When a Deed Is Not a Market Sale: Foreclosure Transfers, Statutory Consideration, and Repeat-Sales House-Price Measurement

Pablo Loschi, independent researcher, Berlin, Germany. ORCID 0009-0004-9455-4713. Corresponding author: loschi.pablo@gmail.com

19 November 2026 · Measurement Note

Abstract

I flag 16,655 transfers to lenders in 1.47 million New York City deeds. Roughly 62% clear standard price and parcel screens, and 94–95% of screened filings since 2016 appear in the city's sales file at the same dollar amount without a flag. Because transfer-tax rules allow the bid or discharged debt to serve as consideration, these amounts sit 24–34 log points above the lender's later resale. Retained in a house repeat-sales index with fixed labels, they raise 2010Q3 by 3.69 log points and lower 2011–12 by 2.28, shifting crisis timing citywide while masking sharp submarket heterogeneity: 2010Q3 distortions reach +10.92 in the Bronx, +10.34 ([+7.21, +13.46]) in bottom pre-2007 ZIP price tertiles and +10.70 in top-quartile foreclosure ZIPs (where the peak-to-trough drop is understated by 5.89 log points), against +0.10 in top-tertile ZIPs. An unpaired hedonic index shows the 2010Q3 excess (+2.07) without the 2011–12 shortfall (+0.14). In Philadelphia and New Jersey, comparable statutes yield below-market foreclosure amounts that deepen measured crashes. Across 40 blind human readings, weighted precision is 0.82 ([0.40, 0.99]) and recall is 0.71 ([0.02, 0.95]), widening 2010Q3 to +2.25 [−0.21, +4.53] under an uninformative Beta(1/2, 1/2) prior on `S4` (0/11). Pure registry structure separates corporate/entity grantees—99.98% of house-index `S1` deeds and 13/13 human non-market calls versus 20.0% of house-index `S4` (11.8% unfinanced)—yielding +2.13 [+0.15, +3.86] on unfinanced entity `S4` deeds and +2.11 [−0.01, +3.87] across all entity `S4` deeds without arbitrary numerical caps, while fixed-prevalence bounds cross zero only above 4.76% `S4` false negatives.

1. Introduction

Studies of house prices drop transfers that are not arm's-length sales, such as gifts, transfers between relatives and repossessions by lenders. Commercial data vendors flag transaction types for this purpose, and Kermani and Wong (2026), for example, keep only the arm's-length transactions in vendor deed data, which excludes deeds in lieu of foreclosure. Careful studies also make the separation by hand, as when Campbell et al. (2011) treat the lender's resale after an unsuccessful foreclosure auction as the forced sale. Researchers working from free public files have only what the record shows. Repeat-sales indices in the tradition of Bailey et al. (1963) and Case and Shiller (1987), hedonic models, studies of foreclosure discounts and local price series all begin from a registry field labelled "consideration" or "sale price". That field comes from transfer-tax administration, which was never built to record market prices. Where tax law defines consideration differently from a negotiated price, the published amount inherits that legal rule. The Case–Shiller screen drops lender repossessions where the deed record identifies them (S&P Dow Jones Indices 2025), and vendor flags do similar work for subscribers. I ask what happens when neither screen is available. In New York's public records, many lender acquisitions follow statutory tax formulas yet still pass ordinary price and parcel filters.

Under New York Tax Law §1401(d) and New York City Administrative Code §11-2101(9), consideration can include "the cancellation or discharge of an indebtedness or obligation". A lender may take title as plaintiff at its own foreclosure sale, through a deed in lieu of foreclosure or via a transfer between servicer and investor. In those cases, the amount recorded in New York City's Automated City Register Information System (ACRIS) may reflect a bid, judgment or debt instead of a price agreed by a buyer and seller. I keep the lender's later resale as a potential market transaction, separate from its acquisition, so a repeat-sales estimator that pairs the two amounts risks treating a non-price as a market price.

These transfers survive the simple price and parcel filters that remove many non-arm's-length sales from public data. Minimum-price filters catch nominal transfers, yet a lender-acquisition amount sits well inside the range of market prices. It tracks the Department of Finance (DOF) market value closely, and 62.2% of flagged deeds survive a $10,000 minimum alongside full-interest, single-parcel and residential-class filters. In the city's free DOF sales file, a standard source for New York City price series (e.g., Schwartz et al. 2003; Ellen et al. 2007), 94–95% of screened filings recorded since 2016 appear at the exact same amount with no transaction-type

flag. Users of that file need a transaction-type screen beyond its price and parcel fields, and the party-based classifier provides one reproducible candidate.

The issue first appeared in a pilot for a companion study of financing-associated price gaps (Loschi 2026). Among 8,208 one-to-three-family repeat pairs from 2016–2025, 232 had a lender as grantee at the first sale. These pairs fell in nominal terms 47.8% of the time in a rising market, against 8.5% of other pairs, and their first-sale amounts were multiples of $1,000 only 22% of the time, against 90%. Here I scale that observation to every deed recorded in four boroughs from 2003 to 2025, every residential property type conveyed by deed and the full 2008 housing cycle. Detailed foreclosure studies already separate lender acquisition from resale (Campbell et al. 2011), while recorded prices can also depart from negotiated ones through financing concessions (Ben-David 2011) or changing property quality (Nowak and Smith 2020). Dropping flagged deeds addresses the first source of bias and leaves the others untouched.

I flag 16,655 deed-family documents in events where law can set the amount, or 1.1% of deeds overall and 3.7–4.2% of priced deeds in 2008–09. Seven in eight are 1–3 family houses, mostly in Queens, Brooklyn and the Bronx. Their share rose almost ninefold from 2005–06 to 2009 before dropping below 1% in 2011–12 when new court rules slowed residential foreclosures. Flagged amounts do not behave like negotiated prices. Among screened repeat-pair sales, 69.9% carry non-zero cents and only 17.5% are multiples of $1,000, compared with 6.3% and 81.2% for ordinary sales. They match DOF market values and prior ordinary sales carried forward, yet sit 24–34 log points above the parcel's next sale. Because that next sale is the lender's resale 94% of the time, the gap likely mixes statutory take-title valuations with forced-sale discounts (Pennington-Cross 2006; Clauretie and Daneshvary 2009; Conklin et al. 2023) and neighborhood or condition changes (Harding et al. 2009; Gerardi et al. 2015; Anenberg and Kung 2014; Fisher et al. 2015) that registry fields alone cannot isolate.

Keeping flagged pairs raises a 1–3 family repeat-sales index by 3.69 log points in 2010Q3 and lowers it by 2.28 over 2011–12 under fixed classifier labels, leaving the citywide peak-to-trough decline virtually unchanged (−0.64 log points). Submarket disaggregation reveals that the citywide mean obscures steep cross-sectional concentration: 2010Q3 distortions reach +10.92 log points in the Bronx, +10.34 ([+7.21, +13.46]) in bottom pre-2007 ZIP price tertiles and +10.70 ([+7.57, +14.17]) in top-quartile foreclosure ZIPs (understating the local peak-to-trough drop by 5.89 log points), versus +0.10 in top-tertile ZIPs. On the matched DOF sample, an unpaired hedonic index recovers the 2010Q3 excess (+2.07) without the 2011–12 shortfall (+0.14), pointing to the repeat-sales pairing identity as the source of the rebound. Comparable administrative rules in Philadelphia and New Jersey move indices in the opposite direction by recording below-market foreclosure amounts that deepen measured crashes instead of shifting timing (Doerner and Leventis 2015). Against 40 blind human readings from a 400-deed stratified audit, weighted precision is 0.82 [0.40, 0.99] and foreclosure recall is 0.71 [0.02, 0.95], leaving a 2010Q3 uncertain-rate range of +2.25 [−0.21, +4.53] under an uninformative Beta(1/2, 1/2) prior on `S4` (crossing zero only above a 4.76% `S4` false-negative prevalence, and staying strictly positive under unfinanced entity-grantee bounds, ordered-strata monotonicity and truncated or crisis-restricted priors). More broadly, Feng and Hu (2013) show in labor-market data that classification errors need not cancel out in aggregate levels when error rates vary across latent states, while Bailey et al. (2020) document how automated record algorithms introduce systematic false links relative to human audits that attenuate downstream elasticities.

2. Institutions and the meaning of recorded amounts

2.1 Consideration under New York transfer-tax law

Every conveyance of real property in New York City is accompanied by state (TP-584 or TP-584-NYC) and city (NYC-RPT) transfer-tax filings. ACRIS, operated by the Office of the City Register within the New York City Department of Finance, publishes a document amount for each deed. ACRIS connects deed cover pages to transfer-tax filings, though public documentation I reviewed does not establish which form field supplies each published document amount, so the exact provenance of a given record remains ambiguous. Section 3 tests whether those amounts behave as the tax definitions imply. State and municipal law both include discharged indebtedness in consideration, while Schedule E of state form TP-584.1 (New York State Department of Taxation and Finance 2019) and Rules of the City of New York 19 RCNY §23-03(d) (City of New York 2024) specify different calculations across foreclosure and deed-in-lieu filings.

- **State foreclosure calculation (Schedule E, Part 1).** When the grantee is the mortgagee, its agent or nominee, or an entity it wholly owns, the starting amount is the foreclosure judgment (including interest accrued to the date of transfer) or the bid price, whichever is higher, plus continuing liens. For an unrelated buyer, it is the bid plus continuing liens.
- **City foreclosure calculation (19 RCNY §23-03(d)(2)).** A referee's deed is taxed on the bid, senior liens not canceled by the sale, and specified purchaser-paid costs, regardless of whether the buyer is the mortgagee.
- **State deed in lieu calculation (Schedule E, Part 2).** The starting amount is the unpaid secured debt, plus continuing liens and other amounts paid by the grantee.
- **City deed in lieu calculation (19 RCNY §23-03(d)(1)).** Tax is computed on outstanding mortgage debt and unpaid accrued interest, whether or not the mortgagor is personally liable for that debt.
- **State recourse debt (Schedule E, Parts 1 and 2).** For a qualifying recourse debt, the state form uses the lower of the calculated amount and the property's fair market value. This state-form cap is not stated in the cited city rule.

These legal rules generally tie the recorded amount to the bid, judgment or debt instead of a negotiated exchange price. Which calculation produced a specific ACRIS amount cannot be inferred from the registry field alone. New York permits deficiency judgments on residential mortgages, so some loans may qualify as recourse under the state form, and registry filings do not disclose whether a filer applied the state recourse cap or how a fair-market-value appraisal was constructed. The state form cited is the July 2019 revision, and I have not verified that earlier revisions computed consideration identically. I do not treat any recorded amount as evidence of legal non-compliance or assume it equals debt. Instead I examine empirically whether flagged amounts behave like negotiated prices.

2.2 Judicial foreclosure in New York

Because New York forecloses judicially, the court enters a judgment of foreclosure and sale and appoints a referee, who auctions the property and conveys it by referee's deed to the successful bidder. When no third party outbids the plaintiff, the plaintiff, or its designee, servicer or securitization trustee, takes title. The lender then markets the property as real-estate-owned (REO) and sells it, usually within one to two years. In my data the median interval between a legal-rule deed and the next sale of the same parcel is 498 days.

Foreclosure timelines in judicial states such as New York are among the longest in the country (Cordell et al. 2015). New York added two procedural requirements during the crisis. Mandatory settlement conferences (CPLR 3408), introduced in 2008 for subprime and high-cost loans, were extended in December 2009 to all residential foreclosures on one-to-four family homes. Administrative Order AO/548/10 of 20 October 2010 required plaintiff's counsel in residential foreclosures to affirm that papers and notarisations had been checked for accuracy, and AO/431/11 replaced it in March 2011, retroactive to November 2010. Both lengthened pending cases, and the timing appears directly in the registry (Section 3.3).

2.3 What ACRIS does and does not record

ACRIS publishes linked tables through NYC Open Data, with a master record per document (type, dates, amount, percentage conveyed), one row per parcel conveyed, one row per party, cross-references to earlier documents, occasional free-text remarks entered at recording, and code lists. No document type distinguishes a referee's deed or a deed in lieu from an ordinary sale, since all are recorded as `DEED`. In the structured extract the event has to be inferred from its parties, although the deed image itself identifies the instrument. A referee's deed names the referee as grantor ("SMITH, ESQ., AS REFEREE"), and a lender acquisition names a bank, servicer, trustee or federal agency as grantee. Staten Island deeds are recorded by the Richmond County Clerk and are absent from ACRIS, so the analysis covers the other four boroughs. Cooperative apartments convey by share transfer and lie outside the deed scope.

Table 1 sets out the six-class taxonomy separating who conveys to whom from the basis of the recorded amount. Three events are likely to carry statutory consideration, namely a referee's deed to a lender, other transfers to a lender (principally deeds in lieu) and transfers between lenders or to a federal agency. Third-party purchases at foreclosure auctions represent competitive bids and are tracked separately, while REO resales remain in the target market population.

Table 1. Deed events and the meaning of the recorded amount (doc type DEED, four boroughs, 2003–2025)

Event	Parties	Amount	Deeds	Amount ≥ $10k	Median (≥ $10k)
Referee's deed to lender	referee → lender or agency	bid, legal rule	6,742	77.3%	$455,726
Other transfer to lender	owner → lender (deed in lieu etc.)	debt, legal rule	7,119	75.7%	$477,067
Lender to lender	lender → lender or agency	legal rule	2,207	10.6%	$283,436
Third-party auction	referee → non-lender	competitive bid	7,882	88.5%	$400,530
REO resale	lender → non-lender	negotiated price	14,272	94.0%	$321,000
Ordinary	neither party a lender or referee	negotiated price	1,354,845	62.5%	$658,000

3. Measuring legal-rule deeds in ACRIS

3.1 Data

I retrieve ACRIS master, parcel, party, cross-reference and remark tables along with DOF annualized sales and assessment rolls from the NYC Open Data API for documents identified from January 2003 to December 2025, when positive amounts appear systematically. To prevent silent API truncation, extraction runs in calendar-month chunks asserted equal to live `count(1)` totals before and after each pull, reconciling all seven sources with zero drift across 3,579,646 deed and mortgage documents, 12,289,847 parcel rows, 24,885,207 party rows, 8,007,274 cross-reference rows, 217,958 remark rows, 845,607 DOF sales and 8,600,667 assessment records.

The deed file contains 1,470,324 deed-family documents in the four boroughs, of which 1,393,082 are of type `DEED`. The percentage-conveyed field is blank (recorded as 0) for 97.7% of 2003 deeds and 39.6% of 2004 deeds and almost never afterwards. Because those records resemble ordinary sales (74% are multiples of $1,000), I treat 0 as unrecorded and drop only partial interests strictly between 0 and 100%. A buyer is coded as financed if a mortgage on the same parcel is dated between 7 days before and 30 days after the deed, which likely captures most contemporaneous purchase loans. Property class comes from ACRIS property codes (1–3 family houses, 4–6 family buildings, residential condominium units and apartment buildings).

3.2 A party-based classifier

Each deed is classified solely from normalized grantor and grantee strings. A party is coded as a lender if its normalized name matches an explicit institutional phrase across four families spanning federal agencies (Fannie Mae, Freddie Mac, HUD, VA, FDIC), banks and servicers (`BANK`, `SAVINGS`, `NATIONAL ASSOCIATION`, `MORTGAGE`, `SERVICING`, `HOME LOANS`), securitization vehicles (`CERTIFICATEHOLDERS`, `PASS-THROUGH`, `REO`), and a fixed list of named servicers and bank REO subsidiaries (Appendix A). A grantor is a referee if the name contains `REFEREE` or `SHERIFF`, and NYC tax-lien trusts or condominium boards count as lenders when taking title on a referee's deed.

No ambiguous single word acts unconditionally as a token. In the pilot, bare `N.A.` matched 19 Chinese and Korean surnames (`"CHEN, NA"`), and at full scale `"LOAN"` is a Vietnamese given name (`"NGUYEN, LOAN"`), `"CHASE"`, `"SANTANDER"`, `"REO"` and `"BANK"` are surnames, and `"GREENPOINT"` is a Brooklyn neighborhood common in developers' LLC names. The person guard rejects `"SURNAME, GIVEN"` strings lacking a corporate suffix. After reviewing frequent hits, person-guard rejections, and unmatched referee's-deed grantees with non-zero cents, I added five missed REO subsidiaries (HomeSales Inc., Liquidation Properties, Property Asset Management, Retained Realty and Residential Funding). Across 1,694,771 distinct grantee names on deed-family documents, 4,396 match the narrow rule, flagging 16,655 deed-family documents (16,068 of type `DEED`).

3.3 Incidence

Figure 1 shows the annual count and share of legal-rule deeds recording at least $10,000. Their share of priced deeds rose nearly ninefold from 0.5% in 2005–06 to 3.7% in 2008 and 4.2% in 2009 before dropping below 1% in 2011–12 when AO/548/10 slowed referee take-backs from 959 in 2009 to 110 in 2011 and 81 in 2012. A second wave of deeds in lieu and post-foreclosure transfers followed in 2016–19, peaking at 1.7%. Across 2003–2025, priced legal-rule deeds total 10,838 of 877,751 priced deeds (1.2%), comprising 5,865 in Queens, 2,706 in Brooklyn, 1,839 in the Bronx and 428 in Manhattan. By property class, 9,466 are 1–3 family houses and 698 are condominium units.

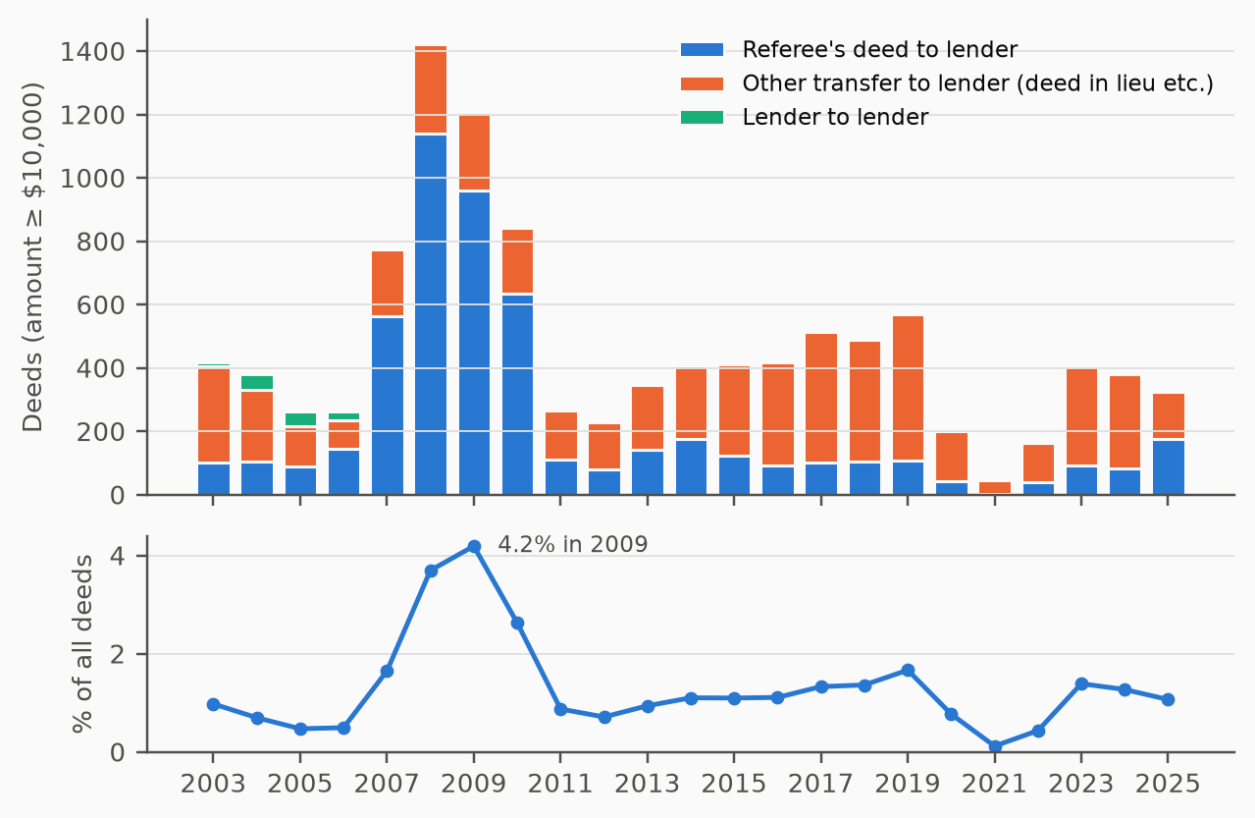

Figure 1. Legal-rule deeds by year (doc type DEED, amount ≥ $10,000). Top: counts by event. Bottom: share of all deeds with an amount of at least $10,000.

3.4 Digit patterns and resale trajectories

Negotiated prices round frequently, whereas judgment balances and debt payoffs accumulate accrued interest and fees to the cent. Because non-zero cents helped identify missed subsidiaries during rule development, digit distributions are a development diagnostic instead of an independent validation and cannot classify an individual deed (Figure 2 and Table 2). Among screened DEED sales entering repeat pairs, 69.9% of legal-rule amounts carry non-zero cents (74.1% on referee's deeds to lenders) versus 6.3% of ordinary sales, while only 17.5% are multiples of $1,000 versus 81.2%. Third-party auction bids and REO resales sit in between, while ordinary cents concentrate in condominium sponsor sales.

Subsequent resale paths broadly mirror the pilot, as repeat pairs starting at a legal-rule deed fell in nominal terms 73.9% of the time (89.9% for 2008–12 acquisitions), against 18.2% of pairs with no lender on either side. Median consideration drops from $449,490 at take-title to $336,648 at resale, with lenders acquiring without a new mortgage in 6,845 of 6,883 such pairs and reselling to financed buyers in 3,592 cases.

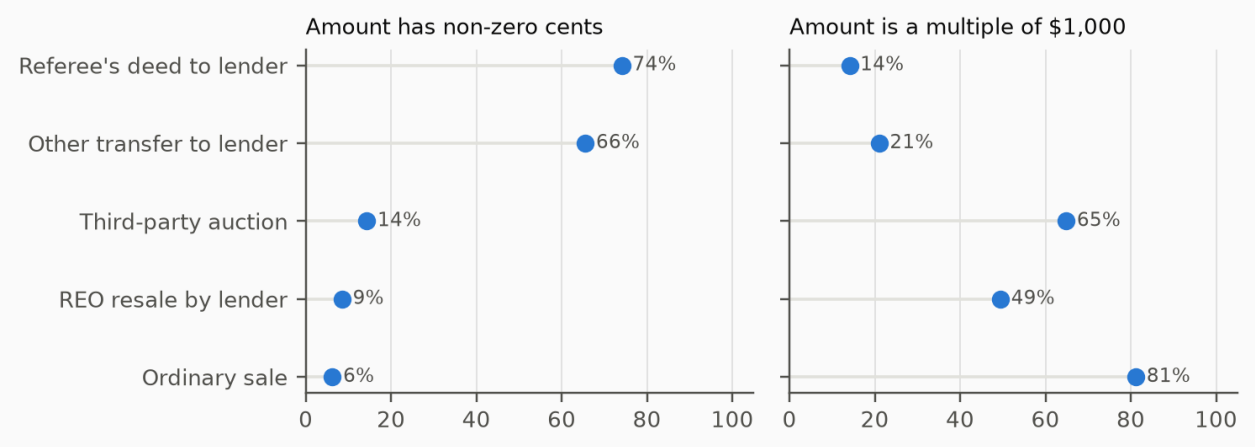

Figure 2. Digit patterns of recorded amounts, screened sales in repeat pairs.

Table 2. Forensic tests (screened DEED sales in repeat pairs; 2003–2025)

	Legal-rule	Referee's deed to lender	Other to lender	Third-party auction	REO resale	Ordinary
Sales in repeat pairs	9,091	4,624	4,312	4,096	10,054	387,872
Amount has cents	69.9%	74.1%	65.6%	14.3%	8.5%	6.3%
Multiple of $1,000	17.5%	14.1%	21.0%	64.9%	49.5%	81.2%
Multiple of $10,000	7.5%	5.3%	9.8%	32.5%	21.2%	44.3%
Pairs starting here, nominal fall	73.9%					18.2%

Notes: "Legal-rule" pools the three events of Table 1 whose amount is set by a legal rule. The last row compares pairs whose first sale is a legal-rule deed with pairs that have no lender or referee on either side.

3.5 External benchmarks

Figure 3 tests where 1–3 family legal-rule amounts sit relative to three benchmarks (median log gap ×100). First, against January tax-class 1 DOF final-roll market values (222,849 matched out of 226,687 eligible sales over 2010–18 and 2022–25, net of borough-year ordinary medians), referee's deeds to lenders sit +4.5 log points overall (+13.1 in 2010–12, +1.1 in 2013–18 and −13.3 in 2022–25) and other lender acquisitions sit −1.5 (+7.5 in 2010–12). Shifting assessment valuation dates by ±1 year moves referee's deeds between −4 and +8 log points, suggesting that the small magnitude persists even though the sign appears sensitive to valuation timing. By contrast, REO resales trade 33.7 log points below benchmark, close to the −0.314 forced-sale hedonic coefficient estimated for Massachusetts by Campbell et al. (2011), and third-party auctions sit 26.2 below.

Second, carrying the parcel's prior ordinary sale forward with the clean index places referee's deeds at +1.4 log points and other acquisitions at +1.3 (versus +7.9 ordinary). Third, carrying the next sale (the lender's resale 94% of the time) backward places referee's deeds +34.2 and other lender acquisitions +23.9 log points *above* next-sale implied values, versus −11.0 for ordinary sales. Because statutory amounts resemble nearby assessments and prior sales instead of liquidation proceeds, ratio plausibility filters cannot catch them.

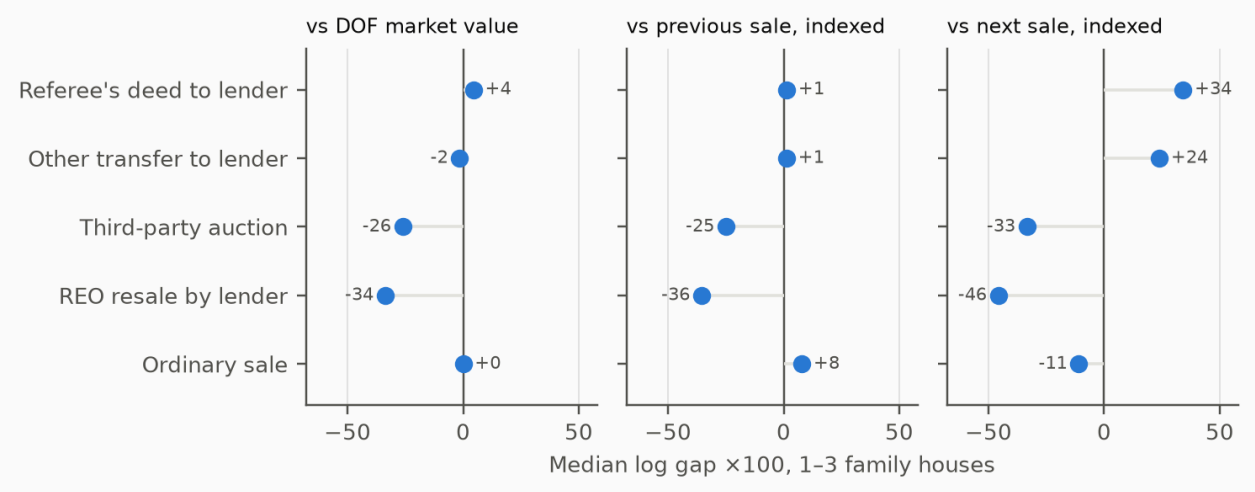

Figure 3. Recorded amount against three external benchmarks, 1–3 family houses. Median log gap ×100. Left: relative to DOF market value net of borough-year ordinary medians. Centre and right: relative to prior ordinary sale indexed forward and next sale indexed backward with the clean (drop) index.

3.6 Conventional screens and DOF sales-file pass-through

Table 3 applies cumulative repeat-sales screens. Removing $0 deeds drops 13% of legal-rule deeds and a $10,000 minimum removes another 22% (chiefly nominal lender-to-lender transfers). Full-interest, single-parcel and residential-class screens exclude almost none, leaving 62.2% of legal-rule deeds surviving (versus 51.6% of ordinary deeds) and 54.6% entering a repeat-sales pair (87.7% of screened legal-rule deeds versus 52.5% of screened ordinary deeds).

Critically, these amounts carry directly into NYC's standard public sales file (Schwartz et al. 2003; Ellen et al. 2007). Among screened legal-rule deeds recorded in 2016–25, 93.8% of referee's deeds to lenders and 95.3% of other lender acquisitions appear in the DOF annualized sales file on the same borough-block-lot within

seven days at the exact same dollar amount (versus 98.9% for ordinary deeds), with no transaction-type flag distinguishing them from market trades. Researchers using that public file without a party-based screen could therefore treat these statutory transfers as arm's-length market trades. Because DOF labels condominium units by building billing lot, matching across files requires joining on borough, block and lot.

Table 3. Share of deeds surviving cumulative screens (deed-family documents, four boroughs, 2003–2025)

Screen (cumulative)	Legal-rule	Share	Ordinary	Share
All deed-family documents	16,655	100.0%	1,430,816	100.0%
Doc type `DEED`	16,068	96.5%	1,354,845	94.7%
Amount > 0	14,424	86.6%	872,576	61.0%
Amount ≥ $10,000	10,838	65.1%	846,528	59.2%
Full interest conveyed	10,832	65.0%	840,379	58.7%
One parcel, no easement/air/partial lot	10,667	64.0%	795,142	55.6%
Residential class	10,361	62.2%	738,334	51.6%
In a screened repeat-sales pair	9,091	54.6%	387,872	27.1%

Notes: Columns exclude third-party auctions and REO resales and restrict document date (recording date when missing) to 2003–2025.

4. Index consequences and institutional contrast

4.1 Repeat-sales timing distortion versus unpaired hedonic estimates

I estimate the quarterly Bailey et al. (1963) repeat-sales index by ordinary least squares,

$$\log R_{ib} - \log R_{ia} = \sum_{t=1}^{T} \delta_t (\mathbb{1}[t = b] - \mathbb{1}[t = a]) + u_q, \qquad \delta_{2003Q1} = 0,$$

across 238,290 consecutive screened pairs (762,062 sales, spanning 140,119 1–3 family house pairs and 86,943 condominium pairs), requiring $\geq$ 180-day holding periods and $|\Delta \log R/\text{year}| \leq 1$. If recorded amount R_L differs from market price P_L by $e_L = \log R_L - \log P_L$, deed L enters $A \to L$ with $+e_L$ and $L \to B$ with $-e_L$. Under **drop**, every pair touching a flagged deed is excluded under fixed pairing, whereas **bridge** strips flagged deeds before forming pairs and adds 5,794 foreclosure-spanning pairs that appear to absorb REO discounts and physical deterioration. Intervals condition on fixed classifier labels via 300 paired Poisson(1) parcel-bootstrap draws.

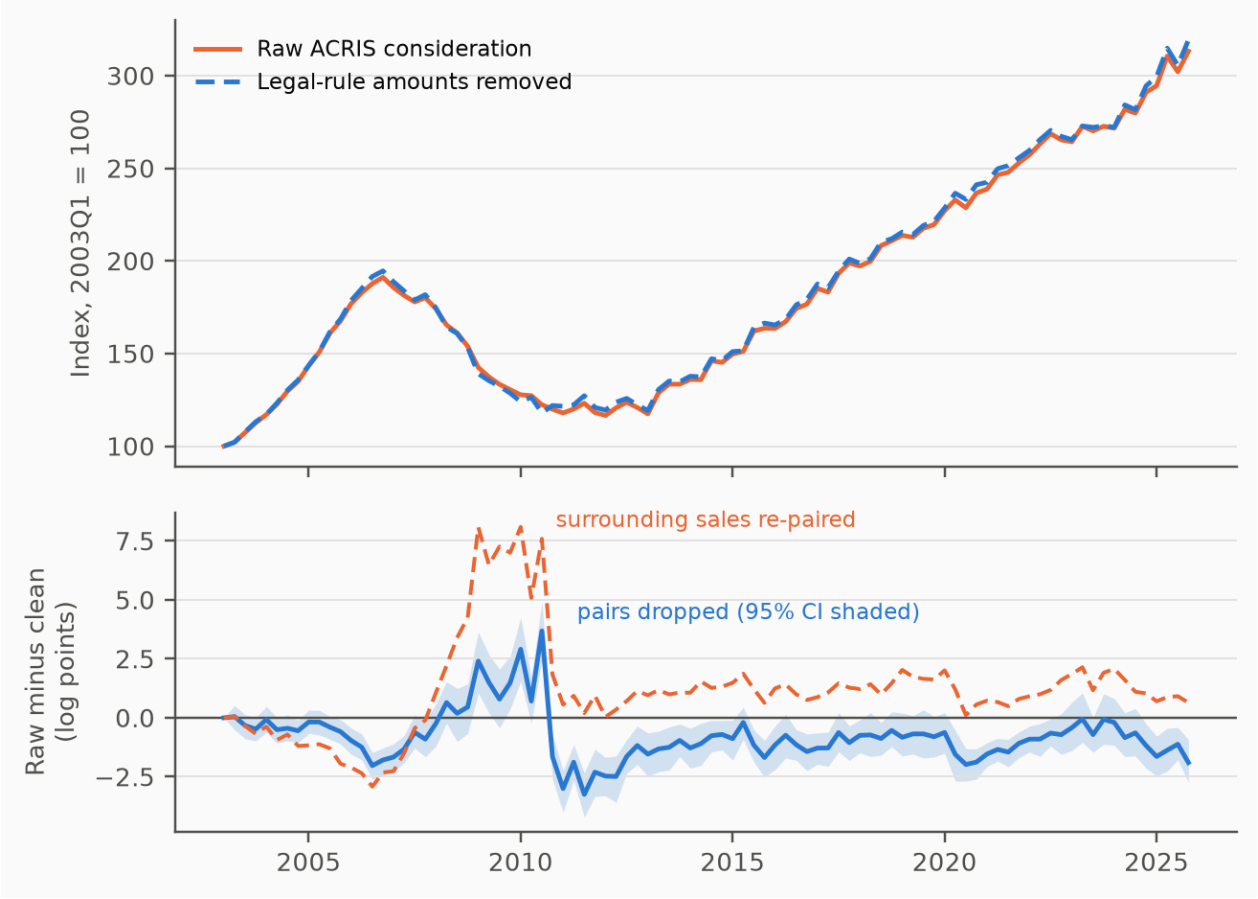

Figure 4. Top: quarterly repeat-sales index for 1–3 family houses in four boroughs, raw and with pairs that touch a flagged deed dropped. Bottom: raw minus clean, log points; solid with pointwise 95% bootstrap band: drop; dashed: bridge.

Figure 4 and Table 4 summarize the repeat-sales and hedonic contrasts for 1–3 family houses. Under the published screen (`D0`, 12,483 pairs dropped), the raw series exceeds the screened index by +1.07 log points over 2008–10 (95% CI [+0.55, +1.59]) and peaks at +3.69 log points in 2010Q3 ([+2.48, +4.92] pointwise). Its simultaneous sup-t band of [+1.59, +5.78] excludes zero in 28 quarters, with a maximum aligned borough $\times$quarter placebo draw of 3.41 across 200 draws and Bronx houses alone reaching +10.92. The gap then drops to −2.28 over 2011–12 ([−2.76, −1.72]) and −1.00 over 2013–25 ([−1.43, −0.57]). Because the two legs largely offset between the 2005–08 peak and 2008–13 trough, cumulative peak-to-trough decline moves by only −0.64 log points ([−3.03, +1.25], raw −49.4 vs screened −50.1). Under the narrowest definition `D4` (3,309 referee's deeds to non-securitization lenders carrying non-zero cents, or 4,754 pairs), 2010Q3 remains +2.63 ([+1.83, +3.67] pointwise, [+0.88, +4.37] simultaneous, aligned placebo max 1.89) and 2011–12 averages −1.09 ([−1.27, −0.91]), indicating that ambiguous REO names do not drive the result. This contrast is inherently retrospective. Truncating pairs at 2010Q3 reverses 2010Q3 to −4.00, dating sales by recording date instead of document date gives +1.75 at 2010Q3, and bridging increases peak-to-trough decline by 5.07 log points ([2.11, 7.28]). For condominiums (1% of pairs flagged), the maximum difference is −0.82 [−2.29, +0.14].

To separate statutory consideration from repeat-sales pairing, I fit an unpaired hedonic regression joining screened 1–3 family sales within 45 days to DOF annual attributes (435,486 sales on 284,880 parcels with 8,341 `D0` deeds retained, where prices agree within 0.5% in 433,219 rows). Controls cover quarter, ZIP code, building class, log gross and land area with squares, age and age squared, 2/3-unit indicators and commercial units. On this common sample (Table 4), the raw hedonic series sits +2.07 ([+1.27, +2.87] joint simultaneous band, and +1.89 with class and location controls alone) above the screened index in 2010Q3 and +1.39 over 2008–10, confirming that elevated take-title consideration does not require repeat-sales weighting. Over 2011–12, however, the hedonic gap vanishes (+0.14 [+0.02, +0.27]), whereas repeat sales fitted strictly on pairs whose endpoints lie inside the hedonic sample retain −2.31 (+2.99 in 2010Q3). The 2011–12 shortfall therefore appears to arise specifically from pairing L negatively into $L \to B$.

Table 4. Summary of repeat-sales and hedonic index differences (raw minus screened, 1–3 family houses, log points)

Estimator / specification	Sample	Dropped	2010Q3 [95% CI]	2008–10	2011–12	Peak–trough
BMN repeat sales (`D0` drop)	140,119 pairs	12,483 pairs	+3.69 [+2.48, +4.92]^{p}	+1.07 [+0.55, +1.59]	−2.28 [−2.76, −1.72]	−0.64 [−3.03, +1.25]
BMN (`D0` drop, sup-t)	140,119 pairs	12,483 pairs	+3.69 [+1.59, +5.78]^{s}	placebo max 3.41	28 qtrs ≠ 0	RMS 1.33 (max 1.27)
BMN (`D4` narrowest)	140,119 pairs	4,754 pairs	+2.63 [+1.83, +3.67]^{p}	+1.05 [+0.83, +1.24]	−1.09 [−1.27, −0.91]	+0.63 [−0.77, +0.98]
BMN (`D0` bridge)	+5,794 pairs	–	+7.59 (+8.09 Q1)	–	+0.61	−5.07 [−7.28, −2.11]
Hedonic, all controls	435,486 sales	8,341 deeds	+2.07 [+1.27, +2.87]^{s}	+1.39 [+1.25, +1.53]	+0.14 [+0.02, +0.27]	–
Hedonic, class + loc.	435,486 sales	8,341 deeds	+1.89 [+1.12, +2.66]^{s}	–	–	–
BMN on hedonic sample	124,768 pairs	10,480 pairs	+2.99	–	−2.31	–

Notes: Superscript ^{p} denotes pointwise 95% paired Poisson(1) parcel-bootstrap intervals (300 draws), and ^{s} denotes simultaneous 95% sup-t bands across all 91 non-base quarters (and across both models and definitions for hedonic rows). Labels are held fixed, and the aligned placebo matches borough, first-sale quarter and second-sale quarter (200 draws).

4.2 Submarket concentration across boroughs and pre-crisis price tiers

Citywide averages dilute statutory-consideration bias across neighborhoods with vastly different judicial foreclosure intensity. Re-estimating the exact same `D0` BMN repeat-sales specification separately by borough and by pre-crisis neighborhood price tier (`scripts/entity_and_submarket_v1_0_19.py`, 999 paired Poisson(1) parcel-bootstrap draws, reproducing the citywide +3.69 and −2.28 estimates to $< 10^{-10}$) exposes sharp cross-sectional concentration (Figure 5 and Table 5). Across boroughs, 2010Q3 inflation reaches +10.92 log points ([+6.28, +15.15]) in the Bronx (11.0% of pairs dropped) and +4.01 ([+2.39, +5.66]) in Queens (10.4% dropped, with a −2.79 [−3.49, −2.14] shortfall over 2011–12). Brooklyn's borough-wide 2010Q3 gap is modest (+0.23 [−1.80, +2.15]; peak +0.98 in 2010Q1, 2011–12 −2.13 [−3.23, −1.17]) because affluent brownstone ZIPs dilute high-foreclosure neighborhoods in eastern Brooklyn, while Manhattan has only 1,836 1–3 family pairs (3.4% dropped; −0.43 [−2.02, +0.89] in 2010Q3).

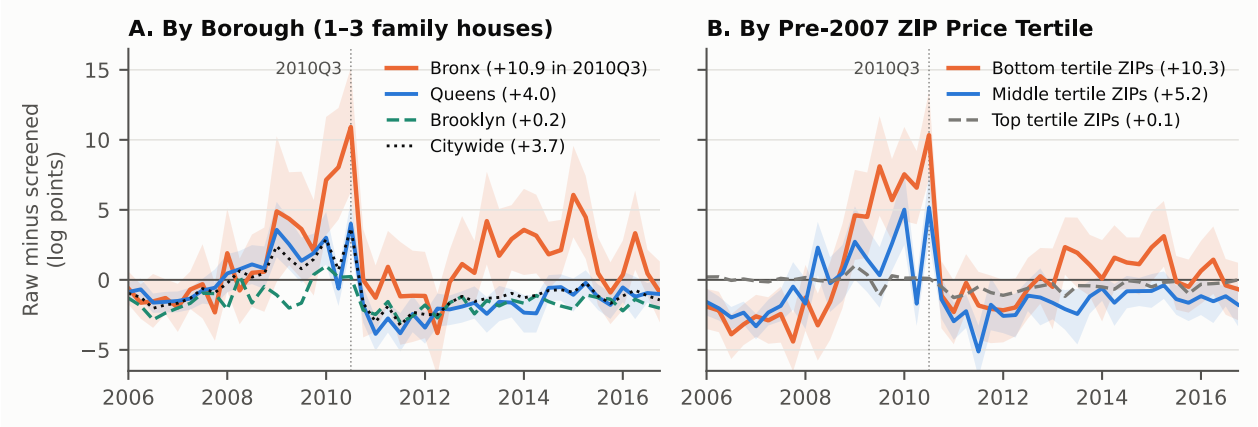

Figure 5. Submarket BMN index differences (raw minus screened, 1–3 family houses, 2006–2016) by borough (left) and pre-2007 ZIP median house price tertile (right), with 95% paired parcel-bootstrap bands.

Grouping the 140,118 house pairs matched to 160 ZIP codes with pre-2007 ordinary house sales into balanced pair-weighted tertiles of 2003–2006 median sale price separates foreclosure-exposed submarkets cleanly without look-ahead bias. Bottom-tertile ZIPs (34 ZIPs, pre-2007 medians $210k–$389.5k; 14.3% of pairs dropped) exhibit a +10.34 log-point 2010Q3 excess ([+7.21, +13.46], 2.8× the citywide effect) and a strictly negative peak-to-trough difference of −2.84 ([−7.27, −1.23]), indicating that unscreened statutory deeds understate the local housing crash. Middle-tertile ZIPs (36 ZIPs, $392k–$475k; 9.5% dropped) show +5.16 ([+2.83, +7.69])

in 2010Q3 and −2.45 ([−3.27, −1.66]) in 2011–12, whereas top-tertile ZIPs (90 ZIPs, ≥$475k; 3.2% dropped) sit virtually at zero (+0.10 [−0.73, +1.14]). In the 26 top-quartile foreclosure-exposure ZIPs (17.0% of pairs dropped, led by Southeast Queens and the South Bronx), 2010Q3 bias hits +10.70 ([+7.57, +14.17]), 2011–12 drops to −2.71 ([−4.34, −1.24]), and the peak-to-trough drop is understated by 5.89 log points ([−11.67, −1.38]). Empirical studies of neighborhood foreclosure contagion or price-tier dynamics (Campbell et al. 2011; Anenberg and Kung 2014; Gerardi et al. 2015) built from unscreened public registries therefore inherit over 10 log points of differential tier bias between lower- and higher-price ZIPs during the crisis trough.

Table 5. Submarket repeat-sales index differences (raw minus `D0`-screened, 1–3 family houses, log points)

Submarket	Pairs (dropped %)	2010Q3 [95% CI]	Peak 09–11 (qtr)	Mean 2011–12 [95% CI]	Peak–trough [95% CI]
Citywide (4 boroughs)	140,119 (8.9%)	+3.69 [+2.53, +4.91]	+3.69 (10Q3)	−2.28 [−2.77, −1.76]	−0.64 [−3.16, +1.16]
Outer boroughs (Bx, Bk, Qns)	138,283 (9.0%)	+3.77 [+2.48, +5.10]	+3.77 (10Q3)	−2.28 [−2.81, −1.77]	−0.50 [−2.99, +1.07]
Bronx	18,925 (11.0%)	+10.92 [+6.28, +15.15]	+10.92 (10Q3)	−0.83 [−2.16, +0.69]	−2.43 [−10.44, +0.69]
Queens	73,010 (10.4%)	+4.01 [+2.39, +5.66]	+4.01 (10Q3)	−2.79 [−3.49, −2.14]	−0.15 [−3.06, +2.70]
Brooklyn	46,348 (6.0%)	+0.23 [−1.80, +2.15]	+0.98 (10Q1)	−2.13 [−3.23, −1.17]	+0.06 [−2.66, +1.49]
Bottom price tertile ZIPs	45,545 (14.3%)	+10.34 [+7.21, +13.46]	+10.34 (10Q3)	−1.37 [−2.85, +0.03]	−2.84 [−7.27, −1.23]
Middle price tertile ZIPs	46,693 (9.5%)	+5.16 [+2.83, +7.69]	+5.16 (10Q3)	−2.45 [−3.27, −1.66]	−1.32 [−3.37, +0.67]
Top price tertile ZIPs	47,880 (3.2%)	+0.10 [−0.73, +1.14]	+1.02 (09Q1)	−0.85 [−1.13, −0.55]	+0.97 [−1.01, +1.52]
Top-quartile foreclosure ZIPs	34,961 (17.0%)	+10.70 [+7.57, +14.17]	+10.70 (10Q3)	−2.71 [−4.34, −1.24]	−5.89 [−11.67, −1.38]

Notes: Exact `D0` BMN estimator with 999 paired Poisson(1) parcel-bootstrap draws (fixed classifier labels). Price tertiles rank 160 ZIPs (140,118 pairs; one single-pair ZIP lacks pre-2007 sales) by 2003–2006 ordinary house median price into equal-pair thirds; Manhattan (`1,836` pairs, 3.4% dropped) is omitted from the table due to thin quarterly house counts (2010Q3 −0.43 [−2.02, +0.89]).

4.3 Statutory comparisons in Philadelphia and New Jersey

Whether statutory consideration inflates crisis levels or deepens crashes depends on local tax rules. Philadelphia publishes transfers with explicit document types (City of Philadelphia, Department of Records 2026). Applying a condensed lender-name classifier (`analysis/philadelphia_check.py`) flags 25,400 sheriff's deeds to lenders, of which 43% record at least $10,000, sitting 10 log points below city fair market value and 30 log points below the parcel's next sale. On a 2003–2016 repeat-sales index (`analysis/philadelphia_index.py`, 44,562 pairs with 2,515 lender-acquisition pairs), retaining those deeds pulls the raw index down by 7.0 log points over 2008–10 (95% CI [5.5, 8.7]) and 5.8 over 2011–12 ([4.3, 7.5]), deepening the peak-to-trough decline from 36.1 to 46.3 log points (+10.2 difference, CI [6.3, 13.0]).

New Jersey provides direct assessor transaction codes across 2003–2009 statewide Mod IV assessment rolls (Rutgers University, Bloustein School and New Jersey Agricultural Experiment Station 2026). There, 29% of 17,750 coded sheriff's sales and 80% of 1,514 deeds in lieu record at least $10,000, sitting 38 and 27 log points below ordinary municipal sales relative to assessed value. Across 32,166 repeat pairs (`analysis/nj_coded_deeds_statewide.py`, 794 coded endpoints), keeping coded pairs increases the 2006Q1–2009Q2 crash from 36.5 to 39.6 log points (+3.1 difference, CI [1.5, 4.7]). Across New York, Philadelphia and New Jersey alike, legal-rule amounts clear conventional $10,000 price floors while reflecting local consideration bases, suggesting that the sign of index distortion in other states could vary sharply with local transfer-tax statutes alongside statutory foreclosure rules that Mian et al. (2015) show roughly double foreclosure propensity in non-judicial states relative to judicial states.

5. Filtering protocol, validation and classification uncertainty

A reproducible screen for registry-based price work rests on five practices implemented in `acris_consideration_filter.py` with standard-library code and optional pandas vectorization, while a `broad` mode adds generic finance terms that can also capture investors and relocation firms.

1. **Classify deeds by parties rather than document type**, joining ACRIS parties onto `DEED` rows.
2. **Flag potential legal-rule acquisitions** while keeping lender resales as potentially arm's-length market trades and separating third-party auction bids.
3. **Drop flagged pairs under fixed pairing**, reporting bridged foreclosures only as a population-changing sensitivity.
4. **Audit name matches on the target registry**, inspecting high-frequency matches, person-guard rejections, unmatched grantees and non-zero cents among referee transfers.
5. **Do not rely on price-plausibility ratios**, because statutory amounts mimic assessed values and prior sales.

Preferring explicit party-role rules over complex price or machine-learning predictors guards against structural breaks across housing cycles, since Dunn et al. (2025) show that flexible models winning pooled cross-validation can break down during volatile periods relative to transparent benchmarks.

Independent non-name fields corroborate the screen deterministically. Across 542 deeds whose free-text ACRIS remarks describe a foreclosure sale, the party-role rule flags 428 (79.0%) as foreclosure-related transfers—207 (38.2%) as legal-rule lender take-backs and 221 (40.8%) as third-party foreclosure auctions—leaving 111 (20.5%) in ordinary transfers and 3 in REO resales (largely where filers omit "Referee" from the structured grantor field, which appears on 60.7% of these records). Among 21 deeds whose remarks specify a "deed in lieu", the rule flags 12 (57.1%) as legal-rule lender acquisitions and 1 as a third-party auction, missing 8 private entities or individuals. Flagged deeds also cross-reference a prior mortgage in 7.3% of non-referee lender transfers and 4.2% of referee transfers (plus 2.0% of third-party auctions), versus 0.03% of ordinary deeds.

5.1 Human-reference validation

Because remarks form a convenience sample, I validate accuracy on a stratified probability sample of 400 deeds drawn from the 411,113 screened sales entering repeat pairs. The sample allocates 150 deeds to `S1` flagged, 50 to `S2` third-party auctions, 100 to `S3` suspicious unflagged (identified via cents and nominal falls or broad matches), and 100 to `S4` other unflagged, with inclusion probabilities fixed ex ante. Forty randomly drawn deeds were read blindly against original ACRIS deed images by an independent reader (Matt Ma), and all 40 calls matched my subsequent verification.

Table 6. Human-reference sample and classification targets

Classifier stratum	Population deeds	Human readings	Non-market calls	Foreclosure-target calls
S1 flagged	9,091	11	9	9
S2 third-party auction	4,096	4	3	3
S3 suspicious unflagged	6,187	14	1	0
S4 other unflagged	391,739	11	0	0

Notes: Random 40-deed subsample weighted by population-to-sample ratios. Results condition on human calls, and the 11 zero-miss `S4` readings carry wide finite-population uncertainty over rare false negatives.

Against this human reference (Table 6), weighted precision is 0.82, foreclosure recall is 0.71, and all-non-market recall is 0.68. Inverting exact hypergeometric tails within each stratum under a Bonferroni split ($\alpha = 0.05$) and monotone ratio bounds (`validation_inference.py`) gives conservative joint 95% finite-population intervals of [0.40, 0.99] for precision and [0.02, 0.95] for foreclosure recall. Two of 11 flagged deeds were negotiated bank purchases (indistinguishable by name alone), while three of four unflagged non-market deeds were referee's deeds to lender-side parties (plus one nominal transfer outside the foreclosure target). This pattern echoes Bailey et al. (2020), who find that automated record-linking rules introduce systematic errors relative to trained human reviewers (15 to 37 percent in historical linking) that correlate with baseline characteristics. At $n = 40$, human evidence exposes real error mechanisms while leaving population recall intervals wide, so substantive false-negative shares in `S4` remain plausible. Exact hypergeometric power calculations (`power_curve_360_audit.py`) show why `S4`, holding 391,739 of 411,113 population deeds (95.3%), drives that

width. Reading only the remaining 89 pre-sampled S4 deeds (audit_batch/s4_blind_batch_89.csv, $n = 129$) lifts the 95% joint foreclosure-recall lower bound from 0.02 to 0.13 (0.10 if one miss occurs), while completing all 400 audit deeds narrows joint precision to [0.73, 0.89] and recall to [0.22, 0.77] if zero S4 misses occur.

5.2 Classification uncertainty in the index

To move beyond fixed-label bootstrap bands, Table 7 propagates human-sample error rates alongside parcel sampling in classification_uncertainty.py, updating independent $\mathrm{Beta}(1/2, 1/2)$ stratum posteriors on human counts, assigning statuses across unread deeds while holding each deed consistent across adjacent pairs, and applying paired Poisson(1) parcel weights in the unchanged BMN estimator.

Allowing classification rates to vary across all four strata under the uninformative Beta(1/2, 1/2) baseline shifts 2010Q3 to +2.25 [−0.21, +4.53] (spanning zero), while the 2011–12 mean stays negative at −1.82 [−3.04, −0.56] except under uniform priors or when rates rely solely on the 18 human-read deeds entering the house index. Panel B of Table 7 diagnoses *why* the baseline 2010Q3 interval crosses zero (scripts/s4_prior_sensitivity_v1_0_18.py; scripts/entity_and_submarket_v1_0_19.py). With 0/11 misses in S4 ($N_{S4} = 391,739$; 227,576 in the house index), the Beta(0.5, 11.5) posterior has mean 4.17% (median 2.00%, 95% interval [0.004%, 20.00%]), placing a posterior mean of 16,322 unflagged non-market deeds inside S4 (median 7,839; 95% interval [17, 78,346])—0.98× all 16,655 flagged legal-rule deeds in the entire 1.47-million-deed registry, and 4.70× at the 97.5th percentile (under Beta(1, 1), mean 30,134 and 97.5th percentile 103,673).

Table 7. Classification and parcel uncertainty, raw minus screened house index, log points

Specification	2010Q3	Mean 2011–12
Panel A: Baseline classification uncertainty		
Fixed labels, parcel bootstrap	+3.71 [+2.53, +4.91]	−2.26 [−2.77, −1.76]
Fixed error shares, parcel bootstrap	+2.36 [+0.98, +3.81]	−1.89 [−2.52, −1.34]
Uncertain rates, S4 fixed at zero	+2.34 [+0.63, +3.98]	−1.84 [−2.61, −1.04]
Uncertain rates, all four strata (untruncated Beta(1/2, 1/2))	+2.25 [−0.21, +4.53]	−1.82 [−3.04, −0.56]
Uniform-prior sensitivity (Beta(1, 1))	+2.18 [−0.54, +4.85]	−1.73 [−3.31, +0.02]
Foreclosure target only	+2.31 [+0.04, +4.55]	−1.81 [−3.07, −0.54]
Rates from house-index deeds only	+2.58 [−1.05, +5.98]	−1.96 [−4.01, +0.35]
Panel B: Pure registry-structure entity bounds and S4 prior sensitivity		
S4 entity grantees (Beta(1/2, 1/2), $p_{\text{ind}} = 0$; 20.0% house)	+2.11 [−0.01, +3.87]	−2.05 [−3.57, −1.13]
S4 entity (Beta(1/2, 1/2)) + $p_{\text{ind}} \leq 0.1\%$	+2.11 [+0.00, +3.87]	−2.04 [−3.57, −1.14]
S4 unfinanced entity grantees (Beta(1/2, 1/2); 11.8% house)	+2.13 [+0.15, +3.86]	−1.99 [−3.12, −1.10]
S4 conditional 0/1 entity read, $p_{S4,\text{ent}} \leq p_{S3}$	+2.13 [+0.11, +3.90]	−2.01 [−3.21, −1.14]
S4 truncated Jeffreys ($p_{S4} \leq 2.0\%$, mean 2,387 deeds)	+2.35 [+0.56, +4.09]	−1.84 [−2.67, −0.95]
S4 weak Beta(0.5, 22.1) (S1 2.21% base rate)	+2.32 [+0.40, +4.19]	−1.83 [−2.74, −0.85]
S4 crisis-only 2007–12 (untruncated Jeffreys)	+2.24 [+0.25, +4.10]	−1.70 [−2.57, −0.49]
Fixed $p_{S4} = 4.5\%$ / 5.0% (tipping point $p^*_{S4} = 4.76\%$)	+2.28 [+0.07, +4.54] / +2.27 [−0.07, +4.58]	−1.82 / −1.83

Notes: Medians and central 95% ranges come from 999 draws in Panel A rows 1–2, 9,999 in row 5, and 1,999 in all remaining rows (Panel B retains exact Jeffreys posteriors on S1–S3 and paired Poisson(1) parcel weights, reproducing Panel A baseline rows to $< 10^{-10}$). In Panel B rows 1–4, S4_entity covers corporate, LLC, trust, estate, and non-person grantees (20.02% of house-index S4; 11.82% also unfinanced). Ranges condition on human calls without modelling reader error.

Pure registry structure sharpens this bound without imposing an arbitrary numerical cap. Statutory debt-discharge lender take-backs convey title to institutional or corporate entities rather than natural-person homebuyers: deterministically, 99.96% of all repeat-pair `S1` deeds (9,087 of 9,091) and 99.98% of house-index `S1` deeds (8,325 of 8,327; 99.23% also unfinanced)—alongside 13 of 13 human-confirmed non-market deeds across `S1`–`S3`—name a corporate, LLC, trust, agency, or non-person grantee. Inside `S4`, by contrast, 79.08% of all repeat-pair deeds (309,801) and 79.98% of house-index deeds (182,014) name natural-person grantees (`"SURNAME, GIVEN"`), leaving only 20.92% overall (81,938) and 20.02% in the house index (45,562) in `S4_entity` (13.23% overall and 11.82% = 26,906 house-index deeds in unfinanced `S4_entity`; 1 of the 11 human-read `S4` deeds fell in `S4_entity` and 10 in `S4_indiv`, both with 0 non-market calls). Applying the untruncated `Beta(1/2, 1/2)` posterior strictly to unfinanced entity grantees (`S4_entity_unfin`, matching 99.2% of `S1`) yields a strictly positive 2010Q3 shift of +2.13 [+0.15, +3.86] ($P(>0) = 0.983$) and −1.99 [−3.12, −1.10] over 2011–12. Applying untruncated `Beta(1/2, 1/2)` across all `S4_entity` deeds sits at the zero boundary (+2.11 [−0.01, +3.87], $P(>0) = 0.973$; +2.11 [+0.00, +3.87] allowing $p_{\text{ind}} \leq 0.1\%$) because `S4_entity` contains distressed-home LLC investors whose crisis-discount purchases push the screened index up when randomly dropped. Conditioning strictly on the single 0/1 entity read inside `S4` (`Beta(0.5, 1.5)`) remains strictly positive under ordered-strata monotonicity $p_{S4,\text{ent}} \leq p_{S3}$ (+2.13 [+0.11, +3.90]), while dropping monotonicity widens 2010Q3 to +1.16 [−5.39, +3.60] (2011–12: −3.01 [−10.09, −1.32]).

Principled `S4` prior-sensitivity diagnostics reinforce this picture: truncated Jeffreys caps from $p_{S4} \leq 0.25\%$ (+2.33 [+0.62, +4.03]) through 2.0% (+2.35 [+0.56, +4.09]) and 5.0% (+2.32 [+0.44, +4.15], $P(>0) = 0.993$), weakly informative Beta priors anchored to 1.0% (+2.35 [+0.54, +4.11]) or the `S1` base rate (2.21%, +2.32 [+0.40, +4.19]), ordered strata $p_{S4} \leq p_{S3}$ (+2.31 [+0.25, +4.23]), and crisis-only 2007–12 `S4` misses (+2.24 [+0.25, +4.10]) all keep 2010Q3 strictly positive. Along a fixed-p_{S4} tipping-point grid, the 2010Q3 95% lower bound crosses zero only at $p^*_{S4} = 4.76\%$ (18,640 implied unflagged deeds in `S4`, or 1.12× all 16,655 flagged legal-rule deeds). Nonetheless, neither structural bounds nor prior diagnostics replace direct documentary evidence: removing prior assumptions completely still requires completing the ready-to-score +89 blind `S4` packet (`audit_batch/s4_blind_batch_89.csv` and `merge_completed_reads.py`). Taken together, documentary, digit, submarket and benchmark checks indicate that New York's consideration statutes inject non-price amounts into ACRIS and DOF sales files and shift crisis timing, aligning with Christensen and Hansen (2026), who show that two-step regressions using binary classifiers carry first-order bias proportional to error rates scaled by $\sqrt{n}$ and require external validation even when misclassification appears below one percent.

6. Conclusion

Standard price and parcel screens miss statutory lender-acquisition amounts because those amounts sit inside ordinary market ranges. In New York City, a party-based screen flags 16,655 potential legal-rule deeds, 62.2% survive full price and parcel screens, and 94–95% of screened 2016–25 filings pass straight into the city's annualized sales file at identical amounts without flags. When these transfers remain in a 1–3 family repeat-sales index under fixed labels, the citywide index shifts up by 3.69 log points in 2010Q3 and down by 2.28 in 2011–12 while leaving the citywide crash nearly unchanged, whereas an unpaired hedonic regression captures only the 2008–10 elevation. Disaggregating by geography and pre-2007 neighborhood price tier reveals far steeper localized distortions—+10.92 in the Bronx, +10.34 in bottom-tertile ZIPs and +10.70 in top-quartile foreclosure ZIPs (understating the local peak-to-trough drop by 5.89 log points), against +0.10 in top-tertile ZIPs. In Philadelphia and New Jersey, where statutes yield lower foreclosure amounts, comparable deeds clear $10,000 floors too but deepen measured crashes instead of shifting timing.

Two practical lessons follow for empirical housing work built on public registries. Researchers should inspect local transfer-tax consideration rules before trusting price floors or ratio checks, and screen transfers by party role where document codes collapse foreclosures and arm's-length sales into `DEED`—especially when studying lower-price or foreclosure-distressed submarkets. Classifier precision and recall also belong inside index inference: with 40 blind human reads (11 in `S4`), an uninformative `Beta(1/2, 1/2)` prior puts ~16,322 unflagged deeds into `S4` on average and leaves citywide 2010Q3 spanning zero (+2.25 [−0.21, +4.53]). Pure registry structure eliminates 80% of house-index `S4` deeds as natural-person buyers and keeps 2010Q3 strictly

positive on unfinanced entity grantees (+2.13 [+0.15, +3.86]) and under ordered-strata monotonicity (+2.13 [+0.11, +3.90]), while fixed-prevalence bounds cross zero only above 4.76% S4 misses; yet removing prior dependence altogether still calls for reading the remaining 89 pre-sampled S4 deeds.

Appendix A. Classifier rule (module version 1.1.0)

Names are upper-cased, & is replaced by AND, periods are removed and remaining punctuation is replaced by spaces. A name matches as a *lender* if it falls into any of four token families. The agency family covers FANNIE MAE, FEDERAL NATIONAL MORTGAGE, FREDDIE MAC, FEDERAL HOME LOAN MORTGAGE, GOVERNMENT NATIONAL MORTGAGE, SECRETARY OF (THE DEPARTMENT OF) HOUSING/HUD, HOUSING AND URBAN DEVELOPMENT, SECRETARY OF VETERANS, VETERANS AFFAIRS, and FEDERAL DEPOSIT INSURANCE. The bank and servicer family matches BANK, BANCORP, BANKING, SAVINGS, NATIONAL ASSOCIATION, FSB, CREDIT UNION, MORTGAGE, MTG, SERVICING, SERVICER, HOME LOAN(S), LOAN SERVIC..., LOAN TRUST, SAVINGS/INVESTMENT/BUILDING AND LOAN, LENDING, and LOAN(S) followed by a corporate suffix. Securitization tokens comprise CERTIFICATEHOLDERS, PASS-THROUGH, ASSET-BACKED, LEGAL TITLE TRUST, PARTICIPATION TRUST, REMIC, SECURITIZATION, and REO. The fourth family checks a fixed list of named servicers and REO subsidiaries (including Nationstar, Ocwen, Selene Finance, HomeSales, Liquidation Properties, Property Asset Management, Retained Realty and Residential Funding).

For the person guard, a match is rejected when formatted "SURNAME, GIVEN", the one-word surname equals the matched token, the given-name segment has ≤ 4 words and no digits, and no corporate marker (INC, LLC, CORP, TRUST, NA, SUCCESSOR, MERGER, …) appears. Referee tokens include REFEREE, REFERE, REFREE, AS REF and SHERIFF, while lien-holders (NYCTL, TAX LIEN, BOARD OF MANAGERS) count as lenders only on a referee's deed. Letting g indicate a grantee lender or referee lien-holder, r a grantor lender and f a grantor referee, events partition into six mutually exclusive categories. Specifically, $f \wedge g$ codes a referee's deed to a lender, $f \wedge \neg g$ codes a third-party auction, $\neg f \wedge g \wedge \neg r$ codes other transfers to a lender, $\neg f \wedge g \wedge r$ codes lender-to-lender transfers, $\neg f \wedge \neg g \wedge r$ codes lender resales, and remaining pairs are ordinary transfers.

Appendix B. Replication and companion full paper

This measurement note condenses version 1.0.15 of the full working paper (doi:10.5281/zenodo.24119841), which also provides two-leg plots, Case–Shiller/FHFA comparisons, momentum estimates and strict two-endpoint appendices. Baseline retrieval scripts, acris_consideration_filter.py and the frozen audit sample reside in version 1.0.11 (doi:10.5281/zenodo.23414919), with hedonic, uncertainty and human-reference scripts archived across versions 1.0.14–1.0.15 (10.5281/zenodo.23761413 and 10.5281/zenodo.24119841). analysis/philadelphia_check.py, analysis/philadelphia_index.py, analysis/nj_coded_deeds_statewide.py, validation_inference.py, classification_uncertainty.py, power_curve_360_audit.py, scripts/s4_prior_sensitivity_v1_0_18.py, scripts/entity_and_submarket_v1_0_19.py and audit_batch/merge_completed_reads.py reproduce the cross-jurisdiction, submarket, power, entity-bound and uncertainty tables.

Declarations

Data availability. Administrative records originate from NYC Open Data, OpenDataPhilly, NJ MOD IV rolls and FRED as archived in Zenodo deposits 10.5281/zenodo.23414919, 10.5281/zenodo.23761413 and 10.5281/zenodo.24119841. Human-reference reading files are available on request.

Funding and competing interests. This research received no external grant funding. The author founded HolaCasaClara, a free comparator of Argentine mortgage loans unrelated to U.S. housing records, and declares no other competing interests.

Generative AI disclosure. All research design, analytical decisions, empirical validation and conclusions are the author's own. During manuscript and code preparation, generative AI tools (OpenAI ChatGPT/Codex, Anthropic Claude and Google Gemini) assisted with copyediting, software engineering and replication checks. All Section 5 validation relies strictly on deterministic registry cross-checks and 40 blind human-reference readings of original ACRIS deed images by Matt Ma verified by the author; no AI document readings enter any estimate, bound or table in this note.

References

Anenberg, Elliot, and Edward Kung. 2014. "Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures." *American Economic Review* 104 (8): 2527–51. https://doi.org/10.1257/aer.104.8.2527.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. 2020. "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature* 58 (4): 997–1044. https://doi.org/10.1257/jel.20191526.

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse. 1963. "A Regression Method for Real Estate Price Index Construction." *Journal of the American Statistical Association* 58 (304): 933–42. https://doi.org/10.1080/01621459.1963.10480679.

Ben-David, Itzhak. 2011. "Financial Constraints and Inflated Home Prices during the Real Estate Boom." *American Economic Journal: Applied Economics* 3 (3): 55–87. https://doi.org/10.1257/app.3.3.55.

Campbell, John Y., Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices." *American Economic Review* 101 (5): 2108–31. https://doi.org/10.1257/aer.101.5.2108.

Case, Karl E., and Robert J. Shiller. 1987. "Prices of Single-Family Homes since 1970: New Indexes for Four Cities." *New England Economic Review*, 45–56.

Christensen, Timothy, and Stephen Hansen. 2026. "Performing Valid Inference with AI/ML-Generated Covariates: A Guide for Empirical Practice." *AEA Papers and Proceedings* 116: 92–97. https://doi.org/10.1257/pandp.20261020.

City of New York. 2024. *New York City Administrative Code §§ 11-2101 to 11-2117 and Rules of the City of New York (19 RCNY § 23-03)*.

City of Philadelphia, Department of Records. 2026. *Real Estate Transfers (Rtt_summary)*. OpenDataPhilly, retrieved 16 October 2026. https://opendataphilly.org/datasets/real-estate-transfers/.

Clauretie, Terrence M., and Nasser Daneshvary. 2009. "Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time." *Real Estate Economics* 37 (1): 43–67. https://doi.org/10.1111/j.1540-6229.2009.00234.x.

Conklin, James N., N. Edward Coulson, Moussa Diop, and Nuno Mota. 2023. "An Alternative Approach to Estimating Foreclosure and Short Sale Discounts." *Journal of Urban Economics* 134: 103546. https://doi.org/10.1016/j.jue.2023.103546.

Cordell, Larry, Liang Geng, Laurie S. Goodman, and Lidan Yang. 2015. "The Cost of Foreclosure Delay." *Real Estate Economics* 43 (4): 916–56. https://doi.org/10.1111/1540-6229.12107.

Doerner, William M., and Andrew Leventis. 2015. "Distressed Sales and the FHFA House Price Index." *Journal of Housing Research* 24 (2): 127–46. https://doi.org/10.1080/10835547.2015.12092100.

Dunn, Abe, Eric English, Kyle Hood, Lowell Mason, and Brian Quistorff. 2025. "Economic Measurement Lost in a Random Forest? A Case Study of Employment Data." *AEA Papers and Proceedings* 115: 68–72. https://doi.org/10.1257/pandp.20251103.

Ellen, Ingrid Gould, Amy Ellen Schwartz, Ioan Voicu, and Michael H. Schill. 2007. "Does Federally Subsidized Rental Housing Depress Neighborhood Property Values?" *Journal of Policy Analysis and Management* 26 (2): 257–80. https://doi.org/10.1002/pam.20247.

Feng, Shuaizhang, and Yingyao Hu. 2013. "Misclassification Errors and the Underestimation of the US Unemployment Rate." *American Economic Review* 103 (2): 1054–70. https://doi.org/10.1257/aer.103.2.1054.

Fisher, Lynn M., Lauren Lambie-Hanson, and Paul Willen. 2015. "The Role of Proximity in Foreclosure Externalities: Evidence from Condominiums." *American Economic Journal: Economic Policy* 7 (1): 119–40. https://doi.org/10.1257/pol.20130102.

Gerardi, Kristopher, Eric Rosenblatt, Paul S. Willen, and Vincent Yao. 2015. "Foreclosure Externalities: New Evidence." *Journal of Urban Economics* 87: 42–56. https://doi.org/10.1016/j.jue.2015.02.004.

Harding, John P., Eric Rosenblatt, and Vincent W. Yao. 2009. "The Contagion Effect of Foreclosed Properties." *Journal of Urban Economics* 66 (3): 164–78. https://doi.org/10.1016/j.jue.2009.07.003.

Kermani, Amir, and Francis Wong. 2026. “Racial Disparities in Housing Returns.” *American Economic Review* 116 (1): 287–331. https://doi.org/10.1257/aer.20240327.

Loschi, Pablo. 2026. *Who Gets the House? Related-Party Transfers and the Cash–Mortgage Price Gap in New York City*. Zenodo. https://doi.org/10.5281/zenodo.22421850.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2015. “Foreclosures, House Prices, and the Real Economy.” *Journal of Finance* 70 (6): 2587–634. https://doi.org/10.1111/jofi.12310.

New York State Department of Taxation and Finance. 2019. *Form TP-584.1: Real Estate Transfer Tax Return Supplemental Schedules, Schedule E and Instructions (Revision 7/19)*. https://www.tax.ny.gov/pdf/current_forms/property/tp584_1.pdf.

Nowak, Adam D., and Patrick S. Smith. 2020. “Quality-Adjusted House Price Indexes.” *American Economic Review: Insights* 2 (3): 339–56. https://doi.org/10.1257/aeri.20190337.

Pennington-Cross, Anthony. 2006. “The Value of Foreclosed Property.” *Journal of Real Estate Research* 28 (2): 193–214. https://doi.org/10.1080/10835547.2006.12091177.

Rutgers University, Bloustein School and New Jersey Agricultural Experiment Station. 2026. *N.j. MOD IV Historical Database*. Https://modiv.rutgers.edu/, retrieved 16 October 2026.

Schwartz, Amy Ellen, Scott Susin, and Ioan Voicu. 2003. “Has Falling Crime Driven New York City’s Real Estate Boom?” *Journal of Housing Research* 14 (1): 101–35. https://doi.org/10.1080/2167034x.2003.12461362.

S&P Dow Jones Indices. 2025. *S&P Cotality Case-Shiller Home Price Indices Methodology*. https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-cotality-cs-home-price-indices.pdf.