History Dependence in the Housing Market*

Philippe Bracke[a,b] Silvana Tenreyro[a,c]

[a] Bank of England,[b] SERC [c]London School of Economics, CfM, CEPR

March 2017

Abstract

Using the universe of housing transactions in England and Wales in the last twenty years, we document a robust pattern of history dependence in housing markets. Sale prices and selling probabilities today are affected by aggregate 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 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

*For helpful comments, we would like to thank Francesco Caselli, Andreas Fuster, Per Krusell, Benedikt Vogt and conference and seminar participants at the Bank of England, Bureau for Economic Policy Analysis (The Hague), European Economic Association Annual Congress (Geneva), Ghent Workshop on Empirical Macroeconomics, and LSE. Tenreyro acknowledges financial support from the ERC Consolidators Grant 2016.

1 Introduction

This paper documents a novel pattern of history dependence in house prices and transactions. Specifically, aggregate house prices in the year a house was previously bought influence the individual price at which the house sells next, as well as the probability that the transaction takes place. The results are based on twenty million housing transactions from England and Wales 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. On average, all else equal, the house bought in 2007 will carry a price premium of over 10 percent over the one bought in 2001. Moreover, the house bought in 2007 will have, on average, 15 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 contributes 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 friction can help us explain the presence of history dependence.

Cognitive frictions constitute the first group of explanations 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.

The second group of explanations 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 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.

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 strong evidence of history dependence in this cash-only sample both on prices and selling probabilities. Loss aversion, however, does not appear to have played a role over and above history dependence. First, only a small fraction of properties experienced losses during this period. Second, for those properties that did lose value, no asymmetric effect is apparent in the data: the effects of past prices on current prices and selling probabilities are similar for gains and losses measured around the previous price benchmark.

We also find that leverage accentuates history dependence. 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-

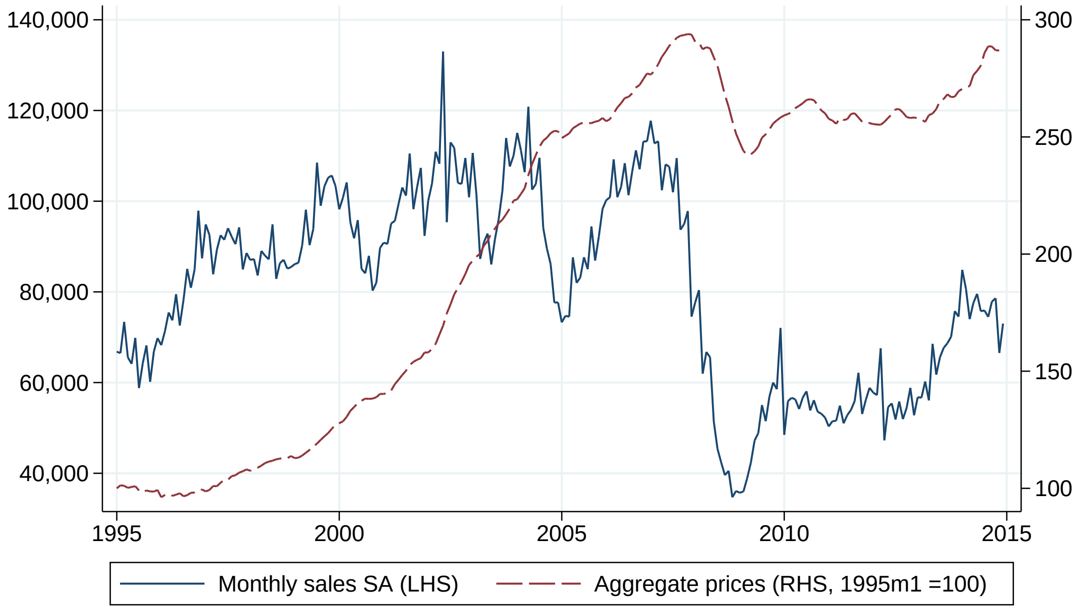

Figure 1: Monthly house prices and sales, England and Wales

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 4.

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

The rest of paper is organized as follows. Section 2 discusses the relation with the existing literature. Section 3 describes the methodology. Section 4 presents the data and documents the patterns of history dependence. It next studies the potential channels underlying history dependence and their quantitative relevance across regions and over time. Section 5 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 6 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 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,[1] and more recently by 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 no significant effects on transacted prices. They find a small role for down-payment effects. Relatedly, Anenberg (2011) analyzes the San Francisco Bay Area housing market and in contrast to Genesove and Mayer (2001), reports significant effects of loss aversion on transacted prices. Unlike these two studies, we find that loss aversion played only a muted role in the England and Wales' housing markets, not least because the overall gains in values for most properties were positive during the period analyzed. Moreover, for properties that registered losses, there is no evidence of asymmetric effects on prices or selling probabilities vis-à-vis gains. Also differently from these studies, we investigate the quantitative implications of history dependence and its underlying channels on the aggregate volume transactions.

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 this paper, we study the effect of history dependence on aggregate outcomes such as prices and number of transactions. In another recent

[1] Ortalo-Magne and Rady (2006) also explore the consequences of down-payment constraints in a theoretical model.

paper, Hong et al. (2016) find some suggestive evidence in the Singaporean condominium market of a kink in the selling probability at zero gains consistent with realization utility (Barberis and Xiong, 2012). Unlike these studies, we do not find evidence for a kink in selling probabilities around zero gains. 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 probabilities.). 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.

In identifying history dependence, the paper relates to Beaudry and DiNardo (1991), who document history dependence in the labor market. The authors take a standard wage equation and show that the unemployment rate when the contract started is a significant determinant of today's wages. They interpret their findings as a result of wage stickiness and insurance contracts (firms insure workers against fluctuations in income over the business cycle). Their results have been replicated in a number of studies and for different countries: for instance, Grant (2003) shows that the results hold for a different period; McDonald and Worswick (1999) show they hold for Canada; and Devereux and Hart (2007) for the United Kingdom.

A closely related set of studies in this literature focuses on the effect of market conditions at the time of labor market entry. Kahn (2010) uses the National Longitudinal Survey of Youth, whose respondents graduated from college between 1979 and 1989. She estimates the effects of both national and state economic conditions at time of college graduation on labor market outcomes for the first two decades of a career. Oreopoulos et al. (2012) also shows that initial labor market conditions have long-term effects on the earnings of college graduates and (less) on the earnings of noncollege workers. Contemporaneously, Moreira (2016) has documented history dependence in firms' performance: firms born during a boom tend to grow persistently faster.

3 Identifying history dependence

The (log) house price is usually modeled as:

$$p_{it} = X_i\beta + \delta_t + w_{it}, \tag{1}$$

where p_{it} is the transaction price of house i sold at time t, X_i is a vector of housing characteristics, δ_t is the aggregate house price level at time t, and w_t is an idiosyncratic error component which contains both unobserved property characteristics (time-varying or time-invariant) and idiosyncratic price effects due to the features of specific transactions.

To study history dependence we start by augmenting the standard hedonic regression above with the house's previous transaction price p_{is}:

$$p_{ist} = X_i\beta + \delta_t + \gamma p_{is} + e_{it}, \tag{2}$$

where s denotes the period when the house was previously purchased. Clearly, in such regression the coefficient γ is not informative about history dependence per se, as it may be capturing unobservable property characteristics of the house not contained in X_i. To isolate the effect of previous aggregate market conditions we decompose p_{is} into $\hat{\delta}_s$, the price index at time s, and $\hat{p}_{i0} = X_i\beta + e_{is}$, the imputed price of the house at time 0, the baseline period (1995 in our dataset);[2] and include both terms in the equation. (To simplify notation, we omit the subscript i and focus on a house evaluated at times t, s, and 0, with $t > s > 0$.) The estimated equation becomes:

$$p_t = X\beta + \delta_t + \gamma_1\hat{\delta}_s + \gamma_2\hat{p}_0 + e_t. \tag{3}$$

By focusing on the aggregate component of past prices ($\hat{\delta}_s$), we sidestep the problem that p_s contains time-invariant unobservable characteristics that could bias our estimation; these characteristics are now captured by the term $\hat{p}_0$.[3]

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 past aggregate prices ($\hat{\delta}_s$). For instance, a property that has been only two years with an owner will often have a higher $\hat{\delta}_s$ than a property that has been eight years with the same owner. We therefore also control for the duration of the tenure (DUR_t), measured as the number of years between two sales. Such variable has the added advantage of controlling for some time-varying unobserved property characteristics such as depreciation. It is likely that depreciation follows a nonlinear pattern; hence we allow for DUR_t to enter the regression non-parametrically through a third-degree polynomial:

$$p_t = X\beta + \delta_t + \gamma_1\hat{\delta}_s + \gamma_2\hat{p}_0 + f(DUR_t) + \varepsilon_t, \tag{4}$$

where the error is now denoted as ε_t to indicate that some time-varying characteristics are controlled for.

[2] We compute $\hat{p}_{i0}$ by simply subtracting $\hat{\delta}_s$ from the previous purchase price, p_{is}. The term $\hat{p}_{i0}$ represents the price the house would have fetched in 1995 assuming the same idiosyncratic term (e_{is}) as the one at the time of the previous purchase (s). The term $\hat{p}_{i0}$ can be interpreted as a time-invariant measure of house quality.

[3] In the Appendix we also show results from a specification with full (6-digit) postcode fixed effects instead of $\hat{p}_0$—results are very similar. In the UK a full postcode corresponds to 10-15 housing units.

Our coefficient of interest, γ_1, could still be biased by other time-varying property characteristics not captured by $f(DUR_t)$, for instance if the likelihood of home improvements and renovations is correlated with aggregate house prices (as in Choi et al., 2014). To address this remaining threat, in the Appendix we show results where we restrict the sample to (a) flats, as flats are less likely to change their value by a lot after a renovation (their size, a critical determinant of price, usually cannot be altered) and (b) properties that were bought new, because this greatly reduces the need for renovations.

When exploring the mechanisms behind history dependence, equation (4) can be rewritten with a measure of gains (or losses) as the variable of interest:

$$p_t = X\beta + \delta_t + \gamma_1 \widehat{GAIN}_t + \gamma_2 \hat{p}_0 + f(DUR_t) + e_t, \tag{5}$$

where $\widehat{GAIN}_t = \hat{\delta}_t - \hat{\delta}_s$ is the (log) difference in aggregate house prices between time t and when the property was bought. Notice that these are expected, rather than realized, gains. Not only does the inclusion of $\widehat{GAIN}_t$ allow us to distinguish between expected gains and losses in the estimating equation—separating pure anchoring or learning from loss aversion—, it also provides a way to estimate the effect of gains and losses in a non-linear, non-parametric way. We do so by splitting $\widehat{GAIN}_t$ 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 transaction probabilities, we start from an equation similar to (4) 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 do not observe DUR_t before this first sale).

4 History dependence in transaction prices and selling probabilities

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 probabilities. We then show the results for history dependence and explore its quantitative relevance.

4.1 Data and summary statistics

The LR records all residential property transactions, with few exceptions:[4] 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 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).[5] The variable 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 $\widehat{GAIN}_t$ variable. We do so at the local authority level, by running a regression such as (1) for each local authority (LA) in England and Wales. Our dataset contains 348 LAs in England and Wales; LAs are larger than the

[4] 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.

[5] 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. See Giglio et al. (2015) and Bracke et al. (2017).

typical American municipality but smaller than the typical metropolitan area (Hilber and Vermeulen, 2016). Figure A1 in the Appendix plots each of these indices grouped by region.

Analysis of transaction prices The analysis of transaction prices presented in this paper relies on the identification of repeat sales—we need information on the previous purchase of a property to make inference about history dependence. 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 the regression analysis ($\widehat{GAIN}_t$). Figure A2 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.

Table 1 displays statistics for the entire LR (first column) and the three samples used in the analysis. The first sample, *Sample 1*, spans all the years from 1995 to 2014. Moving to the right columns of Table 1 means restricting attention to sales that happened in later years. We use these more restricted samples for some of the analyses presented in the paper because more information is available in later years. Since 2002, the LR dataset includes a variable ('charge') which indicates the use of a mortgage to purchase the property[6]—hence we label as *Sample 2* the subset of transactions whose previous purchase happened after 2001. Since 2005, the UK Financial Conduct Authority (FCA) has been recording information on all owner-occupier mortgages into the Product Sales Database (PSD)—hence we label as *Sample 3* the subset of transactions that can be matched into the PSD. These more restricted samples contain more flats and, therefore, 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

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

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 1* 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, as for this sample we can tell whether the property was purchased with a mortgage. 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)	*Sample 1* (sales with previous purchase in 1995-2014)	*Sample 2* (sales with previous purchase in 2002-2014)	*Sample 3* (sales with previous purchase in 2005-2014)
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
Previous purchase price (p_s)				
Mean		122,338	170,955	202,007
p1		16,000	22,500	42,200
p25		55,000	95,000	120,000
p50		90,000	142,900	166,000
p75		154,000	205,000	235,000
p99		540,000	676,000	800,000
Expected log capital gains ($\widehat{GAIN}_t$)				
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

not part of the analyzed data (it is used to construct the history dependence variable).

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 $\widehat{GAIN}_t$ in Table 1. Additional calculations, not reported in the table, reveal that *Sample 1* contains 489,542 sales with an expected loss (a negative $\widehat{GAIN}_t$) out of 7.5 million transactions.

Analysis of selling probabilities To estimate the impact of history dependence on a property's selling probability (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 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, meaning that we can follow it for ten years). To keep the empirical analysis computationally manageable, we extract a 50 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 DUR_t before that observation.

4.2 History dependence measure

Transaction prices Table 3 contains regressions with the current sale price of a house as the dependent variable. All regressions control for property type as measured by the LR (flat, terraced, semi-detached or detached property; new or second-hand property; property sold as leasehold or freehold) as well as the number of years elapsed since the current sellers have bought the property (DUR_t). The regressions include year-by-local authority fixed effects to control for average local prices. Table 3 has three pairs of columns, each pair corresponding to a sample.

The first columns of each pairs show the results of regressing today's prices on the prices of previous purchases of the same properties. This is for descriptive purposes only,

Table 2: Summary statistics, analysis of selling probabilities

Notes: The table shows descriptive statistics of the dataset used to analyze the selling probability 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 probabilities we extract a 50 percent random sample of the data.

	Sample 1 (1995-2014)	*Sample 2* (2002-2014)	*Sample 3* (2005-2014)
Property × year obs (N)	68,925,352	33,828,768	18,170,180
Sales	3,598,666	1,500,362	636,611
Properties	5,838,767	4,304,097	3,174,433
Sell prob (Sales/N)	0.05	0.04	0.04
Purchase price (p_s)			
Mean	122,404	172,412	204,951
p1	16,000	23,000	45,000
p25	55,000	96,000	123,760
p50	90,000	144,950	169,950
p75	154,500	208,000	238,000
p99	540,000	684,995	800,075
Expected log capital gains ($GAIN_t$)			
Mean	0.41	0.14	0.01
p1	-0.17	-0.19	-0.20
p25	0.08	-0.00	-0.06
p50	0.29	0.09	0.01
p75	0.74	0.25	0.07
p99	1.28	0.73	0.39
Years since purchase (DUR_t)			
Mean	5.83	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	0.16	0.19	0.20
Terraced	0.30	0.31	0.31
Semi	0.29	0.28	0.28
Detached	0.25	0.23	0.22
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

Table 3: History dependence regressions

Notes: The upper panel of the table reports results for the transaction price analysis and the bottom half of the table reports results for the selling probability analysis. In each of the two panels, the first row refers to a regression of the form $y_t = X\beta + \delta_t + \gamma p_s + f(DUR_t) + \varepsilon_t$ whereas the other two rows refer to the regression $y_t = X\beta + \delta_t + \gamma_1 \hat{\delta}_s + \gamma_2 \hat{p}_0 + f(DUR_t) + \varepsilon_t$, where y_t is either the transaction price or a binary indicator of whether a transaction is taking place for a given property in any given year (we omit the individual subscript i for simplicity). In the first type of regression, the variable of interest is the previous purchase price of the property (p_s). In the second type of regression, the variable of interest is the level of aggregate local house prices at the time of purchase ($\hat{\delta}_s$) and the imputed 1995 value of the property ($\hat{p}_0$) is used as an additional control for housing quality (computed as $\hat{p}_0 = p_s - \hat{\delta}_s$). 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 of years between sales (DUR_t). 'Y×LA' indicates year-by-local authority fixed effects (δ_t in the regression formula). Standard errors (in parentheses) are double-clustered by year and local auhority.

Dependent variable:	**Transaction price** (p_t)					
	Sample 1 (1995-2014)		*Sample 2* (2002-2014)		*Sample 3* (2005-2014)	
	(1)	(2)	(3)	(4)	(5)	(6)
Previous price (p_s)	0.687 (0.017)		0.708 (0.017)		0.825 (0.016)	
Idiosyncratic factor ($\hat{p}_0$)		0.755 (0.014)		0.761 (0.022)		0.844 (0.018)
Previous aggr. factor ($\hat{\delta}_s$)		0.090 (0.021)		0.129 (0.018)		0.185 (0.021)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA
N	7,527,731	7,527,731	3,199,389	3,199,389	1,385,653	1,385,653

Dependent variable:	**Selling probability** (q_t)					
	Sample 1 (1995-2014)		*Sample 2* (2002-2014)		*Sample 3* (2005-2014)	
	(1)	(2)	(3)	(4)	(5)	(6)
Previous price (p_s)	-0.008 (0.002)		-0.009 (0.003)		-0.008 (0.002)	
Idiosyncratic factor ($\hat{p}_0$)		-0.009 (0.003)		-0.008 (0.003)		-0.007 (0.002)
Previous aggr. factor ($\hat{\delta}_s$)		0.001 (0.004)		-0.014 (0.005)		-0.052 (0.006)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA
N	68,925,353	68,925,353	33,828,766	33,828,766	18,170,179	18,170,179

since any coefficient on previous prices may be capturing the effect of unobserved property characteristics rather than pure history dependence. As expected, the regressions yield a large and significant correlation between current and past prices of the property.

The other columns explore the effect of past aggregate prices ($\hat{\delta}_s$). Columns (2), (4), and (6) split the previous sale price (p_s) into a part not related to the aggregate price level (the imputed baseline price, denoted as $\hat{p}_0$—which can be interpreted as the price the house would have fetched in the baseline year, 1995)—and $\hat{\delta}_s$. While the imputed baseline price retains a large and significant coefficient, the effect on $\hat{\delta}_s$ is also positive and significant.

The coefficient on $\hat{\delta}_s$ in the regression on *Sample 1* indicates that an 10 percent increase of the aggregate price level at the time of purchase raises the subsequent selling price of a house by 0.9 percent.

Selling probabilities We aim at investigating whether the purchase price of a property affects the probability that a house sells in any subsequent period. As anticipated in the methodology section, we use a linear model analogous to equation (4) but with a binary dependent variable indicating whether the property was sold in any given year. The lower panel of Table 3 shows the results for history dependence in selling probabilities.

The coefficient on the previous price (p_s) is -0.008 or -0.009 for all samples. These are substantial effects since the average selling probability in the sample is 0.05 as shown in Table 2. The coefficients on past aggregate prices ($\hat{\delta}_s$) indicate no significant effect in *Sample 1*, but negative and significant effects in the more recent samples.

Robustness checks The two panels of Table A1 in the Appendix replicate the results of the initial history dependence regressions for price and quantities using two subsamples: flats and properties which were bought new. If anything, history dependence coefficients are larger than in Table 3 for these more homogeneous subsamples.

4.3 Nonlinear effects and mechanisms

We now use the $\widehat{GAIN}_t$ variable instead of $\hat{\delta}_s$ and split this variable into different bins to capture possibly nonlinear effects of history on current prices and transactions. (Negative bin values indicate losses.) The upper half of Figure 2 shows the effect of gains and losses on transaction prices.[7] A loss is associated to a higher sale price and, in a symmetric way, gains are associated to lower price sales. Interestingly, after a 35 percent gain the effect stabilizes. Standard errors get bigger for larger gains because there are fewer properties with such a long holding period. Moreover, for long tenures the collinearity between $\widehat{GAIN}_t$ and DUR_t increases substantially (only properties with a long holding period experience capital gains of more than 100 percent).

The lower half of the Figure shows the effect of expected gains and losses on selling probabilities. For losses and gains up to 35 percent we have a similar picture to the one above, albeit with the sign reversed. Losses induce lower selling probabilities and gains higher selling probabilities. Once again the effect flattens out and in fact diminishes for large expected gains (and longer durations). Appendix Figure A7 replicates the same analysis with a probit regression (rather than an OLS regression) and displays similar results.[8]

Figure 2 contains coefficients from regressions on all three samples. All samples display the same pattern, but larger and older samples have more coefficients because they span a longer time period. This consistency between samples is in apparent contrast with the different coefficients shown in Table 3. In fact Figure 2 makes it clear that the discrepancies in Table 3 are due to restricting the effect of history dependence to be linear. When the effect is estimated non-parametrically the inconsistencies disappear.

[7] Table A2 and Table A3 in the Appendix show the regression coefficients.

[8] The probit specification is:

$$\text{Prob}(q_t = 1) = \Phi\left[X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2\hat{p}_0 + f(DUR_t) + e_t\right]$$

For computational reasons, the probit regression is esitmated on a 10 (rather than 50) percent random sample of the LR and does not include local-authority fixed effects.

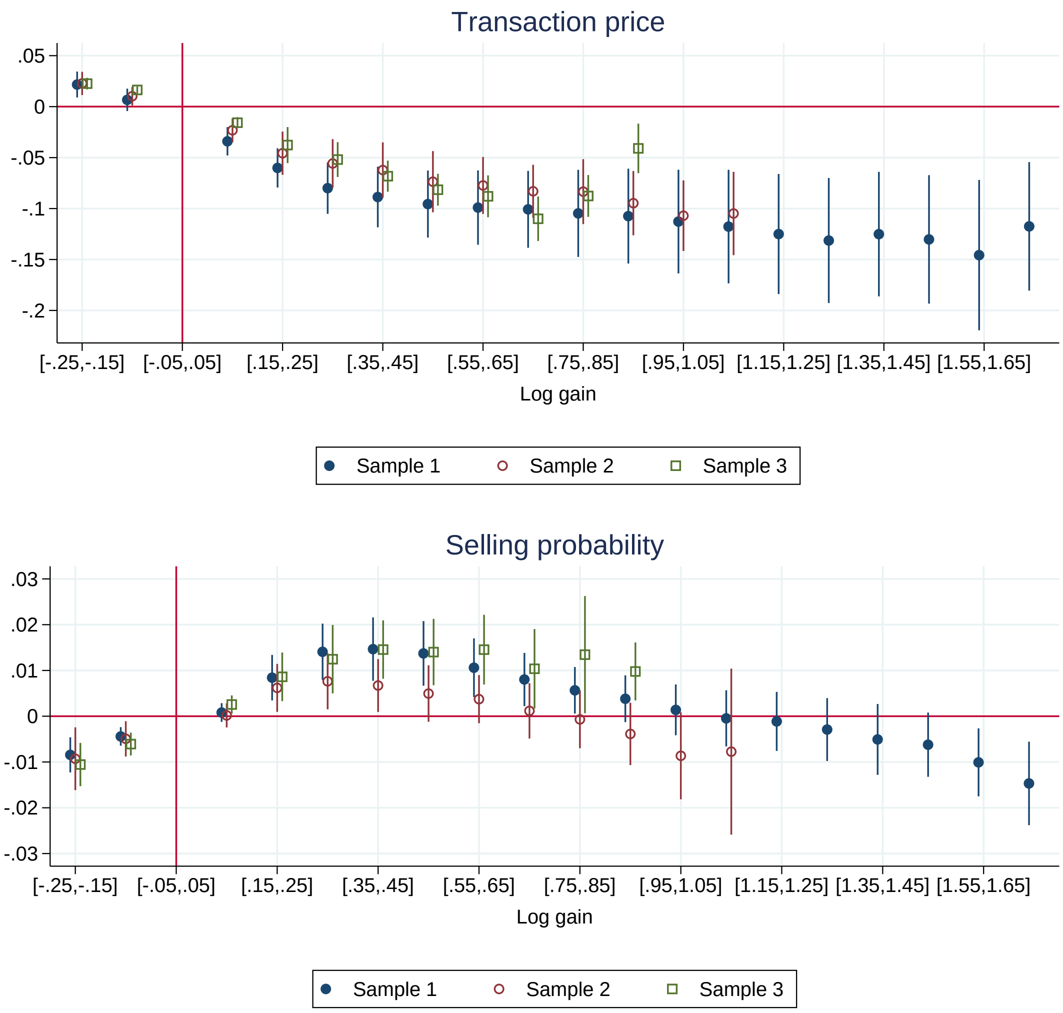

Figure 2: Nonlinear effects of gains and losses

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 $y_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2\hat{p}_0 + f(DUR_t) + e_t$, 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 A2 and A3 in the Appendix. As for the regressions reported in Table 3, 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 of years between sales (DUR_t). Regressions have year-by-local authority fixed effects (δ_t in the regression formula) and standard errors are double-clustered by year and local auhority.

Figure A5 and A6 in the Appendix replicate Figure 2 for each region using *Sample 1*. The pattern of transaction price and selling probability effects appears to be very similar across regions.

Alternative regression specifications are shown in Figures A8 and A9 in the Appendix, for transaction and selling probability analysis respectively. Not including the imputed baseline price $\hat{p}_0$ or the holding period DUR_t in the regressions has a limited impact on results. Using full (6-digit) postcode fixed effects rather than p_0 yields equivalent regression coefficients. Restricting the sample to apartments or properties bought new does not alter the nonlinear effects of $\widehat{GAIN}_t$ on transaction prices and selling probabilities.

The role of credit vs 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 restrict our attention to *Sample 2*—where we can distinguish between properties purchased with cash and properties purchased with a mortgage—and *Sample 3*—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).[9]

We show results graphically in Figure 3 and 4 and in tabular form in Table A2 and A3 in the Appendix.[10]

Both for transaction prices and selling probabilities, Figure 3 shows that the effect on properties bought with a mortgage is not statistically different from the effect on properties bought with cash in all the intervals of $\widehat{GAIN}_t$ considered. In both analyses

[9]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.

[10]Regressions are run on the different subsamples separately. Nearly identical results are obtained by running the regressions on a stacked dataset where the subsamples are distinguished by a dummy variable and the effect of control variables are constrained to be the same on the two subgroups.

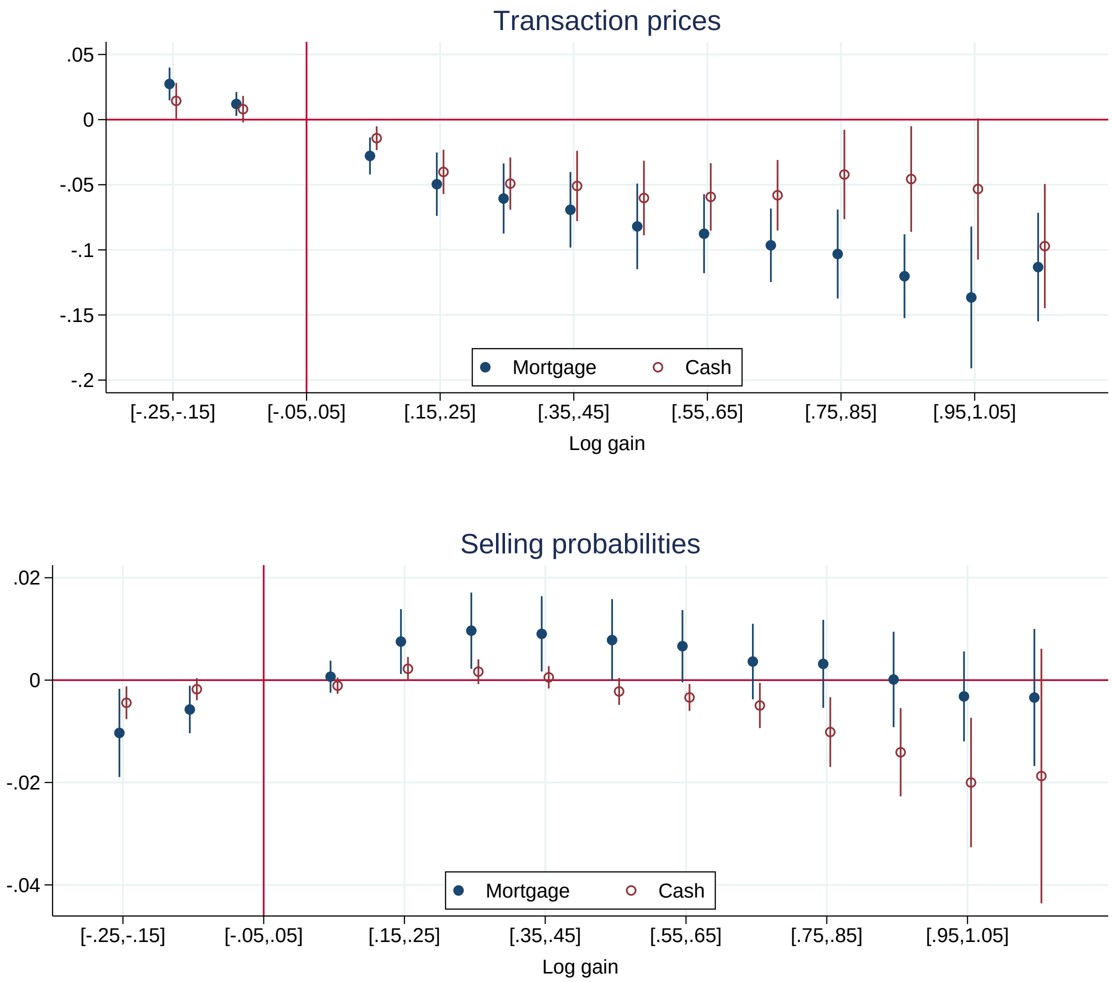

Figure 3: Nonlinear effects of expected gains and losses in *Sample 2*

Notes: The charts replicate the analysis of Figure 2 but uses only *Sample 2* observations and runs the regression $y_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2\hat{p}_0 + f(DUR_t) + e_t$ separately for properties that were bought with a mortgage and properties that were bought with cash. (Information on whether the buyer used a mortgage to finance the transaction is available from the Land Registry since 2002.) The precise values of the coefficients are reported in Table A2 and A3 in the Appendix.

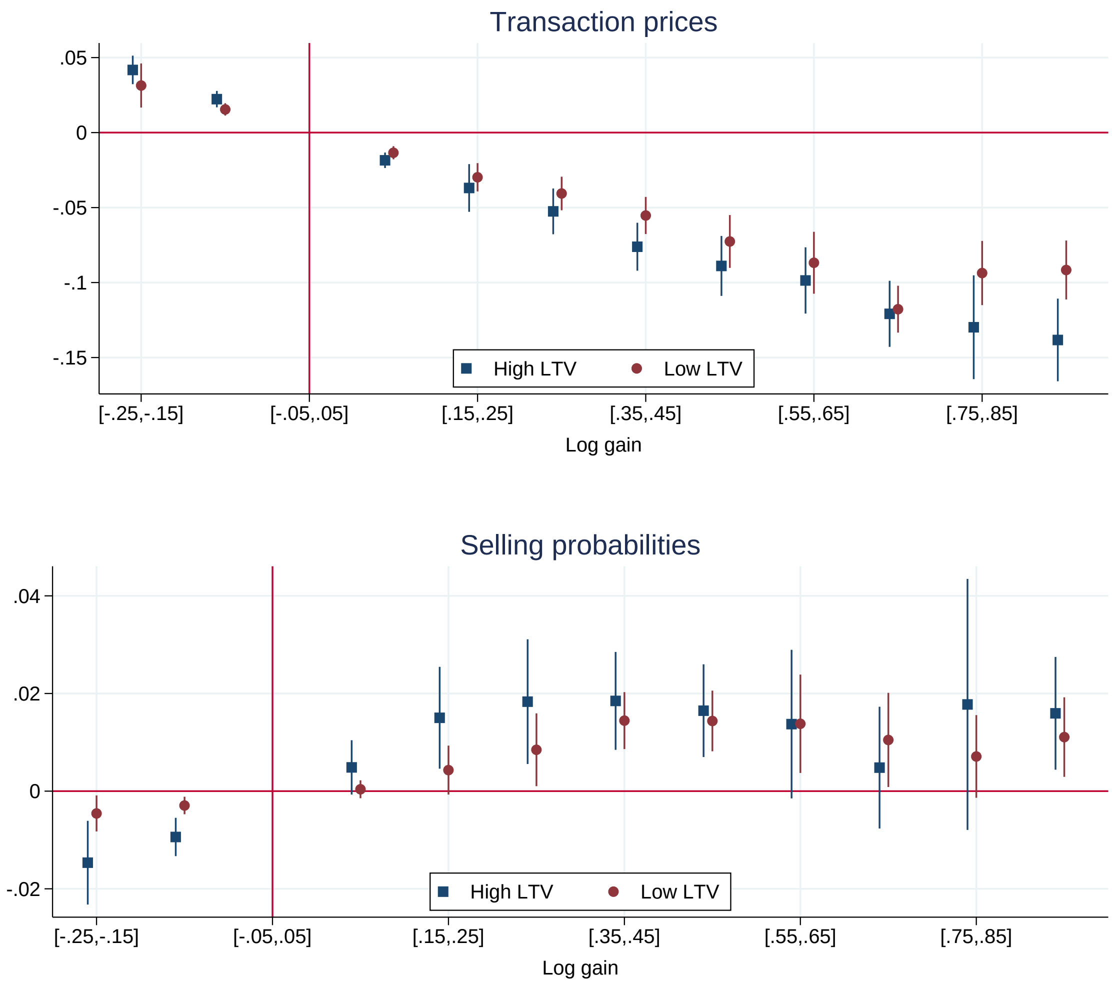

Figure 4: Nonlinear effects of expected gains and losses in *Sample 3*

Notes: The charts replicate the analysis of Figure 2 but uses only *Sample 3* observations and runs the regression $y_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2\hat{p}_0 + f(DUR_t) + e_t$ separately for properties that were bought with a high-LTV or a low-LTV mortgage, where the threshold LTV ratio is 80 percent. 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 *Sample 3*: matched properties bought with a high LTV, matched properties bought with a low LTV, properties that were bought with a mortgage according to the LR but do not match with the PSD, and properties that were bought with cash according to the LR. For the sake of clarity this figure only shows the coefficients on high- and low-LTV properties, but Table A2 and A3 in the Appendix report the exact regression coefficients for all four groups.

however the point estimates for properties bought with a mortgage are always further from the zero line than the coefficients for properties bought with cash. In the regression on selling probabilities, most of the the coefficients corresponding to properties bought with cash are not statistically different from zero. Appendix Figure A10 and A11 confirm that running separate regressions for each region yields similar results, both for transaction prices and selling probabilities.

The analysis on *Sample 3* allows us to highlight the effect of properties bought with a high leverage (properties bought with an LTV higher than 80 percent). Similar to the analysis of *Sample 2*, the effect on highly leveraged properties is larger than on properties bought with a low LTV across the whole range of possible capital gains. However, for individual coefficients across the distribution of gains and losses, we cannot significantly reject the null of equal effects.

The post-2007 fall in transactions Can the results on history dependence be related to the fall in housing market activity that occurred in the England and Wales after 2007? As shown in Figure 1, the aggregate number of transactions did not return to its pre-crisis level even after seven years, in 2014. 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 probability for a property was 7.7 percent; this probability 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 probabilities and shown in the lower half of Figure 2. By summing all these numbers we get the total contribution, in

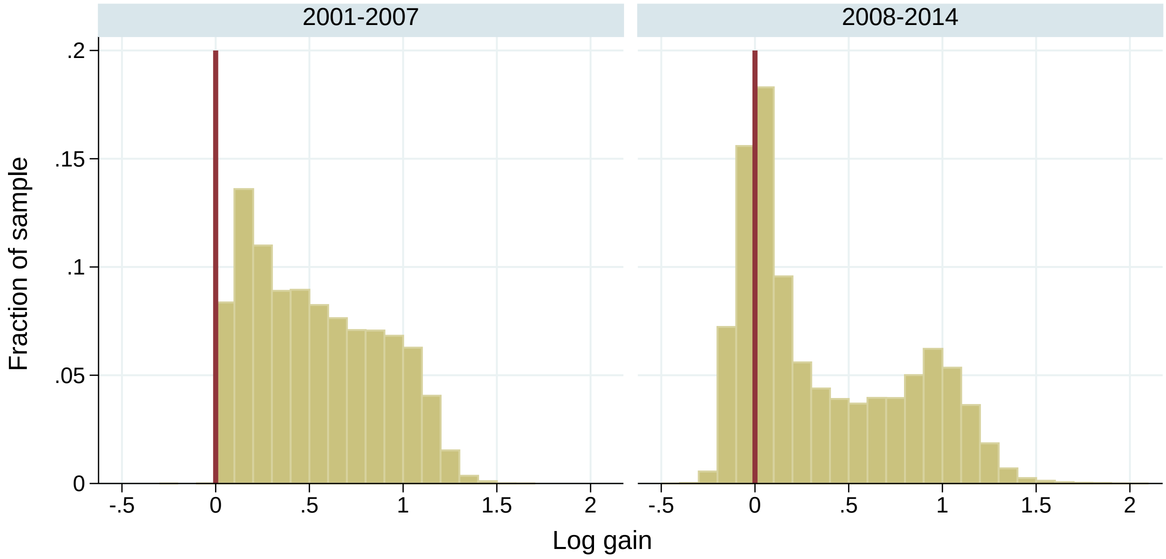

Figure 5: Distribution of ongoing capital gains, pre and post crisis

Notes: The charts show the distribution of the $\widehat{GAIN}_t$ variable in two subperiods: 2001-2007 and 2008-2014. The bin width replicates the allocation of dummy variables used to split $\widehat{GAIN}_t$ and compute the coefficients shown in Figure 2, 3, and 4. For each property, $\widehat{GAIN}_t$ 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 analyisis of selling probabilities and hence $\widehat{GAIN}_t$ 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.

percentage points, of history dependence to the fall in transactions: -0.4. Since the total fall in transactions between the two periods was 4.4 percentage points, history dependence explains around 10 percent of the fall. If we repeat the analysis using the results from the probit regression we get a 13 percent explanatory power.

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 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. We find that England and Wales house prices would have been 4 percent lower in the absence of history dependence.

5 Extensions: list prices and time on the market

In this section we study history dependence in the selling decision process, not just on the outcomes. 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.

5.1 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).

Similar to Tables 1 and 2, 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 selling probability 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 probabilities in the previous section,

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). *Sample Z1* refers to this entire sample whereas *Sample Z1 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		Selling probabilities	
	Sample Z1 (previous LR record in 1995-2014)	*Sample Z1 sold* (matched with LR record in after listing)	*Sample Z1* (previous LR record in 1995-2014)	*Sample Z1 sold* (matched with LR record in 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

the analysis of conditional selling probabilities 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.

5.2 History dependence in list prices and time on the market

In this part of the paper we directly analyse the nonparametric results displayed in Figure 6, which mirrors the way results were presented in Figure 2, 3, and 4 in the previous section.

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 quite small 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 hazard 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 selling probabilities vary significantly between properties with different expected gains. By contrast, when analysing the sample of sold properties, selling probabilities 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

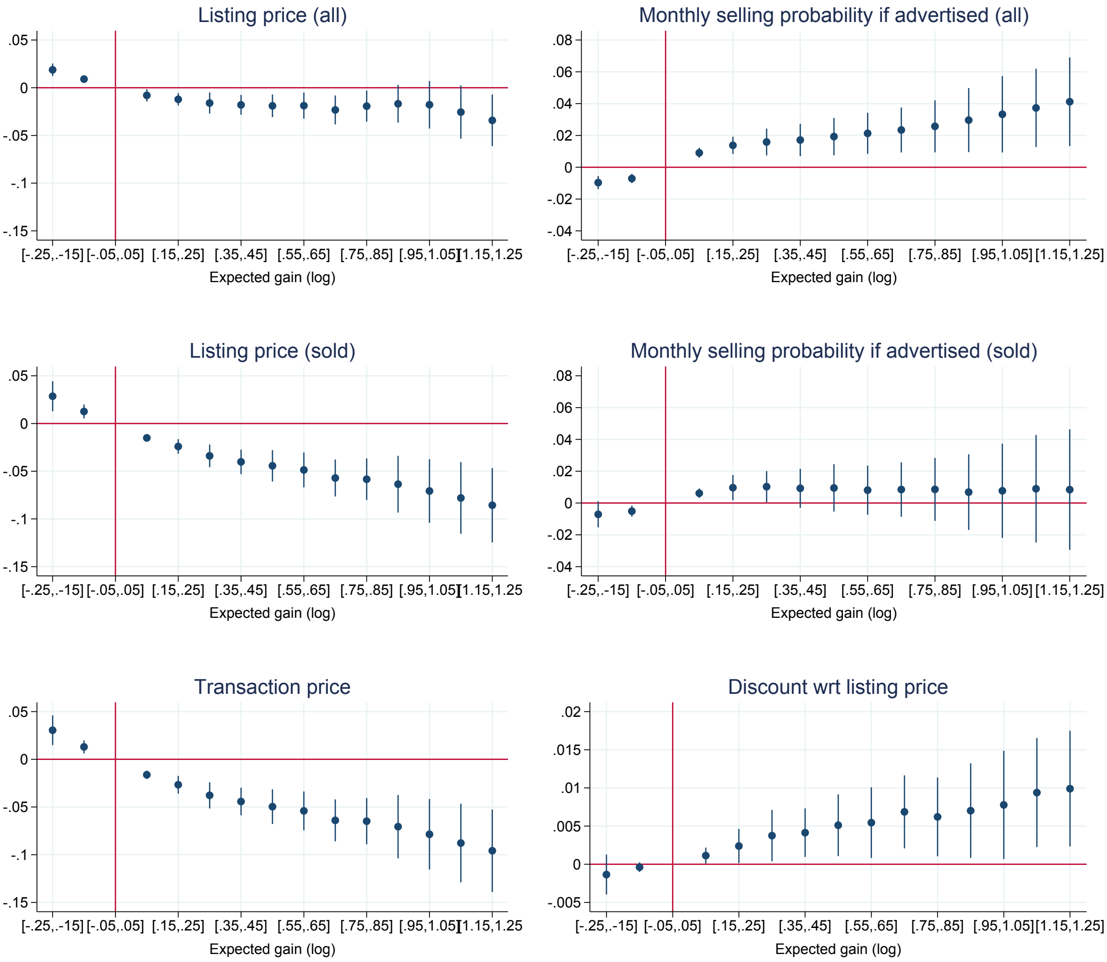

Figure 6: Effects of gains and losses on list prices and time on the market

Notes: The charts report the coefficients and associated 95-percent confidence bands on the $\widehat{GAIN}_{kt}$ dummy variables in the regression $y_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2 \hat{p}_0 + f(DUR_t) + e_t$. 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 Z1*, 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 *Sample Z1 sold*, made of the subset of listings in *Sample Z1* that can be matched with a subsequent sale in the LR, provided that the sale occurs withn 12 months of the listing. Also the bottom row shows results estimated from *Sample Z1 sold*. 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 ($l - p$).

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 prices seems to indicate substantial seller bargaining power.

Comparing the effect on properties bought with a mortgage with properties bought with cash, or the effect on properties bought with a high-LTV mortgages with other properties, yields similar results to the analysis shown in the previous section. Leveraged properties show larger effects on the whole range of expected gains, but the effects are never statistically different from those on non-leveraged properties. The down-payment effect does not seem to be the main driver of history dependence. In the interest of space, we put the relevant charts in the Appendix.

6 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 aggregate house prices in the year a house was previously bought influence the individual price at which the house sells next, as well as the probability that the transaction takes place. The evidence appears to be consistent with the presence of cognitive frictions, either in the form of anchoring or learning. 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), a large part of the effect is independent of leverage and seems to be driven by simple cognitive frictions. The evidence points to significant nominal anchoring or reference dependence without asymmetries.

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 people have nominal anchors. 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.

Beaudry, P. and J. DiNardo (1991): "The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data," *Journal of Political Economy*, 665–688.

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 (2017): "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.

Devereux, P. J. and R. A. Hart (2007): "The spot market matters: evidence on implicit contracts from Britain," *Scottish Journal of Political Economy*, 54, 661–683.

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.

Grant, D. (2003): "The effect of implicit contracts on the movement of wages over the business cycle: Evidence from the national longitudinal surveys," *Industrial & Labor Relations Review*, 56, 393–408.

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

Hilber, C. A. and W. Vermeulen (2016): "The impact of supply constraints on house prices in England," *The Economic Journal*, 126, 358–405.

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

Kahn, L. B. (2010): "The long-term labor market consequences of graduating from college in a bad economy," *Labour Economics*, 17, 303–316.

McDonald, J. T. and C. Worswick (1999): "Wages, Implicit Contracts, and the Business Cyle: Evidence from Canadian Micro Data," *Journal of Political Economy*, 107, 884–892.

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.

Moreira, S. (2016): "Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles," mimeo.

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

Oreopoulos, P., T. von Wachter, and A. Heisz (2012): "The short-and long-term career effects of graduating in a recession," *American Economic Journal: Applied Economics*, 4, 1–29.

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

A.1 Additional figures

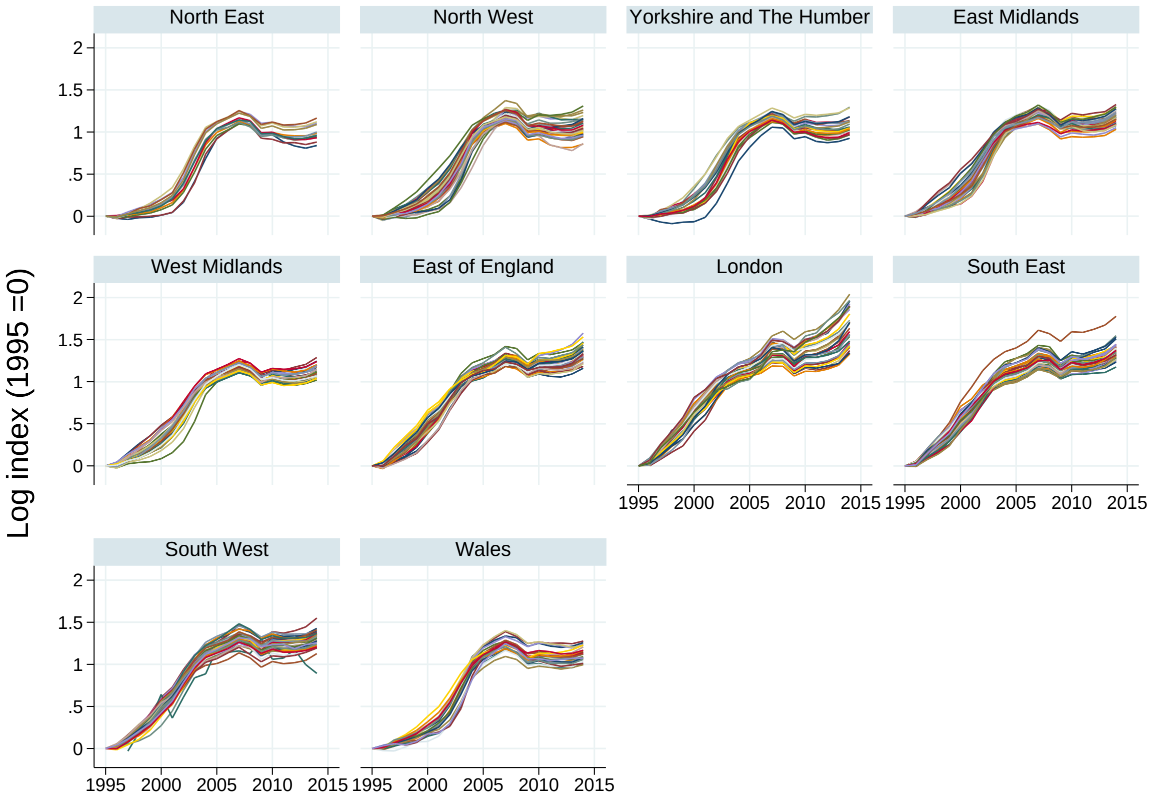

Figure A1: Local house prices

Notes: Quality-adjusted aggregate house prices for this paper are estimated using 1995-2014 Land Registry data from England and Wales. The lines in the charts plot the δ_t coefficients from the regression $p_{it} = X_i\beta + \delta_t + e_{it}$ run for each local authority in England and Wales, where X_i is a vector of housing characteristics included in the Land Registry: type of property, whether the property is new, and whether the property is sold under a leasehold arrangement. In the figure local aggregate house prices are grouped by region to highlight within-region variation.

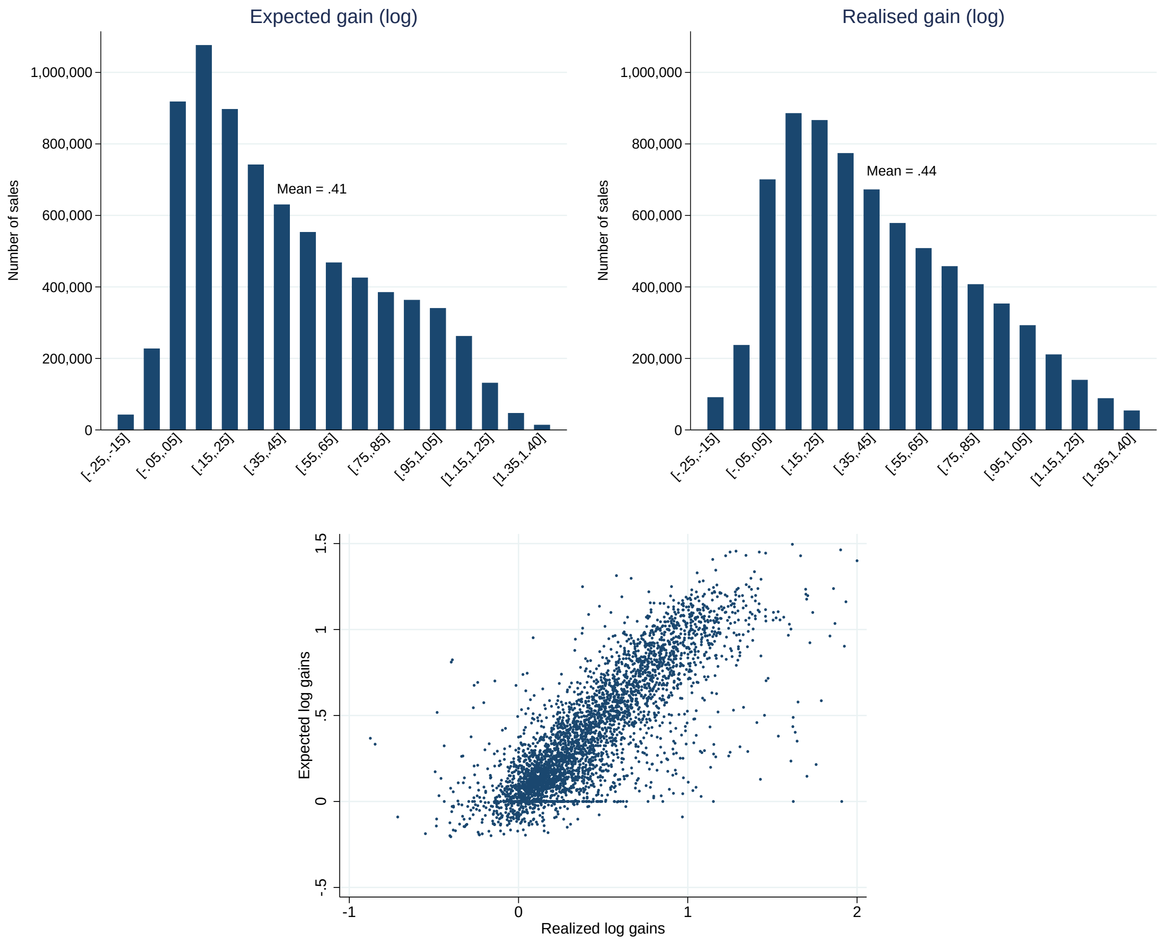

Figure A2: Distribution of gains, 1995-2014

Notes: The upper left chart shows the distribution of expected gains, $\widehat{GAIN}_t$ in *Sample 1*. Expected gains are computed as the change in the local-authority 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 A3: Estimation samples for the analysis of transaction prices

Notes: The charts show graphically the composition of the samples used in the analysis of transaction prices and described in Table 1. The samples are overlapped on light grey bars which represent all the sales included in the England and Wales Land Registry. *Sample 1* 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 is used to include the previous price or the previous aggregate price index in the main regression.) *Sample 2* is a subset of *Sample 1* and refers to properties whose first sale took place after 2001. For this sample we can tell whether the property was purchased with a mortgage. *Sample 3* is a subset of *Sample 2* and refers to 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.

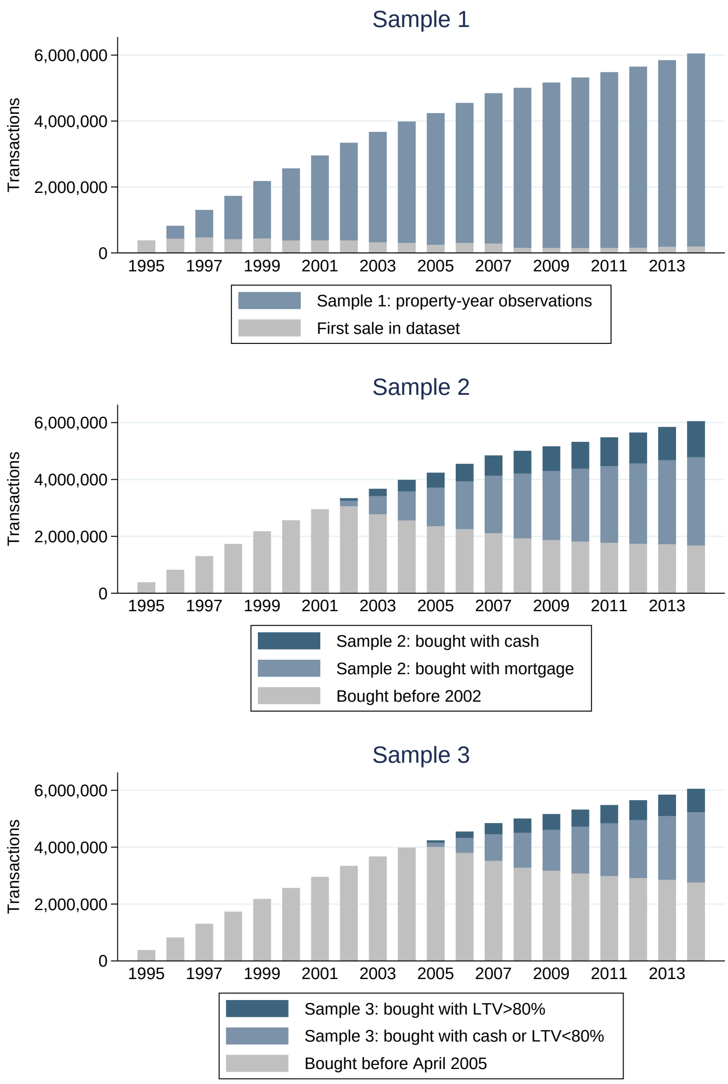

Figure A4: Estimation samples for the analysis of selling probabilities

Notes: The charts show graphically the composition of the samples used in the analysis of selling probabilities and described in Table 2. The samples are made of property-by-year observations. The light grey bars represent sales included in the England and Wales Land Registry (LR) where a property appears for the first time. After that event the dataset 'follows' the property in each year, assigning it a binary variable which indicates whether the property was sold ($q_t = \{0, 1\}$). *Sample 1* is made of property-by-year observations generated by every property which appeared in the LR. The first sale (observation) for each property does not belong to the sample (conceptually it belongs to a previous unobserved spell that ended with the sale). *Sample 2* is a subset of *Sample 1* and refers to properties whose LR first sale took place after 2001. For this sample we can tell whether the property was purchased with a mortgage. *Sample 3* is a subset of *Sample 2* and refers to properties whose LR 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. The property-by-year format of the dataset requires increasing computational resources; this figure refers to the 50% random sample of the LR that is used in the analysis of the paper.

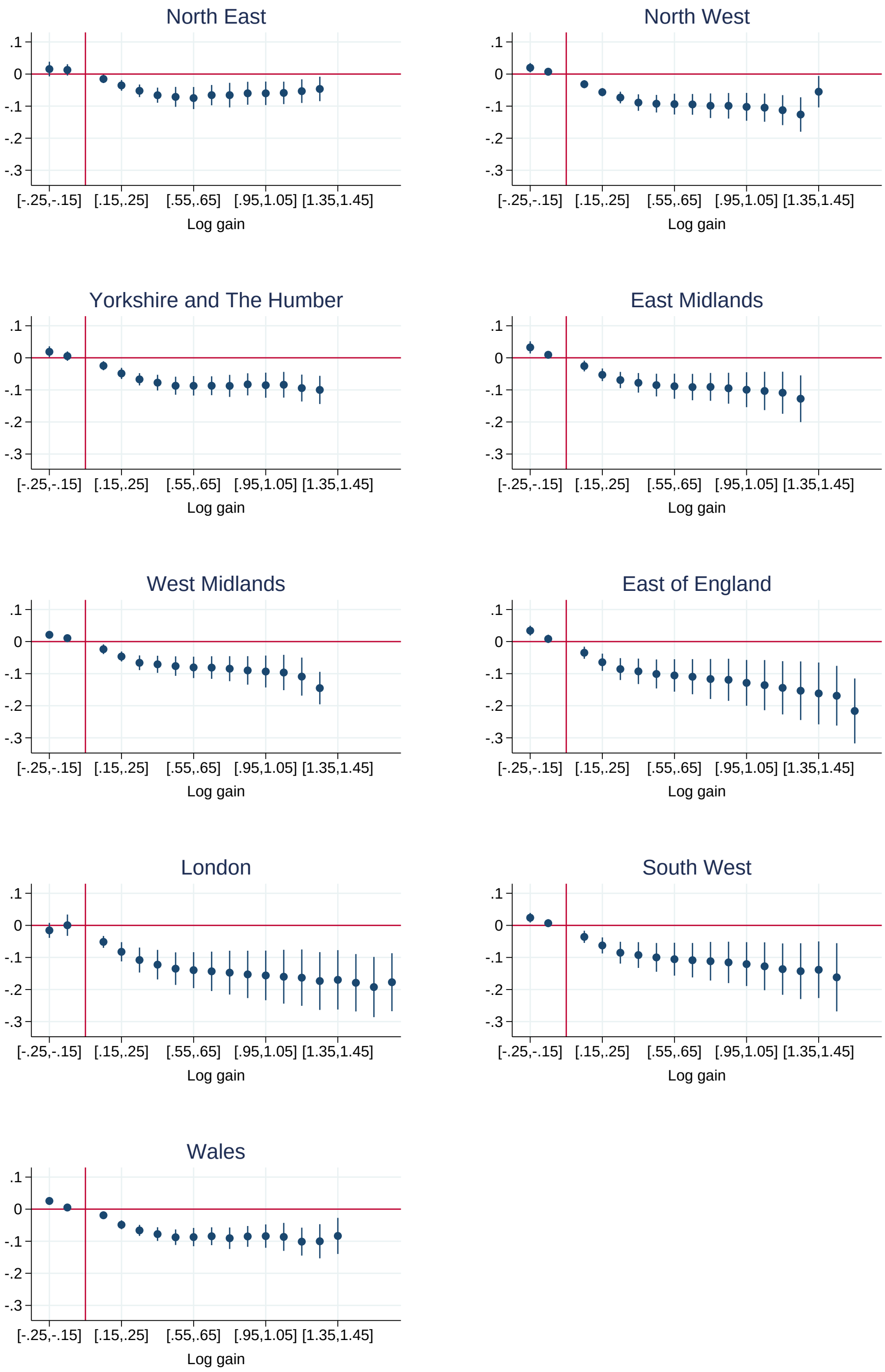

Figure A5: Nonlinear effects of expected gains and losses on transaction prices, by region

Notes: The charts replicate the *Sample 1* analysis of the upper half of Figure 2 for each region in England 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 $p_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2 \hat{p}_0 + f(DUR_t) + e_t$, run separately for each region. Regressions have year-by-local authority fixed effects and standard errors are double-clustered by year and local auhority.

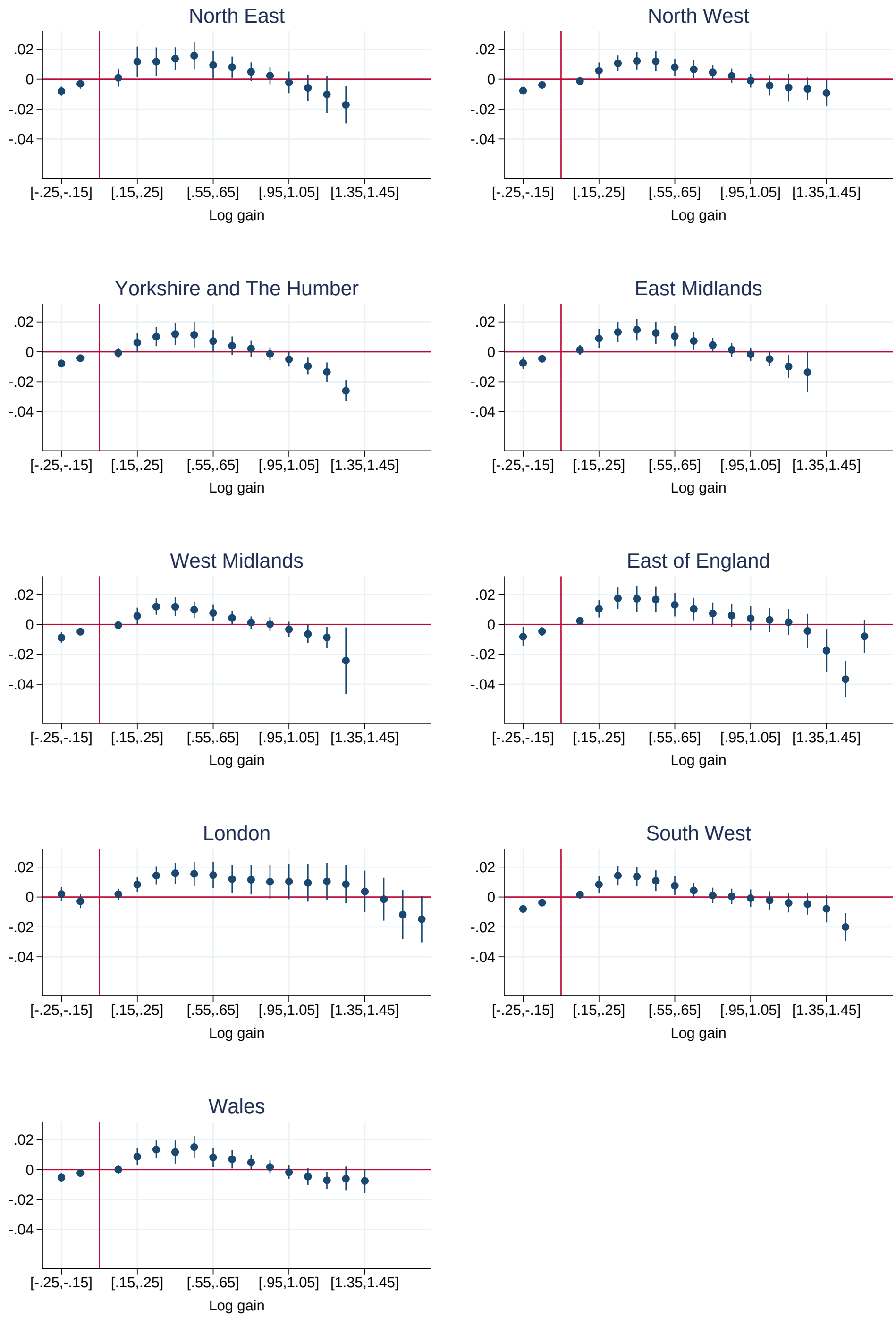

Figure A6: Nonlinear effects of expected gains and losses on selling probabilities, by region

Notes: The charts replicate the *Sample 1* analysis of the bottom half of Figure 2 for each region in England 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 $q_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2 \hat{p}_0 + f(DUR_t) + e_t$, run separately for each region. Regressions have year-by-local authority fixed effects and standard errors are double-clustered by year and local auhority.

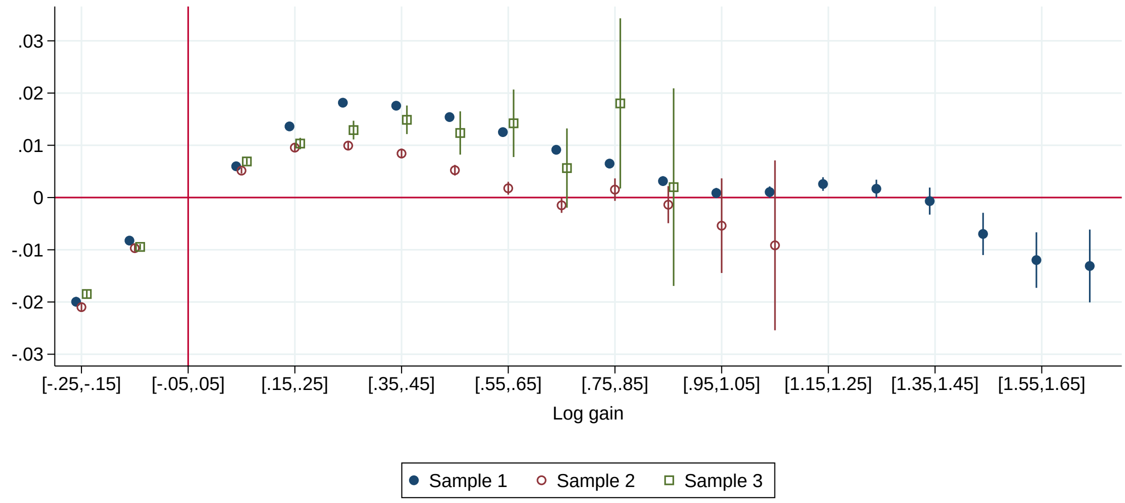

Figure A7: Nonlinear effects of expected gains and losses on selling probabilities, probit regression

Notes: The chart replicates the lower half of Figure 2 using a probit (Prob$(q_t = 1) = \Phi\left[X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2\hat{p}_0 + f(DUR_t) + e_t\right]$) instead of an OLS regression, and shows marginal effects estimated at the means of regression variables. For computational reasons, the probit regression is esitmated on a 10 (rather than 50) percent random sample of the LR, does not include local-authority fixed effects, and standard errors are computed without clustering.

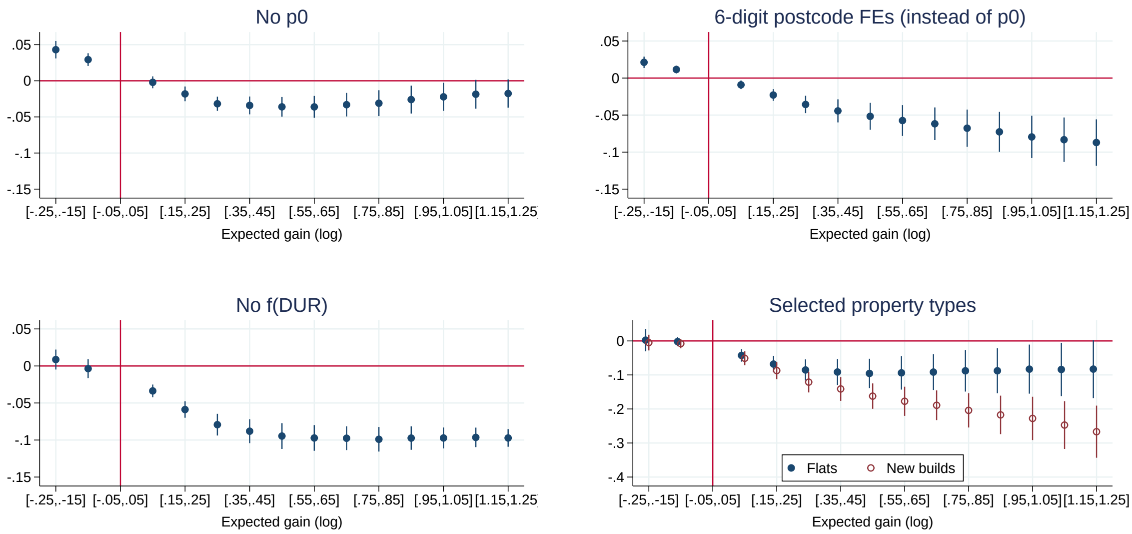

Figure A8: Robustness, transaction prices

Notes: The four charts display the results of alternative specifications in the main regression for nonlinear effects of gains and losses on transaction prices shown in the upper half of Figure 2. Top-left chart: excluding the imputed baseline price $\hat{p}_0$; top-right chart: using full (6-digit) postcode fixed effects rather than the imputed baseline price; bottom-left chart: excluding the third-degree polynomial in holding period; bottom-right chart: separate regressions on samples of only apartments and properties bought new.

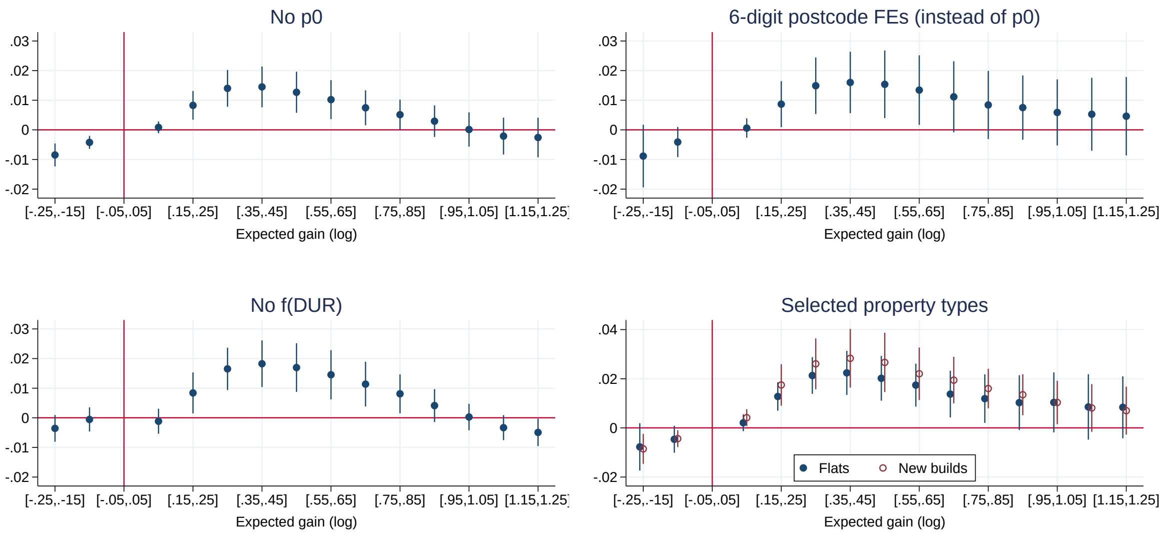

Figure A9: Robustness, selling probabilities

Notes: The four charts display the results of alternative specifications in the main regression for nonlinear effects of gains and losses on selling probabilities shown in the lower half of Figure 2. Top-left chart: excluding the imputed baseline price $\hat{p}_0$; top-right chart: using full (6-digit) postcode fixed effects rather than the imputed baseline price; bottom-left chart: excluding the third-degree polynomial in holding period; bottom-right chart: separate regressions on samples of only apartments and properties bought new.

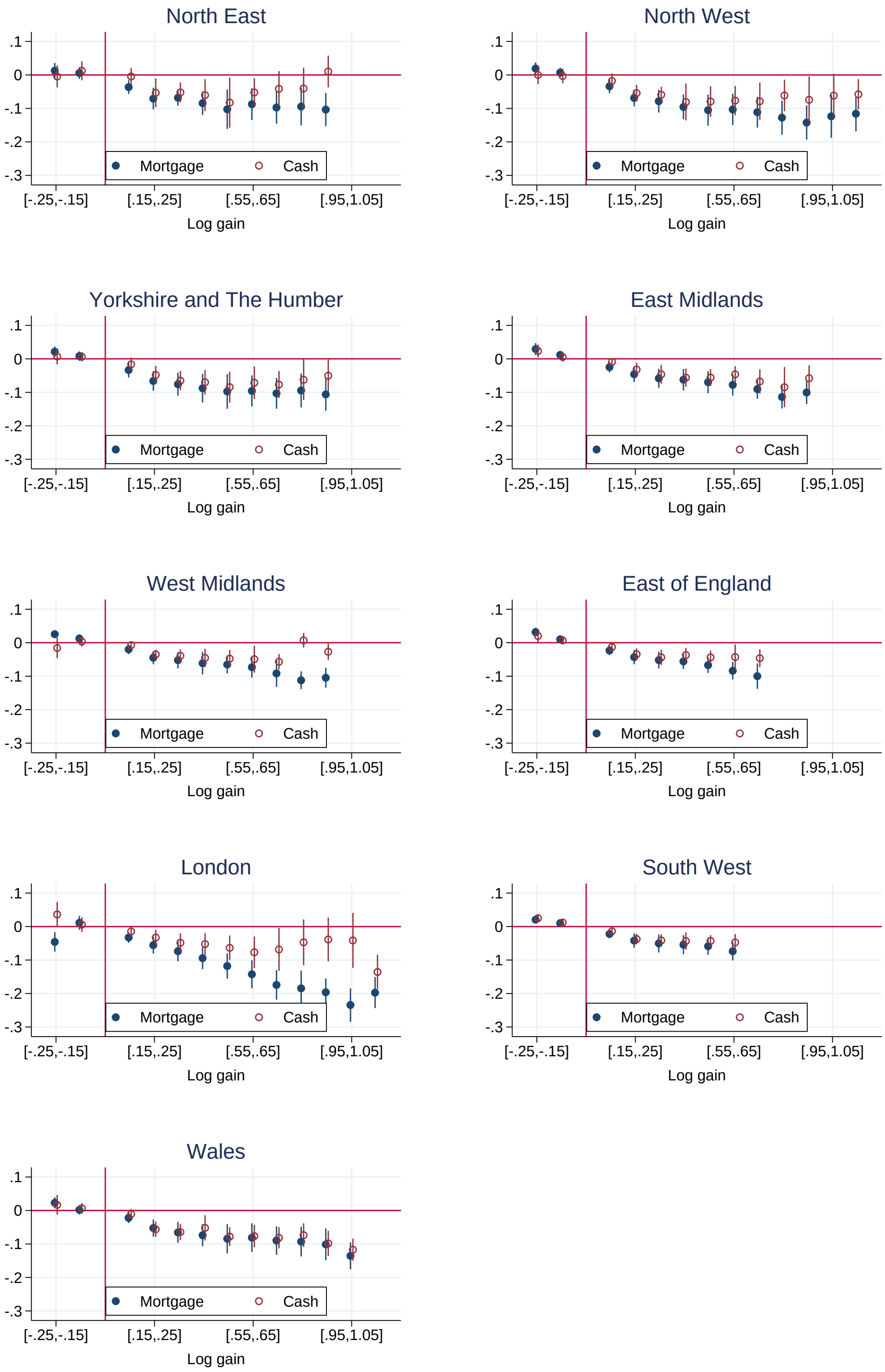

Figure A10: Nonlinear effects of expected gains and losses on transaction prices: properties bought with cash vs properties bought with a mortgage, by region

Notes: The charts replicate the analysis of the upper half of Figure 3 for each region in England Wales. Regressions have year-by-local authority fixed effects and standard errors are double-clustered by year and local auhority.

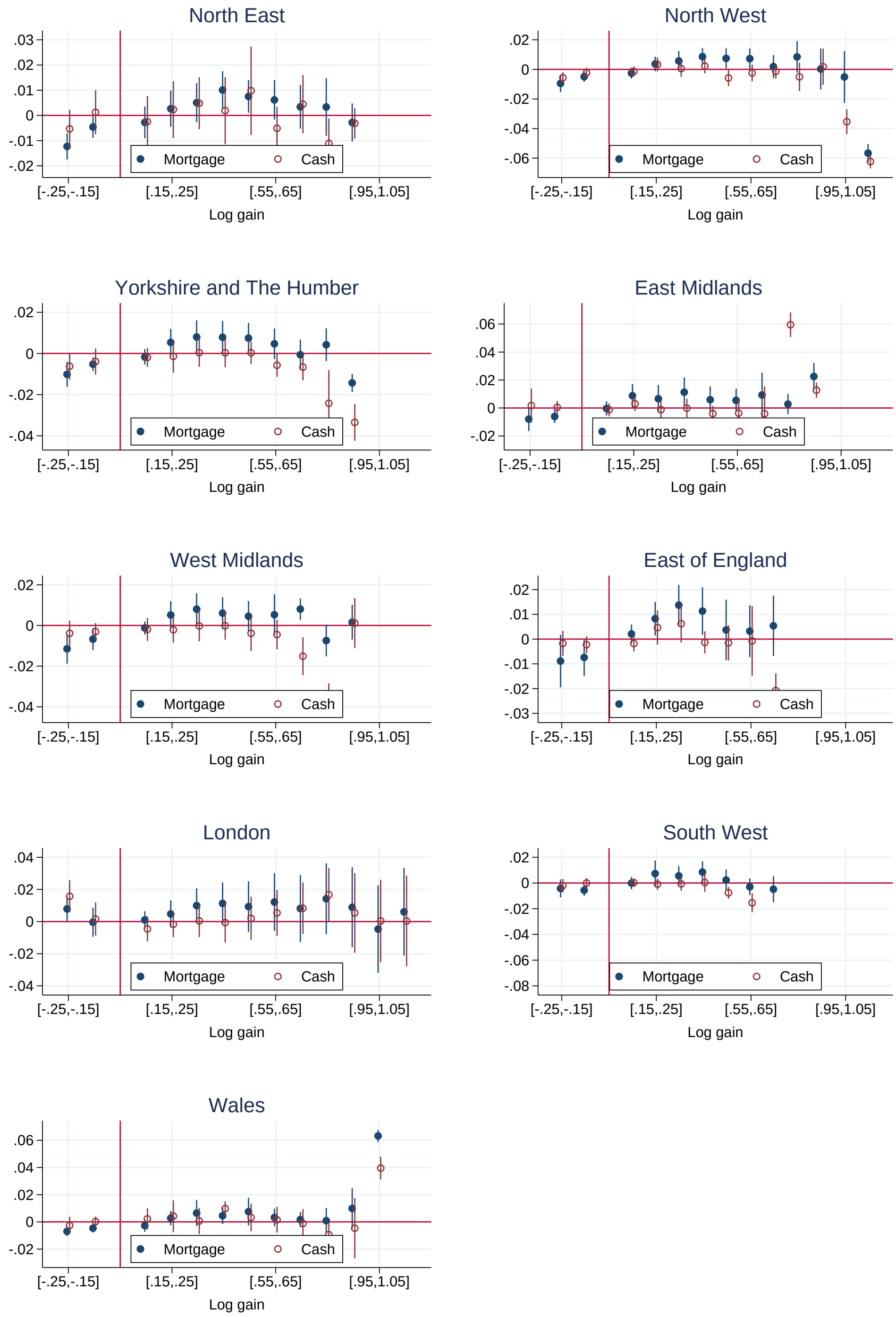

Figure A11: Nonlinear effects of expected gains and losses on selling probabilities: properties bought with cash vs properties bought with a mortgage, by region

Notes: The charts replicate the analysis of the bottom half of Figure 3 for each region in England Wales. Regressions have year-by-local authority fixed effects and standard errors are double-clustered by year and local auhority.

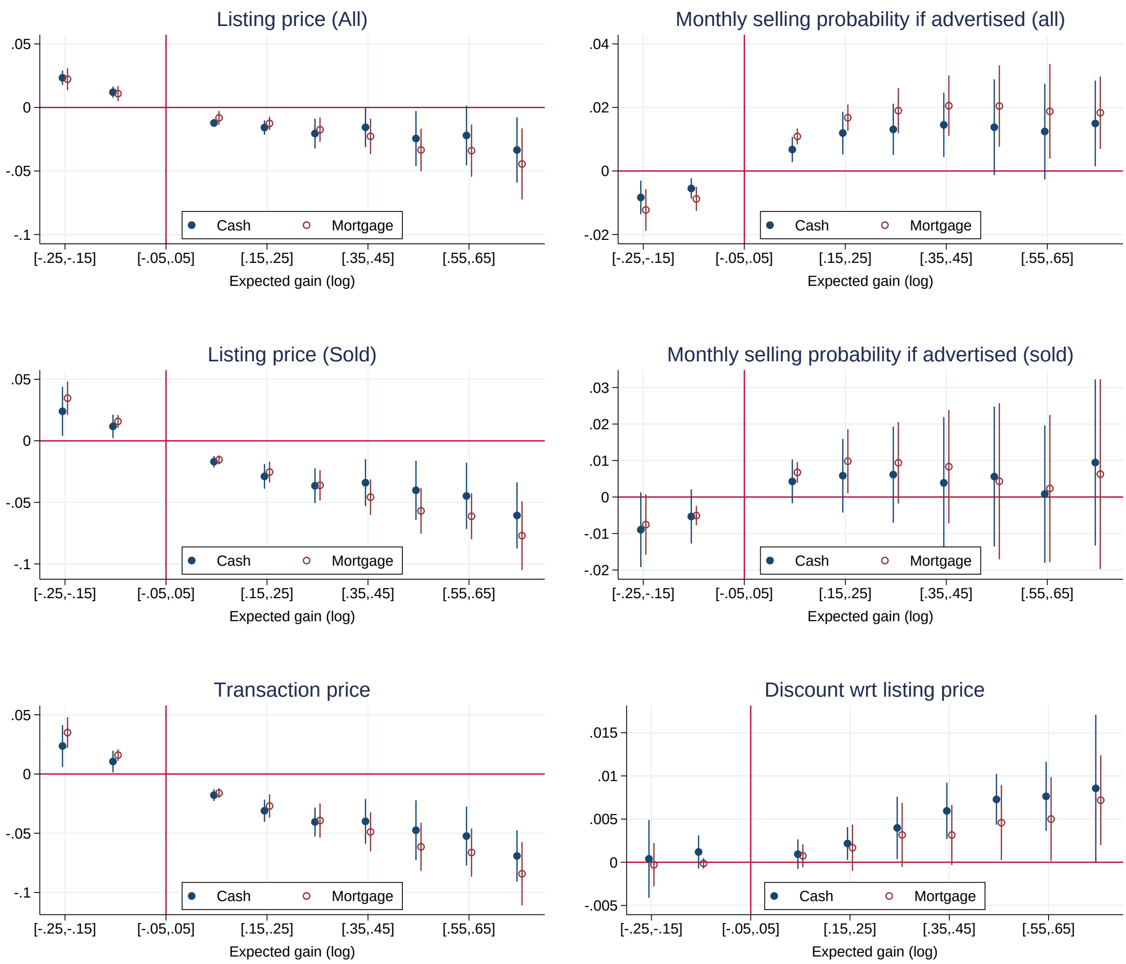

Figure A12: Effects of expected gains and losses on list prices and time on the market: cash vs mortgage-funded properties

Notes: The charts replicate the analysis of Figure 6 for properties that appeared on the Zoopla property portal after 2008 and were previously bought after 2001. For these properties we know the source of funding of the original purchase—mortgage or cash. Distinguishing between these two groups of properties allows us to assess the relative importance of cognitive and credit frictions in generating history dependence in the housing market.

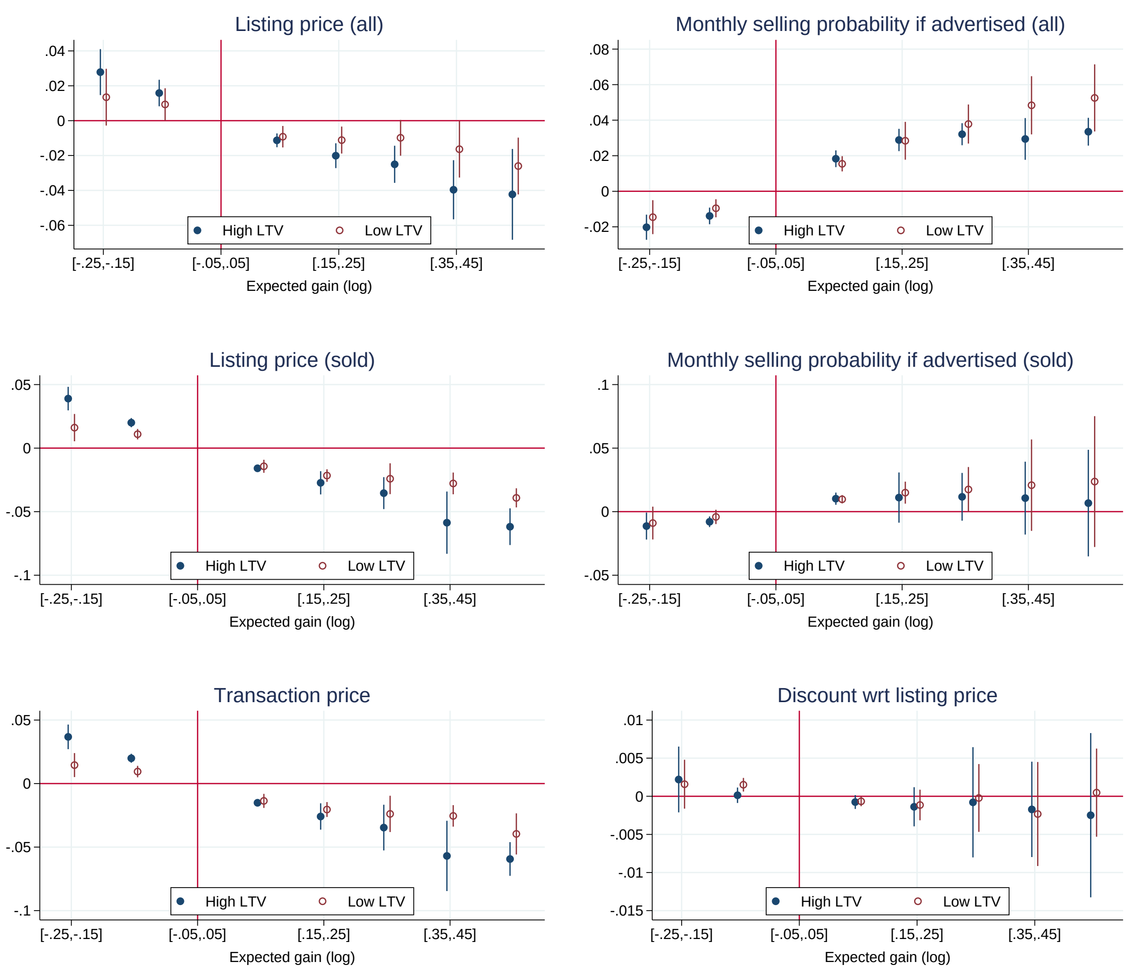

Figure A13: Effects of expected gains and losses on list prices and time on the market: properties bought with high or low LTV

Notes: The charts replicate the analysis of Figure 6 for properties that appeared on the Zoopla property portal after 2008 and were previously bought after 2004. For these properties we know detailed mortgage information thanks to the Product Sales Database of the Financial Conduct Authority. We distinguish between properties with high and low LTV (where the threshold is 80 percent) to assess the relative importance of cognitive and credit frictions in generating history dependence in the housing market.

A.2 Additional tables

Table A1: Robustness regressions

Notes: This table replicates the analysis of Table 3 for two subsets of the data: flats and properties which were bought new. The upper panel of the table reports results for the transaction price analysis and the bottom half of the table reports results for the selling probability analysis. In each of the two panels, coefficients refer to a regression of the form $y_t = X\beta + \delta_t + \gamma_1 \hat{\delta}_s + \gamma_2 \hat{p}_0 + f(DUR_t) + \varepsilon_t$, where y_t is either the transaction price or a binary indicator of whether a transaction is taking place for a given property in any given year (we omit the individual subscript i for simplicity). Standard errors in parentheses are double-clustered by local authority and year.

Dependent variable:	**Transacted price** (p_t)					
	Sample 1 (1995-2014)		*Sample 2* (2002-2014)		*Sample 3* (2005-2014)	
	Flats	Bought new	Flats	Bought new	Flats	Bought new
	(1)	(2)	(3)	(4)	(5)	(6)
Previous aggr. factor ($\hat{\delta}_s$)	0.045 (0.030)	0.207 (0.028)	0.142 (0.033)	0.234 (0.026)	0.109 (0.030)	0.248 (0.048)
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	1,445,912	838,622	697,269	323,861	326,547	133,630

Dependent variable:	**Selling probability** (q_t)					
	Sample 1 (1995-2014)		*Sample 2* (2002-2014)		*Sample 3* (2005-2014)	
	Flats	Bought new	Flats	Bought new	Flats	Bought new
	(1)	(2)	(3)	(4)	(5)	(6)
Previous aggr. factor ($\hat{\delta}_s$)	-0.003 (0.008)	-0.004 (0.005)	-0.032 (0.011)	-0.025 (0.007)	-0.061 (0.012)	-0.054 (0.010)
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	10,686,747	7,055,918	6,273,728	3,364,024	3,647,561	1,847,214

Table A2: Effects of expected gains and losses on transaction prices

Notes: 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 $p_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2\hat{p}_0 + f(DUR_t) + e_t$, where p_t is the transaction price. The coefficients are displayed graphically with their 95 percent confidence bands in the upper half of Figure 2 (column 1, 2, and 5) , 3 (column 3 and 4), and 4 (column 6 and 7). Column 3 and 4 show regression results separately for properties that were bought with a mortgage and properties that were bought with cash. (Information on whether the buyer used a mortgage to finance the transaction is available from the Land Registry since 2002.) Column 6-9 show regression results for properties that were bought with a loan-to-value ration (LTV) greater than 80, properties that were bought with an LTV lower than 80, properties that were bought with a mortgage according to the Land Registry (LR) but do not match with the PSD, and properties that were bought with cash according to the LR. The latter division of *Sample 3* into subgroups depends on the match between LR and PSD which is described in Appendix B.2. Standard errors double-clustered at the year and local-authority level in parentheses.

Dependent variable:	Transaction price (p_t)								
	Sample 1 (1995-2014)	*Sample 2* (2002-2014)			*Sample 3* (2005-2014)				
	All	All	Cash	Mortgage	All	High-LTV	Low-LTV	Mortgage no match	Cash
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gain [-.25,-.15]	0.022	0.023	0.027	0.014	0.023	0.042	0.031	0.003	0.017
	(0.006)	(0.005)	(0.006)	(0.006)	(0.003)	(0.004)	(0.006)	(0.010)	(0.003)
Gain [-.15,-.05]	0.007	0.010	0.012	0.008	0.016	0.022	0.015	0.016	0.016
	(0.005)	(0.004)	(0.004)	(0.005)	(0.001)	(0.002)	(0.002)	(0.005)	(0.002)
Gain [.05,.15]	-0.034	-0.023	-0.028	-0.014	-0.016	-0.018	-0.013	-0.016	-0.011
	(0.007)	(0.005)	(0.007)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Gain [.15,.25]	-0.060	-0.046	-0.050	-0.040	-0.038	-0.037	-0.030	-0.041	-0.041
	(0.009)	(0.010)	(0.011)	(0.008)	(0.008)	(0.007)	(0.004)	(0.008)	(0.006)
Gain [.25,.35]	-0.080	-0.056	-0.061	-0.049	-0.052	-0.053	-0.041	-0.066	-0.057
	(0.012)	(0.011)	(0.012)	(0.009)	(0.008)	(0.007)	(0.005)	(0.009)	(0.006)
Gain [.35,.45]	-0.089	-0.062	-0.069	-0.051	-0.068	-0.076	-0.055	-0.092	-0.051
	(0.014)	(0.012)	(0.013)	(0.012)	(0.007)	(0.007)	(0.005)	(0.010)	(0.015)
Gain [.45,.55]	-0.096	-0.074	-0.082	-0.060	-0.082	-0.089	-0.073	-0.105	-0.060
	(0.016)	(0.014)	(0.015)	(0.013)	(0.007)	(0.009)	(0.008)	(0.017)	(0.011)
Gain [.55,.65]	-0.099	-0.077	-0.088	-0.059	-0.088	-0.099	-0.087	-0.097	-0.069
	(0.017)	(0.013)	(0.014)	(0.012)	(0.009)	(0.010)	(0.009)	(0.013)	(0.007)
Gain [.65,.75]	-0.101	-0.083	-0.096	-0.058	-0.110	-0.121	-0.118	-0.135	-0.061
	(0.018)	(0.012)	(0.013)	(0.012)	(0.010)	(0.010)	(0.007)	(0.017)	(0.048)
Gain [.75,.85]	-0.105	-0.083	-0.103	-0.042	-0.088	-0.130	-0.094	-0.080	-0.010
	(0.020)	(0.015)	(0.016)	(0.016)	(0.009)	(0.015)	(0.009)	(0.012)	(0.013)
Gain [.85,.95]	-0.107	-0.095	-0.120	-0.046	-0.041	-0.138	-0.092	-0.093	0.141
	(0.022)	(0.014)	(0.015)	(0.019)	(0.011)	(0.012)	(0.009)	(0.017)	(0.026)
Gain [.95,1.05]	-0.113	-0.107	-0.137	-0.053					
	(0.024)	(0.016)	(0.025)	(0.025)					
Gain [1.05,1.15]	-0.118	-0.105	-0.113	-0.097					
	(0.027)	(0.019)	(0.019)	(0.022)					
Gain [1.15,1.25]	-0.125								
	(0.028)								
Gain [1.25,1.35]	-0.131								
	(0.029)								
Gain [1.35,1.45]	-0.125								
	(0.029)								
Gain [1.45,1.55]	-0.130								
	(0.030)								
Gain [1.55,1.65]	-0.146								
	(0.035)								
Gain [1.65,1.75]	-0.117								
	(0.030)								
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idiosyncratic factor ($\hat{p}_0$)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA
N	7,525,439	3,197,753	2,298,449	899,304	1,384,017	376,802	365,966	236,794	404,455

Table A3: Effects of expected gains and losses on selling probabilities

Notes: The table is analogous to Table A2 but refers to the regression $q_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2\hat{p}_0 + f(DUR_t) + e_t$, 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-local authority fixed effects (δ_t in the regression formula) and standard errors are double-clustered by year and local auhority.

Dependent variable:	Selling probability (q_t)								
	Sample 1 (1995-2014)	*Sample 2* (2002-2014)			*Sample 3* (2005-2014)				
	All	All	Cash	Mortgage	All	High-LTV	Low-LTV	Mortgage no match	Cash
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gain [-.25,-.15]	-0.008	-0.009	-0.010	-0.004	-0.011	-0.015	-0.005	-0.008	-0.007
	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)
Gain [-.15,-.05]	-0.004	-0.005	-0.006	-0.002	-0.006	-0.009	-0.003	-0.005	-0.003
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Gain [.05,.15]	0.001	0.000	0.001	-0.001	0.003	0.005	0.000	0.003	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Gain [.15,.25]	0.008	0.006	0.008	0.002	0.009	0.015	0.004	0.009	0.006
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.005)	(0.002)	(0.001)	(0.001)
Gain [.25,.35]	0.014	0.008	0.010	0.002	0.012	0.018	0.008	0.012	0.008
	(0.003)	(0.003)	(0.003)	(0.001)	(0.003)	(0.006)	(0.003)	(0.002)	(0.002)
Gain [.35,.45]	0.015	0.007	0.009	0.001	0.015	0.018	0.014	0.012	0.009
	(0.003)	(0.003)	(0.003)	(0.001)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)
Gain [.45,.55]	0.014	0.005	0.008	-0.002	0.014	0.016	0.014	0.012	0.008
	(0.003)	(0.003)	(0.004)	(0.001)	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)
Gain [.55,.65]	0.011	0.004	0.007	-0.003	0.015	0.014	0.014	0.012	0.010
	(0.003)	(0.002)	(0.003)	(0.001)	(0.003)	(0.007)	(0.004)	(0.003)	(0.006)
Gain [.65,.75]	0.008	0.001	0.004	-0.005	0.010	0.005	0.011	0.004	0.012
	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Gain [.75,.85]	0.006	-0.001	0.003	-0.010	0.013	0.018	0.007	0.011	0.012
	(0.002)	(0.003)	(0.004)	(0.003)	(0.006)	(0.011)	(0.004)	(0.010)	(0.002)
Gain [.85,.95]	0.004	-0.004	0.000	-0.014	0.010	0.016	0.011	0.001	0.010
	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)	(0.005)	(0.003)
Gain [.95,1.05]	0.001	-0.009	-0.003	-0.020					
	(0.003)	(0.004)	(0.004)	(0.006)					
Gain [1.05,1.15]	-0.000	-0.008	-0.003	-0.020					
	(0.003)	(0.008)	(0.006)	(0.011)					
Gain [1.15,1.25]	-0.001								
	(0.003)								
Gain [1.25,1.35]	-0.003								
	(0.003)								
Gain [1.35,1.45]	-0.005								
	(0.004)								
Gain [1.45,1.55]	-0.006								
	(0.003)								
Gain [1.55,1.65]	-0.010								
	(0.004)								
Gain [1.65,1.75]	-0.015								
	(0.004)								
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idiosyncratic factor ($\hat{p}_0$)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA
N	68,872,541	33,788,935	24,693,241	9,095,694	18,130,348	4,815,413	5,241,343	3,433,296	4,640,296

Table A4: Effects of expected gains and losses on list prices

Notes: The regressions are similar to those in Table A2 but with WhenFresh/Zoopla list prices (l_t) as the dependent variable: $l_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2 \hat{p}_0 + f(DUR_t) + e_t$. The coefficients are displayed graphically with their 95 percent confidence bands in the top-left chart of Figure 6 (column 1) , **??** (column 3 and 4), and A13 (column 6 and 7). Column 3 and 4 show regression results separately for properties that were bought with a mortgage and properties that were bought with cash. Column 6-9 show regression results for properties that were bought with a loan-to-value ration (LTV) greater than 80, properties that were bought with an LTV lower than 80, properties that were bought with a mortgage according to the Land Registry (LR) but do not match with the PSD, and properties that were bought with cash according to the LR. Standard errors in parentheses are double-clustered at the year and local-authority level.

Dependent variable:	Listing price (l_t)								
	Sample Z1 (Previous LR record in 1995-2014)	*Sample Z2* (Previous LR record in 2002-2014)			*Sample Z3* (Previous LR record in 2005-2014)				
	All	All	Cash	Mortgage	All	High-LTV	Low-LTV	Mortgage no match	Cash
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gain [-.25,-.15]	0.019	0.018	0.022	0.023	0.015	0.028	0.013	-0.004	0.021
	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)	(0.005)	(0.006)	(0.012)	(0.006)
Gain [-.15,-.05]	0.009	0.009	0.011	0.012	0.009	0.016	0.009	0.003	0.013
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)
Gain [.05,.15]	-0.008	-0.008	-0.008	-0.012	-0.009	-0.011	-0.009	-0.009	-0.012
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)
Gain [.15,.25]	-0.012	-0.012	-0.013	-0.016	-0.012	-0.020	-0.011	-0.018	-0.017
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)	(0.002)
Gain [.25,.35]	-0.016	-0.016	-0.018	-0.021	-0.011	-0.025	-0.010	-0.013	-0.025
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.007)	(0.005)
Gain [.35,.45]	-0.018	-0.018	-0.023	-0.016	-0.019	-0.040	-0.016	-0.042	-0.015
	(0.004)	(0.006)	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)	(0.011)	(0.011)
Gain [.45,.55]	-0.019	-0.028	-0.033	-0.024	-0.025	-0.042	-0.026	-0.074	-0.013
	(0.005)	(0.007)	(0.007)	(0.008)	(0.005)	(0.010)	(0.006)	(0.009)	(0.013)
Gain [.55,.65]	-0.019	-0.027	-0.034	-0.022	-0.034	-0.063	-0.032	-0.060	-0.050
	(0.005)	(0.008)	(0.008)	(0.009)	(0.007)	(0.009)	(0.011)	(0.015)	(0.008)
Gain [.65,.75]	-0.023	-0.037	-0.044	-0.033	-0.039	-0.081	-0.032	-0.083	-0.005
	(0.006)	(0.010)	(0.011)	(0.010)	(0.006)	(0.010)	(0.008)	(0.021)	(0.013)
Gain [.75,.85]	-0.019	-0.046	-0.065	-0.006	-0.052	-0.117	-0.042	-0.110	-0.004
	(0.006)	(0.012)	(0.012)	(0.013)	(0.005)	(0.046)	(0.013)	(0.010)	(0.033)
Gain [.85,.95]	-0.017	-0.072	-0.098	-0.016	-0.111	-0.107	-0.124	-0.116	-0.072
	(0.008)	(0.024)	(0.026)	(0.018)	(0.006)	(0.006)	(0.007)	(0.014)	(0.016)
Gain [.95,1.05]	-0.018	-0.076	-0.108	-0.012					
	(0.010)	(0.017)	(0.017)	(0.015)					
Gain [1.05,1.15]	-0.026	-0.080	-0.135	-0.004					
	(0.011)	(0.016)	(0.017)	(0.013)					
Gain [1.15,1.25]	-0.034								
	(0.011)								
Gain [1.25,1.35]	-0.042								
	(0.012)								
Gain [1.35,1.45]	-0.039								
	(0.012)								
Gain [1.45,1.55]	-0.045								
	(0.013)								
Gain [1.55,1.65]	-0.055								
	(0.016)								
Gain [1.65,1.75]	-0.045								
	(0.018)								
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idiosyncratic factor ($\hat{p}_0$)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA
N	2,597,866	1,991,038	1,532,353	458,685	1,380,009	446,460	933,549	210,693	315,186

Table A5: Effects on monthly selling probabilities once listed

Notes: The regressions are similar to those in Table A3 but with an indicator of whether the property advertised for sale on Zoopla has been sold in any given month (h_t) as the dependent variable: $h_t = X\beta + \delta_t + \sum_k \gamma_{1k}\widehat{GAIN}_{kt} + \gamma_2 \hat{p}_0 + f(DUR_t) + e_t$. The coefficients are displayed graphically with their 95 percent confidence bands in the top-right chart of Figure 6 (column 1) , **??** (column 3 and 4), and **??** (column 6 and 7). Regressions have year-by-local authority fixed effects (δ_t in the regression formula) and standard errors are double-clustered by year and local authority. Standard errors in parentheses are double-clustered at the year and local-authority level.

Dependent variable:	Monthly selling probability once listed (h_t)								
	Sample Z1 (Previous LR record in 1995-2014)	*Sample Z2* (Previous LR record in 2002-2014)			*Sample Z3* (Previous LR record in 2005-2014)				
	All	Cash	All	Mortgage	All	High-LTV	Low-LTV	Mortgage no match	Cash
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gain [-.25,-.15]	-0.010	-0.012	-0.012	-0.008	-0.017	-0.020	-0.015	-0.012	-0.015
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)
Gain [-.15,-.05]	-0.007	-0.009	-0.009	-0.005	-0.012	-0.014	-0.010	-0.009	-0.009
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Gain [.05,.15]	0.009	0.010	0.011	0.007	0.016	0.018	0.015	0.013	0.015
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Gain [.15,.25]	0.014	0.016	0.017	0.012	0.029	0.029	0.028	0.023	0.027
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
Gain [.25,.35]	0.016	0.018	0.019	0.013	0.035	0.032	0.038	0.024	0.035
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.003)
Gain [.35,.45]	0.017	0.019	0.021	0.015	0.041	0.029	0.048	0.029	0.048
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.008)	(0.005)
Gain [.45,.55]	0.019	0.019	0.020	0.014	0.044	0.034	0.053	0.027	0.049
	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)	(0.003)	(0.007)	(0.011)	(0.006)
Gain [.55,.65]	0.021	0.018	0.019	0.012	0.043	0.037	0.047	0.018	0.054
	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)	(0.008)	(0.007)	(0.015)	(0.011)
Gain [.65,.75]	0.023	0.018	0.018	0.015	0.046	0.029	0.057	0.025	0.047
	(0.005)	(0.005)	(0.004)	(0.005)	(0.008)	(0.011)	(0.015)	(0.010)	(0.007)
Gain [.75,.85]	0.026	0.020	0.020	0.019	0.031	0.036	-0.017	0.008	0.062
	(0.006)	(0.005)	(0.004)	(0.010)	(0.017)	(0.017)	(0.014)	(0.032)	(0.013)
Gain [.85,.95]	0.030	0.026	0.029	0.012	0.053	0.007	0.018	0.063	0.063
	(0.008)	(0.010)	(0.009)	(0.013)	(0.011)	(0.013)	(0.016)	(0.013)	(0.010)
Gain [.95,1.05]	0.033	-0.005	0.001	-0.028					
	(0.009)	(0.011)	(0.012)	(0.013)					
Gain [1.05,1.15]	0.037	0.025	0.039	-0.006					
	(0.010)	(0.012)	(0.013)	(0.011)					
Gain [1.15,1.25]	0.041								
	(0.011)								
Gain [1.25,1.35]	0.040								
	(0.011)								
Gain [1.35,1.45]	0.038								
	(0.013)								
Gain [1.45,1.55]	0.037								
	(0.009)								
Gain [1.55,1.65]	0.029								
	(0.012)								
Gain [1.65,1.75]	0.033								
	(0.019)								
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idiosyncratic factor ($\hat{p}_0$)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA	Y×LA
N	13,778,554	10,470,053	8,030,645	2,439,408	7,138,462	2,305,677	2,088,638	1,102,618	1,641,529

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 A14 shows that the total number of cash purchases in England and Wales is less cyclical than the number of mortgages.

Table A6 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).

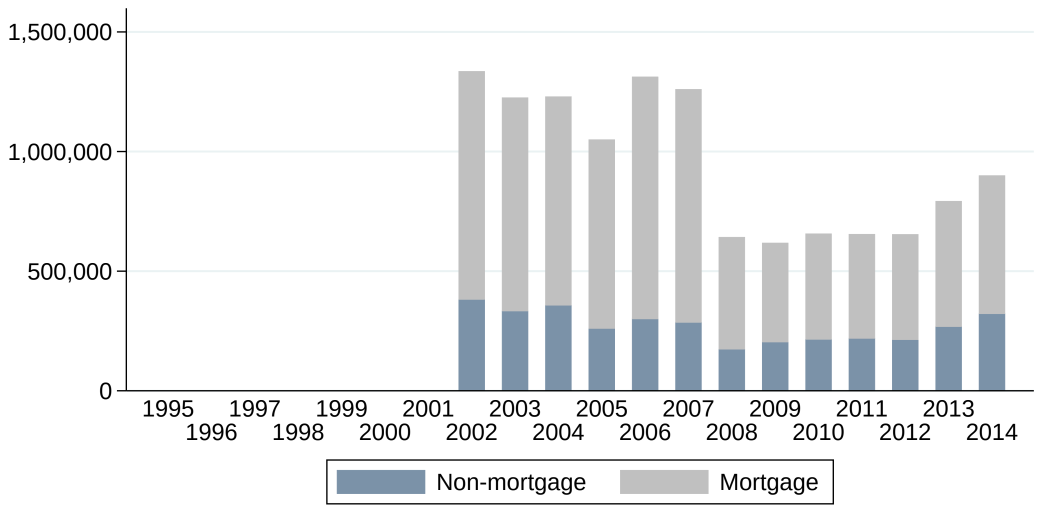

Figure A14: Mortgage vs non-mortgage purchases, 2002-2014

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 A6: 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.

	Sample 2 (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 ($\widehat{GAIN}_t$)		
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.[11]

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.[12]

The date in which a bank transfer the mortgage amount to the buyer is the completion date or a few days before. Figure A15 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.[13] 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

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

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

[13] 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.

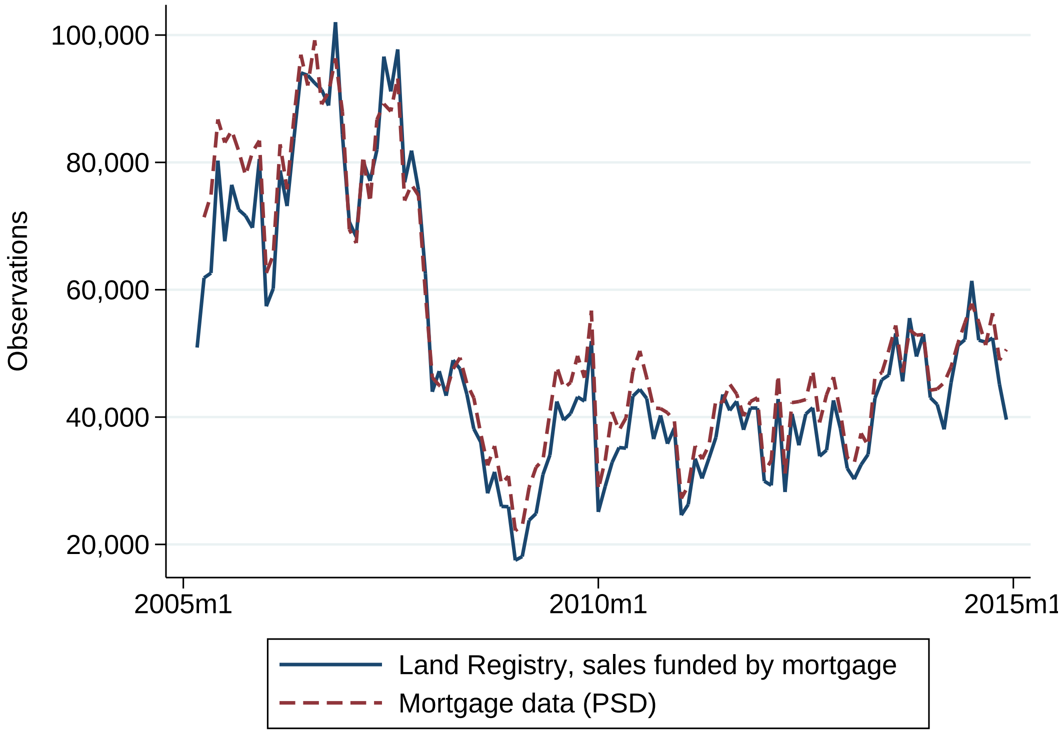

Figure A15: Number of observations by month in the Land Registry and Product Sale Data before matching

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.[14]

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 A7 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 A16 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

[14]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 A7: 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 ($\widehat{GAIN}_t$)				
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

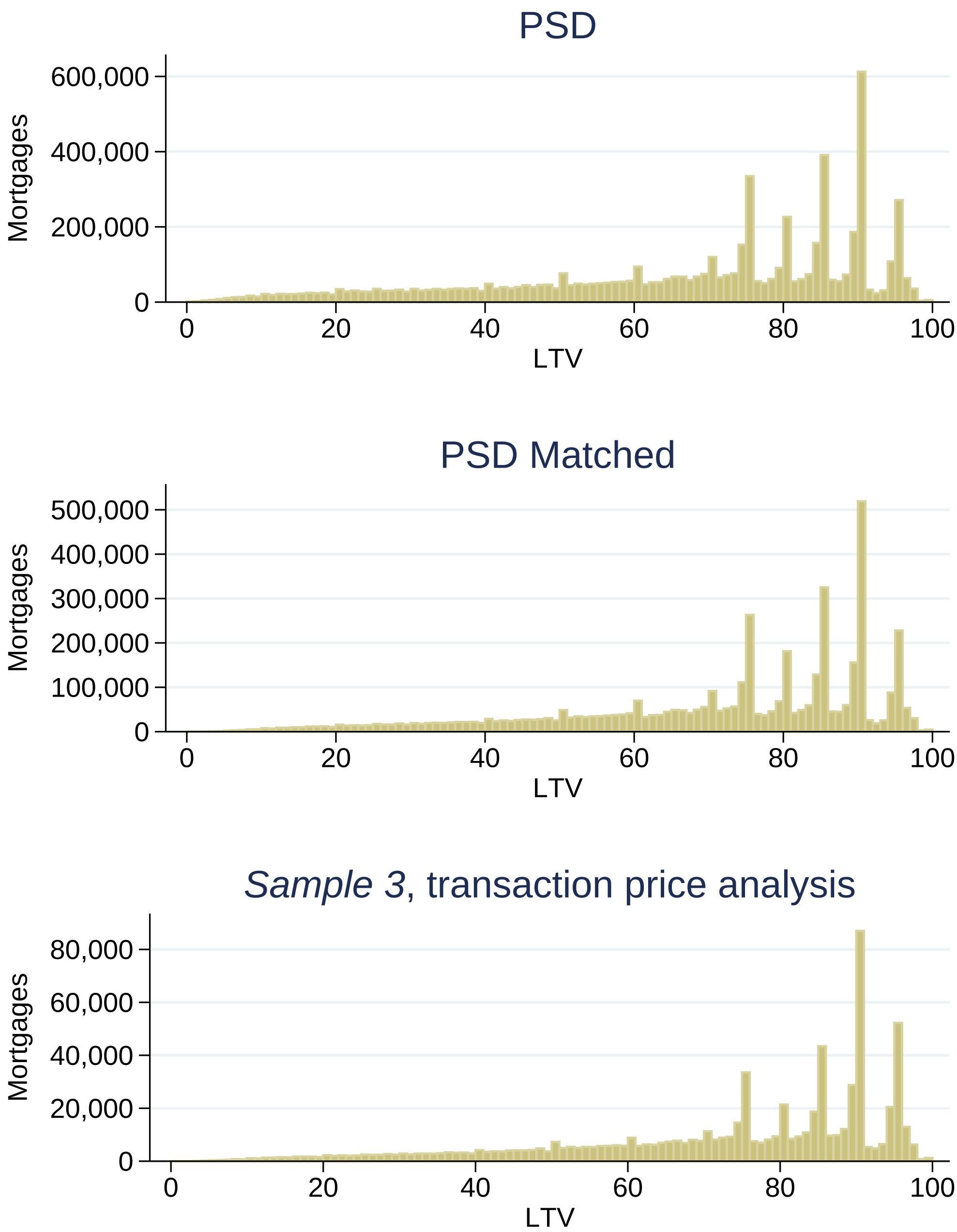

Figure A16: LTV distributions in the Product Sales Data and the matched observations

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[15]—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

[15]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,[16] 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.

[16] 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.