samples (whose descriptive statistics are shown in Table 1) and expanding them so that each house has an observation in each year since its first appearance in the LR. (For the empirical analysis we create a variable which equals one if property i sells in year t, and zero otherwise.) To keep the computational burden manageable, for the analysis of selling propensities we extract a ten percent random sample of the data.

	Properties bought in 1995-2014	Properties bought in 2002-2014	Properties bought in 2005-2014
Property × year obs (N)	13,788,911	6,766,378	3,629,414
Sales	721,385	300,962	128,121
Properties	1,167,571	860,635	634,300
Sell prob (Sales/N)	0.05	0.04	0.04
Purchase price (p_s)			
Mean	122,400	172,495	204,461
p1	16,000	23,500	45,000
p25	55,000	96,000	123,239
p50	90,000	144,000	168,500
p75	154,000	207,000	237,000
p99	535,000	675,000	785,000
Expected log capital gains ($GAIN_t$)			
Mean	0.43	0.15	0.02
p1	-0.17	-0.20	-0.23
p25	0.08	0.01	-0.05
p50	0.30	0.10	0.02
p75	0.76	0.26	0.08
p99	1.33	0.78	0.41
Years since purchase (DUR_t)			
Mean	5.82	4.48	3.67
p1	1	1	1
p25	2	2	2
p50	5	4	3
p75	8	6	5
p99	17	12	9
Property type (proportion)			
Flat	0.15	0.19	0.20
Terraced	0.30	0.31	0.31
Semi	0.29	0.28	0.28
Detached	0.25	0.23	0.21
Lease	0.21	0.24	0.25
New	0.10	0.10	0.10
Matched-in variables (averages)			
Bought with mortgage		0.73	0.74
Bought with LTV>80%			0.48

the regression analysis ($\widehat{GAIN}_{jst}$). Figure A1 in the Appendix, shows the two similar distributions of realized and expected gains.

Table 1 shows descriptive statistics for the analysis of transaction prices and distinguishes between 'sales' and 'properties' to highlight the presence of repeat sales. The first column displays statistics for the entire LR while the other columns only include properties that appear at least twice in the LR. Since 2002, the LR dataset can be augmented with an additional variable ('charge') that indicates the use of a mortgage to purchase the property.[9] Since 2005, the UK Financial Conduct Authority (FCA) has been recording information on all owner-occupier mortgages into the Product Sales Data (PSD).[10] These more restricted samples contain more flats and more leasehold properties. There are no new properties in these samples, since transactions are part of repeat-sale pairs and the first purchase (which could potentially refer to a new build) is not part of the analyzed data (it is used to construct the $\widehat{GAIN}_{jst}$ variable). The sales descibed in the third and fourth column of Table 1 allow us to investigate the mechanisms behind history dependence thanks to the additional information on how housing purchases were financed.

Given the aggregate movement in house prices shown in Figure 1, for most households in England and Wales homeownership has produced gains rather than losses—as shown by the descriptive statistics on $\widehat{GAIN}_{jst}$ in Table 1. Additional calculations, not reported in the table, reveal that the second column of the table contains half a million sales with an expected loss (a negative $\widehat{GAIN}_{jst}$) out of 7.5 million transactions.

Analysis of selling propensities To estimate the impact of history dependence on a property's selling propensity (and, in aggregate, on the number of transactions) we reshape and expand the dataset so that each house has an observation in each year since

[9]This variable is not available in the public dataset but can be purchased from the Land Registry.

[10]The PSD have been provided to the Bank of England under a data-sharing agreement. The PSD include regulated mortgage contracts only, and therefore exclude other regulated home finance products such as home purchase plans and home reversions, and unregulated products such as second charge lending and buy-to-let mortgages.

its first appearance in the LR (its first sale after 1995). With 12 million properties and 20 years, the final extended datasets has over 120 million rows (the average property appears for the first time in the middle of the sample and we can follow it for ten years). To keep the empirical analysis computationally manageable, we extract a ten percent random sample of the properties. We create a variable, q_{it}, which equals one if property i sells in year t, and zero otherwise. We treat the first sale as missing because we do not observe $\widehat{GAIN}_{jst}$ before that observation.

3.2 History dependence

The left hand side part of Figure 2 shows the effect of gains and losses on transaction prices. The analysis is based on regression (4).[11] 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. The regressions include year-by-postcode district fixed effects to control for average local prices.

In the charts, negative bin values indicate losses. A loss between 25 and 15 percent is associated to a three percent increase in the transaction price; a loss between 15 and 5 percent is associated to a one and a half percent increase. By contrast, gains are associated to lower transaction prices (as compared to the baseline category of properties that expect to break even). The most populated bin (gains between 35 and 45 percent) is associated with a 4 percent decrease in the transaction price.

Standard errors get bigger for larger gains because there are fewer properties with long holding periods. Moreover, for long tenures the collinearity between year of purchase-sale and $\widehat{GAIN}_{jst}$ increases substantially (only properties with a long holding period experience capital gains of more than 100 percent).

The right hand side part of the Figure shows the effect of expected gains and losses on selling propensities. We aim at investigating whether the purchase price of a property

[11] Table A1 and Table A2 in the Appendix show the regression coefficients.

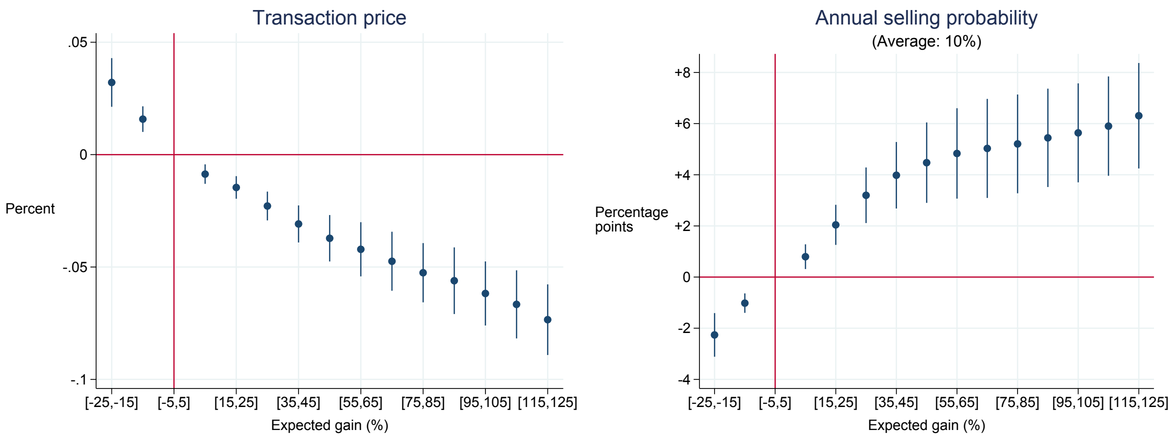


Figure 2: Effects of gains and losses

Notes: The charts show the coefficients and corresponding 95-percent confidence bands for the k dummy variables associated with different expected gains/losses ($\widehat{GAIN}_{kt}$'s) in the regression $y_t = v_i + X\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN}_{jkst} + \lambda_t \otimes \lambda_s + w_t$, where y_t is the transaction price (p_t) in the upper chart and an binary indicator of sale (q_t) in the bottom chart (we omit the individual subscript i for simplicity). 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) and standard errors are double-clustered by year and postcode district.

affects the likelihood that a house is traded in any subsequent period. As explained in the methodology section, we use a linear model—equation (4)—with a binary dependent variable indicating whether the property was sold in any given year. We get a similar picture to the one for transaction prices, albeit with the reversed sign. Losses induce lower selling propensities and gains higher selling propensities. While the unconditional annual transaction probability of a house in the sample is 5 percent as indicated by Table 2, properties with repeat sales are traded more often by construction. For those properties, the unconditional transaction probability is 10 percent, and we should compare the magnitude of the effects against this number. Properties expecting a capital loss between 25 and 15 percent have a selling propensity which is two percentage points lower in any given year. From there, the effect on selling propensities is gradually increasing with expected gains, reaching a positve four percentage points for gains between 35 and 45 percent. Once again the effect flattens out slightly for large—above 35 percent—expected gains.

Figure A2 and A3 in the Appendix replicate Figure 2 for each region in England and Wales. The pattern of transaction prices and selling propensities appears to be similar across different parts of England and Wales.

Alternative specifications Regression (4) on which we base our main analysis is designed to control for as many confounding factors as possible. It is instructive to check whether the patterns identified here are also found with less stringent specifications. The results of this analysis are presented in Appendix Figures A4-A6.

In our first alternative specification we do not include purchase- and sale-year combinations ($\lambda_s \otimes \lambda_t$). This is equivalent to not controlling for holding period and other time-varying factors that are homogeneous across England and Wales. The results on transaction prices are similar to before, although with larger standard errors, indicating that including tenure increases the precision of estimates. The results on selling propensities display the same increasing pattern but magnitudes and standard errors are much larger, implying—as expected—that it is necessary to control for holding period when analysing house selling propensities.

In another specification we do not include individual-property fixed effects. The overall pattern of history dependence appears to be the same, but effects tend to revert back to zero for larger gains. One possible explanation for this result is that neglecting granular fixed effects makes estimates less precise especially when focusing on large expected gains. A related, alternative regression uses full-postcode rather than individual-property fixed effects. Since in the UK a full postcode corresponds to 15 properties on average, this specification allows us to still control for granular effects (albeit not property-specific) while avoiding the reduction in sample that comes with the use of repeat sales. Results are similar to the baseline case but a little smaller quantitatively.

Finally, we run a separate regression for properties which have been flagged as extended, to check that our results are not driven by time-varying property improvements. The England and Wales Energy Performance of Buildings Data[12] lists for each individual property whether an extension was added. The data is drawn from Energy Performance Certificates issued for domestic and non-domestic buildings constructed, sold or let since 2008. Figure A7 in the Appendix shows that there is no significant difference in the history dependence displayed by these two groups of properties.

Linearity and prospect theory Prospect theory predicts a value function which is steeper for losses than for gains (Tversky and Kahneman, 1991; Genesove and Mayer, 2001). At first sight Figure 2 suggests that the effect on transaction prices and selling propensities over the range of gains and losses could be approximated by a linear function. At the same time, a slight change in slope is apparent in the chart, especially for gains above 30-40%. To study this more formally, we run a restricted version of our baseline regression (4):

$$y_{it} = v_i + X_i\beta + \delta_{jt} + \gamma_0 \,\widehat{GAIN}_{jst} + \gamma_1 \,\widehat{GAIN}_{jst}\, \mathbb{I}\{\widehat{GAIN}_{jst} < x\} + \lambda_t \otimes \lambda_s + w_{it}. \quad (6)$$

[12]Available at https://epc.opendatacommunities.org.

Table 3: Linearity and loss aversion

Notes: The table shows selected coefficients from regressions of the form $y_{it} = v_i + X_i\beta + \delta_{jt} + \gamma_0\ \widehat{GAIN}_{jst} + \gamma_1\ \widehat{GAIN}_{jst}\ \mathbb{I}\{\widehat{GAIN}_{jst} < x\} + \lambda_t \otimes \lambda_s + w_{it}$, where $\widehat{GAIN}_{jst}$ enters linearly and interacted with a dummy variable indicating that $\widehat{GAIN}_{jst}$ is below a threshold x. This specification is used to test whether the reaction of prices and selling propensities is steeper for negative gains, indicating loss aversion. The other variables and the sample are as in Figure 2.

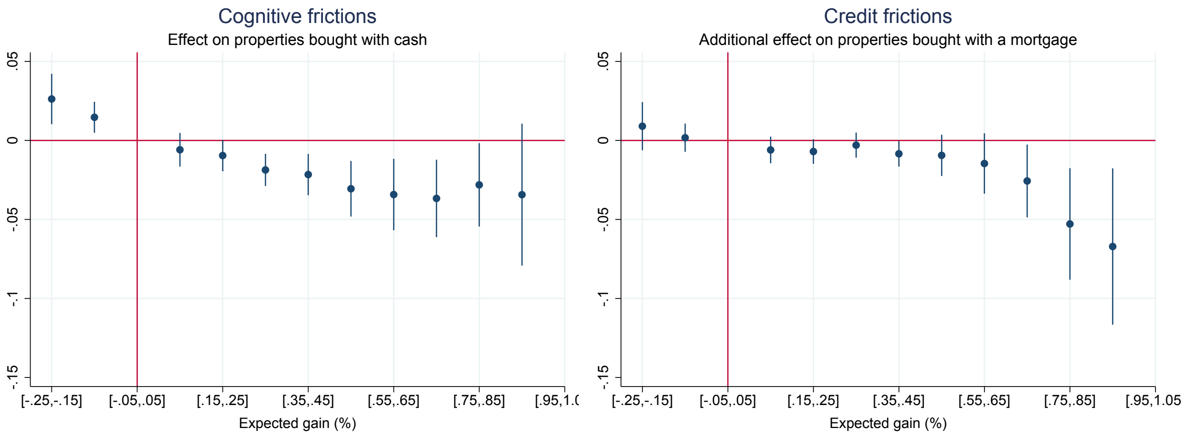
Dependent variable:	**Transaction price** (p_t)				
	No break	x = 0	x = 0.2	x = 0.4	x = 0.6
	(1)	(2)	(3)	(4)	(5)
$\widehat{GAIN}_{jst}$	-0.073	-0.068	-0.073	-0.074	-0.075
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
$\widehat{GAIN}_{jst} \times \mathbb{I}(\widehat{GAIN}_{jst} < x)$		-0.144	-0.014	-0.007	-0.009
		(0.031)	(0.007)	(0.003)	(0.002)
N	4,285,851	4,285,851	4,285,851	4,285,851	4,285,851

Dependent variable:	**Selling propensity** (q_t)				
	No break	x = 0	x = 0.2	x = 0.4	x = 0.6
	(1)	(2)	(3)	(4)	(5)
$\widehat{GAIN}_{jst}$	0.081	0.074	0.081	0.083	0.084
	(0.009)	(0.010)	(0.009)	(0.009)	(0.008)
$\widehat{GAIN}_{jst} \times \mathbb{I}(\widehat{GAIN}_{jst} < x)$		0.094	0.004	0.022	0.021
		(0.027)	(0.012)	(0.005)	(0.003)
N	13,745,918	13,745,918	13,745,918	13,745,918	13,745,918

In this specification capital gains enter in a linear fashion, augmented by an additional effect for gains lower than x, which is meant to capture loss aversion. The regression coefficients for different levels of x are reported in Table 3; the first column contains the estimated γ_0 coefficient when no loss aversion term is added to the regression. Compared to this benchmark, the specification that yields that most notable change in γ_0 while at the same time producing the most statistically significant γ_1 has $x = 0$, consistent with loss aversion. Moreover, the additional contribution to the slope of the $\widehat{GAIN}_{jst}$ effect becomes smaller, but mostly still significant, for expected gains above 0.2, 0.4 or 0.6, revealing a declining marginal utility of price consistent with prospect theory.

The role of credit and cognitive frictions Mortgage debt increased in the UK up to the financial crisis in parallel with house prices (Bunn and Rostom, 2015). Is there a relation between history dependence and household leverage? To answer this question, we have to focus our attention on post-2001 transactions—where we can distinguish between properties purchased with cash and properties purchased with a mortgage—and

Transaction prices



Selling propensities

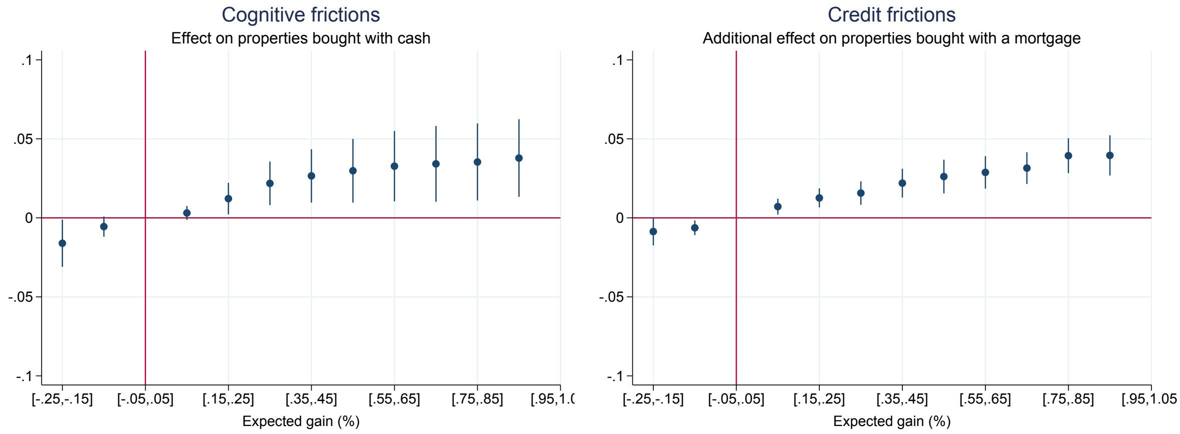


Figure 3: Credit vs cognitive frictions (2002–2014)

Notes: The charts replicate the analysis of Figure 2 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 $y_t = X\beta + \delta_{jt} + \sum_k \gamma_{1k}\left(\widehat{GAIN}_{jkst} \times post2001\right) + \sum_k \gamma_{2k}\left(\widehat{GAIN}_{jkst} \times mortgage\right) + \lambda_t \otimes \lambda_s + w_t$. 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.

post-2005q1 transactions—where we can distinguish, among the mortgaged properties, properties purchased with a LTV greater than 80 (the median LTV in the Product Sales Data) from other properties. Because our attention is on history dependence, in both cases this funding information refers to the previous purchase of the property (at time s), not to the current period being analyzed (t).[13]

We show results graphically in Figure 3 and 4 and in tabular form in Table A1 and A2 in the Appendix. The regressions are run on the same sample as before, to preserve the property fixed effects. However, $\widehat{GAIN}_{jst}$ is interacted with the relevant subsample (post-2001 or post-2005q1 transactions) so that we can focus on the additional information available.

The analysis brings forward potentially different mechanisms for history dependence in transaction prices and selling propensities. In Figure 3, the baseline effect on transaction prices (top-left chart) is reminiscent of the result on the entire sample (Figure 2) whereby transaction prices decline with higher gains. The results are however noisier, with larger standard errors. The top-right chart shows that the additional effect of $\widehat{GAIN}_{jst}$ for properties bought with a mortgage are limited, except for large gains.

In the regression on selling propensities, both the baseline category and properties bought with a mortgage display a pattern similar to the main result, with selling propensity increasing with expected gains. However, effects are sharper for properties bought with a mortgage. Given that the coefficients on properties bought with a mortgage represent the additional effect of leverage on top of the baseline effect showed in the bottom-left chart, results suggest that credit friction play an important role in this type of history dependence.

We run the same analysis focusing on post-2015q1 transactions, where we can distinguish the effect of properties bought with a high leverage (i.e. with an LTV higher than 80 percent) from the effect on other mortgage-funded properties. Similar to the analysis of post-2001 transactions, the top row of the figure shows that most of the effect of expected

[13] Hence we do not attempt to estimate the current LTV for the properties in our sample, but focus exclusively on the LTV at the time of purchase.

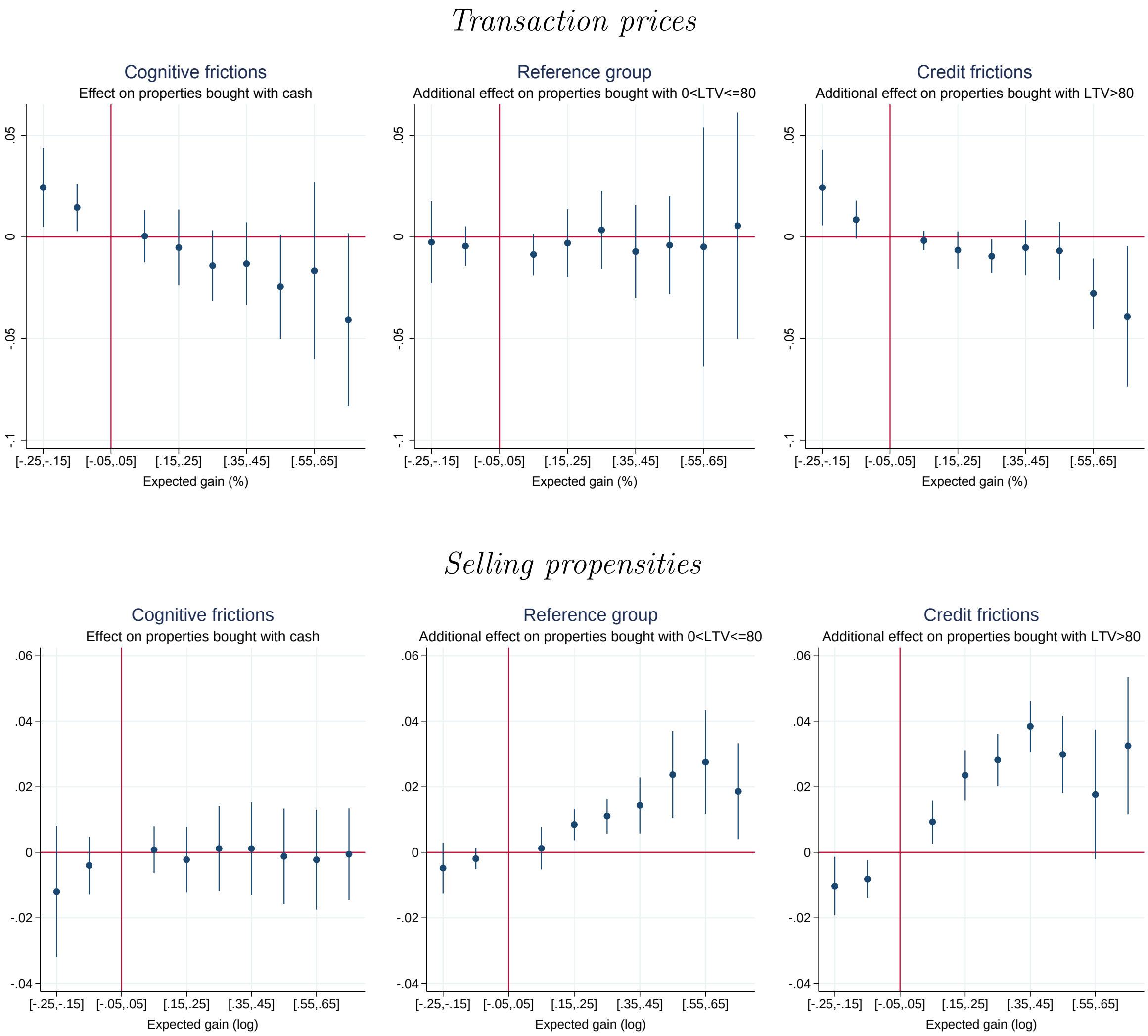


Figure 4: Nonlinear effects of expected gains and losses after 2005q1

Notes: The charts replicate the analysis of Figure 2 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 $y_t = X\beta + \delta_{jt} + \sum_k \gamma_{1k} \left(\widehat{GAIN}_{jkst} \times post2005q1\right) + \sum_k \gamma_{2k} \left(\widehat{GAIN}_{jkst} \times mortgage\right) + \sum_k \gamma_{3k} \left(\widehat{GAIN}_{jkst} \times ltv80\right) + \lambda_t \otimes \lambda_s + w_t$. 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.

gains on prices is present for properties bought with cash, but standard errors are large. The bottow row of the figure shows that all the effect of history dependence on selling propensities comes from properties bought with a mortgage. Within this group, there is both an effect on properties bought with a low leverage and an additional effect on properties bought with an LTV greater than 80 percent. This is slightly different from Figure 3, where an effect on properties bought with cash was apparent. However, both figures confirms the importance of down-payment effects in explaining selling propensities.

The post-2007 fall in transactions As shown in Figure 1, after 2007 the aggregate number of housing transactions in England and Wales did not return to its pre-crisis level even after seven years, in 2014. Can the results on history dependence be related to this fall in housing market activity? To answer this question, we first compare the distribution of ongoing expected capital gains in the two periods, 2001-2007 and 2008-2014. Figure 5 shows there were practically no losses in the 2001-2007 period, and the bulk of properties was in the 0-100 percent capital gain interval. By contrast, in 2008-2014 a few properties were experiencing potential losses and many other properties had expected gains close to zero.

In 2001-2007 the average annual selling propensity for a property was 7.7 percent; this propensity fell to 3.3 percent in the 2008-2014 period. To compute the contribution of history dependence to this fall, we first calculate the change in each of the bins of the expected gain distribution between the two periods, then multiply these differences by the coefficients obtained from the regression on selling propensities and shown in the lower half of Figure 2. By summing all these numbers we get the total contribution, in percentage points, of history dependence to the fall in transactions: -1.4. Since the total fall in transactions between the two periods was 4.4 percentage points, history dependence explains around one third of the fall.

The fall in transactions in the post-crisis period happened in conjunction with house price resilience: without history dependence house prices in England and Wales would

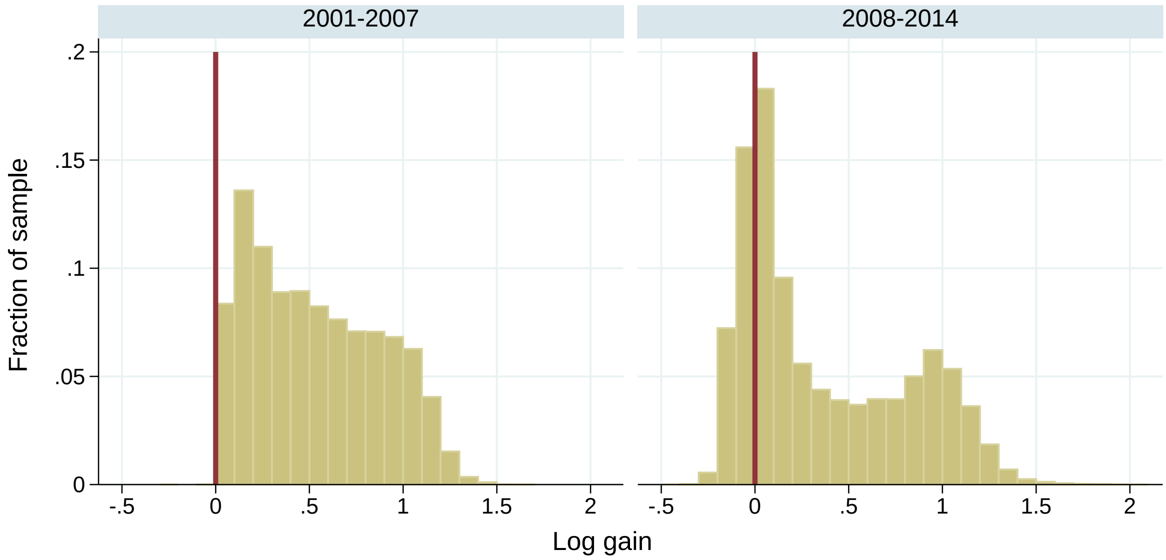


Figure 5: Distribution of ongoing capital gains, pre and post crisis

Notes: The charts show the distribution of the $\widehat{GAIN}_{jst}$ variable in two subperiods: 2001-2007 and 2008-2014. The bin width replicates the allocation of dummy variables used to split $\widehat{GAIN}_{jst}$ and compute the coefficients shown in Figure 2, 3, and 4. For each property, $\widehat{GAIN}_{jst}$ is computed as the difference between the current estimated log house price index and the log index when the house was purchased. The indices are calculated at the local authority level. The distributions are estimated for the analysis of selling propensities and hence $\widehat{GAIN}_{jst}$ is computed for each property in each year since it first appeared in the Land Registry—these are current expected gains rather than realized gains.

have experienced a larger fall. To estimate the size of this counterfactual drop we employ the same method as above: we multiply the changes in the bins that make up the distribution of expected gains by the coefficients shown in the upper half of Figure 2. The overall effect on prices is more modest: we find that England and Wales house prices would have been one percent lower in the absence of history dependence.

4 Extensions: list prices and time on the market

In this section we study history dependence in the selling decision. The analysis is based on data from WhenFresh, a company that collects all daily listings from Zoopla, a major UK property portal. Using this source allows us to study list prices and time on the market for properties that were advertised for sale in England and Wales after 2008.

Many of these properties can be matched back to a previous purchase on the LR. Some of these properties were later sold and recorded again on the LR.

4.1 Data and summary statistics

Zoopla is the second UK property portal in terms of traffic. Its dataset starts in November 2008. In this paper we restrict our attention to sale listings where an address can be precisely identified. The dataset contains information on the address of properties, list prices, and property attributes (such as property type and number of bedrooms).

Zoopla collects data only from estate agents, not individual sellers. In the UK, most transactions occur via estate agents (in 2010, only 11 percent of homes were sold privately—see Office of Fair Trading, 2010).

Table 4 shows the descriptive statistics for the WhenFresh/Zoopla dataset. The table contains information on both the dataset used to analyze list prices (the first two columns) and the dataset used to study the monthly probability of sale once advertised (the last two columns). In both cases, the table shows separate statistics for the entire sample of advertised properties and the sample of properties that were actually sold (as indicated by a match between the listing data in WhenFresh/Zoopla and the transaction data in the Land Registry). Because of the way history dependence is measured, all samples are restricted to those properties for which a *previous* sale was identified in the Land Registry.

Similar to the analysis of unconditional selling propensities in the previous section, the analysis of conditional probabilities of sale is performed on an expanded dataset where each row corresponds to a property-time observation. In this case, the time dimension is monthly; we allow for properties to stay on the market for up to 12 months, as in Anenberg (2015)—in this way we avoid cases in which property listings are simply 'forgotten' on the website.

Table 4: WhenFresh/Zoopla summary statistics

Notes: The table contains statistics for the subset of WhenFresh/Zoopla listings for which it was possible to retrieve a *previous* purchase in the Land Registry (LR) (the matching procedure is described in Appendix B.3). *All* refers to this entire sample whereas *Sold* contains listings that match a *subsequent* sale in the LR. The first two columns report statistics for the analysis of list prices; the third and the fourth column describe the dataset used to analyze the time on market of listed properties. The latter dataset is created by expanding the original sample for list price analysis so that each advertised property has an observation in each month since its appearance on Zoopla until its sale or withdrawal. (We truncate the number of month at 12 when there is no sale.)

	Prices		Probabilities of sale	
	All (previous LR record)	*Sold* (matched with LR record after listing)	*All* (previous LR record)	*Sold* (matched with LR record after listing)
Listings (N)	2,601,406	1,127,866	2,601,406	1,127,866
Properties	2,040,936	1,079,646	2,040,936	1,079,646
Monthly observations			13,800,249	5,261,150
List price (l_t)				
Mean	232,658	236,199	228,792	236,315
p1	59,950	64,950	60,000	64,950
p25	130,000	139,950	129,950	139,950
p50	185,000	189,995	180,000	189,995
p75	275,000	275,000	270,000	275,000
p99	925,000	900,000	899,950	899,950
Property type (proportion)				
Flat	0.16	0.15	0.16	0.15
Terraced	0.32	0.33	0.31	0.32
Semi	0.29	0.31	0.29	0.31
Detached	0.23	0.21	0.24	0.22
Bedrooms	2.84	2.81	2.85	2.82
Lease	0.21	0.19	0.22	0.20
New	0.10	0.10	0.11	0.10
Capital gains ($GAIN_t$)				
Mean	0.28	0.31	0.28	0.30
p1	-0.19	-0.17	-0.20	-0.18
p25	-0.00	0.01	-0.01	0.00
p50	0.11	0.13	0.10	0.13
p75	0.54	0.59	0.56	0.59
p99	1.27	1.29	1.26	1.27
Years since last purchase (DUR_t)				
Mean	6.68	6.97	6.73	6.94
p1	0	0	0	0
p25	3	4	4	4
p50	6	6	6	6
p75	9	10	9	10
p99	17	17	17	17
Months since listing (TOM_t)				
Mean			4.40	3.57
p1			1	1
p25			2	2
p50			4	3
p75			6	5
p99			12	10

4.2 History dependence in list prices and time on the market

The nonparametric results on the effect of $\widehat{GAIN}_{jst}$ are displayed in Figure 6. Because the WhenFresh/Zoopla data start in 2008, using individual-property fixed effects as in the first part of this paper would restrict the sample to properties that transacted multiple times in a time window of only a few years. For this analysis, we use full-postcode fixed effects instead.

The top-left chart of Figure 6 is derived from the sample of all listings; the chart shows that sellers who expect a loss tend to post higher list prices; whereas properties that are experiencing a gain tend to post a lower price. This is consistent with the analysis on actual prices in the previous section, although the effect appears smaller when compared to Figure 2. The chart below, on the left-hand side of the medium row, shows the results for the sample of properties that were eventually sold. The effects, especially the discounts on properties that enjoy substantial expected gains, are larger and comparable to Figure 2. This intriguing difference seems to suggest that discounts associated with large expected gains help the selling process.

The results on the rate at which a house sells once it has been advertised on the property portal (top- and medium-right charts) are consistent with this interpretation When analysing the sample of all listings, for which price effects are muted, monthly probabilities of sale vary significantly between properties with different expected gains. By contrast, when analysing the sample of sold properties, probabilities of sale are relatively homogeneous.

The bottom-left chart in Figure 6 reports the effect on transaction prices, for properties advertised on Zoopla that were actually sold. The effects of expected gains are similar to the ones on list prices and reminiscent of the results for the entire LR sample in Figure 2. The effects on implied discounts, defined as the difference between list and transaction price, are relatively small, reaching around 1 percent for properties with large expected gains, but consistent with the idea that sellers expecting large gains are more willing to accept lower offers. The similarity between effects on listing and transaction

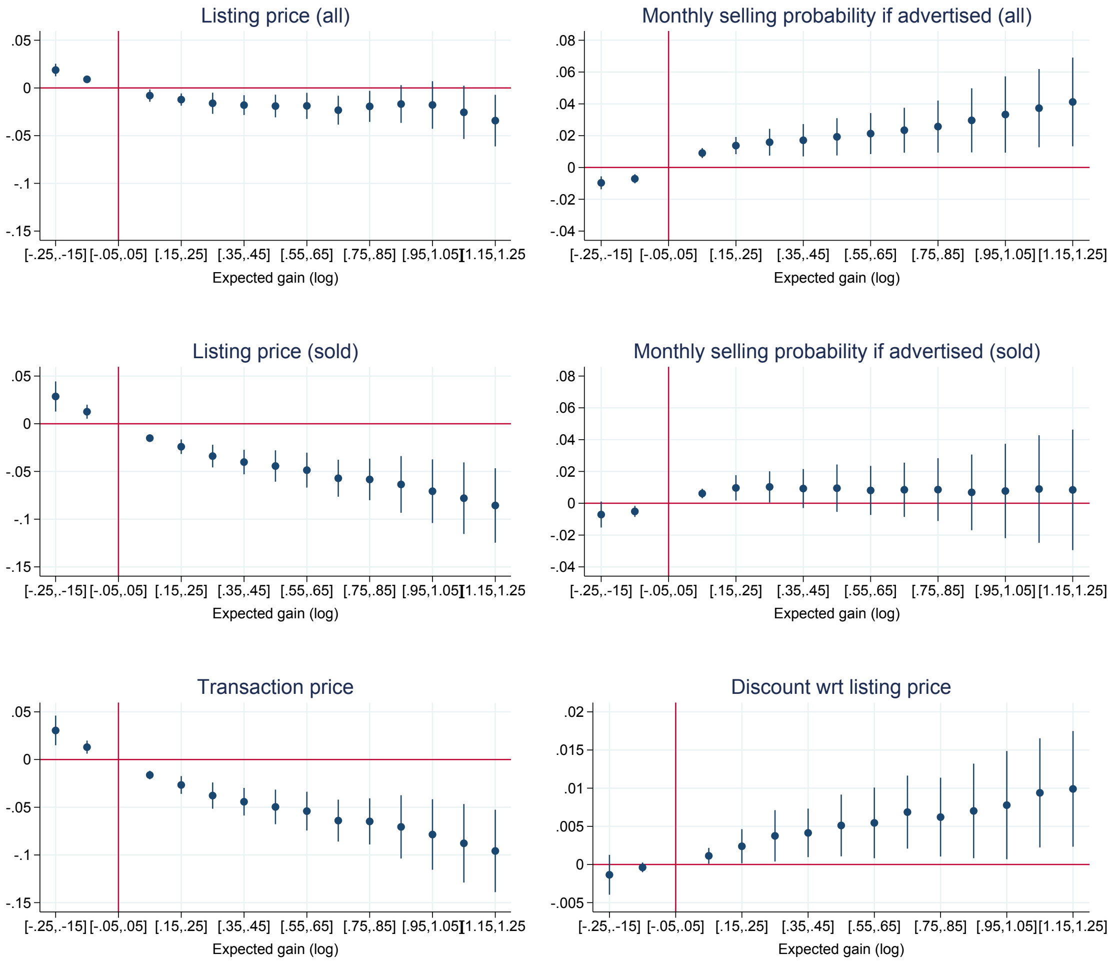


Figure 6: Effects of gains and losses on list prices and time on the market

Notes: The charts report the coefficients and associated 95-percent confidence bands on the $\widehat{GAIN}_{kt}$ dummy variables in the regression $y_t = \phi_h + X\beta + \delta_t + \sum_k \gamma_k \widehat{GAIN}_{kt} + \lambda_s \otimes \lambda_t + w_t$, where ϕ_h represents full-postcode fixed effects (in contrast to the individual-property fixed effects in the first part of the paper). The confidence bands in the chart are computed through standard errors double clustered by year and local authority. The two charts in the upper row refer to the entire sample, made of all listings that have appeared on the Zoopla property portal since 2009, provided that a previous sale of the same property can be retrieved from the Land Registry (LR). The dependent variables are the property list price (l_t) in the first chart and a monthly selling indicator (h_t) in the second chart. The middle row replicates the analysis of the upper row on the *Sold* subsample, made of the subset of listings that can be matched with a subsequent sale in the LR, provided that the sale occurs within 12 months of the listing. Also the bottom row shows results estimated from the *Sold* subsample. The bottom left chart is based on a regression where the dependent variable is the final transaction price (p) of properties, whereas the bottom right chart reports results of a regression on the discount between listing and transaction price ($l - p$).

prices seems to indicate substantial seller bargaining power.

5 Conclusions

This paper investigates history dependence in the housing market using the universe of housing transactions in England and Wales in the last twenty years. We find that house prices in the year a house was previously bought influence the price at which the house sells next, as well as the likelihood that a transaction takes place. Our data allow us to separate properties which were bought with a mortgage and properties which were bought with cash. For a subsample of the data, we can also separate out properties which were bought with a high-LTV mortgage.

While point estimates of the history dependence effects are larger for houses financed through a mortgage and in particular high-LTV ones, consistent with downpayment effects as in Stein (1995), part of the effect on transaction prices (but not on selling propensities) is independent of leverage and seems to be driven by cognitive frictions.

We find similar evidence of history dependence for advertised prices; sellers appear to have enough bargaining power to pass through a significant part of their history premia to transaction prices.

Our findings raise interesting trade-offs in an environment in which housing market activity is history dependent. In particular, while higher house price growth could spur more housing market activity today, it raises the need to sustain this growth in the future, feeding in the unsettling need for potentially spiraling house prices.

References

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

——— (2015): "Information frictions and housing market dynamics," *International Economic Review*, forthcoming.

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

Beggs, A. and K. Graddy (2009): "Anchoring Effects: Evidence from Art Auctions," *American Economic Review*, 99, 1027–39.

Best, M. C., J. Cloyne, E. Ilzetzki, and H. J. Kleven (2015): "Interest rates, debt and intertemporal allocation: evidence from notched mortgage contracts in the United Kingdom," Staff Working Paper 543, Bank of England.

Bracke, P., T. Pinchbeck, and J. Wyatt (2018): "The Time Value of Housing: Historical Evidence on Discount Rates," *The Economic Journal*, forthcoming.

Bunn, P. and M. Rostom (2015): "Household debt and spending in the United Kingdom," Staff Working Paper 554, Bank of England.

Choi, H.-S., H. Hong, and J. Scheinkman (2014): "Speculating on home improvements," *Journal of Financial Economics*, 111, 609 – 624.

Davis, M. A. and E. Quintin (2016): "On the Nature of Self-Assessed House Prices," *Real Estate Economics*, forthcoming.

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

Ferreira, F., J. Gyourko, and J. Tracy (2012): "Housing busts and household mobility: an update," *Economic Policy Review*, 1–15.

Genesove, D. and C. Mayer (2001): "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *The Quarterly Journal of Economics*, 116, 1233–1260.

Giglio, S., M. Maggiori, and J. Stroebel (2015): "Very Long-Run Discount Rates," *The Quarterly Journal of Economics*, 130, 1–53.

Guren, A. M. (2017): "House Price Momentum and Strategic Complementarity," *Journal of Political Economy*, forthcoming.

Head, A. C., H. Sun, and C. Zhou (2018): "Indebted Sellers, Liquidity and Mortgage Standards," Tech. rep., mimeo.

Hong, D., R. Loh, and M. Warachka (2016): "Realization utility and real estate," mimeo.

Mian, A. and A. Sufi (2009): "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *The Quarterly Journal of Economics*, 124, 1449–1496.

Ngai, L. R. and S. Tenreyro (2014): "Hot and Cold Seasons in the Housing Market," *American Economic Review*, 104, 3991–4026.

Office of Fair Trading (2010): "Home buying and selling: A Market Study," Tech. rep.

Ortalo-Magne, F. and S. Rady (2006): "Housing market dynamics: On the contribution of income shocks and credit constraints," *The Review of Economic Studies*, 73, 459–485.

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

Tversky, A. and D. Kahneman (1982): "Judgements of and by Representativeness," in *Judgement under Uncertainty: Heuristics and Biases*, ed. by D. Kahneman, P. Slovic, and A. Tversky, Cambridge: Cambridge University Press.

——— (1991): "Loss aversion in riskless choice: A reference-dependent model," *The quarterly journal of economics*, 106, 1039–1061.

A Appendix For Online Publication: Figures and Tables

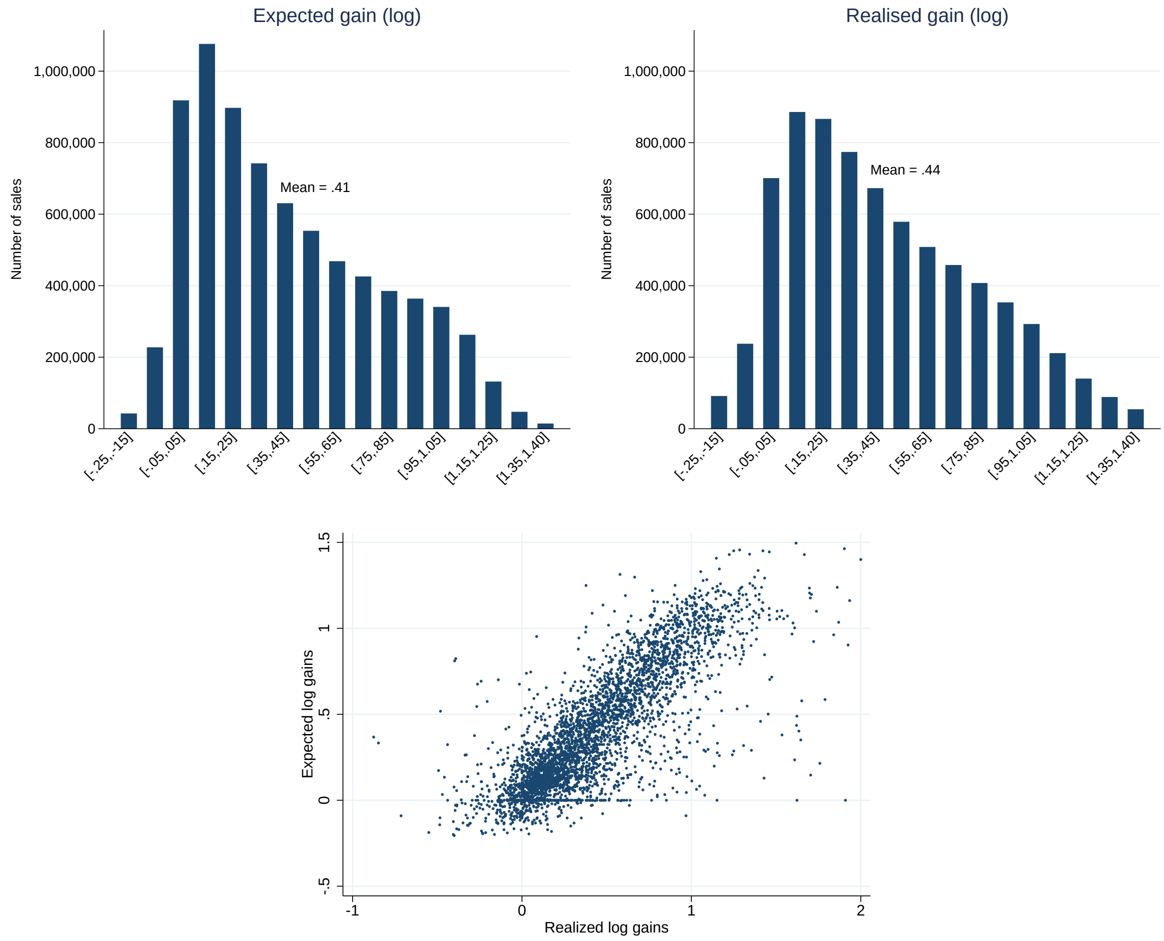


Figure A1: Distribution of gains, 1995-2014

Notes: The upper 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 upper right chart shows the distribution of actual gains, $GAIN_t$, where actual gains are computed as the log house price difference between two pairs of repeat sales. The relation between expected and actual gains is plotted in the bottom chart, which reports results for 0.05 percent random sample of the data.

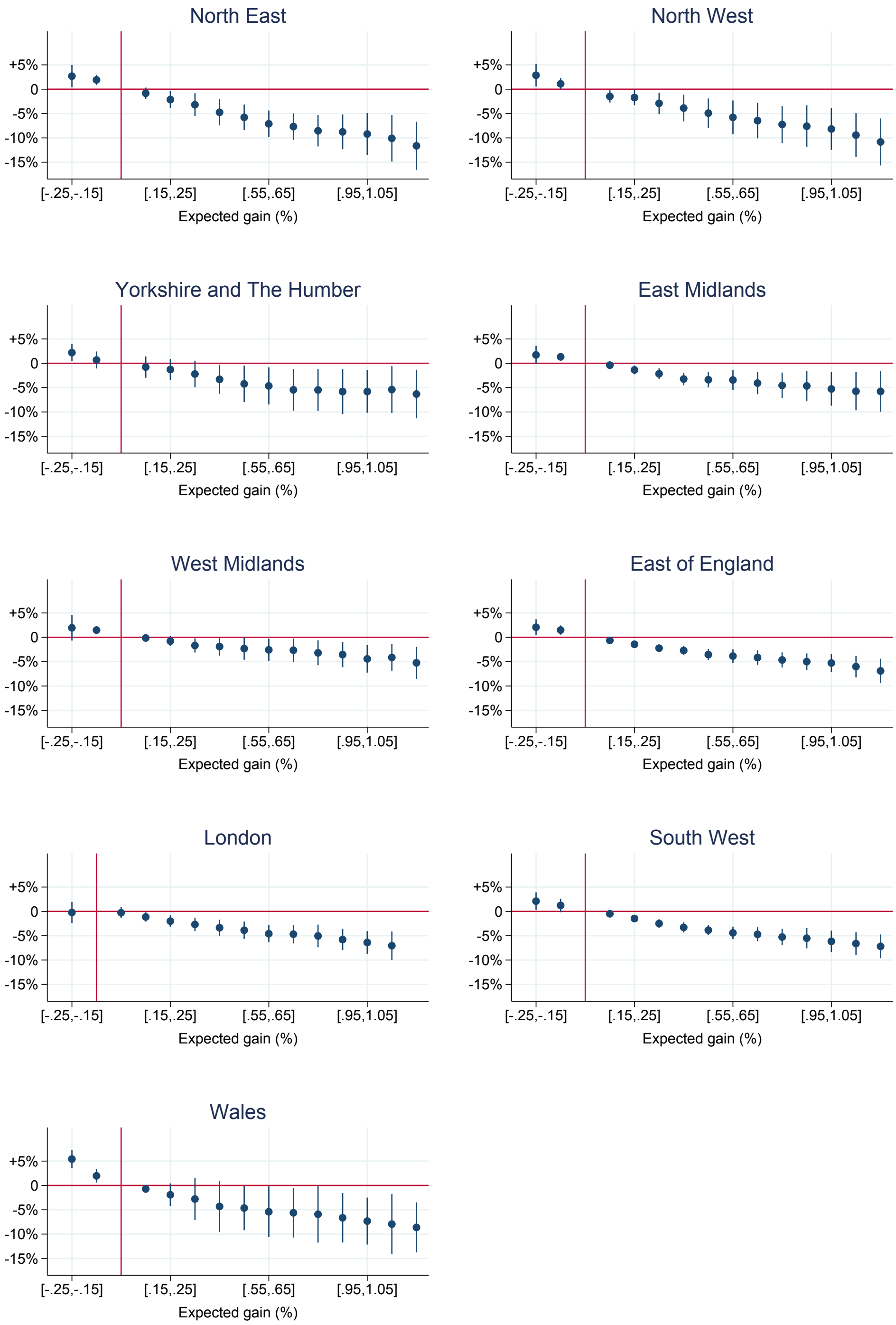


Figure A2: Effects of expected gains and losses on transaction prices, by region

Notes: The charts replicate the analysis of the upper half 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 ($\widehat{GAIN}_{kt}$'s) in the regression $p_t = v_i + X\beta + \delta_t + \sum_k \gamma_k \widehat{GAIN}_{kt} + \lambda_s \otimes \lambda_t + w_t$, run separately for each region. Regressions have year-by-postcode district fixed effects and standard errors are double-clustered by year and postcode district.

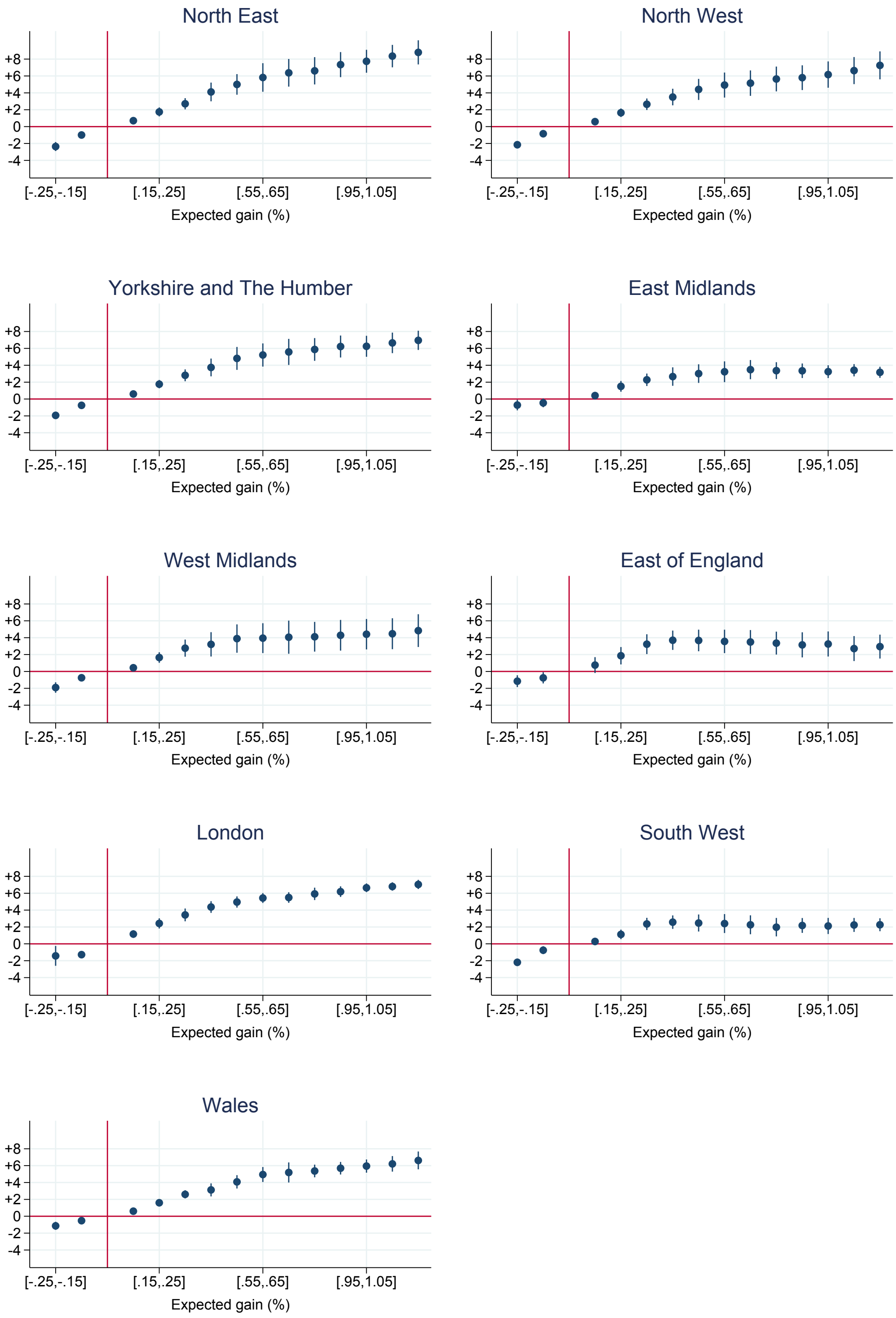


Figure A3: Effects of expected gains and losses on selling propensities, by region

Notes: The charts replicate the analysis of the bottom half 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 ($widehatGAIN_{kt}$'s) in the regression $q_t = v_i + X\beta + \delta_t + \sum_k \gamma_k \widehat{GAIN}_{kt} + \lambda_s \otimes \lambda_t + w_t$, run separately for each region. Regressions have year-by-postcode district fixed effects and standard errors are double-clustered by year and postcode district.

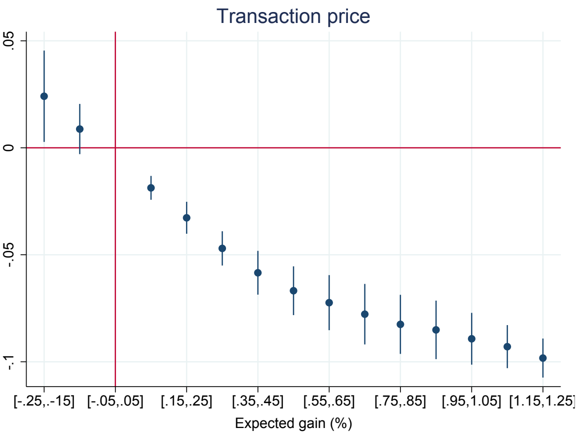


Figure A4: Robustness: No control for holding period

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).

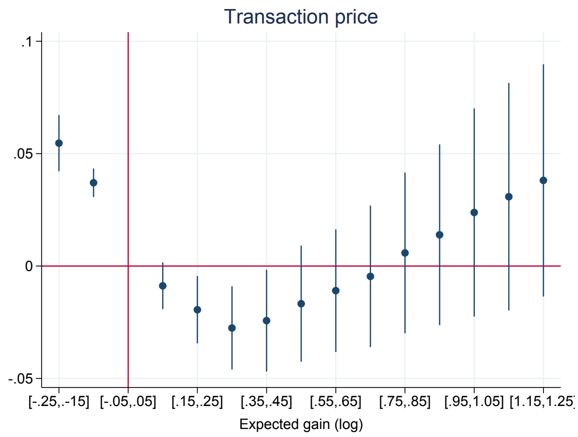


Figure A5: Robustness: No individual-property fixed effects

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_i).

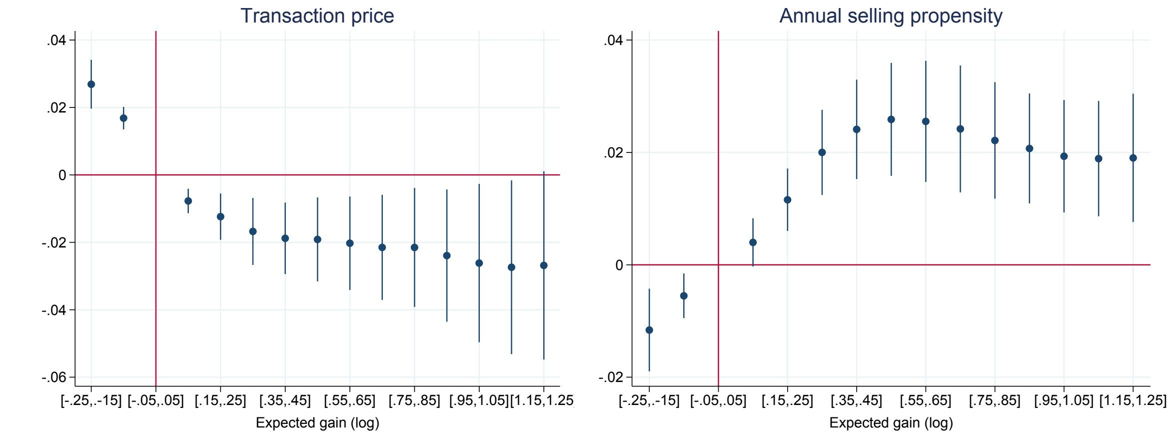


Figure A6: Robustness: full-postcode fixed effects

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 indivdual-property fixed effects (v_i).

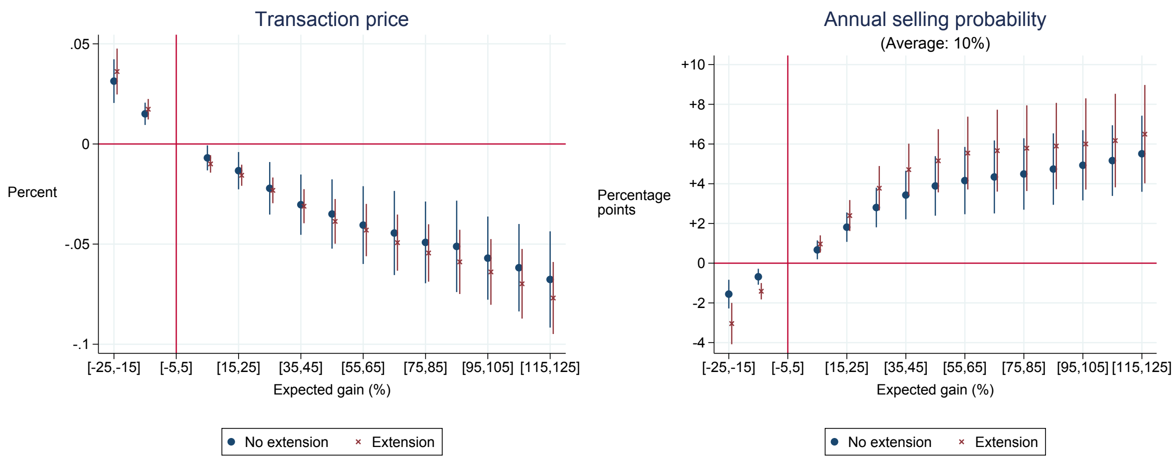


Figure A7: Robustness: No houses with extensions

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

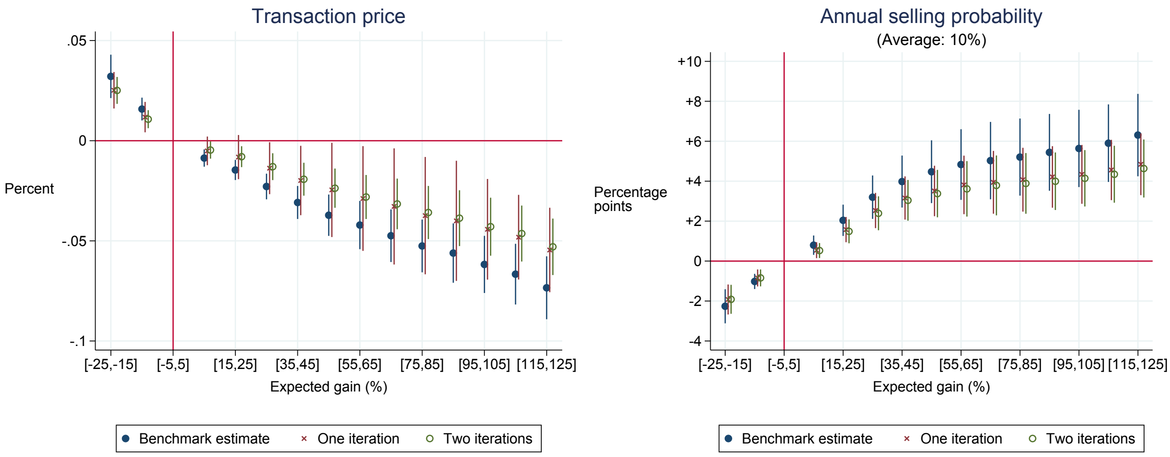


Figure A8: Robustness: Iterating the expected gain measure

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 ($\widehat{GAIN}_{kt}$'s) in the regression $y_t = v_i + X\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN}_{jkst} + \lambda_t \otimes \lambda_s + w_t$. The crosses show the coefficients of a similar regression, $y_t = v_i + X\beta + \delta_{jt} + \sum_k \gamma_k \widehat{GAIN'}_{jkst} + \lambda_t \otimes \lambda_s + w_t$, 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 $\widehat{GAIN'}_{jkst}$.

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

Notes: The first column of the table contains the coefficients and standard errors for the k dummy variables associated with different gains/losses ($\widehat{GAIN}_{kt}$'s) in the regression $p_t = v_i + X\beta + \delta_t + \sum_k \gamma_k \widehat{GAIN}_{kt} + \lambda_s \otimes \lambda_t + w_t$, where p_t is the (log) transaction price. The coefficients are displayed graphically with their 95 percent confidence bands in the left hand side part of Figure 2 (column 1).

Column 2 shows the coefficient on the interaction $\widehat{GAIN}_t \times post2001$, where $post2001$ indicates sales whose previous purchase took place after 2001, in the regression $p_t = v_i + X\beta + \delta_t + \sum_k \gamma_{1k}(\widehat{GAIN}_{kt} \times post2001) + \sum_k \gamma_{2k}(\widehat{GAIN}_{kt} \times Mortgage) + \lambda_s \otimes \lambda_t + w_t$. Column 3 shows the coefficient on $\widehat{GAIN}_{kt} \times Mortgage$ 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 $\widehat{GAIN}_t \times post2005q1$, where $post2005q1$ indicates sales whose previous purchase took place after March 2005, in the regression $p_t = v_i + X\beta + \delta_t + \sum_k \gamma_{1k}(\widehat{GAIN}_{kt} \times post2005q1) + \sum_k \gamma_{2k}(\widehat{GAIN}_{kt} \times Mortgage) + \sum_k \gamma_{3k}(\widehat{GAIN}_{kt} \times HighLTV) + \lambda_s \otimes \lambda_t + w_t$, 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. Standard errors double-clustered at the year and postcode district (PCD) level are in parentheses.

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	0.026	0.009	0.024	-0.003	0.024
	(0.005)	(0.008)	(0.007)	(0.009)	(0.009)	(0.009)
Gain [-.15,-.05]	0.016	0.015	0.002	0.015	-0.004	0.009
	(0.003)	(0.005)	(0.005)	(0.006)	(0.005)	(0.004)
Gain [.05,.15]	-0.009	-0.006	-0.008	0.000	-0.009	-0.002
	(0.002)	(0.005)	(0.002)	(0.006)	(0.005)	(0.002)
Gain [.15,.25]	-0.015	-0.010	-0.007	-0.005	-0.003	-0.006
	(0.002)	(0.004)	(0.003)	(0.009)	(0.008)	(0.004)
Gain [.25,.35]	-0.023	-0.019	-0.005	-0.014	0.003	-0.009
	(0.003)	(0.004)	(0.003)	(0.008)	(0.009)	(0.004)
Gain [.35,.45]	-0.031	-0.022	-0.009	-0.013	-0.007	-0.005
	(0.004)	(0.005)	(0.003)	(0.010)	(0.011)	(0.006)
Gain [.45,.55]	-0.037	-0.031	-0.016	-0.025	-0.004	-0.007
	(0.005)	(0.006)	(0.005)	(0.012)	(0.012)	(0.007)
Gain [.55,.65]	-0.042	-0.034	-0.022	-0.017	-0.005	-0.028
	(0.006)	(0.011)	(0.006)	(0.021)	(0.028)	(0.008)
Gain [.65,.75]	-0.047	-0.037	-0.033	-0.041	0.006	-0.039
	(0.006)	(0.011)	(0.008)	(0.020)	(0.027)	(0.017)
Gain [.75,.85]	-0.053	-0.028	-0.048	-0.082	0.057	-0.036
	(0.006)	(0.015)	(0.015)	(0.037)	(0.036)	(0.022)
Gain [.85,.95]	-0.056	-0.034	-0.067	-0.055	-0.170	0.168
	(0.007)	(0.024)	(0.015)	(0.052)	(0.096)	(0.082)
Gain [.95,1.05]	-0.062	-0.028	-0.099	-0.105	0.240	-0.175
	(0.007)	(0.036)	(0.024)	(0.064)	(0.081)	(0.100)
Gain [1.05,1.15]	-0.067	-0.035	-0.124	-0.295	0.091	-0.012
	(0.007)	(0.041)	(0.031)	(0.161)	(0.190)	(0.109)
Gain [1.15,1.25]	-0.073	-0.056	-0.150	-0.276	0.000	0.259
	(0.008)	(0.080)	(0.053)	(0.132)	(0.000)	(0.008)
Gain [1.25,1.35]	-0.083	-0.063	-0.212	-0.053	-0.379	
	(0.008)	(0.086)	(0.051)	(0.057)	(0.031)	
Gain [1.35,1.45]	-0.089	0.088	-0.027	0.741	0.000	
	(0.010)	(0.149)	(0.193)	(0.060)	(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 $q_t = v_i + X\beta + \delta_t + \sum_k \gamma_k \widehat{GAIN}_{kt} + \lambda_s \times \lambda_t + w_t$, where q_t 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). All regressions control for property type as measured by the Land Registry (X: flat, terrached, semi-detached or detached property; new or second-hand property; property sold as leasehold or freehold) and for a nonparametric function (a third-degree polynomial) of the number ofyears between sales (DUR_t). Regressions have year-by-postcode district (PCD) fixed effects (δ_t 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	-0.009	0.015	-0.016	-0.007
	(0.004)	(0.007)	(0.004)	(0.003)	(0.005)	(0.004)
Gain [-.15,-.05]	-0.010	-0.006	-0.006	0.000	-0.005	-0.007
	(0.002)	(0.003)	(0.002)	(0.000)	(0.002)	(0.003)
Gain [.05,.15]	0.008	0.003	0.007	0.000	0.002	0.010
	(0.002)	(0.002)	(0.002)	(0.000)	(0.003)	(0.004)
Gain [.15,.25]	0.020	0.012	0.013	0.000	0.006	0.023
	(0.004)	(0.005)	(0.003)	(0.000)	(0.005)	(0.004)
Gain [.25,.35]	0.032	0.022	0.016	0.000	0.012	0.026
	(0.005)	(0.007)	(0.004)	(0.000)	(0.006)	(0.003)
Gain [.35,.45]	0.040	0.027	0.022	0.016	0.000	0.037
	(0.006)	(0.008)	(0.004)	(0.005)	(0.000)	(0.004)
Gain [.45,.55]	0.045	0.030	0.026	0.020	0.000	0.034
	(0.007)	(0.010)	(0.005)	(0.005)	(0.000)	(0.004)
Gain [.55,.65]	0.048	0.033	0.029	0.027	0.000	0.012
	(0.008)	(0.011)	(0.005)	(0.005)	(0.000)	(0.007)
Gain [.65,.75]	0.050	0.034	0.032	0.016	0.000	0.028
	(0.009)	(0.011)	(0.005)	(0.006)	(0.000)	(0.010)
Gain [.75,.85]	0.052	0.035	0.039	0.020	0.000	0.027
	(0.009)	(0.012)	(0.005)	(0.009)	(0.000)	(0.008)
Gain [.85,.95]	0.054	0.038	0.040	0.016	0.000	0.094
	(0.009)	(0.012)	(0.006)	(0.007)	(0.000)	(0.021)
Gain [.95,1.05]	0.056	0.040	0.043	0.043	0.000	-0.041
	(0.009)	(0.012)	(0.005)	(0.039)	(0.000)	(0.042)
Gain [1.05,1.15]	0.059	0.043	0.044	0.034	0.000	-0.029
	(0.009)	(0.012)	(0.007)	(0.010)	(0.000)	(0.013)
Gain [1.15,1.25]	0.063	0.047	0.046	0.000	0.000	1.003
	(0.010)	(0.012)	(0.009)	(0.000)	(0.000)	(0.024)
Gain [1.25,1.35]	0.067	0.051	0.054	0.003	0.000	0.017
	(0.011)	(0.013)	(0.017)	(0.012)	(0.000)	(0.027)
Gain [1.35,1.45]	0.071	0.056	0.055	0.000	0.008	0.021
	(0.012)	(0.014)	(0.019)	(0.000)	(0.009)	(0.007)
N	13,704,178	13,704,178	13,704,178	13,749,301	13,749,301	13,749,301

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).
Colums 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 colums 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_i 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)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Idiosyncratic factor ($\hat{p}_0$)	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA
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 A9 shows that the total number of cash purchases in England and Wales is less cyclical than the number of mortgages.

Table A4 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).

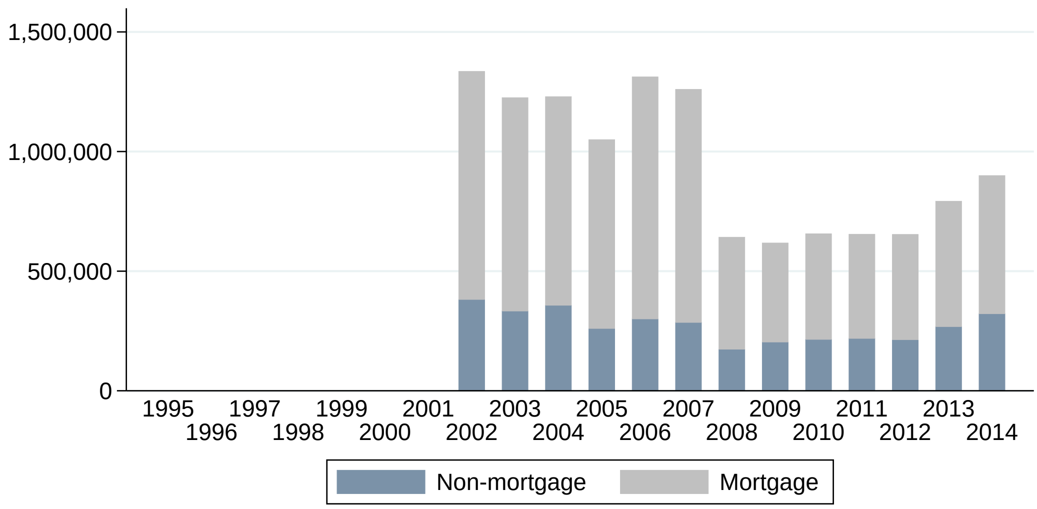


Figure A9: Mortgage vs non-mortgage purchases, 2002-2014

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 ben 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 A4: Summary statistics: bought with a mortgage vs bought with cash

Notes: This table repeats the analysis of the upper half 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_t)		
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 analysed 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.[14]

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.[15]

The date in which a bank transfer the mortgage amount to the buyer is the completion date or a few days before. Figure A10 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.[16] 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

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

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

[16]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.

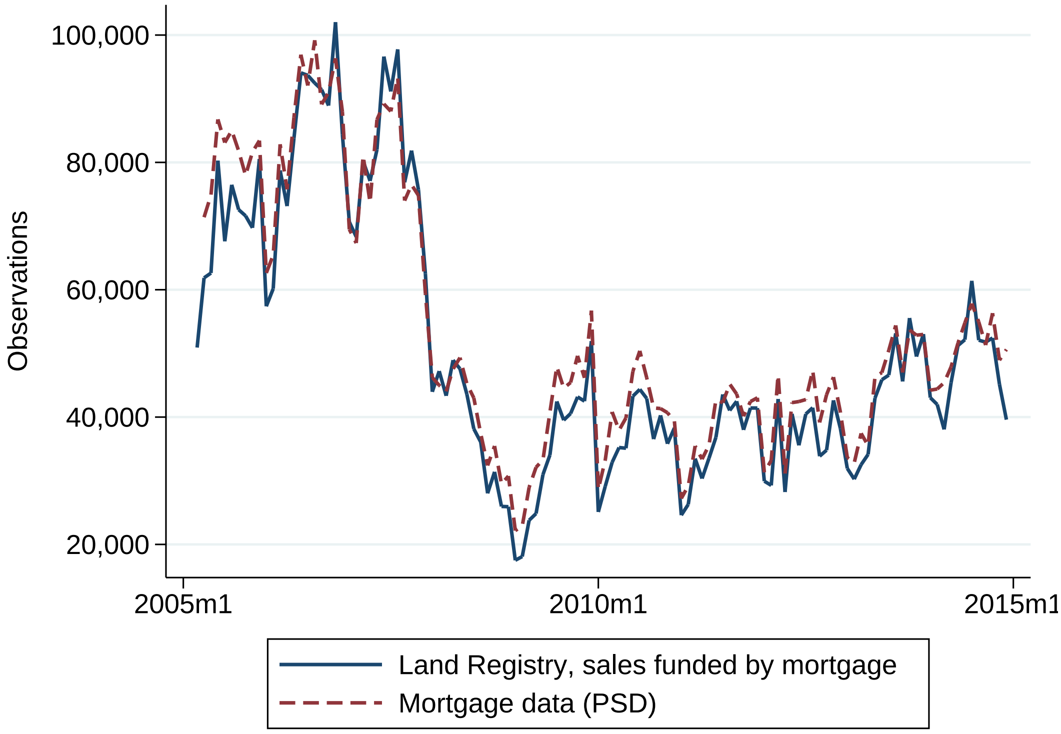


Figure A10: Number of observations by month in the Land Registry and Product Sale Data before matching

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.[17]

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 A5 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 A11 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

[17] 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., 2015).

Table A5: 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_t)				
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

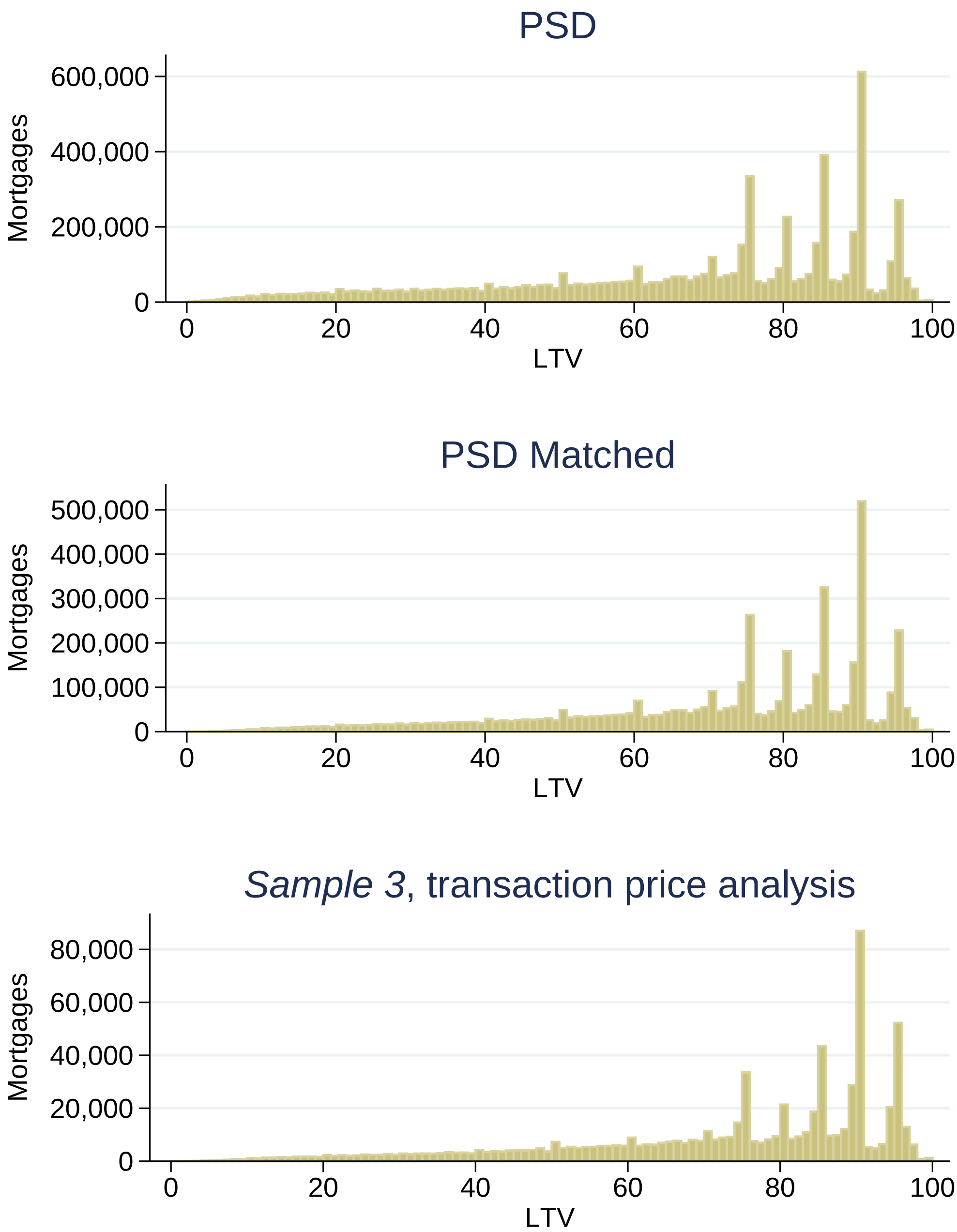


Figure A11: LTV distributions in the Product Sales Data and the matched observations

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[18]—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 the

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

WhenFresh/Zoopla dataset) and (2) the first part of the address, which corresponds to the street number for a house and the appartment number for a flat. The combination of these two variables is likely to identify a unique property,[19] 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.

[19] 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.