Who Gets the House?

Related-Party Transfers and the Cash–Mortgage Price Gap in New York City

Pablo Loschi*

Abstract

Using 6,006 repeat-sale pairs for 5,957 New York City houses in 2016–2025, I estimate a financed-minus-cash price contrast of 9.25 log points (95 percent interval [7.17, 11.35]). I classify sales without a matched mortgage as cash. The contrast falls to 4.87 [2.69, 6.76] after I exclude 358 pairs with a same-surname buyer and seller at either sale. In repeated deletions of comparable unflagged pairs, the estimate changes little. In a sample of cash deeds, transfer reports show non-market signs on 16 of 40 flagged deeds and 2 of 20 unflagged deeds. With common borough-quarter controls, the house contrast exceeds the condominium contrast by 5.81 log points (95 percent interval [4.00, 7.77]) after I apply the surname screen, require at least five years between sales and trim both tails of annualized price growth within each property type. The sensitivity to transaction composition limits what the recorded price gap can tell us about the value of closing certainty.

JEL: G21, G51, R21, R31. *Keywords:* cash purchases; mortgage finance; related-party transfers; execution risk; repeat sales; New York City.

*Independent researcher, Berlin, Germany. Correspondence: loschi.pablo@gmail.com. ORCID: 0009-0004-9455-4713. I thank Mercedes Arana Puchuri, who read the register documents as a research assistant. This research received no external funding. The author founded HolaCasaClara, a free comparator of Argentine mortgage loans, which has no connection to New York housing records or this study. No other competing interests or outside rights to review the research before dissemination are reported.

1 Introduction

Buyers who pay cash often pay less. Asabere et al. (1992) document a cash discount among comparable row homes. Reher and Valkanov (2024) estimate an 11 percent mortgage premium across US markets, and Han and Hong (2024) use Los Angeles listing and closing records to examine how financing risk shapes the discount. In May 2026 New York legislators considered a 1 percent tax on cash purchases of homes for $1 million or more in New York City. The enacted budget omitted it (Bloomberg News, 2026a,b; Office of the Governor, State of New York, 2026). A tax intended to offset a cash buyer's closing advantage needs a measure of that advantage. In public records, the measured price gap also reflects which buyers and transactions enter each group.

I link sales in Manhattan, the Bronx, Brooklyn and Queens from 2016 to 2025 to deeds and mortgages. I compare the appreciation of properties that switch from an unmatched to a matched mortgage with those that switch in the other direction. For houses the financed-minus-cash contrast is 9.25 log points [7.17, 11.35], with ZIP-year effects and buyer and seller type controls. It stays positive when I restrict the sample to consistent records, omit any borough or change the window used to match mortgages.

I flag a deed when a natural-person seller and buyer share a surname. Flagged deeds account for 5.7 percent of cash-labelled endpoints and 2.3 percent of financed endpoints. Excluding 358 pairs with a flagged deed lowers the contrast to 4.87 log points, a paired change of -4.38 [-5.63, -3.18]. A low recorded price at the cash sale raises measured appreciation in one switching arm and lowers it in the other, which can widen the contrast even when the two arm effects nearly offset in a symmetry check. Across 999 draws that delete the same number of unflagged pairs matched on observables, the median change is +0.44 log points. A shared surname is an imperfect proxy. The observed change combines the effects of dropping pairs with different transaction circumstances and refitting the model.

Transfer reports help explain what the surname rule picks up. A research assistant read the reports for a random sample of flagged and unflagged cash deeds, and I checked the readings against the same images. In that sample, 40 percent of flagged deeds and 10 percent of unflagged deeds show a non-market sign, such as a declared transfer between relatives. Estimates that limit the influence of extreme appreciation give 2.83 to 4.87 log points after the surname screen. The company comparison varies with the buyer-type controls.

Using prior municipal market value and pooled controls, the screened house contrast at or above $1 million is 4.33 log points [−1.10, 9.31]. The difference from the lower band is −0.94 log points [−6.55, 4.45], with too little precision to distinguish the bands. Under common borough-quarter controls, the condominium contrast moves from 0.83 to -0.37 log

points after the surname screen. Among pairs held at least five years, after trimming the growth tails within each property type, the screened contrast is 4.78 for houses and −1.03 for condominiums, a difference of 5.81 log points [4.00, 7.77].

Reher and Valkanov (2024) discuss why family transfers could inflate a cash comparison and use national data that exclude flagged intrafamily transfers and foreclosures. Han and Hong (2024) restrict their estimation to arm's-length sales, and Campbell et al. (2011) remove apparent intrafamily transfers when studying house prices in Massachusetts. I measure how much an implementable screen changes a repeat-sale financing contrast in public records. Matched deletions show how much the contrast typically moves when comparable unflagged observations are dropped, and the checked transfer reports show what the surname flag catches. The design controls for fixed property attributes (Bailey et al., 1963). Selection into financing and changes in property condition can still matter (Tidwell et al., 2018; Nowak and Smith, 2020). Bargaining varies with buyer and seller characteristics, and survey evidence shows that down-payment requirements can change what buyers are willing to pay (Harding et al., 2003; Fuster and Zafar, 2021).

I do not observe contracts, contingencies or failed closings, and "cash" here means no matched mortgage (Appendix A). Nor do I estimate the incidence of the mortgage recording tax.

2 Financing institutions in New York City

2.1 The recording tax and ownership form

New York Tax Law Article 11 taxes mortgages of real property (New York State, 2026; New York City Department of Finance, 2026e). Sections 253 and 253-a, together with the city schedule, produce standard borrower rates of 1.80 percent below $500,000 of mortgage principal and 1.925 percent at or above that amount for the covered residential categories. A qualifying lender generally pays the separate 0.25 percent special additional tax, and exemptions and instrument-specific credits can change the amount due. Section 253(2) excludes the first $10,000 of principal for qualifying one- and two-family residences, reducing that component by $30. These rules are distinct from transfer taxes on the sale price.[1]

For illustration, a standard $720,000 mortgage used to buy a $900,000 condominium implies a borrower recording charge of $13,860 before any applicable instrument-specific credits. A cash purchase records no purchase mortgage, and a co-operative share loan is

[1]Primary sources: New York Tax Law §§253, 253-a, and the NYC Department of Finance mortgage recording tax schedule.

secured by personal property rather than a mortgage on an individually owned real-property unit, so it does not attract the same tax.

The distinction is statutory, and because sellers' prices, financing terms and ownership choices can all respond to the tax, statutory borrower liability is unlikely to equal the ultimate economic burden. Even a cut in a buyer-paid transaction tax is partly passed to sellers: Ke et al. (2025) estimate that buyers kept about 80 percent of the surplus from the United Kingdom's 2020–21 stamp duty holiday. The same boundary creates a measurement problem, since a search of real-property mortgages alone cannot identify co-operative share financing, and ACRIS personal-property filings have to be examined separately.

The proposal considered in May 2026 would have taxed purchases without a mortgage. These records cannot establish its full tax base: an unmatched purchase may involve unobserved borrowing, and the condominium linkage leaves some sales unclassified.

3 Data and record linkage

3.1 Samples and the boundary of replication

The combined five-borough house input (208,352 sales) and four-borough rebuilt condominium input (86,381 sales) total 294,733 sale rows. Excluding the 41,445 Staten Island house sales, whose deeds are recorded by the Richmond County Clerk outside ACRIS, leaves 166,907 house and 86,381 condominium sales (253,288 four-borough purchases in all). The condominium sample requires an original parcel identifier and a recorded price strictly between $300,000 and $6,000,000.

I combine NYC Department of Finance annualized sales with ACRIS real-property master and legal records. Because condominium billing lots can differ from the unit lots on which deeds and mortgages are recorded, I recover a unique deed through borough, block, consideration and date before assigning the unit parcel.

Public records add DOB permit dates, PLUTO geography, ACRIS deed parties and consideration, co-operative transactions and personal-property filings (New York City Department of Finance, 2026c,b,a; New York City Department of Buildings, 2026; New York City Department of City Planning, 2026; Freddie Mac, 2026). The house analysis starts from its original panel. I rebuilt the condominium panel from the sale and document extracts, using unit lots 1001–6999. The online appendix documents the computational boundary and the corrected inputs.

Linkage quality has several dimensions: Bailey et al. (2020) distinguish match rates and sample representativeness from false-link and missed-link errors, but their error rates

come from historical census data and are unlikely to carry over to ACRIS. In this paper, agreement with existing flags checks the computation, and readings of the original register images, made without AI by a research assistant and checked by me, test the financing labels and the surname flag on random samples (Appendices C and D.3).

3.2 Financing and repeat-sale construction

For houses and condominiums, financing is the recorded indicator of a matched purchase-window mortgage. The strict, base and wide windows are respectively $[-15, +30]$, $[-15, +90]$ and $[-15, +180]$ days around the sale, using document dates with recording-date fallback in the panel-construction code. I keep the narrow, base and wide labels, and for brevity I call the absence of a matched mortgage "cash", although it can include unobserved borrowing, missed linkage or instruments outside the rule. Only new mortgages count, so assignments and satisfactions are excluded.

Within each eligible house or condominium parcel, I sort sales by date and keep adjacent observations at least 1,095 days apart, which yields 8,208 house pairs and 6,836 condominium pairs. All sales fall between 2016 and 2025, and in the headline house sample first sales run from January 2016 to December 2022 and resales from March 2019 to December 2025. Adjacency is established within the eligible, already restricted panel, so a transfer excluded from that panel can still lie between two adjacent sales. Repeated unit sales in co-operatives require an additional unit identifier and follow the broader adjacency audit described below.

The permit restriction removes a house pair if a recorded permit issuance lies between its two sale dates, including endpoints, and leaves 6,316 pairs. Permit issuance does not establish that work was completed or how much the property's condition changed, so I treat the restriction as a sensitivity check for changes in condition and make no claim that it holds quality constant.

3.3 Parties, consideration and geography

I assign a deed to each house endpoint only when the closest candidate document date within 45 days is unique, and leave ties unresolved. A consideration sensitivity additionally requires a positive deed amount within the greater of one dollar or one percent of the DOF sale price. This checks that the records agree with each other, which is weaker than showing that consideration equals a negotiated retail value.

Grantor and grantee names supply entity, trust/estate and unknown-party indicators, with explicit company forms taking precedence over trust tokens. The narrower institutional rule identifies named financial institutions, and a broader alternative also uses generic lending

and servicing terms. I exclude a pair when either party is flagged at either date. These rules pick out institution-associated transactions, which overlap with, but are broader than, legally adjudicated foreclosure, deed-in-lieu or REO status. ACRIS records referee's deeds and other foreclosure instruments under the ordinary `DEED` type, so the rules work from party names. The same party records supply the same-surname flag of Section 5.

PLUTO version 26v2 supplies a fixed geographic crosswalk for 8,167 house pairs. After the preceding cumulative restrictions, geography reduces the sample from 6,033 to 6,006 pairs. The full specification spans 137 ZIP codes and 52 community districts. Because the identifiers are current, they control for location but cannot capture past building condition. Of 1,227 ZIP-year endpoint cells, 353 have fewer than five endpoints, so fine geographic adjustment rests on limited support in some cells.

4 Measuring financing-associated price gaps

4.1 The joint estimator

For pair i with sale dates a and z, define $y_i = \log P_{iz} - \log P_{ia}$. Let Q_i contain indicators for cash-to-financed (cf) and financed-to-cash (fc) transitions. The nuisance design X_i contains changes between sale dates in borough-quarter indicators, with optional local-year and party-type terms. The joint regression is

$$y_i = X_i\gamma + Q_i\beta + \varepsilon_i, \qquad \widehat{\pi}_J = \tfrac{1}{2}(\widehat{\beta}_{cf} - \widehat{\beta}_{fc}). \tag{1}$$

Same-financing pairs help estimate the market controls and provide the reference group for the switching coefficients. All fixed effects enter in first differences, and I attach no interpretation to individual coefficients in what is a redundant nuisance design. Sparse least squares projects onto that design, with numerical convergence checks.

A simpler statistic residualizes only the outcome and compares switching-group means. With M_X the residual maker and

$$c_i = \tfrac{1}{2}\left(\frac{1\{cf\}_i}{N_{cf}} - \frac{1\{fc\}_i}{N_{fc}}\right), \tag{2}$$

it is $\widehat{\pi}_R = c'M_X y$. The joint coefficients instead satisfy $\widehat{\beta} = (Q'M_X Q)^{-1}Q'M_X y$. Since residualizing y leaves the financing indicators untouched, the two estimands can respond differently when market controls change. Table 1 reports both, but only equation (1) gives the financing coefficient in a jointly controlled regression, and I use it for all comparisons.

Neither estimator is causal without further assumptions, because within-property differencing removes fixed attributes but leaves time-varying condition, credit access, bargaining circumstances and selection into financing. Permits, recorded party types and local appreciation address the observed parts of those differences, which falls well short of making financing exogenous.

4.2 What symmetry diagnoses

Suppose the two residual switching means can be written as

$$\mu_{cf} = \pi + u_{cf}, \qquad \mu_{fc} = -\pi + u_{fc}. \tag{3}$$

Then their half-difference is $\pi + (u_{cf} - u_{fc})/2$, while their sum $A = u_{cf} + u_{fc}$. A small A restricts the sum of the omitted components and leaves their difference free, so opposing condition or composition effects can cancel in A while biasing the half-difference.

This point connects the permit and institution analyses: permits address recorded authorisations to change the property, whereas party records address who transacts and under what observable institutional circumstances. An institution acquiring title without a new mortgage can enter a cash group even when the transfer arises from debt resolution. If the two arms move in opposite directions, either source of bias can leave the sum close to zero. For that reason I evaluate each restriction through a refitted model with paired uncertainty and do not let either one serve as validation of the other.

4.3 Cumulative evidence and neighbourhood appreciation

In the specification sequence of Table 1, the institutional exclusion increases the permit-free gap, whereas buyer and seller type controls reduce it. The name-based exclusion therefore leaves the positive house contrast in place. I read the change from party controls as descriptive, since the sequence of steps is one of many and does not amount to a causal decomposition.

The geography-linked baseline holds the sample fixed before adding local-year controls. Community-district-year and ZIP-year effects are alternatives, each supplementing borough-quarter effects. The final joint estimate is 9.25 log points, with bootstrap interval [7.17, 11.35]. The corresponding residual statistic is smaller, but it is not the coefficient of financing in the same jointly controlled regression.

I evaluate each change with its covariance in Table 2, where ZIP-year effects change the joint estimate by 0.39 log points, with interval [-1.02, 1.60]. Local appreciation therefore

Table 1: Cumulative house robustness

Specification	Pairs	Residual	Joint	Joint 95% CI
Original sample	8,208	11.14	11.42	[9.60, 13.19]*
Permit-free	6,316	9.34	9.63	[7.80, 11.56]*
+ Named institutions excluded	6,033	10.48	10.82	[8.96, 12.67]
+ Buyer/seller type controls	6,033	7.23	8.50	[6.58, 10.42]
Same geography-linked sample	6,006	7.56	8.87	[6.81, 10.84]*
+ Community-district/year	6,006	7.47	9.50	[7.66, 11.33]
+ ZIP/year	6,006	6.39	9.25	[7.17, 11.35]*

All estimates and intervals are log points. Every row includes borough-quarter effects. The last two rows are alternative geographic specifications added to the same linked baseline. Institutional exclusions use grantor and grantee name rules at both dates and do not verify distress. Stars denote percentile intervals from 999 common parcel-bootstrap refits; other intervals use joint parcel influence functions (CR0). Intervals are pointwise. The residual and joint columns are different estimands.

explains no detectable part of the joint house gap, although current geography and sparse cells limit how small a contribution I can rule out.

Table 2: Paired changes in the joint house contrast

Comparison	Change	95% CI
Party controls	-2.32	[-2.98, -1.65]
Community-district/year	0.63	[-0.12, 1.38]
ZIP/year	0.39	[-1.02, 1.60]
Original to final	-2.17	[-4.02, -0.36]

Log points. Differences use cross-specification covariance on original parcel clusters. ZIP/year and original-to-final intervals use the common full-refit bootstrap; the others use the joint analytic covariance. Party and geographic comparisons hold the relevant estimation sample fixed. Original-to-final also changes the sample.

For every specification I sum observation influences within the original parcel and stack them before computing the covariance, giving excluded parcels zero influence in that specification, which preserves the covariance induced by overlapping samples. The 999-draw bootstrap resamples original parcel clusters and refits all nuisance controls in four principal specifications using the same draws, and it confirms the main joint-coefficient comparison. The high-dimensional residual statistic has a shifted resampling distribution, which is another reason to avoid reading its attenuation as an explained component. All intervals are

pointwise and leave out unobserved classification error.

4.4 Outlier-robust estimates

Least squares gives unusually large price changes substantial influence, and repeat-sale growth has a long upper tail. In the headline sample, 15 percent of cash-to-financed pairs more than doubled in price, against 4 percent of pairs financed at both sales. Among cash-to-financed pairs, those that more than doubled in price are about 5 times as likely to include a same-surname deed (17 percent against 3 percent). The point estimate is smaller for pairs resold within 4 years (8.45 log points [4.32, 12.11]) than for those held at least 5.5 years (10.62 [7.20, 14.33]), although the paired short-minus-long interval ([-6.53, 2.42]) is wide and admits both a negative and a positive difference, so holding horizon alone does not isolate short-term flips. As Appendix Table 9 shows, removing the 60 parcels with the largest positive influence, one percent of the sample, lowers the estimate to 4.68 log points. Because that deletion selects on the fitted outcome, it has no valid sampling interpretation. Table 3 reports Huber estimates and growth-trimmed fits. The trimming cutoffs use growth percentiles for the sample as a whole.

Table 3: Outlier-robust estimates of the financing contrast

	Houses, full controls (6,006 pairs)			Borough-quarter only	
Estimator	Contrast	95% CI	Change from LS	Houses	Condos
Least squares (reference)	9.25	[7.17, 11.35]		11.42	0.83
Huber M-estimator	6.73	[5.32, 8.76]	-2.52 [-3.31, -1.14]	8.83	-0.51
Trim growth at 1st/99th pct.	7.66	[5.77, 9.53]	-1.59 [-2.76, -0.45]	9.23	-0.46
Trim growth at 2.5th/97.5th pct.	5.15	[3.55, 6.78]	-4.10 [-5.64, -2.53]	6.91	-0.59

Log points. Columns 2–4 use the headline specification of Table 1 (ZIP-year and borough-quarter effects, buyer and seller type controls). Columns 5–6 use the common specification of Table 4. The Huber estimator ($k = 1.345$; MAD scale from least-squares residuals, updated once) refits all nuisance controls by iteratively reweighted least squares. Trimming removes pairs whose annualized log growth lies outside the stated percentiles of the estimation sample; cutoffs ignore financing labels. Intervals are percentile intervals from the same 999 parcel-bootstrap draws as Table 1, re-estimating the Huber scale and trimming cutoffs in every draw; the change column is the paired difference within draws. Tuning constants are conventional defaults and were not varied.

The Huber estimator, which down-weights 34 percent of cash-to-financed pairs, gives 6.73 log points [5.32, 8.76]. Trimming annualized growth at the 1st and 99th percentiles gives 7.66, and at the 2.5th and 97.5th percentiles 5.15. For houses, every robust estimate is below least squares and every interval excludes zero, so the house contrast remains positive under these estimators even though its magnitude depends on how much weight extreme appreciation

receives. Under the common borough-quarter specification, the Huber estimate is 8.83 [7.39, 10.28] for houses and -0.51 [-1.00, 0.02] for condominiums. The condominium contrast is much smaller and moves with the estimator, while the divergence between property types survives.

4.5 Property types and classification sensitivity

When I put houses and condominiums under common borough-quarter controls in Table 4, the condominium estimate is 0.83 log points, with interval [0.06, 1.61]. Even with common controls, the two property types differ in ownership form, in the properties themselves and in matching coverage, so I would not attribute the contrast to standardisation or tax status alone.

Table 4: House and condominium contrasts under common controls

Type	Pairs	N_{cf}	N_{fc}	Joint	95% CI
House	8,208	1669	1035	11.42	[9.72, 13.13]
Condo	6,836	1276	1737	0.83	[0.06, 1.61]

Both rows use borough-quarter effects and adjacent eligible sales at least 1,095 days apart. All available eligible pairs are used; the cumulative house restrictions are not imposed on the condominium sample. Parcel-clustered CR0 intervals, in log points. Differences in coverage, ownership and financing measurement prevent a causal property-type interpretation.

The house findings survive narrower and wider financing windows, stable-label restrictions, broader institutional exclusions and consideration-consistent deed matching (Table 5). Disagreement across windows gives a measure of internal consistency: among house sales, strict and base labels differ for 1,020 observations and wide and base labels for 2,193.

Lender take-title deeds. Distressed sales can substantially reduce realised housing returns (Kermani and Wong, 2026). The institutional rule removes pairs with a named financial institution at either sale, but a lender that takes title through foreclosure or a deed in lieu can also appear under a servicer, trust or REO-vehicle name. In a companion paper (Loschi, 2026) I show that such deeds can record an amount set by transfer-tax rules instead of a negotiated price. Applying its party-name classifier, unchanged, to the deeds linked to the headline pairs flags 93 of the 6,006 pairs. Of their 96 flagged endpoints, 93 are unfinanced, which means that lenders taking title enter the analysis as cash buyers, and 38 of the flagged pairs survive even the broader institution exclusion. Excluding the flagged pairs raises the

Table 5: Classification and linkage sensitivity

Restriction	Pairs	Joint	95% CI
Financing window: strict	6,006	9.56	[7.76, 11.36]
Financing window: base	6,006	9.25	[7.44, 11.07]
Financing window: wide	6,006	8.60	[6.77, 10.42]
Stable labels at both dates	5,864	9.03	[7.18, 10.88]
Broader institution exclusion	5,918	9.31	[7.48, 11.13]
Unique deed / amount match	5,798	9.23	[7.44, 11.02]

All rows use the ZIP-year model, with borough-quarter and buyer/seller controls. Log points; parcel CR0 intervals. Stable labels agree between strict and wide windows at both dates. Amount consistency requires a unique nearest deed within 45 days and positive consideration within the larger of one dollar and one percent of the DOF price. These tests concern internal consistency; Appendix D reports document checks and the remaining uncertainty in financing labels.

house contrast to 9.78 log points [7.73, 11.84], a paired change of 0.52 [0.19, 0.97]. The classifier's broader option gives 9.76, and combined with the same-surname screen of Section 5 the contrast is 5.35 [3.23, 7.35], against 4.87 for that screen alone. This is the direction legal-rule amounts imply, because a take-title amount above market value lowers measured appreciation in cash-to-financed pairs and raises it in financed-to-cash pairs, and both shrink the contrast. In a human reading of 40 deeds drawn at random from a stratified sample of 400, the companion note reports weighted precision of 0.82 (conservative joint 95 percent interval 0.40 to 0.99) and foreclosure recall of 0.71 (0.02 to 0.95); two of the flagged deeds were negotiated purchases by banks. I therefore treat the exclusion as a sensitivity check.

5 Who pays cash?

5.1 Related-party transfers

Figure 1 summarises the estimates across samples and estimators.

A below-market price at an unfinanced purchase raises measured appreciation in a cash-to-financed pair. At an unfinanced resale it lowers appreciation in a financed-to-cash pair. Both effects widen the half-difference. Their effects on the sum A in equation (3) offset only to the extent that they have similar magnitudes. A small sum can therefore conceal a large effect on the half-difference.

The ACRIS party extract names the grantors and grantees of the deed linked to each endpoint. Linked deeds with party records cover 98 percent of first-sale and 99 percent

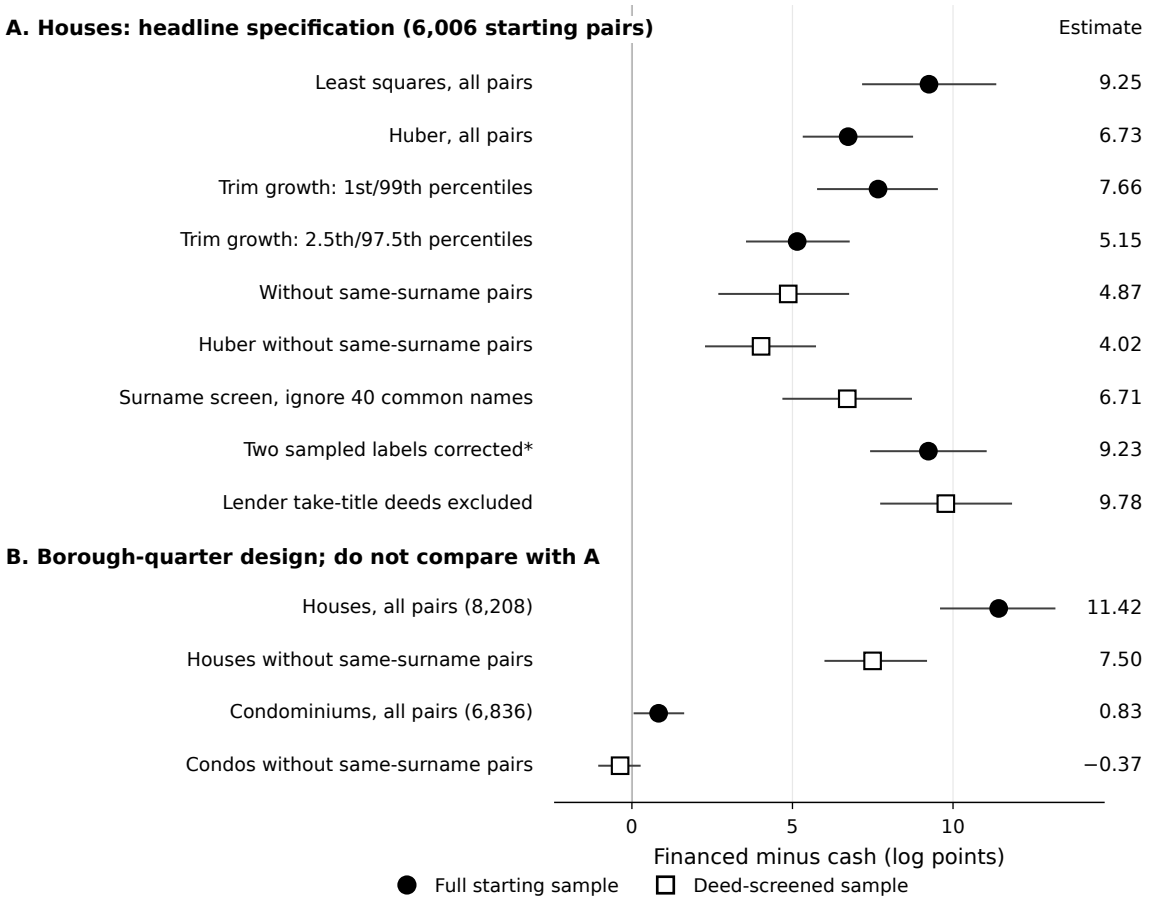

Figure 1: Financed-minus-cash price contrast under alternative samples and estimators

Points are estimates in log points; bars are saved 95% intervals. The two sampled-label corrections use an analytic parcel-clustered interval (asterisk); all other rows use saved parcel-bootstrap percentile intervals. Panel A uses the headline ZIP-year and borough-quarter effects and buyer/seller type controls, starting from 6,006 pairs before screening or trimming. Panel B uses the different common borough-quarter design, starting from 8,208 house and 6,836 condominium pairs; rows must not be compared across panels. Same-surname pairs have a deed at either endpoint on which a natural-person grantor and grantee share a surname. The lender exclusion is a classifier-based sensitivity screen. Rows are alternative estimates, not causal decompositions.

of resale endpoints in the headline sample. I flag a deed when a natural-person grantor and grantee share a surname, using names recorded in the 'surname, given name' form, so company, trust and estate names never match. The flag marks 358 of the 6,006 pairs, and same-surname deeds are 2.5 times as common at unfinanced endpoints (5.7 percent) as at financed ones (2.3 percent). Their prices look like below-market transfers: cash-to-financed pairs with a flagged deed have mean log appreciation of 65.4 log points, against 34.1 log points for other cash-to-financed pairs, whereas flagged financed-to-cash pairs fall by 26.8 log points on average (-26.8 log points) while other financed-to-cash pairs rise by 19.8 log points.

Panel A of Table 6 re-estimates the headline specification without flagged pairs. The house contrast falls to 4.87 log points [2.69, 6.76], a paired change of -4.38 [-5.63, -3.18]. The Huber estimate falls to 4.02 [2.28, 5.73], and every estimator gives between 2.83 and 4.87 log points. The flag makes two kinds of error. Unrelated parties can share a common surname, and ignoring the 40 most frequent surnames in the extract reduces potential coincidences and gives 6.71 [4.69, 8.72], although it also misses genuine relatives with common names. Relatives with different surnames, including many spouses and in-laws, are never flagged, and endpoints without party records cannot be flagged. The contrast that remains probably still contains some related-party transfers.

Removing 358 pairs from 6,006 changes the sample as well as its composition, and the flagged pairs differ from others in location and party records. A matched-deletion exercise asks whether removing any 358 similar pairs would move the contrast as much. Each of 999 draws deletes exactly 358 unflagged pairs, sampled without replacement within cells defined by borough, financing transition, the availability of party records at each endpoint and any entity, trust or unknown party, and refits the full joint model. The design was fixed before fitting and uses no outcome or influence information. The median contrast after deletion is 9.69 log points (a median change of +0.44), the central 95 percent of changes runs from -0.12 to +0.95, and no draw falls by as much as the surname screen's -4.38. Adding first-sale and resale eras (2016–2019 against 2020–2025) to the cells gives a median of 9.68 and a central range of changes from -0.09 to +1.00. Whether kinship itself causes it is a separate question, since flagged pairs may also differ in circumstances the records do not show.

Every sampled deed carries a New York State real property transfer report (form RP-5217NYC), on which the parties declare the price and any condition of the transfer. I drew a random sample of 40 flagged and 20 unflagged cash-endpoint deeds. A research assistant read each one from the ACRIS images without knowing which were flagged, recording the transfer report and the transfer taxes paid, and I checked every reading against the same images (Appendix C). 16 of the 40 flagged deeds, 40 percent [26, 55], declare a sale between

relatives or former relatives (13 deeds), a buyer who is also a seller, or a deed that is not a warranty or bargain-and-sale deed, or report a zero price with no transfer tax, and 10 more combine a shared address or an owner on both sides with no declared condition. Among the unflagged cash deeds the share is 10 percent [3, 30]. Declarations can omit relationships, and the sample is too small to measure the flag's error rates, so I report these as descriptive rates of disclosed signs.

Table 6: Related-party transfers and the identity of the cash buyer

A. Financing contrast by estimator

Sample	Pairs	Least squares	Huber	Trim 1/99	Trim 2.5/97.5
All pairs	6,006	9.25 [7.17, 11.35]	6.73 [5.32, 8.76]	7.66	5.15
No same-surname deeds	5,648	4.87 [2.69, 6.76]	4.02 [2.28, 5.73]	4.03	2.83
Same, strict flag	5,768	6.71 [4.69, 8.72]	4.94 [3.43, 6.79]	5.13	3.06

B. Least-squares contrast by the buyer at the cash sale

Sample	Pairs	Company buyer	Other buyer	Other minus company
All pairs	6,006	1.53 [-1.79, 5.44]	13.77 [10.96, 15.98]	12.24 [7.50, 16.06]
No same-surname deeds	5,648	1.55 [-1.75, 5.49]	7.11 [4.48, 9.12]	5.56 [0.97, 9.07]

Log points; 95% percentile intervals from the paper's 999 common parcel-bootstrap draws, with every nuisance control, Huber scale and trimming cutoff re-estimated in each draw. Headline specification of Table 1. A same-surname transfer is a deed at either endpoint on which a natural-person grantor and grantee share a surname; company, trust and estate names never match. The strict flag ignores the 40 most frequent surnames in the party extract. Panel B splits both switching arms by whether the grantee at the unfinanced sale has an explicit company form; always-financed and always-unfinanced pairs are pooled.

5.2 Company and non-company cash buyers

Panel B of Table 6 splits both switching arms by the buyer at the unfinanced sale. Companies, identified by an explicit corporate form, are the cash buyer in 794 of the 1,781 switching pairs; the remaining 987 non-company cash purchases comprise 867 natural-person deeds, 68 trust or estate deeds and 52 records with missing or unclassified buyer names. When a company pays cash, the contrast is 1.53 log points [-1.79, 5.44]. When another party pays cash it is 13.77 [10.96, 15.98], and 7.11 [4.48, 9.12] after excluding same-surname transfers. Even then the difference between the two groups is 5.56 log points [0.97, 9.07].

Investor activity spans novice speculators and large single-family rental portfolios (Bayer et al., 2021; Mills et al., 2019). In Orange County, the relationship between the purchase price and later foreclosure varied by buyer type (Turnbull and van der Vlist, 2022). Here company form is weak evidence of institutional ownership, since an LLC can represent an

individual buyer. The estimated company-cash contrast is smaller, but its interval admits a positive gap as well as zero. The comparison depends on the buyer-type controls in the nuisance design, which difference out a company indicator at both sales. Dropping the buyer-company control, the company-cash contrast is 7.28 log points (parcel bootstrap 95% interval [4.48, 10.81], 999 draws) against 13.14 [10.30, 15.36] for non-company cash buyers, and without same-surname transfers it is 8.02 [5.48, 11.56] against 6.35 [3.64, 8.34], a non-company-minus-company difference of −1.68 log points [−6.12, 1.57]. With no party controls at all the figures are 9.44 and 13.60 before screening, and 9.92 and 7.02 after screening (difference −2.90 [−7.27, 0.21]). After the surname screen, the specification without the buyer-company indicator does not establish an ordering between company and non-company cash buyers. Even so, the remaining gap among non-company buyers need not be a premium for certainty, because non-company cash purchases also include estate sales, sales to tenants and other circumstances that names do not reveal.

5.3 Property-value splits around $1 million

Testing whether the cash–mortgage contrast disappears at the top of the distribution is a property-value comparison rather than a direct test of a transfer tax defined by transaction price, and the answer depends on how properties are grouped and whether market controls are pooled across bands (Table 7). Using the geometric mean of both observed prices $\sqrt{P_a P_z}$ so neither endpoint is picked by financing status, the pooled-control estimate at $1 million or more is 0.50 log points [−2.98, 3.28] before screening and −1.03 [−4.40, 1.64] after dropping same-surname pairs, against 14.16 and 8.43 below $1 million; first-sale price bands give 2.75 [−1.79, 6.28] and −1.25 [−5.23, 2.16] at the top. Under pooled controls, the all-pair upper-band intervals include both zero and the roughly 3 percent rational-seller benchmark of Reher and Valkanov (2024), while the screened upper-band intervals ([−4.40, 1.64] and [−5.23, 2.16]) include a 1 percent transfer-tax equivalent.

The observed-price bands use prices that also enter measured appreciation $y = \log P_z - \log P_a$. Estimates change when I group properties by prior Department of Finance (DOF) market value or fit the controls separately (Online Appendix Table OA5). Fitting all time, geography and party controls separately inside the geometric-mean upper band raises the estimate to 2.72 [−1.75, 6.09] all pairs and 1.27 [−3.27, 4.20] screened.

To group properties without realized sale prices, I match 5,992 of the 6,006 pairs (14 lack pre-sale valuation records; 12 use non-Class-1 valuation records, and dropping those 12 leaves pooled screened estimates of 4.33 and 5.06 log points) to NYC DOF Final Assessment Roll market values (`FullVal`) whose fiscal-year label equals the calendar year of the first sale.

These rolls were published the previous May (New York City Department of Finance, 2026d). Only 770 of the 871 pairs at or above $1 million in prior `FullVal` also meet the geometric-mean $1 million cutoff. Grouping by prior DOF market value gives positive screened point estimates on both sides of $1 million (4.33 [−1.10, 9.31] at or above $1 million against 5.26 [3.03, 7.29] below with pooled controls, and 5.00 [−4.43, 17.68] with separate controls). The upper-band estimates are imprecise, and the data do not distinguish the two bands (pooled screened high-minus-low difference −0.94 [−6.55, 4.45]). As a sensitivity aligned with the 0.86 median ratio of `FullVal` to first-sale price, splitting at $850,000 gives positive screened intervals under both pooled (5.03 [0.83, 8.71]) and separate controls (5.82 [0.31, 11.64]). Neither property-value split isolates an effect of the proposed cash tax.

Table 7: Financing contrast by property-value band (pooled controls, 999 draws)

Band definition	Switching pairs all / screened	All pairs [95% CI]	Surname screen [95% CI]	Paired change [95% CI]
Geometric mean of sale prices ($\geq$ \$1M)				
Below $1 million	1,175 / 1,059	14.16 [11.53, 16.58]	8.43 [5.90, 10.88]	−5.74 [−7.28, −4.13]
At least $1 million	606 / 575	0.50 [−2.98, 3.28]	−1.03 [−4.40, 1.64]	−1.53 [−3.29, 0.19]
First-sale price ($\geq$ \$1M)				
Below $1 million	1,323 / 1,207	9.76 [7.64, 12.22]	5.69 [3.40, 7.92]	−4.07 [−5.43, −2.76]
At least $1 million	458 / 427	2.75 [−1.79, 6.28]	−1.25 [−5.23, 2.16]	−4.00 [−6.06, −1.75]
Prior DOF market value ($\geq$ \$1M, $N = 5{,}992$)				
Below $1 million	1,431 / 1,321	9.13 [6.90, 11.30]	5.26 [3.03, 7.29]	−3.86 [−5.07, −2.68]
At least $1 million	343 / 307	10.22 [4.58, 15.30]	4.33 [−1.10, 9.31]	−5.89 [−9.46, −2.62]
Prior DOF market value sensitivity ($\geq$ \$850k, $N = 5{,}992$)				
Below $850,000	1,254 / 1,161	8.73 [6.44, 10.89]	4.98 [2.77, 7.12]	−3.74 [−4.88, −2.57]
At least $850,000	520 / 467	10.63 [6.13, 14.71]	5.03 [0.83, 8.71]	−5.60 [−8.47, −2.88]

Log points across 999 common parcel-bootstrap draws (8,127 clusters, seed 20260909). Every row passes the pre-specified precision rule ($\geq$ 100 switching pairs and interval width $\leq$ 20 log points; Online Appendix Table OA5 reports separate-band fits without pooled controls). Switching arms are interacted with band indicators while retaining same-financing pairs and pooled ZIP-year, borough-quarter, and party controls. The first two blocks split all 6,006 pairs on observed sale prices ($\sqrt{P_a P_z}$ and P_a). The last two split 5,992 matched pairs on prior NYC Department of Finance Final Assessment Roll market value (`FullVal`, published in May of the calendar year before the first sale; split at $1 million and in a $850,000 sensitivity aligned with the 0.86 median ratio of `FullVal` to P_a).

5.4 Condominiums and a shared-address screen

Condominium endpoint deeds are matched on unit BBL, date and amount in the ACRIS extracts. A unique deed is linked for 13,639 of 13,672 condominium endpoints, and 33 ambiguous matches are left unlinked. The same surname rule flags 195 of 6,836 condominium pairs. Panel A of Table 8 applies the screen to both property types under the common borough-quarter design of Table 4. The house contrast falls from 11.42 to 7.50 log points

[6.00, 9.19], while the condominium contrast moves from 0.83 (bootstrap interval [0.05, 1.63]) to -0.37 [-1.05, 0.27], a paired change of -1.20 [-1.65, -0.78]. Ignoring the 40 most common condominium surnames gives -0.18 [-0.87, 0.49]. If I also keep only pairs held at least five years and trim the top and bottom 2.5 percent of annualized log price growth within each property type, the contrasts are 4.78 log points for houses and -1.03 for condominiums. The difference is 5.81 log points (95 percent parcel-bootstrap interval [4.00, 7.77], Online Appendix OA1.2).

A second proxy uses the mailing addresses recorded for each party. Panel B of Table 8 additionally excludes 1,595 house pairs with a deed on which the grantor and grantee share a normalized street address, city and ZIP code. The contrast falls to 3.76 log points [1.16, 6.05], but the additional change of -1.11 [-2.76, 0.29] cannot be distinguished from zero, and restricting the match to natural persons gives 4.41 [2.14, 6.75]. Shared addresses are about as common at unfinanced endpoints (16.7 percent) as at financed ones (17.2 percent) and include common mailing agents and co-located parties. I therefore treat this screen as a broad sensitivity check that says little about kinship.

Table 8: The same-surname screen across property types, and a shared-address sensitivity

A. Common borough-quarter design (Table 4)				
	Pairs	All pairs	Flagged Unflagged pairs	Change
Houses	8,208	11.42 [9.60, 13.19]	449 7.50 [6.00, 9.19]	-3.93 [-4.81, -2.94]
Condominiums	6,836	0.83 [0.05, 1.63]	195 -0.37 [-1.05, 0.27]	-1.20 [-1.65, -0.78]
B. Headline house specification: excluding pairs with				
	Pairs	Contrast	Change vs. surname screen	
Same-surname deeds	5,648	4.87 [2.69, 6.76]		
or shared address	4,053	3.76 [1.16, 6.05]	-1.11 [-2.76, 0.29]	
or shared address, persons	4,458	4.41 [2.14, 6.75]	-0.46 [-1.43, 0.71]	

Log points; 95% percentile intervals from the paper's 999 common parcel-bootstrap draws, refitting all controls in each draw. Panel A: least squares under borough-quarter effects only, as in Table 4; the same-surname flag of Table 6 is applied to deeds linked to each endpoint. Condominium deeds are linked for 13,639 of 13,672 endpoints (33 ambiguous matches left unlinked). Panel B: headline specification of Table 1. A shared address is a grantor and grantee on the same deed with the same normalized mailing street, city and ZIP code; it can reflect a common mailing agent or co-location and is not evidence of kinship. The natural-person variant counts only addresses of parties with personal names.

6 Interpreting the financing gap

6.1 What the magnitudes imply

Earlier purchase prices can influence sellers' asking prices and willingness to sell (Genesove and Mayer, 2001; Andersen et al., 2022). If I attribute the entire observed contrast to execution risk and assume a 10 percent chance that an accepted financed purchase fails, a seller-indifference calculation requires a loss on failure of about 87 percent of the house price, 45 percent after the surname screen, and about 8 percent for condominiums. These are requirements of that assumed explanation, not estimates of actual losses or failure rates; the derivation and scenarios are in Online Appendix OA2.

6.2 Comparing estimates across samples

The 4.87 log-point screened house contrast is below the roughly 11 percent premium Reher and Valkanov (2024) estimate in their national sample. The two studies differ in periods, samples and price models. I cannot tell how much of their difference reflects the transfer screens. The Los Angeles evidence of Han and Hong (2024) also uses time to close and relisting, which the New York sales records do not observe.

6.3 Building class and mortgage leverage

Figure 2 summarises the building-class, condominium and recorded-LTV mechanism checks (Online Appendix OA7, Tables OA6–OA8). First, NYC Tax Class 1 houses combine single-family homes with income-producing two- and three-family properties, whereas condominiums are single units. Exact matching on parcel, date and price links all 6,006 headline pairs to NYC DOF building-class categories (2,951 one-family and 3,055 two-to-three-family pairs at first sale; 69 cross categories). In the screened headline specification, the contrast is 5.70 log points [3.19, 8.07] for one-family houses and 4.11 [1.12, 6.89] for two-to-three-family houses (difference −1.59 [−5.20, 2.01]), with closely aligned separate-control (4.95 vs. 4.82) and constant-class (5.39 vs. 4.13) estimates. In the long-hold (≥ 5 years) trimmed borough-quarter comparison, one-family houses alone yield 4.92 [2.84, 7.64], exceeding condominiums by 5.95 [3.76, 8.66], against 4.67 [2.17, 7.23] for two-to-three-family houses (difference −0.26 [−3.76, 3.03]). Multi-unit investment houses therefore do not drive the house-versus-condominium divergence.

Second, Ben-David (2011) finds in Cook County that some highly leveraged buyers inflated recorded prices to draw larger mortgages, which would raise recorded financed prices relative to cash ones. Across all 9,349 financed endpoints in the headline sample, recorded

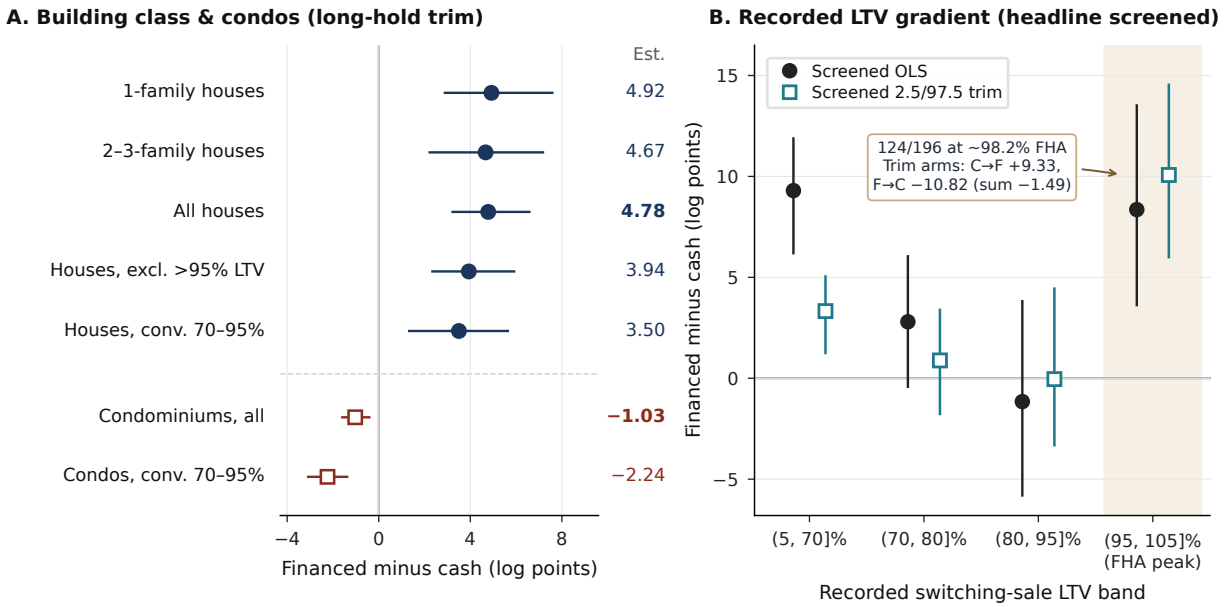

Figure 2: Building-class, condominium, and recorded LTV mechanism comparisons

Points are estimates in log points; bars are 95% parcel-bootstrap percentile intervals (999 draws). Panel A uses the common borough-quarter design restricted to surname-screened pairs held at least five years with within-type 2.5th/97.5th annualized-growth trimming (14,874 union parcels; Online Appendices OA1.2 and OA7). Conventional 70–95% rows restrict switching pairs to $0.70 < \text{LTV} \leq 0.95$ while keeping always-financed and always-cash pairs for market controls. Panel B plots the four-bin switching LTV gradient in the headline surname-screened house sample ($N = 5,648$ OLS; $N = 5,366$ trimmed) with ZIP-year, borough-quarter, and party controls, absorbing out-of-range switching pairs separately. Inside $(95\%, 105\%]$, 124 of 196 screened switching pairs sit in the 97.5–98.5% FHA cluster, with symmetric trimmed arm effects ($C \to F = +9.33$, $F \to C = -10.82$; sum -1.49).

mortgage amounts match 100 percent of records, with 9,172 (98.1 percent) in the standard $0.05 < \text{LTV} \leq 1.05$ band (median 77.4 percent; 3 missing or non-positive, $26 \leq 5$ percent and $148 > 105$ percent). Absorbing the 48 out-of-range screened switching pairs separately, screened OLS gives 6.44 [3.87, 8.51] at LTV $\leq$ 80 percent versus 3.42 [0.27, 7.26] above 80 percent (difference −3.02 [−6.50, 1.73]; linear slope −1.20 [−2.15, −0.10] per 10 percentage points). Part of that negative OLS slope is mechanical, because dividing by sale price puts price noise in the denominator (correlation −0.44 at first sale and −0.44 at resale): within 3,505 always-financed pairs, the slope on ΔLTV is −0.74 [−1.19, −0.43] per 10 percentage points (−0.51 [−0.81, −0.23] trimmed). Under 2.5th/97.5th growth trimming, the binary ordering reverses (4.85 [2.31, 8.53] above 80 percent vs. 2.19 [0.56, 3.75] at or below 80 percent; difference 2.66 [−0.08, 6.66]), and finer bins show a U-shaped profile (Figure 2, Panel B): near-zero trimmed contrasts in conventional 70–80 (2.78 OLS; 0.87 trimmed) and 80–95 percent (−1.16; −0.04) bands alongside a sharp spike above 95 percent LTV (8.36 [3.57, 13.58] OLS; 10.07 [5.94, 14.60] trimmed; 9.19 [4.55, 14.35] above 70–80 percent), where 124 of 196 screened switching pairs cluster around 98.2 percent LTV (FHA 96.5 percent financing plus financed upfront mortgage insurance).

Separating the two switching directions inside each LTV band tests whether the > 95 percent spike reflects one-sided renovation flips into FHA mortgages or a symmetric price-level markup at the FHA endpoint (Online Appendix Table OA7). In the trimmed headline fit, the two arms above 95 percent LTV are nearly symmetric and opposite in sign: cash-to-financed pairs appreciate by 9.33 [2.97, 15.13] while financed-to-cash pairs fall by −10.82 [−17.78, −5.11] (sum −1.49 [−11.05, 6.78]; OLS 9.52 and −7.19, sum 2.33). In conventional 70–80 and 80–95 percent bands, both arms move in the same direction (5.12 and 3.35; 3.22 and 3.30), leaving contrasts near zero. Corporate-entity sellers are more common in > 95 percent pairs on both sides of the FHA segment (63.7 vs. 53.3 percent in $C \to F$, median hold 4.40 vs. 4.84 years; 40.4 vs. 18.6 percent at the initial financed sale in $F \to C$), yet splitting > 95 percent pairs by seller type shows positive symmetric trimmed contrasts for both entity sellers (12.79 [6.29, 18.43]) and non-entity sellers (6.94 [1.88, 14.04]). Symmetry across $C \to F$ and $F \to C$ rules out purely one-sided unrecorded renovation between cash purchase and FHA resale and points instead to an endpoint price markup consistent with FHA financed upfront mortgage insurance, seller-paid closing-cost concessions and leveraged price inflation (Ben-David, 2011).

Linking all 13,672 of 13,672 corrected condominium endpoints (100.0 percent; 99.99 percent financing-flag agreement) to the linked dataset shows that > 95 percent LTV mortgages are almost absent in condominiums: 1.0 percent of valid financed endpoints (0.15 percent in the 97.5–98.5 percent FHA peak) against 16.0 percent (10.3 percent) in the 8,208-pair house

sample, and 0.9 percent (1 pair) versus 12.4 percent (177 pairs) among screened switching pairs (Online Appendix Table OA8). Excluding > 95 percent switches in the paired long-hold (≥ 5 years) trimmed borough-quarter comparison lowers the house contrast from 4.78 to 3.94 [2.29, 5.97] and narrows the house-minus-condominium gap from 5.81 to 4.93 [3.28, 7.12], a paired reduction of -0.88 [-1.51, -0.01]. Restricting switches to conventional 70–95 percent LTV yields 3.50 [1.28, 5.69] for houses versus -2.24 [-3.13, -1.33] for condominiums, a divergence of 5.74 [3.18, 8.20] (5.37 [2.99, 8.18] when absorbing non-conventional switches; Figure 2, Panel A).[2] While the FHA > 95 percent tail accounts for roughly one-sixth of the long-hold trimmed house-versus-condominium gap, a substantial conventional-only divergence remains.

6.4 Co-operative financing and credit conditions

I treat the co-operative comparison as a measurement sensitivity check because timing-unique assignments cannot establish purchase financing (Online Appendix OA3).

The financing-specific associations with mortgage rates and inventory do not survive financing-specific time trends, so I cannot isolate an execution-risk mechanism from these comparisons (Online Appendices OA4 and OA5).

7 Conclusion

The financed-minus-cash contrast in New York City's house records falls from 9.25 to 4.87 log points when I exclude pairs with a same-surname buyer and seller at either sale. The change is -4.38 log points [-5.63, -3.18]. Matched deletions of unflagged pairs usually move the estimate little. The forms associated with flagged deeds more often disclose conditions that can make a price unrepresentative of an open-market sale. How much of the change reflects kinship remains unknown.

A positive house contrast remains after the surname screen and the growth and holding-period checks, while the corresponding condominium contrast is much smaller. Building-class and LTV checks show that one-family houses exhibit at least as large a contrast as multi-unit houses, that the trimmed headline house gap concentrates sharply and symmetrically across switching directions in the > 95 percent FHA-clustered LTV tail, and that excluding > 95 percent mortgages trims only part of the long-hold house-versus-condominium divergence, leaving a large conventional-only gap. Grouping by prior municipal market value gives posi-

[2] In the headline house specification (adding ZIP-year and party controls unavailable for condominiums), excluding > 95 percent switches yields 4.50 [2.25, 6.49] in OLS and 1.88 [0.38, 3.40] trimmed.

tive house point estimates above and below $1 million, with too little precision to distinguish the groups. These results limit the use of the raw cash discount to calibrate a tax intended to offset an advantage at closing.

The records leave two questions open. A missed mortgage can put a financed transaction in the cash group. In the probability sample, the deed and mortgage records supported purchase financing for 2 of 20 initially unmatched endpoints and contradicted none of the matched labels. The sensitivity tests show that the contrast can reverse when errors concentrate at endpoints whose reassignment lowers it (Appendix A). Even with better financing labels, estimating the value of closing certainty would require information on offers, contingencies and failed closings, or variation in financing unrelated to the circumstances of the sale.

Declarations

Data availability. I use public records from New York City agencies and Freddie Mac. The paper deposit (doi:10.5281/zenodo.22421850) includes code, the rebuilt condo panel, and de-identified reading sheets. The README links to the pinned v3.8 replication inputs (doi:10.5281/zenodo.23372552) needed to verify the estimates. I also use linked dataset v1.2.0 (doi:10.5281/zenodo.23557585). The online appendix maps the public and restricted files. The deed classifier comes from Loschi (2026).

Funding and interests. Sole author; independent researcher. There was no external funding. The author founded HolaCasaClara, a free comparator of Argentine mortgage loans, which has no connection to New York housing records or this study. No other competing interests or outside rights to review the research before dissemination are reported.

Prior dissemination. Earlier versions circulated as a working paper (SSRN 7421138) and with the replication deposit.

Declaration of generative AI and AI-assisted technologies in the writing process. I designed the study, specified the empirical and validation designs, matched the pre-sale assessment rolls and rebuilt the condominium linkage, ran and checked all estimation and verification outputs, and inspected every human research-assistant reading against the original register images. No AI document reading is used as evidence anywhere in the paper. During preparation, I used generative AI tools (OpenAI ChatGPT/Codex, Google Gemini, and Anthropic Claude) to assist with coding, literature checks, and language editing, and I take full responsibility for all specifications, interpretations, citations, and results.

Ethics. The research uses administrative public records; no human subjects were involved.

References

Andersen, S., Badarinza, C., Liu, L., Marx, J., and Ramadorai, T. (2022). Reference dependence in the housing market. *American Economic Review*, 112(10):3398–3440.

Asabere, P. K., Huffman, F. E., and Mehdian, S. (1992). The price effects of cash versus mortgage transactions. *Real Estate Economics*, 20(1):141–154.

Bailey, M. J., Cole, C., Henderson, M., and Massey, C. (2020). How well do automated linking methods perform? lessons from US historical data. *Journal of Economic Literature*, 58(4):997–1044.

Bailey, M. J., Muth, R. F., and Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304):933–942.

Bayer, P., Mangum, K., and Roberts, J. W. (2021). Speculative fever: Investor contagion in the housing bubble. *American Economic Review*, 111(2):609–651.

Ben-David, I. (2011). Financial constraints and inflated home prices during the real estate boom. *American Economic Journal: Applied Economics*, 3(3):55–87.

Bloomberg News (2026a). NY plans new tax on NYC homes over $1 million bought with cash. Bloomberg, 14 May 2026.

Bloomberg News (2026b). NYC cash home purchases tax is likely to be dropped from New York state budget. Bloomberg, 21 May 2026.

Campbell, J. Y., Giglio, S., and Pathak, P. (2011). Forced sales and house prices. *American Economic Review*, 101(5):2108–2131.

Freddie Mac (2026). Primary mortgage market survey: Historical archives. Weekly 30-year mortgage rates, 2015–2025; 574 observations parsed from preserved annual archive extracts.

Fuster, A. and Zafar, B. (2021). The sensitivity of housing demand to financing conditions: Evidence from a survey. *American Economic Journal: Economic Policy*, 13(1):231–265.

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

Han, L. and Hong, S.-H. (2024). Cash is king? understanding financing risk in housing markets. *Review of Finance*, 28(6):2083–2118.

Harding, J. P., Rosenthal, S. S., and Sirmans, C. F. (2003). Estimating bargaining power in the market for existing homes. *Review of Economics and Statistics*, 85(1):178–188.

Ke, Q., Zhu, B., White, M., and Chi, B. (2025). Transactions tax change during the pandemic: A study of the UK housing market. *Journal of Housing Economics*, 69:102088.

Kermani, A. and Wong, F. (2026). Racial disparities in housing returns. *American Economic Review*, 116(1):287–331.

Loschi, P. (2026). When a deed is not a market sale: Foreclosure transfers, statutory consideration, and repeat-sales house-price measurement. Working paper and replication package, Zenodo. Concept DOI resolving to the latest version.

Mills, J., Molloy, R., and Zarutskie, R. (2019). Large-scale buy-to-rent investors in the single-family housing market: The emergence of a new asset class. *Real Estate Economics*, 47(2):399–430.

New York City Department of Buildings (2026). DOB permit issuance. NYC Open Data: ipu4-2q9a; preserved original permit snapshot.

New York City Department of City Planning (2026). PLUTO, version 26v2. NYC Open Data: 64uk-42ks; targeted parcel geography extract, September 2026.

New York City Department of Finance (2026a). ACRIS personal property master and legals. NYC Open Data: sv7x-dduq and uqqa-hym2; September 2026 extract.

New York City Department of Finance (2026b). ACRIS real property master, legals, and parties. NYC Open Data: bnx9-e6tj, 8h5j-fqxa, 636b-3b5g; archived panel and September 2026 targeted extracts.

New York City Department of Finance (2026c). Annualized calendar sales. NYC Open Data, w2pb-icbu; archived 2016–2025 price sample and September 2026 co-operative extract.

New York City Department of Finance (2026d). Final assessment rolls (tax class 1, fy2016–fy2022) and property valuation and assessment data. Historical Final Assessment Roll Tax Class 1 files (FY16–FY22) and NYC Open Data yjxr-fw8i; pre-sale market values matched prior to first-sale date.

New York City Department of Finance (2026e). Mortgage recording tax.

New York State (2026). New York Tax Law §§ 253, 253-a, 261, 1402, 1402-a, 1402-b, 1404; New York City Administrative Code §§ 11-2101 to 11-2117. Consolidated Laws of New York, Chapter 60, Articles 11 and 31; Real Property Transfer Tax. §1402-b added by L.2019, c.59, pt. D, effective 1 July 2019, with conveyances under binding contracts entered into on or before 1 April 2019 excluded.

Nowak, A. D. and Smith, P. S. (2020). Quality-adjusted house price indexes. *American Economic Review: Insights*, 2(3):339–356.

Office of the Governor, State of New York (2026). Governor Hochul signs budget that makes New York more affordable, keeps New Yorkers safe and expands opportunity for all. Press release, 28 May 2026 (FY 2027 Enacted Budget).

Reher, M. and Valkanov, R. (2024). The mortgage-cash premium puzzle. *Journal of Finance*, 79(5):3149–3201.

Tidwell, A., Jauregui, A., Sah, V., and Narwold, A. (2018). Cash and distressed house sales price discounts: Dual sample selection spatial interdependence approaches. *Journal of Real Estate Finance and Economics*, 56(1):101–139.

Turnbull, G. K. and van der Vlist, A. J. (2022). The price of ignorance: Foreclosures, uninformed buyers and house prices. *Journal of Housing Economics*, 57:101844.

A Measurement limits

I measure every transaction attribute from records. The document checks cover only a small sample of deeds and financing labels, and I have not checked the rest against the source instruments. A financed label means a purchase-window mortgage was matched and a cash label means none was, and neither confirms how the purchase was paid for. A shared surname or mailing address marks a possible relationship, missing relatives with different names and sometimes matching unrelated people. Transfer reports in a small random sample of deeds show non-market signs in 40 percent of flagged deeds and 10 percent of other cash deeds, but the declarations are self-reported. Permit issuance shows authorised work that may not have been completed, and institutional-name rules mix foreclosure transfers with lender resales. Financing arrangements can also make a recorded price exceed the net price: in Cook County deed records, Ben-David (2011) finds that some highly leveraged buyers inflated recorded prices to draw larger mortgages, which would raise financed prices relative to cash ones. Section 6.3 and Online Appendix OA7 examine recorded loan-to-value ratios directly; the public records used here still do not show seller credits or concessions, so I cannot separate price inflation from borrower selection or underwriting risk. Excluding lender take-title deeds flagged by the companion classifier raises the house contrast by 0.52 log points. Co-operative financing assignments rely on timing and can attach a filing to the wrong unit. I document the linkage rules and recompute the downstream analysis.

A research assistant read the documents behind a random sample of 40 sale endpoints, and I checked the readings (Appendix D.3). The design-weighted discrepancy estimate is 2.19 percent, with conservative 95 percent interval [0.19, 22.96] percent, conditional on reading accuracy. At the sample-calibrated random-recoding rates and upper-bound stress assumptions, the average contrast ranges from 3.13 to 8.63 log points. Errors related to price growth can be more consequential: at the sample point rates (10.0 percent hidden mortgages and 0.0 percent false financing), the zero crossings are 1.16 standard deviations for all pairs and 0.75 after the surname screen. With false financing at its 16.1 percent Wilson upper bound and hidden mortgages still at 10.0 percent, they are 0.19 and 0.12, respectively. All reported intervals condition on these measurements.

B Joint uncertainty and numerical implementation

Write $r = M_X y$ and $m = M_X c$. The derivative of the residual contrast with respect to an observation weight, evaluated at unit weights, is

$$\mathrm{IF}_{R,i} = m_i r_i - \frac{1\{cf\}_i \bar{r}_{cf}}{2N_{cf}} + \frac{1\{fc\}_i \bar{r}_{fc}}{2N_{fc}}. \tag{4}$$

The first term accounts for refitting the nuisance projection and the remaining terms for changes in group denominators. For the joint estimator, let $R = M_X Q$ and $e = M_X y - R\widehat{\beta}$. Its coefficient derivative is

$$\mathrm{IF}_{\beta,i} = (R'R)^{-1} R_i' e_i. \tag{5}$$

I apply the half-difference contrast, sum within parcel and stack across specifications. The covariance is the cross-product of these stacked cluster influences (CR0), and paired differences use the corresponding covariance terms.

A numerical check perturbs five parcel weights in the full ZIP-year model, refits the nuisance parameters and matches both analytic derivatives to finite differences within 10^{-7}. The common bootstrap draws multinomial weights over the 8,127 original house parcels and refits the original, permit-free, linked-party-control and ZIP-year specifications in each of 999 draws. Redundant nuisance columns are projected out. Intervals are pointwise, with no multiple-testing correction, and they condition on the observed linkage and classifications.

C Deed-image review of the same-surname flag

I drew the sample before any deed was read, taking 40 endpoint deeds at random among flagged pairs and 20 among unflagged cash endpoints of the headline sample. The 60 recorded documents run to 684 pages, retrieved from the public ACRIS image service, and each has a cover page, the deed and a New York State real property transfer report (RP-5217NYC). Item 12 of the report gives the full sale price and item 14 lists conditions of the transfer, among them (A) a sale between relatives or former relatives, (C) a buyer who is also a seller and (E) a deed other than a warranty or bargain-and-sale deed. Prices and conditions come from the report, while parties' addresses and the city and state transfer taxes, whose rates imply the taxed consideration, come from the cover pages. Where the report price and the taxed consideration disagree, I record both and class the case as unclear. A research assistant read all 60 documents, without AI tools and in a shuffled order that hid which deeds were flagged, recording the report price, the conditions ticked, whether a buyer is also a seller, whether buyer and seller share a surname or an address, and the taxes on the cover page. I

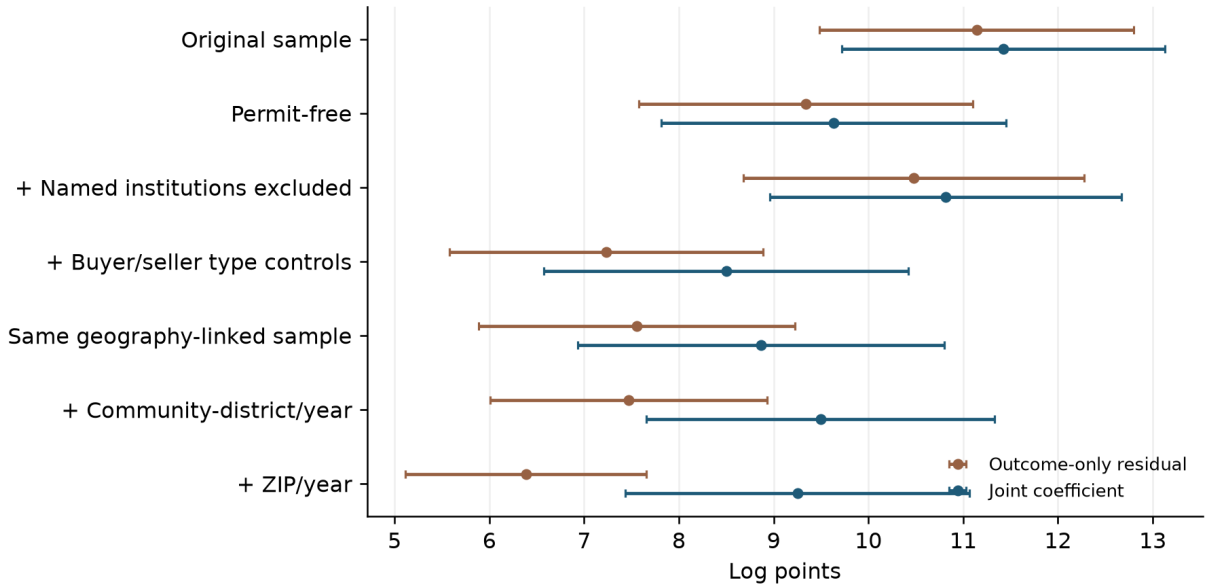

Figure 3: The residual-mean statistic and joint coefficient across controls

Point estimates and pointwise parcel CR0 intervals. The figure uses one analytic interval method throughout; the principal table also reports common full-refit bootstrap intervals. Community-district/year and ZIP/year are alternative final controls.

then checked every reading against the images and agreed with all of them before the flags were revealed. The public source-and-checks bundle includes the scored summaries in `human_scoring/out/*.json`, de-identified reading sheets (`human_scoring/csv_deidentified/*.csv`), and the offline scoring script (`human_scoring/score_deidentified_validation.py`). The underlying sheets with ACRIS document numbers, parcel keys, analyst notes, and raw analyst identifiers are held in a restricted file rather than posted in the public bundle.

A deed shows a non-market sign when it declares condition A, C or E, or reports a zero price with no transfer tax. The rule is conservative because conditions are self-declared. A transfer from a father to a son of the same name, for example, may declare none. Among flagged deeds, 10 more give one address for both sides without declaring a condition, and in three of them a buyer is also among the sellers. For one flagged quitclaim deed the transfer report gives a zero price, although the taxes imply a positive consideration. Among unflagged deeds, three give one address for both sides and two are unclear, one with a zero price and tax paid and one with a price but no tax. One unflagged deed has matching price and tax records and no declared conditions, yet the instrument gives a price for several lots together. The transfer report alone does not reveal that the price cannot be allocated to the house. The review supports the flag as a marker of non-market transfers, but it is too small to measure the flag's sensitivity or specificity.

D Focused validation of the house contrast

A first pass checks the input rows behind the 6,006 pairs, which cover 5,957 parcels and 11,963 distinct sale endpoints. Every pair's price change and financing flags reconstruct from those rows, the three financing windows are nested, and the linear coefficient weights reproduce the joint contrast. This confirms the computation, though it leaves the linkage untested.

Ten pairs have an endpoint with more than one house record on the same parcel and date, and three financed endpoints have no positive mortgage amount. Neither proves a wrong label, and dropping these cases and applying the existing safeguards together leaves 5,579 pairs and a contrast of 9.02 log points. The house panel does not keep the mortgage document identifiers needed to verify loan purpose directly.

Table 9 reports the full refitted sequence.

Influence comes from the fitted joint model, aggregated by parcel, and every deletion refits all nuisance controls. Removing the single most influential parcel leaves 9.07 log points, and removing the 60 with the largest absolute influence leaves 8.51, so no single record drives the estimate. Removing the 60 largest positive influences lowers it to 4.68. Because that deletion selects on the outcome, Section 4.4 handles the same concern with robust estimators at conventional tuning constants. Omitting each borough and each of the three largest ZIP groups is reported in full.

The replication output also reports adversarial perturbations to pair growth with labels and design held fixed, moving the outcomes of the parcels with the largest coefficient weights against the sign of those weights. These are exact linear sensitivity calculations for contaminated outcomes and bound neither misclassification nor unobserved renovation.

D.1 Errors in the financing labels

A recoded label raises or lowers the contrast depending on the pair's growth net of time and location effects. Table 10 starts from the headline sample with the 2 sampled human-supported mortgages corrected, holds the 40 checked endpoints fixed, and recodes shares of the remaining 2,643 cash-labelled and 9,329 financed-labelled endpoint occurrences. The hidden-mortgage scenarios use the human estimate of 10.0 percent and its Wilson upper bound, 30.1 percent. No matched label was contradicted; I therefore use both zero and its 16.1 percent Wilson upper bound for false financing, without interpreting zero detected errors as zero prevalence. Both switching arms are rebuilt after recoding, and the two surname-screen variants use the same draws. The random scenario means range from 3.13 to 8.63 log points; these are sensitivity results, not corrected population estimates.

Table 9: Focused house validation: influence, geography and source consistency

Restriction	Pairs	Contrast	Change
Main reference	6,006	9.25	+0.00
Omit top 1 absolute influences	6,005	9.07	-0.18
Omit top 5 absolute influences	6,001	9.68	+0.42
Omit top 10 absolute influences	5,996	9.79	+0.54
Omit top 30 absolute influences	5,976	9.41	+0.16
Omit top 60 absolute influences	5,946	8.51	-0.74
Omit 60 largest positive influences	5,946	4.68	-4.57
Omit Manhattan	5,962	9.26	+0.01
Omit Bronx	4,962	9.42	+0.17
Omit Brooklyn	4,052	10.39	+1.14
Omit Queens	3,042	7.91	-1.34
Omit ZIP 11234	5,801	9.11	-0.14
Omit ZIP 10465	5,844	9.01	-0.25
Omit ZIP 11357	5,875	9.25	-0.00
Joint record-consistency safeguards	5,585	8.98	-0.27
Omit ambiguous source dates	5,996	9.29	+0.04
Omit missing positive loan amounts	6,003	9.25	-0.00
Joint safeguards and source checks	5,579	9.02	-0.23

Log points. Every row refits the joint switch model and all nuisance controls on its retained sample. The influence exercises remove whole parcels. Absolute- and positive-influence exclusions are selected using the fitted outcome; they are diagnostics and should not be used as preferred samples, and ordinary confidence intervals would not account for that selection. Geographic and source restrictions also change the estimation population. The three ZIPs are selected by pair count. Combined safeguards require stable financing windows, unique amount-consistent deeds at both endpoints and no broad institutional-name flag. Source checks additionally remove ambiguous parcel/date records and financed endpoints without positive recorded loan amounts. These restrictions check internal consistency and do not validate the classifications. The complete model outputs and nominal parcel intervals are retained in the replication files.

Timing and company-pool scenarios. I also recode every eligible cash endpoint with a mortgage in the wide window but not in the base window (92 endpoint occurrences), giving 8.58 log points, and draw hidden mortgages only from company buyers, giving 9.11 to 9.77. These are hypothetical pools defined from metadata; the human sample does not establish the purpose of loans at unsampled endpoints.

Errors related to growth. Panel C uses the sample point rates (10.0 percent hidden mortgages and 0.0 percent false financing) and draws 264 recoded labels with a probability that rises with their effect against the contrast. I summarise that tilt as the mean standardized effect of the chosen labels, zero when errors are unrelated to growth. The contrast reaches zero at 1.16 standard deviations, or 0.75 without same-surname pairs. Panel D retains the upper-bound stress: 10.0 percent hidden mortgages and 16.1 percent false financing, recoding 1,767 labels. Its crossings are 0.19 and 0.12, respectively. At the point rates, ranking recodings by their baseline residual-growth effect gives -5.93 and 20.11 log points under two adverse and favourable assignments. These illustrative assignments do not establish global bounds over all allowable recodings.

Table 11 applies the surname, broad lender-name and legal-rule screens successively to the original headline labels, without an earlier AI-review correction step. The broad lender-name rule changes the contrast by -0.01 log points, and the legal-rule exclusion changes it by $+0.31$ [0.11, 0.54]. These paired comparisons describe the screens; they do not establish that every excluded deed was misclassified.

D.2 Influential endpoints

I selected seven influential sale endpoints for document checking. A research assistant read the deeds, transfer reports and candidate mortgages without AI, and I checked the readings. H016-a, H010-z and P01 have bundle prices with no allocation to the house; H016-a covers 10 lots. P01 also has a purchase mortgage covering both lots, so I exclude its pair under the bundle rule rather than apply the single-parcel mortgage rule. P02–P05 are purchase mortgages whose index dates on the recording covers differ from the dates inside the documents.

Because I selected these cases for influence, Table 12 also applies both rules to every pair. All 3 checked bundle cases satisfy the multi-parcel rule, and all 4 checked date mismatches satisfy the single-parcel index-date rule. Dropping 80 pairs with a multi-parcel endpoint deed gives 9.82 log points, and recoding the 45 endpoints that meet the date rule gives 9.36. Together they give 9.91 [8.09, 11.73], compared with 9.85 after the targeted changes alone. The index-date rule requires a single-parcel mortgage recorded near the acquisition, with its indexed document date outside the purchase window and a recorded borrower matching the

Table 10: Financing-label sensitivity calibrated to the human probability sample

Scenario	Labels recoded	Arms (c-to-f / f-to-c)	All pairs	Without same-surname pairs	Removed by screen
Human-corrected baseline	0	1,044 / 737	9.23	4.85	48%
A. Random recoding, both directions					
Hidden 10.0%, false financed 0.0%	264	979 / 705	8.63 [7.86, 9.46]	4.52 [3.71, 5.34]	48%
Hidden 10.0%, false financed 16.1%	1,767	1,424 / 1,199	4.05 [3.07, 5.09]	2.03 [1.18, 3.06]	50%
Hidden 30.1%, false financed 0.0%	796	829 / 620	7.35 [5.95, 8.92]	3.90 [2.30, 5.74]	47%
Hidden 30.1%, false financed 16.1%	2,299	1,334 / 1,163	3.13 [2.01, 4.27]	1.57 [0.46, 2.83]	51%
B. Timing and company-pool sensitivity					
Wide-window mortgages	92	1,043 / 718	8.58	4.21	51%
Company pool, hidden 10.0%	264	988 / 717	9.11 [8.32, 9.88]	4.79 [3.94, 5.59]	47%
Company pool, hidden 30.1%	796	840 / 647	9.77 [8.77, 10.97]	5.17 [4.00, 6.47]	47%
C. Growth-related recoding: sample point rates					
Mean standardized effect 0.30	264	–	6.43	2.67	–
Mean standardized effect 0.72	264	–	3.32	0.17	–
Mean standardized effect 1.02	264	–	1.06	−1.50	–
Mean standardized effect 1.37	264	–	−1.58	−3.35	–
Ranked assignment against	264	997 / 678	−5.93	−6.12	−3%
Ranked assignment in favour	264	1,025 / 746	20.11	15.60	22%
D. Growth-related recoding: upper-bound stress					
Mean standardized effect 0.29	1,767	–	−2.32	−3.03	–
Mean standardized effect 0.48	1,767	–	−6.53	−6.32	–
Mean standardized effect 0.63	1,767	–	−10.19	−9.27	–
Mean standardized effect 0.81	1,767	–	−14.27	−12.74	–

Log points, headline specification. The baseline corrects only the two sampled human-supported purchase mortgages; the 40 human-read endpoints stay fixed. No older AI corrections or protected AI-review pairs are used. Hidden-mortgage rates are 10.0 and 30.1 percent (sample estimate and Wilson upper bound). No false-financed label was detected: zero and its 16.1 percent Wilson upper bound are the random scenarios. The latter is a stress assumption, not a detected error rate. Panel B uses candidate timing and company status as hypothetical pools, not human findings about unsampled loans. Panel C uses the sample point rates: 10.0 percent hidden and 0.0 percent false-financed labels. Panel D retains the upper-bound stress: 10.0 percent hidden and 16.1 percent false financed. Both sweeps use 96 draws per tilt. The point-rate tilt grid extends to $\kappa = 1.5$ to bracket both crossings; the upper-bound grid remains at 0–1. Random and company rows use 200 draws; brackets are 2.5th–97.5th percentiles of estimates across recodings, not confidence intervals. Random, company and growth-related scenarios recode at most one endpoint per pair; the wide-window row recodes all eligible endpoints.

Table 11: Surname, lender-name and legal-rule screens applied together

Sample	Pairs	Contrast [95% CI]	Change from headline	Change from row above
Headline sample	6,006	9.25 [7.17, 11.35]	–	–
without same-surname pairs	5,648	4.87 [2.69, 6.76]	−4.38 [−5.63, −3.18]	−4.38 [−5.63, −3.18]
and broad lender names	5,561	4.86 [2.73, 6.82]	−4.40 [−5.64, −3.12]	−0.01 [−0.40, 0.45]
and legal-rule deeds	5,523	5.16 [2.98, 7.16]	−4.09 [−5.38, −2.77]	+0.31 [0.11, 0.54]

Original headline labels and model. I apply the stated screens successively; there is no older AI-review correction step. Intervals are pointwise parcel-bootstrap percentiles from 999 common refits; changes are paired with the headline and preceding row. The classifier is a sensitivity screen, not a verified correction. The separate human sample corrections appear in Table 13.

buyer; unsampled cases meeting it have not all been checked against complete instruments.

Table 12: Human-confirmed influential cases and the same rules applied to every pair

Handling	Pairs	Contrast	Analytic 95% interval
Headline sample	6,006	9.25	[7.44, 11.07]
Drop three bundle-price pairs	6,003	9.66	[7.88, 11.44]
Recode four index-date mismatches	6,006	9.44	[7.62, 11.25]
Both targeted changes	6,003	9.85	[8.07, 11.63]
Drop all multi-parcel endpoint deeds	5,926	9.82	[8.03, 11.61]
Recode all endpoints meeting the date rule	6,006	9.36	[7.52, 11.21]
Both general rules	5,926	9.91	[8.09, 11.73]

Log points. A research assistant read the seven targeted endpoints without AI, and I checked the readings. All three bundle cases satisfy the multi-parcel rule; all four purchase mortgages satisfy the single-parcel index-date rule. The P01 purchase mortgage covers both lots, and its sale is excluded as a bundle. No row uses the earlier AI review. Broad rules use the complete metadata trace; unreviewed records are not thereby document-validated. Analytic parcel-clustered intervals condition on each scenario and omit classification uncertainty.

D.3 A random sample of financing labels

I drew 20 endpoints at random from each original financing label among the 11,852 previously unreviewed sale endpoints (starting from the 11,963 unique sale endpoints across the 6,006 headline pairs and excluding 111 previously inspected endpoints: counting distinct endpoints across an earlier superseded queue and the targeted influential cases of Appendix D.2, 55 cash-labelled and 56 financed-labelled endpoints in total). The labels were hidden from the research assistant. The research assistant read the deed and candidate mortgages, checked the borrower, collateral and dates, and searched the no-candidate cases for a purchase mortgage; I checked every reading. The search covers mortgages on the parcel from 15 days before to 90 days after the sale. A no-purchase-found call describes this documented search and does not establish the absence of borrowing outside it.

The checked readings support purchase mortgages for all 22 candidate endpoints and find no purchase mortgage for the 18 no-candidate endpoints. Against the original labels, 2 of 20 unmatched endpoints (N015 and N017) have purchase borrowing, while no matched endpoint is contradicted; 0 cases remain unresolved. This gives a design-weighted discrepancy estimate of 2.19 percent and conservative 95 percent interval [0.19, 22.96] percent. The bound uses exact finite-population hypergeometric inversion within both strata, with a Bonferroni allocation across four one-sided tails. Correcting only those sampled labels changes the contrast from 9.25 to 9.23 log points; no unsampled label is recoded (Table 13).

Table 13: Human document review of financing labels and sample-specific corrections

Recorded label	Sample	Purchase supported	No purchase found
No matched mortgage	20	2	18
Matched mortgage	20	20	0
Regressions correcting only sampled endpoints			
Handling	Pairs	Contrast	Analytic 95% interval
Original headline sample	6,006	9.25	[7.44, 11.07]
Correct two sampled financing labels	6,006	9.23	[7.42, 11.05]
Also exclude sampled bundle	6,005	9.26	[7.44, 11.07]

A research assistant read the original images without AI, and I checked every reading. Each original financing stratum contributes 20 endpoints; the supported and no-purchase-found counts come from the checked readings. The latter is a finite-search result. N015 and N017 are purchase mortgages behind originally unmatched labels; no originally matched label is contradicted. No unresolved case remains. The design-weighted discrepancy estimate is 2.19 percent with conservative simultaneous 95% interval [0.19, 22.96] percent, conditional on reading accuracy. Regression intervals are analytic and parcel-clustered, conditional on each scenario; no unsampled label is corrected.