

## B Data Appendix

**New York City Department of Finance Annualized Rolling Sales.** The New York City Department of Finance (NYCDOF) Annualized Rolling Sales files contain details on real-property transactions for the five boroughs from 2003 to the present (we use the data through 2011). The data are realized by the NYCDOF on a quarterly basis and are derived from the universe of transfer-tax filings (which are mandatory for all residential and commercial sales). Geographic detail for each sale includes the street address (and zip code), the tax lot (borough-block-lot number), and the neighborhood (Chelsea, Tribeca, Upper West Side, etc.). The Rolling Sales files contain limited details about the properties themselves, including square footage, number of units (residential and commercial), tax class (residential, owned by utility co., or all other property), and building class category (a more detailed property code—for example, one-family homes, two-family homes, residential vacant land, walk-up condo, etc.). Transaction details in the data include the sale price and date. A sale price of \$0 indicates a transfer of ownership without cash consideration (ex. from parents to children).

New York City properties are subject to the mansion tax if they are single-, double-, or triple-family homes, or individual condo or co-op units. We define taxable sales as those transactions of a single residential unit (and no commercial units) with a building classification of “one family homes,” “two family homes,” “three family homes,” “tax class 1 condos,” “coops - walkup apartments,” “coops - elevator apartments,” “special condo billing lots/condo-rental,” “condos - walkup apartments,” “condos - elevator apartments,” “condos - 2–10 unit residential,” “condos - 2–10 unit with commercial use,” or “condo coops/condops.” We define co-ops as a building code of “coops - walkup apartments,” or “coops - elevator apartments.” We define a commercial sale to be a transaction with at least one commercial unit (and no residential units) or a tax class of 3 or 4.

**New York State Office of Real Property Service SalesWeb.** The New York State Office of Real Property Service (NYSORPS) publishes sales records for all real-property transactions (excluding New York City) recorded between 2002–2006 and 2008–2010 available through the “SalesWeb” database. Since deeds are recorded after the sale, this data includes a small number of sales from 2007. The database is compiled by ORPS from filings of the State of New York Property Transfer Report (form RP-5217).

The NYS deeds records indicate several details about each transaction and property. Transaction-specific details include the sale price and date, the date the deed was recorded (and recording details such as book and page number), the buyer’s, seller’s, and attorney’s name and address (often missing), the number of parcels included in the transaction, and details about the relationship between the buyer and the seller (whether the sale is between relatives, whether the buyer is also a seller, whether one party is a business or the government, etc.). Of particular interest to us is whether the sale is defined by the state as arms-length. The data dictionary defines an arms-length sale as “a sale of real property in the open market, between an informed and willing buyer and seller where neither is under any compulsion to participate in the transaction, unaffected by any unusual conditions indicating a reasonable possibility that the full sales price is not equal to the fair market value of the property assuming fee ownership”, which excludes sales between current or former relatives, related companies or partners in business, sales where one of the buyers is also a seller, or sales with “other unusual factors affecting sale price.” Property details include the square footage, assessed value (for property-tax purposes), address (including street address, county, zip code, school district), and the property class (one-family home, condo, etc.). We consider as subject to the mansion tax all single-unit sales with property class equal to one-, two-, or three-family residence, residential condo, or a seasonal residence.

**New Jersey Treasury SR1A File.** We make use of sales records from the New Jersey Treasury’s SR1A file for 1996–2011, which contains records of all SR1A forms filed at the time of sale (the form is mandatory in the state for all residential sales). Each record includes the sale price and date the deed was drawn, buyer and seller name and address (often missing), deed recording details (date submitted, date recorded, document number), and whether there are additional lots associated with the sale. Property details include land value, tax lot, square footage, and property class. We define taxable sales as those with a residential property class.

**New York City County Register Deeds Records.** These data are collected from the county registers for the five counties in New York City: Bronx County, Kings County, New York County, Queen’s County, and Richmond County. The records were collected by an anonymous private firm and made available to us by the Paul Millstein Center for Real Estate at the Columbia Graduate School of Business.

These data include additional detail as compared to the Rolling Sales files, although at the expense of precision.

Prices in this data set are rounded to the nearest \$100, which leads to misallocation of sales to one side of a tax notch. Transaction details include the sale price and date, an indicator for whether the unit is newly constructed, the number of parcels being sold, whether the purchase was made in cash (i.e. whether a mortgage is associated with the sale), and indicators for private lenders and within-family sales. Property details are limited to address, zip code, and county.

**Data Cleaning.** We begin by dropping all transactions with a price below \$100 (1,658,639 in NY State, 954,241 in NJ, and 274,118). The bulk of these transactions have a zero price, representing transfers of property between parties not associated with a proper sale (e.g., a gift or inheritance). This restriction is relatively innocuous, as our analysis focuses on sales around each tax notch (although this choice does affect the descriptive statistics). More importantly, we attempt to identify and discard all duplicate records. In New York State, we identify duplicates as sales that occur within 90 days of one another at the same street number in the same grid number (a unique tax lot id). Of these 48,073 duplicates, we always keep the later sale (in case duplicates are representative of updates to the records). For New Jersey, since we do not observe tax lots, we identify all duplicate sales that occur at the same standardized address within 90 days of one another and drop all but the final duplicate (343,221). Finally, for NYC we identify duplicates as properties in the same borough at the same standardized address that sell within 90 days of one another (20,420). While these duplicates represent a large number of sales, and there are several ways one could define duplicates, our estimates are insensitive to whether and how we clean duplicates (e.g., cleaning NY state based on address or NYC based on tax lot).

**Real Estate Board of New York Listings Service.** We have collected residential real-estate listings from the Real Estate Board of New York’s (REBNY) electronic listing service. REBNY is a trade association of about 300 realty firms operating in Manhattan and Brooklyn. REBNY accounts for about 50% of all residential real-estate listings in these boroughs. A condition of REBNY membership is that realtors are required to post all listings and updates to the listing service within 24 hours.

Using the REBNY listing service, we have collected all “closed” (i.e. sold) or “permanently off market” residential listings posted between 2003 (when the electronic listings are first available) and 2010. REBNY listings include the typical details available on a real-estate listing: asking price, address, date on the market and a description of the property. Additionally, we observe all updates to each listing (and the dates of each update), which lets us see how asking prices evolve and determine the length of time a property is on the market. Finally, we observe the final outcome of the listing: whether the property is sold or taken off the market.

We create several variables for each REBNY listing. We define the initial asking price as the first posted price on the listing, and the final asking price as the last posted price while the listing is “active.” We identify the length of time that a listing spends on the market as the number of days between the initial posting and the date that the listing is updated as “in contract.” We define the discount between two prices as the percent drop in price  $-\frac{p_0 - p_1}{p_0}$ , where  $p_0$  and  $p_1$  are prices and  $p_0$  is posted before  $p_1$ .

One caveat to the REBNY listings is that the price is often not updated at the time of sale. To overcome this, we match REBNY listings to the NYCDOF data by address and date. Of the 48,220 closed REBNY listings for Manhattan, we achieve a match rate of 92%. Non-matches fall in a number of categories. Sales in some condop buildings are missing from the DOF data due to a clerical error at the NYC DOF. Some transactions contain only street address or a non-standard way of specifying the apartment number (in particular, commercial units and unusual properties such as storage units fall in this category). Occasionally, the same building may have two different street addresses and a unit may be listed differently in the two databases. At the same time, of the 23,655 Manhattan listings that are not reported as closed in the REBNY listings database, we find 7,425 corresponding sales in the NYCDOF data. We treat such matches as an indication that the property was sold without the REBNY realtor (either sold by the owner or using another realtor).

## C Robustness of incidence estimates

Table A.1 demonstrates that our estimates are quite robust to variety of estimation approaches. Incidence estimates are very consistent, and gap estimates vary somewhat but remain positive and large in most specification checks that

we consider. Intuitively, there are good reasons for why results may vary as one adjusts the order of polynomials and the omitted region. Both incidence and gap estimated using cross-sectional data (the only exception to it our estimates for NJ that rely on pre/post comparison) involve prediction out of sample (into the omitted region). As the size of the omitted region increases, one has to predict far out of sample so that the “forecast” error is bound to increase. Furthermore, very flexible polynomials that can fit data in sample well are not restricted in their behavior in the omitted region and in some cases may generate non-monotonicity or explosive behavior within the omitted region — overfitting is not the right approach for predicting out of sample. On the other hand, the omitted region that is too small generates bias in the estimates of the counterfactual. Nevertheless, our results are robust to reasonable modifications of our baseline specification as discussed below.

While our preferred specification uses a third-order polynomial, our incidence estimates are not too sensitive to this choice. The second through fifth rows of Table A.1 present estimates that we obtain using different orders—the results are similar, although inspection of the fit of the data suggests that very low-order polynomials cannot capture properly the shape of the distribution, while very high-order polynomials (not reported) introduce very unrealistic behavior in the omitted region. As the result, there is a bit of sensitivity to the order of polynomials in the gap estimates, which are positive and significant for all specifications up to the fifth order polynomial, but shrink somewhat for higher orders.

The results are only somewhat sensitive to selecting a narrow omitted region. The estimates in the sixth through eighth rows of Table A.1 illustrate that a smaller omitted region leads to smaller incidence estimate (\$3000 to \$5000 less than the baseline). We do not estimate  $\hat{Z}$  for the narrowest specifications since it does not make sense to restrict the gap to be so small, especially given the visual evidence of the width of the gap. Relatedly, the estimate of  $\hat{Z}$  using the omitted regions through \$1.1M are smaller than the baseline — understandable, given the visual evidence in Figures 1, 2 and 3 indicating that the gap extends further than that (consistently with the theoretical argument as well), and the fact that the counterfactual is bound to be biased downward when part of the “true” gap is relied on in estimation.

On the other hand, our results are robust to extending the omitted region beyond the baseline, as is seen in Rows 10 through 12. The estimates of  $\hat{Z}$  are consistently large, positive, and significant, although less precise as we use less and less data (and need to predict the counterfactual over a larger range). Reassuringly, the incidence estimates change little as we vary the upper bound of the omitted region. Similarly, none of the estimates are too sensitive to expanding the omitted region below the threshold. The results in rows twelve through fourteen of Table A.1 show that both the incidence and gap estimates grow as the bunching region is expanded below the threshold, however differences in estimates are economically small and not statistically distinguishable from the baseline. Naturally, the standard errors also grow as the omitted region is expanded.

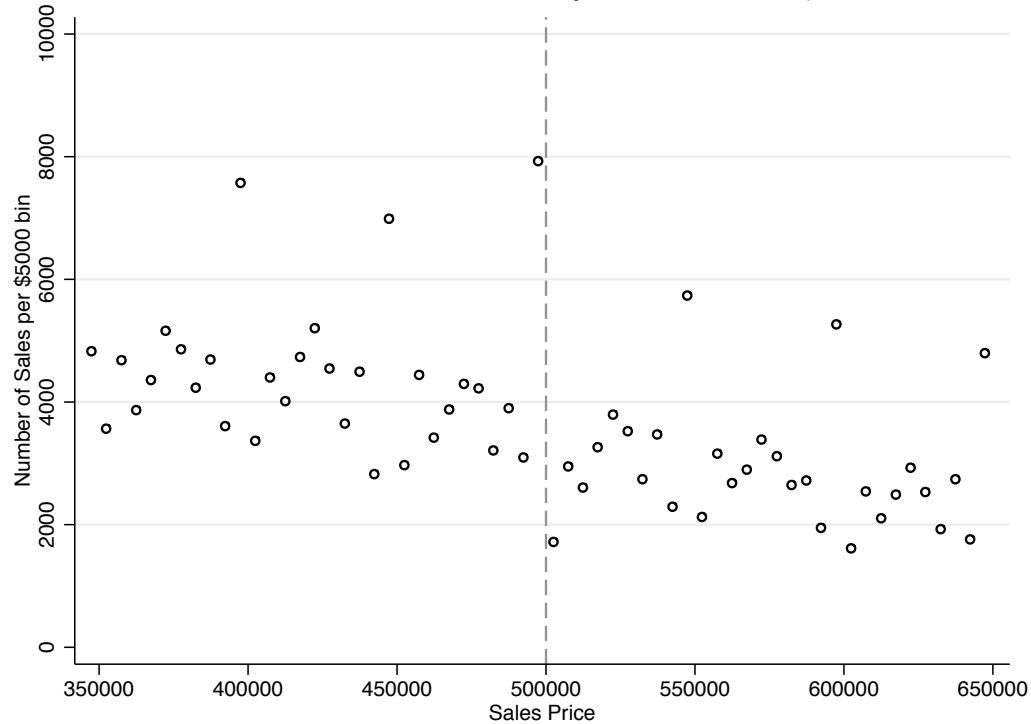
We also estimate our counterfactuals for bunching and gap separately using only data below and above the omitted region (respectively). We present in Row 16 our estimate using a 3rd order polynomial and data below the omitted region for the bunching/incidence counterfactual and a 1st order polynomial using data above the omitted region for the missing mass, and a 2nd order above the omitted region in Row 17. Again, incidence and  $\hat{Z}$  are comparable with our baseline. Furthermore, bootstrapped standard errors increase significantly as the order of polynomial increases, underscoring our earlier point that allowing for overfitting by estimating high order polynomials that are then used to project into the omitted region is a questionable approach. This observation (and visual inspection of the fit) justified our choice of the baseline specification that relies on the 3rd order polynomial and only a level shift at the threshold.

Our baseline estimate is also not sensitive to allowing for a discontinuity at the threshold. The baseline specification relies on the data both below and above the omitted region. Since the latter is distorted by the tax we rudimentarily control for it by allowing for a level shift in the distribution. The estimate in row 18 of Table A.1 demonstrates that incidence and gap increase slightly when we do not allow for this discontinuity.

For completeness, we also estimate analogous specification by OLS—this is the standard approach in the recent public finance work on notches and kinks—but we note that any of these methods involves specifying the parametric density function and the maximum likelihood estimation is a natural choice that guarantees that the estimates satisfy the law of probability rather than the hard-to-interpret mean zero residual restriction. Additionally, by requiring the data to be binned, OLS will throw out information. We report the OLS results obtained by binning into \$5000 and \$10,000 bins in rows 17 and 18 and conclude that they are quantitatively similar to the baseline.

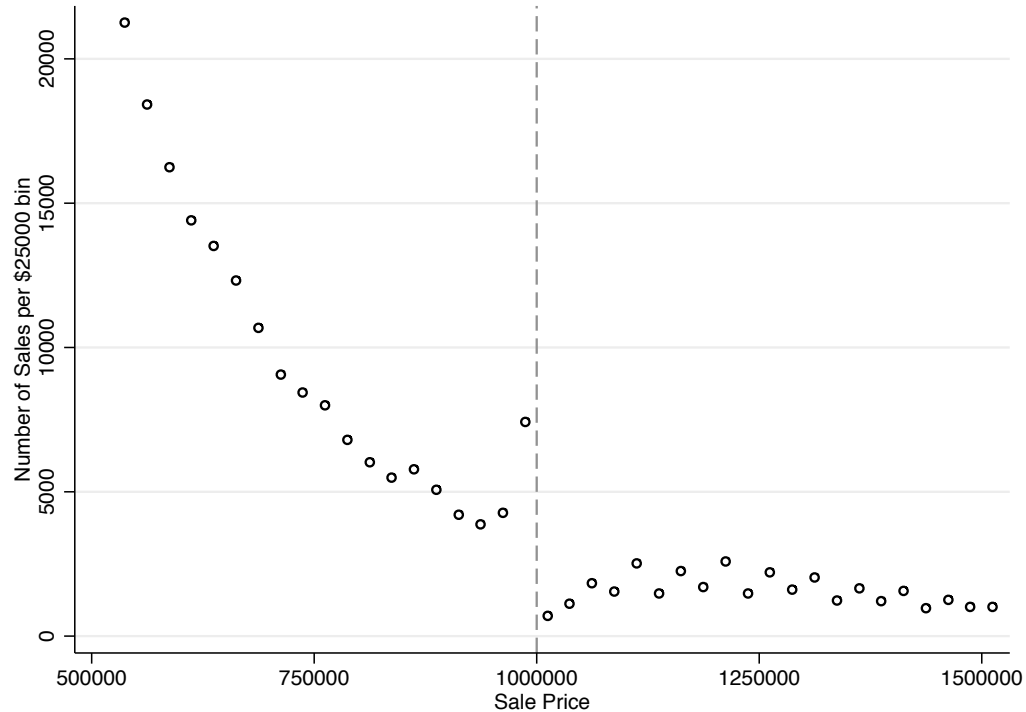
The placebo estimates in Table A.2 show that our estimates for NYC are not spurious. Using the same procedure, we estimate the incidence and gap for all commercial sales (which are not subject to the mansion tax) and for residential sales at other multiples of \$100,000. In all cases, we find small negative incidence estimates and relatively small gap estimates.

Figure A.1: Distribution of Sales in New York City around the \$500,000 RPTT tax notch



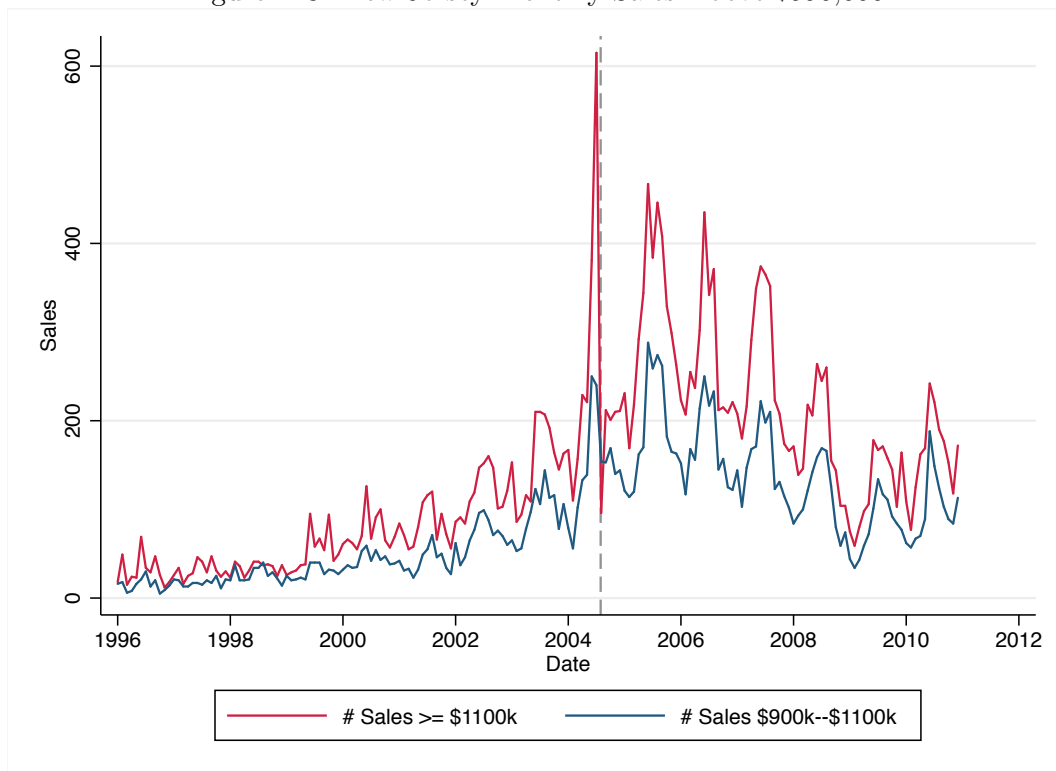
Notes: Plot of the number of sales in each \$5,000 price bin between \$350,000 and \$650,000. Data from the NYC Rolling Sales file for 2003–2011. Both commercial and non-commercial sales are subject to the NYC RPTT.

Figure A.2: Distribution of Taxable Sales in New York State



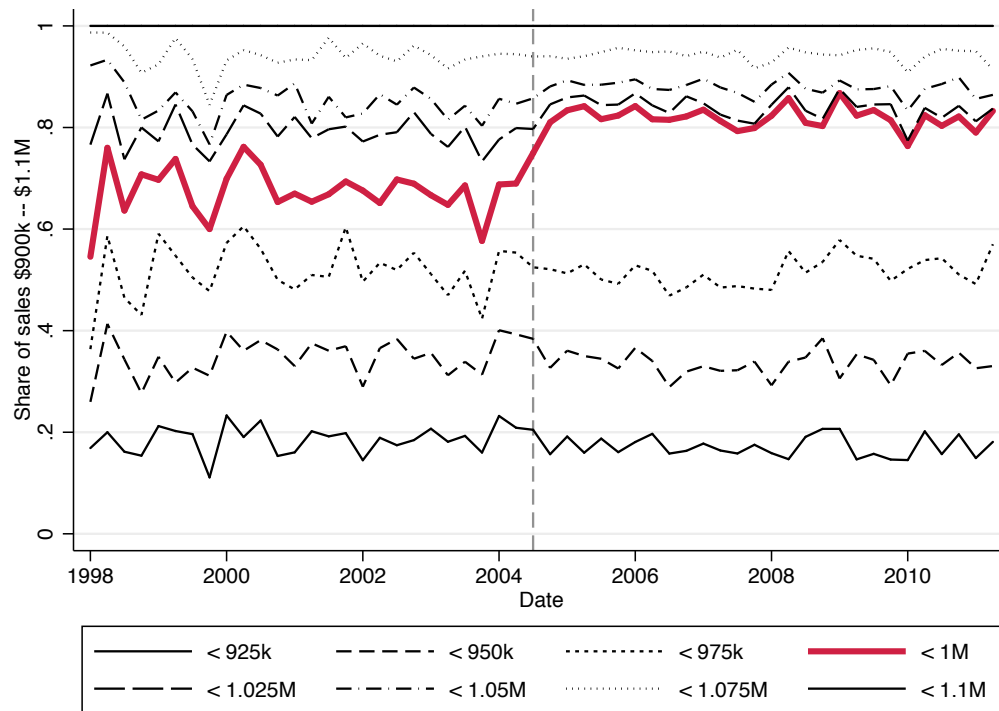
Notes: Plot of the number of mansion-tax eligible sales in each \$25,000 price bin between \$510,000 and \$1,500,000. Data from the NYC Rolling Sales file for 2003–2011 (taxable sales defined as single-unit non-commercial sales of one-, two-, or three-family homes, coops, and condos) and from N.Y. State Office of Real Property Service deeds records for 2002–2006 and 2008–2010 (taxable defined as all single-parcel residential sales of one-, two-, or three-family homes).

Figure A.3: New Jersey Monthly Sales Above \$990,000



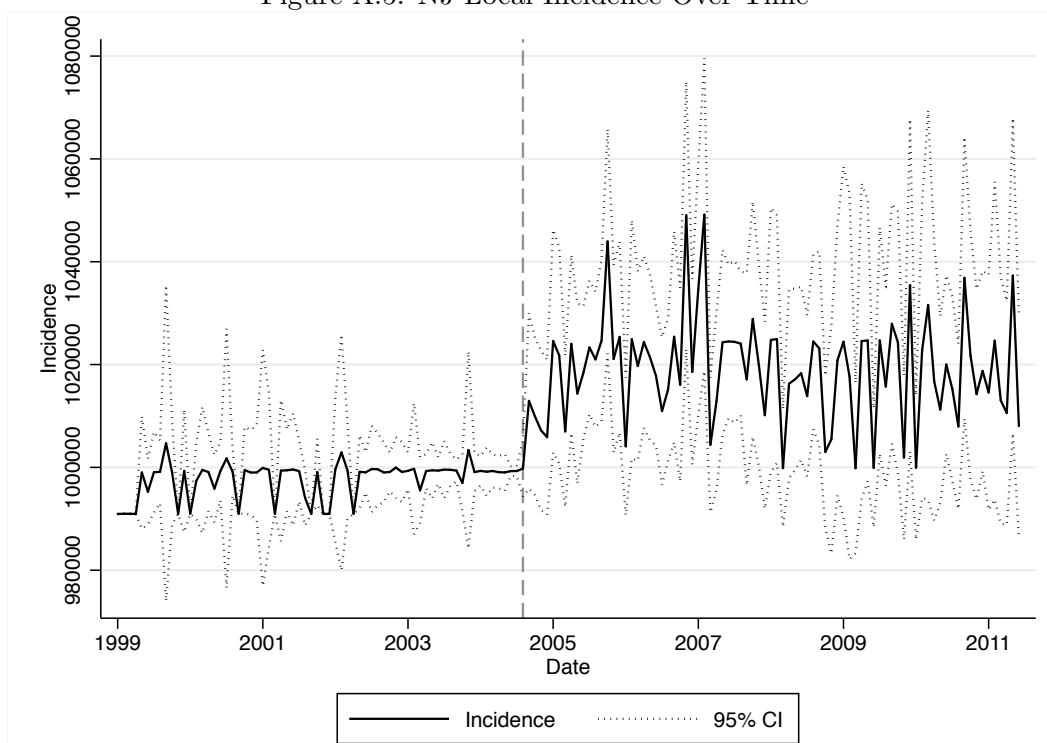
Notes: Total taxable NJ sales in given price range by month. Data from NJ Treasury SR1A file for 1996–2011 (taxable defined as any residential sale). Mansion tax introduced in August, 2004 (indicated by dashed gray line).

Figure A.4: Distribution of Monthly Sales in New Jersey (\$900k – \$1M)



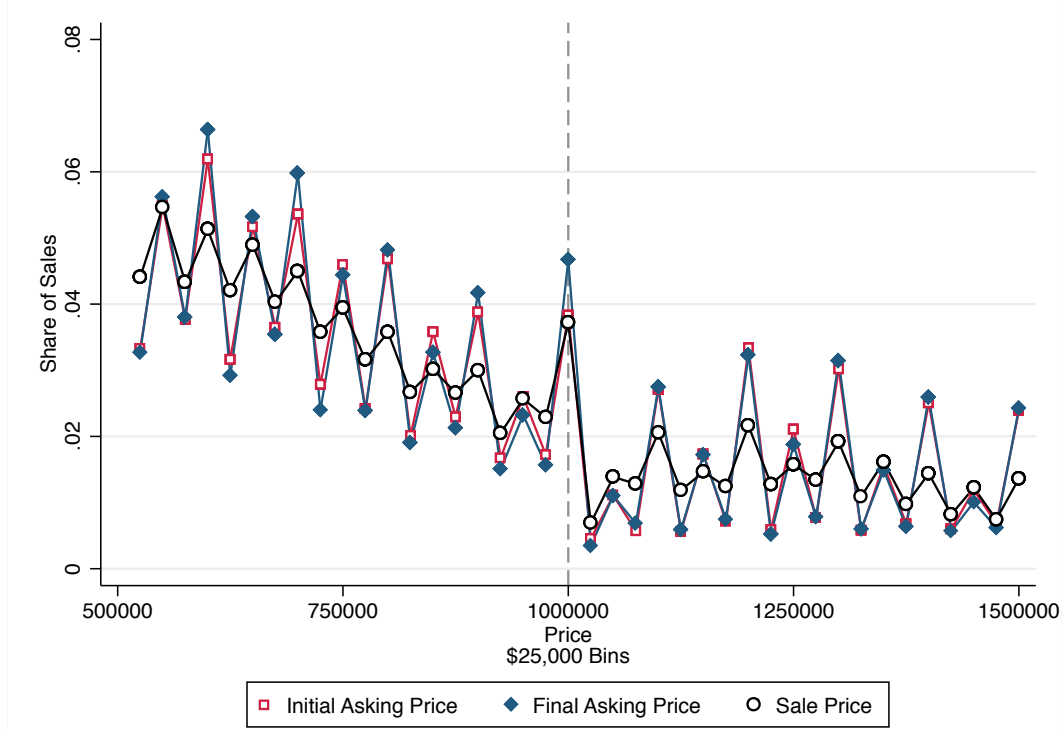
Notes: Number of taxable sales in given range as a share of total sales between \$900,000 and \$1,100,000 by month. Data from NJ Treasury SR1A file for 1998–2011 (taxable defined as any residential sale). Mansion tax introduced in August, 2004 (denoted by gray dashed line).

Figure A.5: NJ Local Incidence Over Time



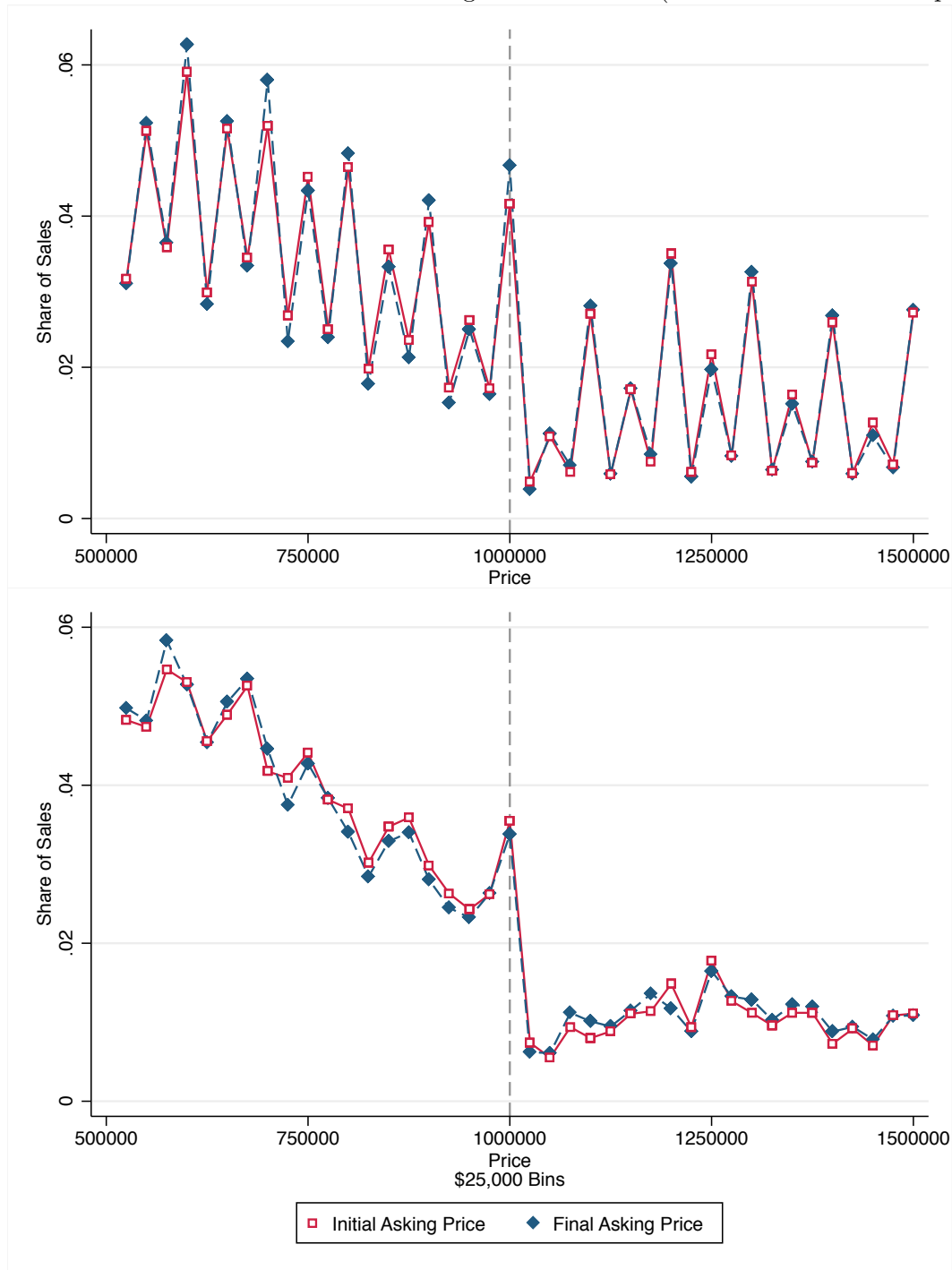
Notes: Monthly baseline local incidence estimates and 95% confidence intervals for NJ. Data from NJ Treasury SR1A file.

Figure A.6: Distribution of Real-Estate Listing Prices in NYC (Sold Properties Only)



Notes: Data from REBNY listings matched to NYC Department of Finance sales records. Sample restricted to “sold” listings: last listing status is “closed” and property can be matched to NYC sales data. Plot of the number of listings per \$25,000 bin as a share of all sales between \$500,000 and \$1,500,000 (bins centered so that the threshold bin spans \$975,001–\$1,000,000).

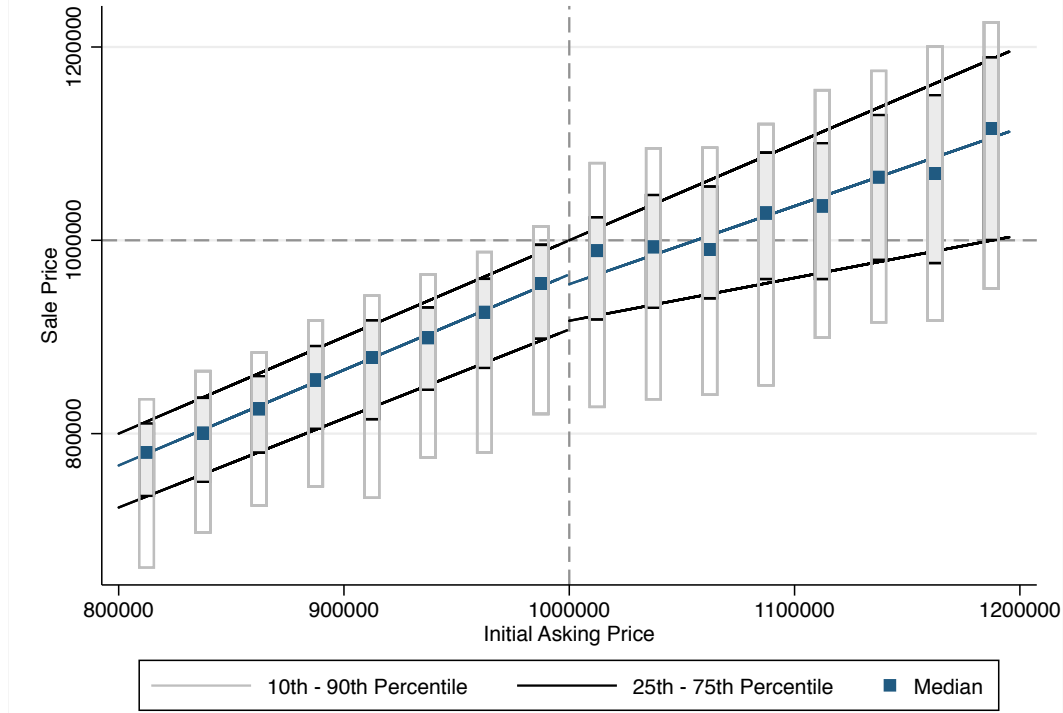
Figure A.7: Distribution of Real-Estate Listing Prices in NYC (All REBNY-Listed Properties)



Notes: Data from REBNY listings. Sample includes all REBNY-listed sales in the given range. Panel (a) presents a plot of the number of listings per \$25,000 bin as a share of all listings between \$500,000 and \$1,500,000 (bins centered so that the threshold bin spans \$975,001–\$1,000,000). Panel (b) presents a smoothed plot of the distribution that accounts for round-number bunching: the log of the per-bin counts from panel (a) are regressed on a cubic in price and dummy variables for multiples of \$50,000 and \$100,000 interacted with the price. Predicted bunching for round-number bins are then subtracted from the corresponding counts.

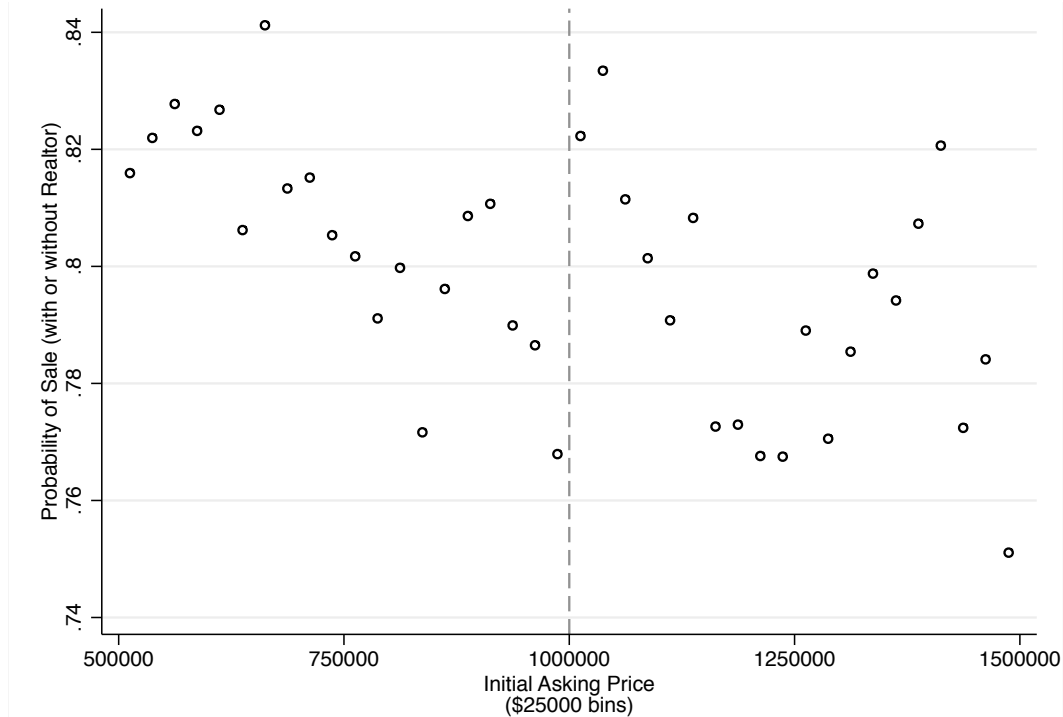


Figure A.8: Distribution of Sale Price by Initial Asking Price (with Quantile Regression)



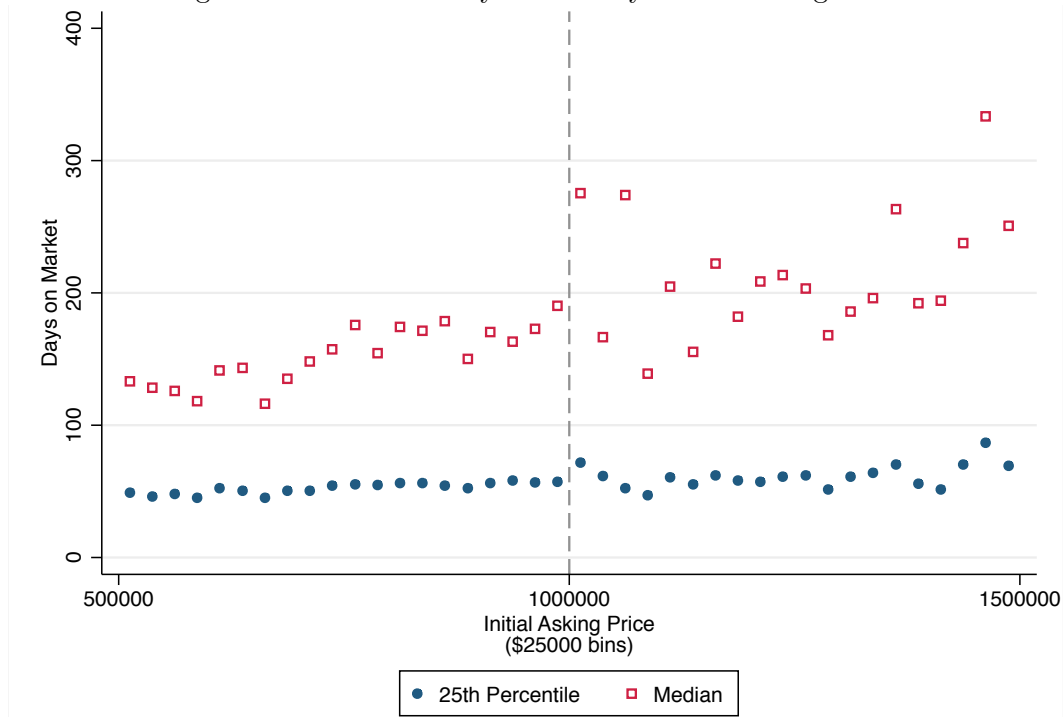
Notes: Plot of the median, 10th, 25th, 75th, and 90th percentiles of sale price per \$25,000 initial-asking-price bin. Data from REBNY listings—sample includes all sold REBNY-listed properties (matched to NYC DOF) in the range \$800,000–1,200,000. Lines represent quantile regressions for the given range (\$800k–\$990k and \$1M – \$1.2M).

Figure A.9: Probability that Listed Property Sells by Initial Asking Price



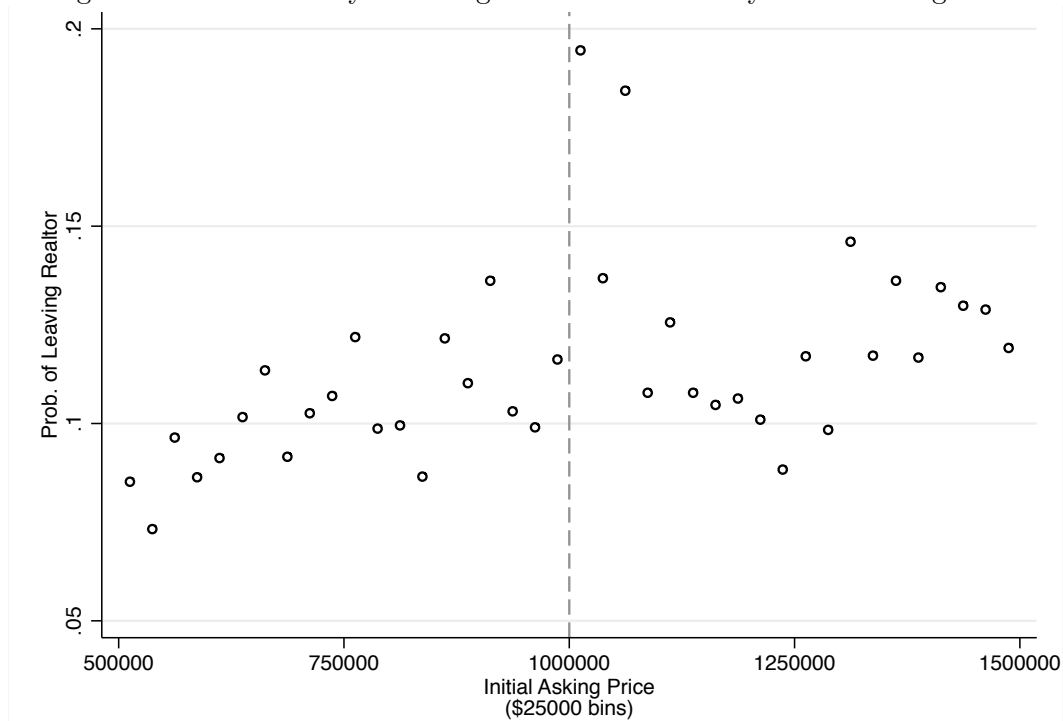
Notes: Plot of the share of REBNY-listed properties that close or are matched to a NYC DOF sale per \$25,000 bin. Data from REBNY listings—sample includes all listed properties in the range \$500,000–1,500,000. “Sold” defined as any property with a final listing status of “closed” or any listing that matches to NYC DOF sales.

Figure A.10: Median Days to Sale by Initial Asking Price

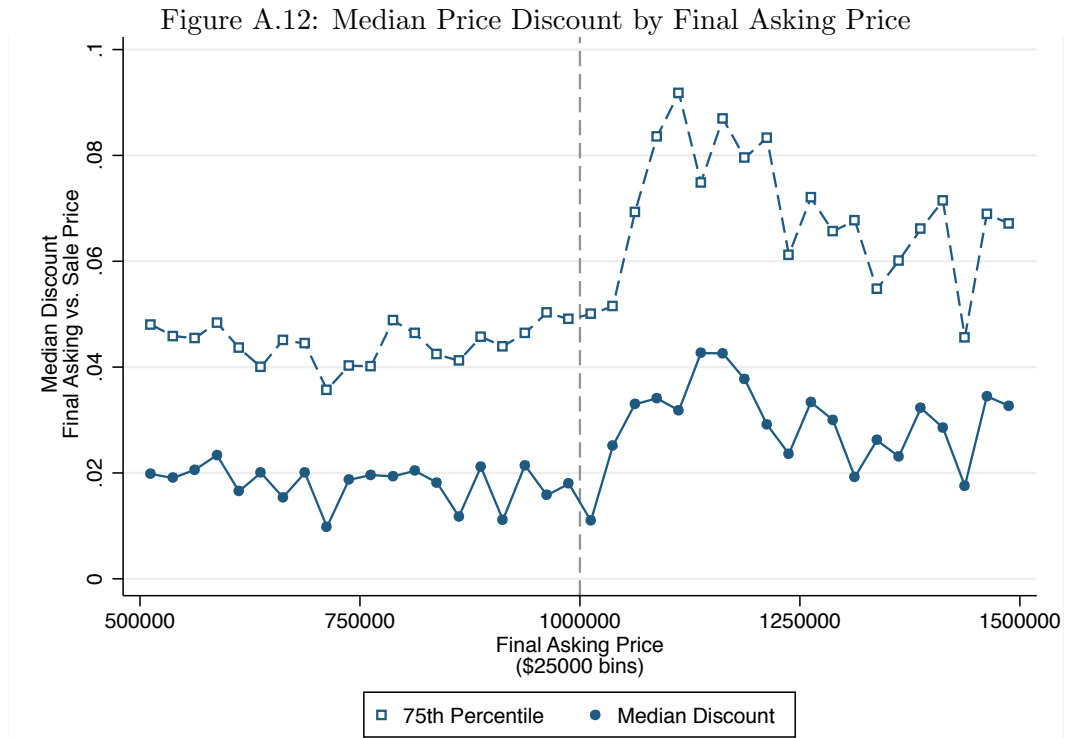


Notes: Plot of the median and 25th percentile of days to sale per \$25,000 initial-asking-price bin. Data from REBNY listings—sample includes all REBNY-listed properties in the range \$500,000–1,500,000. Days to sale defined as the number of days between initial listing of the property and buyer and seller entering into contract (defined as final status = “in contract”). Unsold properties are assigned a value of 999 days.

Figure A.11: Probability of Selling Without REBNY by Initial Asking Price

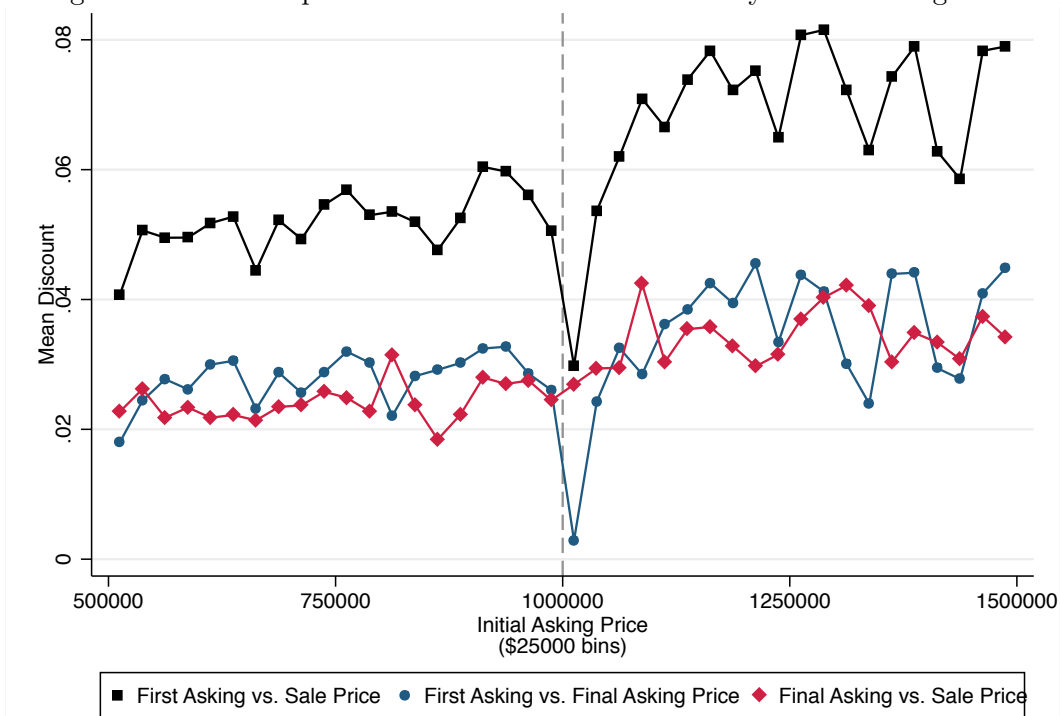


Notes: Plot of the share of REBNY listed properties that are sold in NYC DOF data, but are not listed as closed in the REBNY listing per \$25,000 initial-asking-price bin. Data from REBNY listings—sample includes all REBNY-listed properties in the range \$500,000–1,500,000.



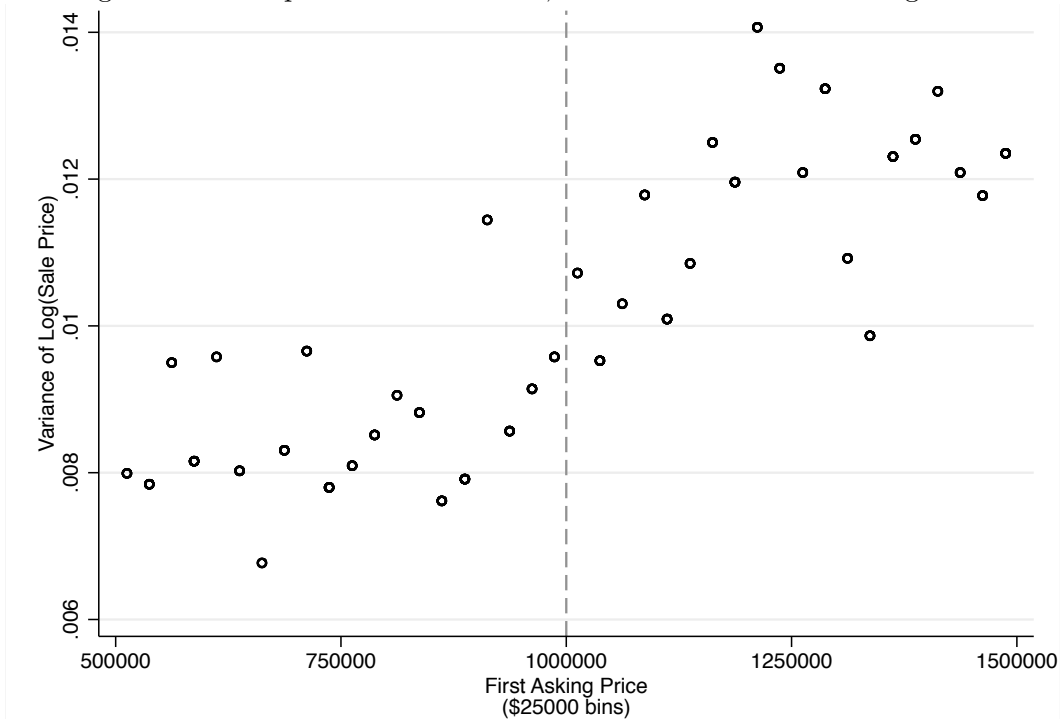
Notes: Plot of the median and 25th percentile discount from final asking price to sale price ( $= 1 - \text{sale}/\text{final}$ ) per \$25,000 final-asking-price bin. Data from REBNY listings—sample includes all closed REBNY-listed properties in the range \$500,000–1,500,000 that match to NYC DOF data.

Figure A.13: Decomposition of Mean Price Discounts by Initial Asking Price

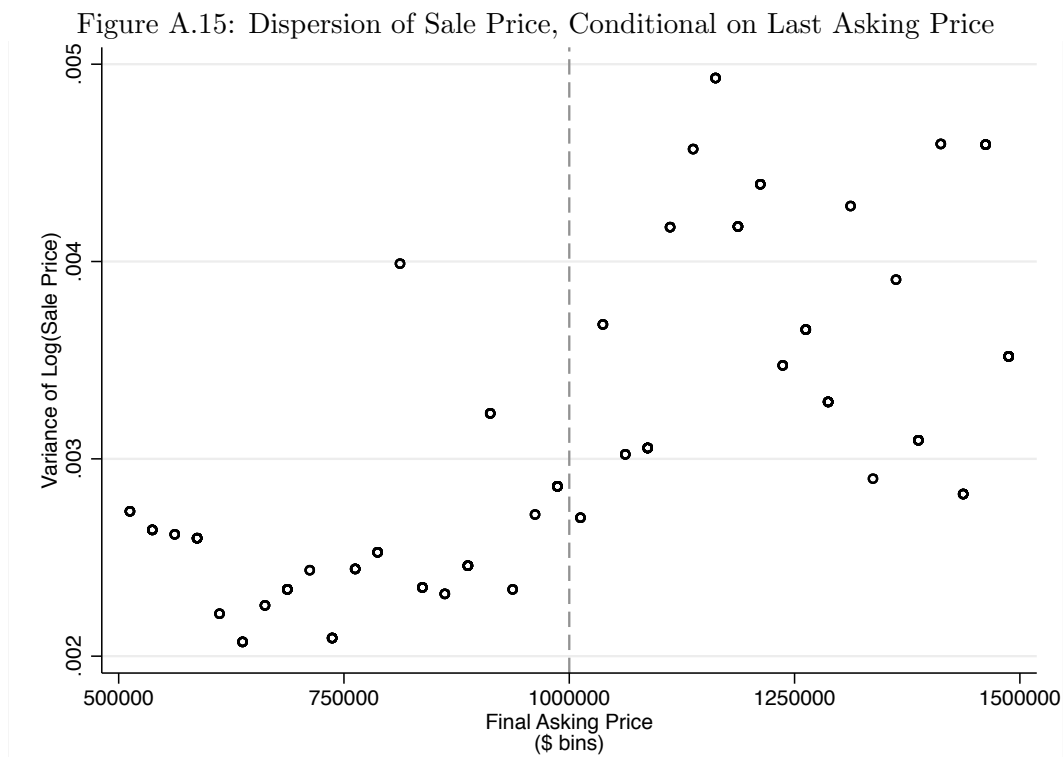


Notes: Plot of the average discount from initial asking to sale price ( $= 1 - \text{sale}/\text{initial}$ ), initial asking to final asking price ( $= 1 - \text{final}/\text{initial}$ ), and final asking price to sale price relative to initial asking price ( $= (\text{final} - \text{sale})/\text{initial}$ ) per \$25,000 initial-asking-price bin. Data from REBNY listings—sample includes all closed REBNY-listed properties in the range \$500,000–1,500,000 that match to NYC DOF data.

Figure A.14: Dispersion of Sale Price, Conditional on First Asking Price

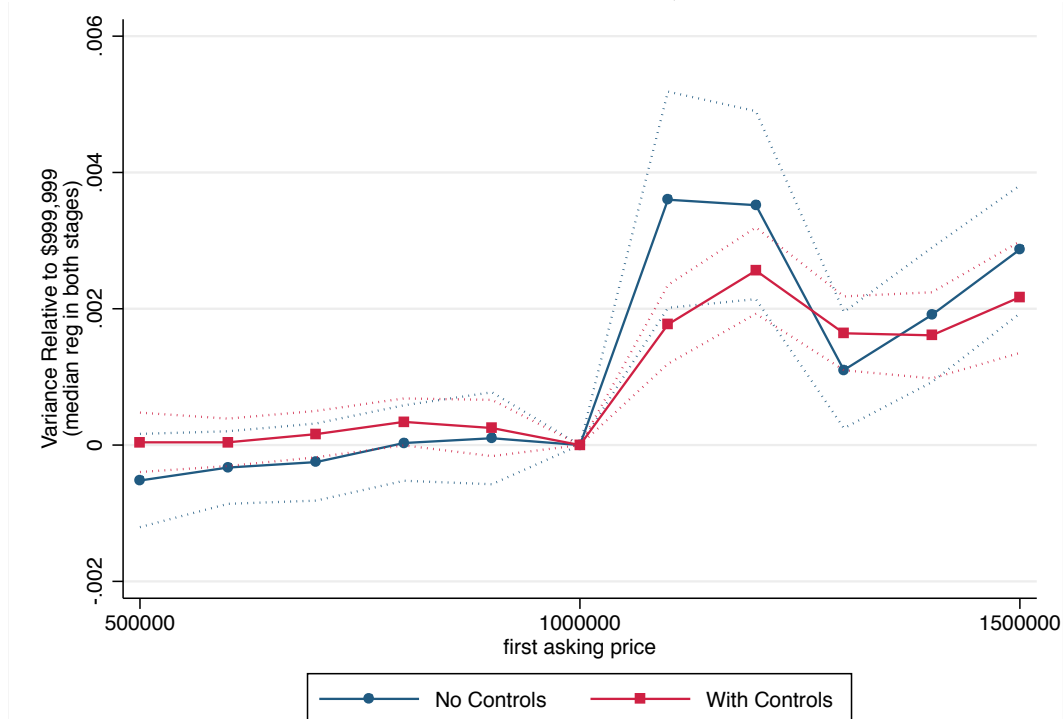


Notes: Plot of the variance of the log of sale price by \$25,000 initial asking price bin. Data from REBNY listings—sample includes all sold REBNY-listed properties (matched to NYC DOF) in the range \$500,000–1,500,000. Observations with price discounts (first asking price to sale price) in the 1st or 99th percentile are omitted.



Notes: Plot of the variance of the log of sale price by \$25,000 final price bin. Data from REBNY listings—sample includes all sold REBNY-listed properties (matched to NYC DOF) in the range \$500,000–1,500,000. Observations with price discounts (first asking price to sale price) in the 1st or 99th percentile are omitted.

Figure A.16: Predicted Dispersion of Log of Sale Price (Median Regression in First Stage)



Notes: Plots of the difference between predicted values at given initial asking price and predicted value at \$1,000,000 from the following procedure. The log of sale price is regressed (median regression) on a linear spline in the log of initial asking price with \$100,000 knots between \$500,000 and \$1,500,000. Squared residuals from this first stage are then regressed on a linear spline in log of initial asking price (using median regression; results are sensitive to outliers). Controls in the indicated results include year of sale, zipcode, building type, whether the sale is of a new unit, and the log of years since construction. Dashed lines represent 95% confidence intervals from 999 wild bootstrap replications of the two-stage procedure, resampling residuals in the first stage by asking-price clusters.

Table A.1: Mansion Tax, NYC

	Specification	Incidence	Std. Error	$\hat{Z}$	Std. Error	$n$
1.	Baseline: 3rd Order, Omit \$990k – \$1.155M	21542.098	1124.005	43861.766	3990.977	102493
2.	1st Order	24095.152	475.318	106372.697	4807.543	102493
3.	2nd Order	22910.857	965.353	46988.518	4118.622	102493
4.	4th Order	21115.344	1160.988	37069.861	5734.794	102493
5.	5th Order	18444.127	1490.541	33652.867	5868.148	102493
6.	Omit \$990k – \$1.01M	16308.129	974.380	.	.	108766
7.	Omit \$990k – \$1.025M	16676.799	967.177	.	.	108378
8.	Omit \$990k – \$1.050M	18168.244	1062.597	2311.782	1280.552	107635
9.	Omit \$990k – \$1.1M	20163.154	1134.501	15145.428	2117.299	105549
10.	Omit \$990k – \$1.2M	21978.363	1079.676	51725.536	5165.912	100846
11.	Omit \$990k – \$1.255M	22122.063	1127.693	53596.361	8617.068	97673
12.	Omit \$990k – \$1.3M	21917.696	1119.279	25178.003	10285.031	96094
13.	Omit \$980k – \$1.155M	24980.343	957.747	43954.765	4051.863	100706
14.	Omit \$970k – \$1.155M	26742.767	1634.857	44726.198	3810.847	99741
15.	Omit \$960k – \$1.155M	37687.334	2374.628	49452.601	4938.454	98915
16.	3rd Order for LHS, 1st Order for RHS	20034.887	1372.811	32006.872	3891.654	102493
17.	3rd Order for LHS, 2nd Order for RHS	20034.887	1372.811	37243.921	9670.122	102493
18.	No Discontinuity at Threshold	21601.728	1048.292	44349.086	2316.284	102493
19.	OLS, \$5000 Bins	19539.143	2093.029	.	.	102493
20.	OLS, \$10000 Bins	20956.556	1539.757	.	.	102493

Notes: Estimates by MLE as described in text (or OLS with binned data where indicated) using data from NYC Department of Finance Rolling Sales file for 2003–2011. Sample is restricted to all taxable sales (single-unit non-commercial sales of one-, two-, or three-family homes, coops, and condos) with prices between \$510,000 and \$1,500,000. We do not estimate  $\hat{Z}$  for specifications 6 and 7 since our method mechanically restricts the missing mass/gap to the range between \$1M and the top of the omitted region. Since the omitted regions above the notch for these two specifications are small, the estimates of the missing mass in the gap are artificially small. Specification 15 estimates the counterfactual separately using data below the omitted region of \$990k–\$1,155,422 to estimate bunching and incidence, and data above the omitted region to estimate the gap and  $\hat{Z}$ .

Table A.2: NYC Mansion Tax: Placebos

Cutoff	Incidence	Std. Error	$Z$	Std. Error	$n$
Commercial	-634.223	69.369	2203.190	512.087	5616
600,000	-689.324	32.958	2430.942	386.221	74477
700,000	-708.755	30.548	1257.927	258.632	86026
800,000	-787.246	29.686	360.535	280.648	93148
900,000	-732.660	32.098	-985.343	197.414	98003
1,100,000	-937.686	141.524	3457.587	559.861	106344
1,200,000	-669.331	42.662	1346.163	591.304	103660

Notes: Data from NYC Department of Finance Rolling Sales file for 2003-2011. Commercial sales are defined as any transaction of at least one commercial unit and no residential units or a NYC tax class of 3 (utility properties) or 4 (commercial or industrial properties) and are not subject to the mansion tax. Placebo estimates are found using the baseline MLE procedure around the given cutoff.

Table A-3: Local Incidence Over Time

Year	NYC				NYS				NJ									
	Incidence	Std. Error	Z	n	Incidence	Std. Error	Z	n	Incidence	Std. Error	Z	Std. Error	n					
1996	.	.	.	.	.	.	.	.	-832.902	2010.303	591.586	4577.626	1929					
1997	.	.	.	.	.	.	.	.	-846.881	892.57	4672.449	3965.021	2308					
1998	.	.	.	.	.	.	.	.	-516.749	325.587	-7.224	2278.698	3176					
1999	.	.	.	.	.	.	.	.	-1651.362	1641.192	5799.470	5644.423	4193					
2000	.	.	.	.	.	.	.	.	-855.516	219.819	1610.774	751.773	5402					
2001	.	.	.	.	.	.	.	.	-890.483	362.37	4516.891	1772.83	6461					
2002	.	.	.	.	.22203.323	3459.236	54264.664	14737.658	9596	739.276	105.021	2921.912	818.467	9606				
2003	14624.035	4208.772	40388.150	15073.276	7310	19003.617	3607.358	63687.589	12761.623	12462	-636.847	95.723	2995.208	668.838	12953			
2004	23429.529	2510.092	19803.960	10841.689	10451	22087.229	2651.167	63869.400	11328.725	17389	-785.303	94.236	3425.059	736.190	10800			
2004 (post)	.	.	.	.	.	.	.	.	.	.	.	.	.	5996.987	4145.917	38365.000	16506.710	8442
2005	32614.473	4526.286	64561.575	12876.311	13554	24014.876	1833.177	47688.156	10156.431	21765	24229.456	1233.458	55087.717	10767.801	24921			
2006	21603.437	2520.204	47836.366	11784.422	14728	24217.067	1562.105	28400.202	10617.074	17958	22080.594	2567.267	34109.311	9633.650	21963			
2007	17487.074	2388.516	42113.910	8840.712	16477	22673.583	10770.933	16409.173	38609.604	948	24119.928	2025.234	27611.911	10546.270	18921			
2008	17685.933	2587.644	34771.013	9461.669	12911	23867.938	2571.473	14847.095	11051.018	11143	19047.947	3479.866	31128.821	10877.795	12913			
2009	20449.578	3434.685	1338.866	10748.730	8682	23930.513	3294.993	24523.627	14400.244	8370	20673.616	3852.975	24910.197	13266.069	10039			
2010	17125.476	3781.441	74753.214	15731.867	8534	22620.920	3170.810	21216.032	12132.485	8831	19520.012	3566.167	37097.427	13552.642	10948			
2011	21739.302	2853.967	70566.195	15725.257	9846	.	.	.	.	.	.	22460.739	5574.445	27266.456	21619.086	3789		

Notes: Baseline local incidence estimates by year for given geography. Data for NYC from Department of Finance Rolling Sales files, restricted to taxable sales (single-unit non-commercial sales of one-, two-, or three-family homes, coops, and condos) in the given year. Data for NYS from the Office of Real Property Services deeds records, restricted to taxable sales (all single-parcel residential sales of one-, two-, or three-family homes) in the given year. NYS ORPS data covers 2002–2007 and 2009–2010; observations in 2007 are from sales made in 2007, but recorded in 2008–2010. Data for NJ from NJ Treasury SR1A file.



Table A.4: Predicted Price Discounts

	Discount: Initial Asking to Sale Price		Discount: Initial Asking to Final Asking	
	Slope	Value at Lower Knot relative to \$1,000,000	Slope	Value at Lower Knot relative to \$1,000,000
<i>Prediction at \$1,000,000</i>		<i>0.05117</i>		<i>0.02464</i>
500000 to 600000	0.12785 (0.6455)	-0.00279 (0.00579)	0.71910 (0.44164)	-0.00355 (0.00410)
600000 to 700000	0.25113 (0.40788)	-0.00151 (0.00441)	-0.02462 (0.24894)	0.00364 (0.00270)
700000 to 800000	0.12379 (0.37161)	0.00100 (0.00416)	0.13650 (0.26585)	0.00340 (0.00266)
800000 to 900000	0.07008 (0.39359)	0.00224 (0.00419)	0.18161 (0.29979)	0.00476 (0.00282)*
900000 to 1000000	-0.29398 (0.46877)	0.00294 (0.00469)	-0.65794 (0.33941)	0.00658 (0.00339)*
1000000 to 1100000	1.95390 (0.57745)***	0 0	1.02142 (0.39059)***	0 0
1100000 to 1200000	0.38174 (0.68556)	0.01954 (0.00577)***	0.77419 (0.47691)	0.01021 (0.00391)***
1200000 to 1300000	0.29954 (0.65681)	0.02336 (0.00550)***	-0.49885 (0.45059)	0.01796 (0.00375)***
1300000 to 1400000	-0.58272 (0.65944)	0.02635 (0.00536)***	0.14506 (0.43487)	0.01297 (0.00350)***
1400000 to 1500000	0.72161 (0.74929)	0.02052 (0.00565)***	0.44734 (0.52392)	0.01442 (0.00367)***
1500000		0.02774 (0.00626)***		0.01889 (0.00431)***
<i>Prediction at \$1,500,000</i>		<i>0.07892</i>		<i>0.04354</i>

Notes: Data from REBNY listings matched to NYC DOF sales records. Estimates from regression of given discount on linear spline in initial asking price (\$100k knots between \$500k and \$1.5M). Slope estimates are for the given interval; predicted values are the difference between the predicted discount at the lower knot of the given interval and the predicted value at \$1,000,000.

Table A.5: Predicted Price Dispersion

	Initial Asking		Final Asking		Initial Asking		Final Asking	
<i>Prediction at \$1,000,000</i>	0.00333	0.00220	0.00099	0.00061	0.00646	-0.00017	0.00004	-0.00115
500000	-0.00023 (0.00065)	-0.00052 (0.00035)	0.00009 (0.00022)	-0.00019 (0.00010)	0.00018 (0.00032)	0.00004 (0.00022)	-0.00002 (0.00010)	0.00001 (0.00006)
600000	0.00038 (0.00056)	-0.00033 (0.00026)	0.00006 (0.00020)	-0.00005 (0.00007)	0.00041 (0.00028)	0.00004 (0.00018)	0.00002 (0.00007)	0.00003 (0.00004)
700000	0.00012 (0.00050)	-0.00025 (0.00027)	0.00008 (0.00023)	-0.00015 (0.00007)**	0.00032 (0.00025)	0.00016 (0.00018)	0.00008 (0.00013)	0.00002 (0.00004)
800000	0.00051 (0.00046)	0.00003 (0.00029)	0.00015 (0.00025)	-0.00015 (0.00008)*	0.00042 (0.00023)*	0.00034 (0.00018)*	0.00010 (0.00015)	0.00002 (0.00005)
900000	0.00084 (0.00071)	0.00010 (0.00034)	-0.00021 (0.00021)	0.00013 (0.00011)	0.00083 (0.00039)**	0.00025 (0.00022)	0.00002 (0.00009)	0.00010 (0.00006)*
1000000	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
1100000	0.00324 (0.00079)***	0.00360 (0.00079)***	0.00184 (0.00067)***	0.00143 (0.00023)***	0.00186 (0.00033)***	0.00177 (0.00030)***	0.00111 (0.00033)***	0.00095 (0.00012)***
1200000	0.00478 (0.00058)***	0.00352 (0.00075)***	0.00139 (0.00030)***	0.00083 (0.00016)***	0.00276 (0.00040)***	0.00256 (0.00033)***	0.00065 (0.00016)***	0.00069 (0.00010)***
1300000	0.00417 (0.00137)***	0.00110 (0.00041)***	0.00101 (0.00061)*	0.00014 (0.00013)	0.00292 (0.00081)***	0.00164 (0.00028)***	0.00089 (0.00049)*	0.00027 (0.00008)***
1400000	0.00235 (0.00082)***	0.00191 (0.00052)***	0.00114 (0.00038)	0.00058 (0.00013)***	0.00201 (0.00044)***	0.00161 (0.00034)***	0.00049 (0.00015)***	0.00056 (0.00009)***
1500000	0.00217 (0.00098)**	0.00287 (0.00046)***	0.00109 (0.00048)***	0.00056 (0.00012)***	0.00236 (0.00045)***	0.00217 (0.00040)***	0.00055 (0.00016)***	0.00049 (0.00012)***
<i>Prediction at \$1,500,000</i>	0.00550	0.00508	0.00208	0.00117	0.00882	0.00200	0.00060	-0.00066
First-Stage $R^2$ /Pseudo $R^2$	0.556660	0.68100	0.59190	0.80230	0.61540	0.71750	0.64260	0.82020
Median Reg. in 1st Stage		X		X		X		X
Property Controls					X	X	X	X

Notes: Data from REBNY listings matched to NYC DOF sales records. The log of sale price is regressed on a linear spline in the log of initial asking price with \$100,000 knots between \$500,000 and \$1,500,000 and a set of controls for property characteristics where indicated (property zipcode, year of sale, property type, and years since construction). Squared residuals from this first stage are then regressed on a linear spline in log of initial asking price and controls where applicable (using median regression; results are sensitive to outliers). Estimates are the difference between the predicted discount at the given knot and the predicted value at \$1,000,000. Standard errors found by 999 bootstraps of the two-stage procedure, resampling first-stage residuals, drawing from a moving block of 50 observations.