

Counting the Six-Figure Renter: Direct Census Estimation, Synthetic-Population Reproduction, and the Move-In Staircase

SparkyData

<https://sparkydata.dev>

Methods note accompanying the article *The six-figure renter*, September 2026 (v3)

Abstract—We estimate how many New York City renter households with incomes of \$100,000 or more pay at least 30% of income in gross rent, separating families and single-person households from nonfamily shared (“roommate”) households, and test what a survey-calibrated synthetic population adds to a question the public census microdata answers directly. From the 2020–24 ACS PUMS with successive difference replication: $67,600 \pm 3,100$ burdened family and single households ($11.6\% \pm 0.5\%$ of the group, 6% of the city’s burdened renters) and $25,600 \pm 1,700$ roommate households ($16.7\% \pm 1.1\%$). Among families and singles, burden falls monotonically with time in the unit, from 22.3% for households that moved in within a year to 3.0% after thirty years, while median gross rent falls from \$3,183 to \$1,522, household composition shifts from young singles in one-bedrooms to older couples, and the share living in pre-1970 buildings of five or more units rises from 41% to 78%. We show that the long-tenure rents match the city’s Housing and Vacancy Survey median for rent-stabilised units (\$1,500) and explain why income cannot end a stabilised tenancy. A 500,000-person synthetic population sampled from the same microdata reproduces all fourteen held-out burden targets within band and every profile statistic within its interval, which we interpret as a calibration check rather than an out-of-sample test. We record a negative result on the population’s imputed wealth layer, which was withdrawn. All code, data and figures reproduce from a fixed seed.

I. INTRODUCTION

Half of New York City’s renter households are rent-burdened by the 30% standard [4], [6], and the city’s own asking rent implies a household needs about \$140,000 to rent at market without being burdened [4]. A recurring question is how many households *with* six-figure incomes are nonetheless burdened, and who they are. No published table answers it: the City Comptroller’s income breakdown stops at \$80,000 and the State Comptroller reports a statewide figure for households above \$75,000 [5].

The public-use microdata does answer it. This note states the direct census estimate with a proper margin, so that the count is attributable to the Census Bureau rather than to a model; separates households where a person or family earns six figures from households that reach six figures by pooling unrelated adults; decomposes the burden by move-in period and describes who occupies each step; checks the

low long-tenure rents against an independent survey and the rent law; says what a synthetic population built from the same microdata contributes; and records a negative result on that population’s modelled wealth layer.

II. DATA

A. Census microdata

We use the 2020–24 five-year American Community Survey Public Use Microdata Sample [1] for a 156-PUMA study region covering the five boroughs, Long Island, the lower Hudson Valley, twelve northern New Jersey counties and southwestern Connecticut: 412,280 housing-unit records and 926,174 person records with their 80 replicate weights. Dollars are adjusted to 2024 with the Bureau’s ADJINC and ADJHSG factors. The analysis universe is occupied rental units in the five boroughs with positive household income and a defined rent-to-income ratio (GRPIP), 2.13 million weighted households.

Variables: gross rent GRNTP; GRPIP, the Bureau’s rent-to-income percentage, top-coded at 101; household income HINCP; the highest personal income PINCP in the household; MV, the householder’s move-in period in seven bands; HHT, from which *families and singles* are family households plus householders living alone (codes 1–4 and 6) and *roommate households* are nonfamily households with the householder not living alone (codes 5 and 7); NP (persons) and BDSP (bedrooms); BLD and YRBLT, from which the *stabilised-stock proxy* is buildings of five or more units ($BLD \geq 6$) built before 1970 ($YRBLT \leq 1960$); OCPIP, selected monthly owner costs as a percentage of income; and TEN, tenure. PUMS does not record rent regulation; the proxy is the closest available cut to the pre-1974, six-plus-unit stock that stabilisation covers, and about half of all NYC renter households live in it.

B. Synthetic population

SparkyData release 0.2 samples whole households with replacement from the study-region PUMS, calibrates the sample weights simultaneously to published household margins (region, tenure, size, income quintile, householder age) and person margins (age, sex, race, nativity, employment, commute mode, NYC residence) by generalised raking in the bounded-logit form of Deville and Särndal [3], then

realises an unweighted population of 500,000 people in 195,799 households by sampling from the calibrated weights. Every variable used in this note is inherited from the sampled census record. The relevant subsample is 52,336 NYC renter households with positive income, of which 18,003 earn \$100,000 or more; 14,216 families and singles (1,659 burdened) and 3,787 roommate households (637 burdened).

III. ESTIMATORS

A. Census side: successive difference replication

For a statistic $\hat{\theta}$ computed with the full weight, the same statistic is recomputed with each of the 80 replicate weights, $\hat{\theta}_r$, and

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{4}{80} \sum_{r=1}^{80} (\hat{\theta}_r - \hat{\theta})^2, \quad (1)$$

the Bureau’s standard estimator [2]. Margins are reported at the Bureau’s 90% level, 1.645 SE. Counts are weighted sums, shares weighted means; medians are weighted medians without a margin.

B. Synthetic side: household bootstrap

The synthetic sample is unweighted. For each statistic the households in the relevant group are resampled with replacement 2,000 times and the 5th and 95th percentiles of the resampled statistic are reported as a 90% interval, matching the census level. The resampling is seeded (20240101). Synthetic counts, where quoted, are scaled by the ratio of census-weighted to synthetic NYC renter households with positive income, 40.8.

C. Reproduction bands

The release’s verification framework compares each held-out target with the synthetic value using a band that combines the published margin with the synthetic sample’s own sampling error in quadrature,

$$B = \sqrt{\text{MOE}_{\text{pub}}^2 + B_{\text{samp}}^2}, \quad B_{\text{samp}} = 1.96 \sqrt{\frac{p(1-p)}{n/\text{deff}}}, \quad (2)$$

where n is the statistic’s own denominator in the synthetic sample and $\text{deff} = 1.96$ is the design effect the calibration reported ($1 + \text{CV}^2$ of the calibrated weights). The median-rent row uses the framework’s quantile band. The two terms are at different confidence levels (90% and 95%), which changes bands by about 2%.

IV. THE REPRODUCTION CHECK

The fourteen renter cost-burden targets in Table I were computed from the PUMS after the build and were never calibration inputs. All fourteen land within band; the median $|d|/B$ is 0.22.

The synthetic households are sampled from the same microdata the targets are computed from, so for any statistic of inherited variables the expected value of the synthetic estimate equals the weighted PUMS estimate by

TABLE I
HELD-OUT REPRODUCTION: FOURTEEN RENTER COST-BURDEN TARGETS, 2024 DOLLARS. $|d|/B$ IS THE ABSOLUTE DIFFERENCE OVER THE BAND.

Statistic	Census	Synthetic	Band	$ d /B$
Renter share, metro	46.1	46.1	0.4	0.14
Renter share, NYC	65.5	65.6	0.6	0.09
Median gross rent, metro (\$)	1,831	1,830	10	0.10
Burdened, metro	52.6	52.6	0.5	0.09
Burdened, NYC	52.6	52.7	0.7	0.09
Severely burdened, NYC	29.3	29.3	0.6	0.13
Burdened, NYC, <\$35k	87.7	87.4	0.9	0.28
Burdened, NYC, \$35–75k	71.5	71.7	1.3	0.14
Burdened, NYC, \$75–100k	40.7	39.9	2.0	0.43
Burdened, NYC, \$100–150k	21.3	21.6	1.5	0.16
Burdened, NYC, \$150–200k	10.6	10.5	1.7	0.07
Burdened, NYC, \$200k+	3.9	4.1	0.8	0.17
Burdened, NYC, \$150k+	6.5	6.5	0.8	0.04
Burdened, metro, \$150k+	5.8	5.7	0.6	0.03

construction. Agreement is therefore not an out-of-sample test. It is a test that calibration to margins that do not include rent did not distort the rent-by-income joint. For a question answerable from inherited variables alone, then, the synthetic population adds convenience rather than information: a joined database of households, members’ incomes, housing and buildings at one fifth the size of the source. The article’s headline counts are census counts, and the synthetic population is presented as reproducing them.

V. FINDINGS

Table II sets the census and synthetic estimates side by side for the statistics the article quotes. Every synthetic value lies within the census margin and every census value within the synthetic interval.

A. The count

$737,600 \pm 7,100$ NYC renter households have income of \$100,000 or more; $93,200 \pm 3,400$ of them (12.6%) are rent-burdened. Families and singles are $584,600 \pm 6,800$ of the households and $67,600 \pm 3,100$ of the burdened (11.6% $\pm 0.5\%$; 1.5% severely burdened); roommate households are $153,000 \pm 3,900$ and $25,600 \pm 1,700$ (16.7% $\pm 1.1\%$). Roommate households are 21% of six-figure renters and 27% of the burdened. At \$150,000 the all-household burdened share is 6.5% $\pm 0.5\%$ and at \$200,000 3.9% $\pm 0.4\%$.

B. The line at this income

Among families and singles the burdened group’s median income is \$123,400 and its median gross rent \$3,923 (35%); the unburdened group’s are \$167,600 and \$2,283 (16%). The income ratio is 1.36, the rent ratio 1.72. In both groups about seven in ten households contain a person with income of \$100,000 or more (72.2% $\pm 2.1\%$ and 70.1% $\pm 0.7\%$), so for this population “six-figure household” and “six-figure earner” largely coincide; among burdened roommate

TABLE II

CENSUS (WITH 90% MOE) AND SYNTHETIC (WITH 90% BOOTSTRAP INTERVAL) ESTIMATES, IN PERCENT. “FAMILIES” ARE FAMILY AND SINGLE-PERSON HOUSEHOLDS WITH INCOME OF \$100,000 OR MORE RENTING IN NYC; “ROOMMATES” ARE NONFAMILY SHARED HOUSEHOLDS AT THE SAME INCOME.

Statistic	Census	Synthetic
Burdened share, all \$100k+ households	12.6 ± 0.5	12.8 [12.4, 13.2]
Burdened share, all \$150k+	6.5 ± 0.5	6.5 [6.1, 6.9]
Burdened share, all \$200k+	3.9 ± 0.4	4.1 [3.5, 4.6]
Burdened share, families and singles \$100k+	11.6 ± 0.5	11.7 [11.2, 12.2]
Burdened share, roommate households \$100k+	16.7 ± 1.1	16.8 [15.9, 17.8]
Families: burdened, moved in ≤12 months	22.3 ± 1.6	22.0 [20.5, 23.6]
Families: burdened, moved in 2–4 years	12.8 ± 1.1	12.6 [11.7, 13.5]
Families: burdened, moved in 10–19 years	6.6 ± 1.3	7.2 [6.2, 8.1]
Families: burdened, moved in 30+ years	3.0 ± 0.8	3.9 [2.9, 4.8]
Families, settled	8.7 ± 0.5	8.9 [8.2, 9.6]
Families, moved in <2 years	19.7 ± 1.2	19.5 [18.5, 20.6]
Roommates, settled	11.8 ± 1.5	12.0 [10.8, 13.3]
Roommates, moved in <2 years	21.4 ± 1.7	21.2 [19.8, 22.7]
Burdened families: Manhattan	53.5 ± 2.5	52.8 [50.9, 54.9]
Burdened families: lives alone	46.3 ± 2.3	45.0 [43.0, 47.0]
Burdened families: has a \$100k earner	72.2 ± 2.1	72.0 [70.2, 73.8]
Burdened families: pre-1970, 5+ units	34.4 ± 2.4	34.8 [32.9, 36.8]
Unburdened families: Manhattan	31.7 ± 0.7	31.6 [31.0, 32.9]
Unburdened families: lives alone	28.4 ± 0.6	28.1 [27.4, 28.8]
Unburdened families: has a \$100k earner	70.1 ± 0.7	70.1 [69.4, 70.8]
Unburdened families: pre-1970, 5+ units	48.2 ± 0.8	48.2 [47.5, 49.0]
Owner cost-burdened, \$100–150k	53.6 ± 1.5	53.6 [52.0, 55.1]
Owner cost-burdened, \$150–200k	27.4 ± 1.6	27.0 [25.3, 28.6]
Owner cost-burdened, \$200k+	8.3 ± 0.6	8.5 [7.7, 9.0]
Renters with less left after rent	67.0 ± 0.3	67.0
Families paying 25–30%	9.3 ± 0.4	9.2

households only $33.7\% \pm 4.0\%$ do. After rent the burdened family’s median residual income is \$78,700; $67.0\% \pm 0.3\%$ of all NYC renter households have less after their own rent. A further $9.3\% \pm 0.4\%$ of families and singles pay between 25% and 30%; raising every rent 10% moves the burdened share from 11.6% to 14.3% [11].

C. The move-in staircase and who is on it

Fig. 1 and Table III give the burdened share by move-in period for families and singles, with the composition of each step. Burden falls from $22.3\% \pm 1.6\%$ (12 months or less) to $3.0\% \pm 0.8\%$ (30 years or more) and median gross rent from \$3,183 to \$1,522. The newest arrivals average 2.1 people, 47% live alone, 55% occupy studios or one-bedrooms and the median householder is 33; at 10–19 years the household averages 3.2 people, 38% have children and the householder is 50; at 30+ years the household averages 2.5, 14% have children, the householder is 68 and $78\% \pm 2\%$ live in pre-1970 buildings of five or more units, against $41\% \pm 2\%$ of the newest arrivals. Households less than two years in the unit are 26% of the group and 45% of the burdened. Settled families (more than two years) are burdened at $8.7\% \pm 0.5\%$, recent ones at $19.7\% \pm 1.2\%$; settled and recent roommate households at $11.8\% \pm 1.5\%$ and $21.4\% \pm 1.7\%$.

D. Sanity check on long-tenure rents

A median gross rent of \$1,522 for six-figure households thirty years in the unit is consistent with independent evidence. The 2023 NYC Housing and Vacancy Survey reports a citywide median rent of \$1,641, a median of \$1,500 for rent-stabilised units, \$2,000 for market rentals and \$988 for rent-controlled units (2023 dollars) [7]; 60% of stabilised units rent below \$1,650. The Furman Center reports an average tenure of eight years in stabilised units against three in unregulated ones, and that about 65% of the stabilised units entered stabilisation because it was built before 1974. In the PUMS, the all-income median rent of households thirty years in the unit is \$1,202 and 77% of them live in the public stock. Rent-stabilised tenants hold a statutory right to renew; the high-income deregulation provision (income above \$200,000 for two consecutive years together with a 5% rent threshold) was repealed by the Housing Stability and Tenant Protection Act of 2019 [9], so income is not a ground for non-renewal, and a \$150,000 household was a ground over the former line. The tenure discount landlords sitting tenants in unregulated units is folded into the staircase and cannot be separated from stabilisation in the PUMS.

E. Owners, same yardstick

Applying OCPIP ≥ 30 to owners with a mortgage: $24.8\% \pm 0.6\%$ of the 410,500 six-figure owner households are cost-burdened, about 102,000 households, against 93,200 burdened six-figure renters of all types. By band: $53.6\% \pm 1.5\%$ at \$100–150k against 21.3% for renters; $27.4\% \pm 1.6\%$ at \$150–200k against 10.6%; $8.3\% \pm 0.6\%$ at \$200k+ against 3.9%. Owner costs include principal repayment and property tax, so the comparison is of positions relative to the same line, not of hardship.

F. What did not replicate

The first version, built on a 100,000-person release at the \$150,000 line, reported that burdened high-income renters were twice as likely to be headed by someone aged 65 or over (from 19 synthetic households) and that 71% had one earner or fewer. Neither survives: at 500,000 people and in the census the 65+ shares are indistinguishable ($8.8\% \pm 1.4\%$ against $11.4\% \pm 0.5\%$ among families and singles at \$100k+, the burdened being if anything younger), and the earner claim came from PUMS WIF (workers in *family*), undefined for nonfamily households and treated as zero.

VI. A NEGATIVE RESULT: THE IMPUTED WEALTH LAYER

The release attaches a balance sheet to every household from the 2022 Survey of Consumer Finances [10]: each household is matched to a donor cell by age band, education, tenure and income band, walking down a fallback ladder when a cell is thin, and its rank within the cell is carried into a net-worth rank through a Gaussian copula. The first version compared liquid assets between burdened and unburdened high earners from this layer. Three defects

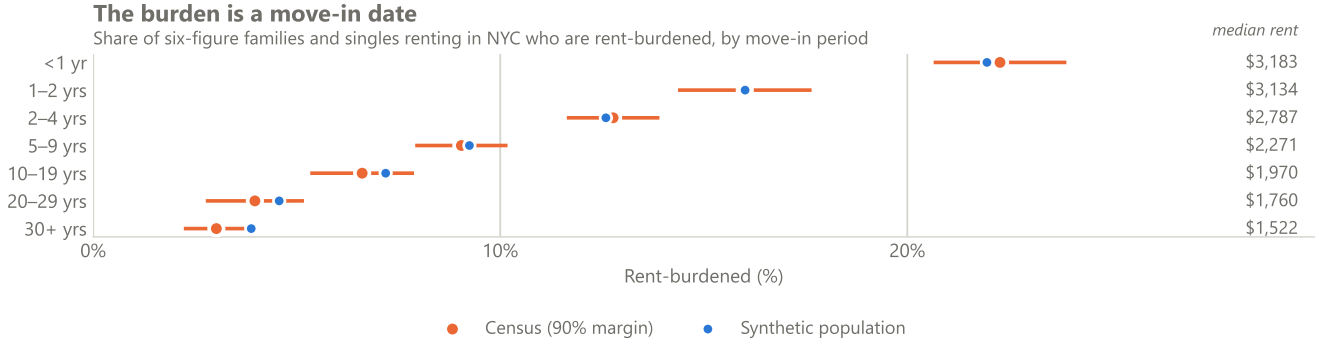


Fig. 1. Burdened share of six-figure family and single-person renter households in NYC by the householder’s move-in period (census with 90% margin in orange, synthetic in blue), with each row’s median gross rent at the right.

TABLE III

SIX-FIGURE FAMILIES AND SINGLES RENTING IN NYC, BY MOVE-IN PERIOD (CENSUS). RENT IS THE MEDIAN GROSS RENT IN 2024 DOLLARS; PEOPLE IS THE MEAN HOUSEHOLD SIZE; AGE IS THE MEDIAN HOUSEHOLDER AGE; OLD 5+ IS THE SHARE IN PRE-1970 BUILDINGS OF FIVE OR MORE UNITS.

Moved in	Burdened (%)	Rent (\$)	People	Alone (%)	Kids (%)	Age	Old 5+ (%)
12 months or less	22.3 ± 1.6	3,183	2.1	47	23	33	41
13 to 23 months	16.0 ± 1.6	3,134	2.2	43	27	35	39
2 to 4 years	12.8 ± 1.1	2,787	2.6	31	33	37	40
5 to 9 years	9.0 ± 1.1	2,271	2.9	23	41	43	43
10 to 19 years	6.6 ± 1.3	1,970	3.2	19	38	50	49
20 to 29 years	4.0 ± 1.2	1,760	2.9	19	25	57	61
30 years or more	3.0 ± 0.8	1,522	2.5	29	14	68	78

make that comparison uninformative. *Rank mapping*: the income band is the household’s income rank *within the metro*, mapped onto *national* SCF income percentiles, so a \$190,000 New York household draws from the national p60–80 cell (median income \$116,000); the cells the burdened families actually drew from are listed in `results.json`. *Coarsened conditioning*: most of the group drew from rung 1 or rung 2 cells, which drop education and, at rung 2, tenure. *Denominator effect*: rent is not a conditioning variable, so a months-of-rent comparison restates the rent gap. The wealth section was withdrawn; the numbers remain in `results.json` under `not_in_article`.

VII. LIMITATIONS

Vintage. The five-year file averages 2020–24. *Regulation*. PUMS does not record stabilisation; the building proxy is coarse in both directions. *Top-coding*. GRNTP is top-coded at \$11,600 and high incomes are component-top-coded. *The threshold*. The 30% standard descends from the 1969 Brooke Amendment’s 25%, raised in 1981 [12]; the residual-income results show it is not a hardship measure at this income. *Composition*. HHT cannot distinguish unmarried partners from roommates, so the roommate share is an upper bound. *Geography*. Nothing below the PUMA is supported.

VIII. REPRODUCIBILITY

`analysis.py`, shipped with the article, computes every number and figure from the PUMS extract and the

release database from seed 20240101; its docstring is the long-form method. `pums_direct.csv`, `reproduction.csv`, `results.json` and `data.csv` (18,003 synthetic six-figure NYC renter households, re-randomised identifiers) accompany it. This note’s tables and figures are generated from those files; no value is typed by hand.

REFERENCES

- [1] U.S. Census Bureau, “American Community Survey 2020–2024 5-Year Public Use Microdata Sample.”
- [2] U.S. Census Bureau, *PUMS Accuracy of the Data (2020–2024)*; R. E. Fay and G. F. Train, *Proc. Govt. Statistics Section, ASA*, 1995.
- [3] J.-C. Deville and C.-E. Särndal, “Calibration estimators in survey sampling,” *J. Amer. Statist. Assoc.*, vol. 87, no. 418, pp. 376–382, 1992.
- [4] NYC Comptroller, “Spotlight: New York City’s Rental Housing Market,” Jan. 2024.
- [5] NYS Comptroller, “Housing Cost Burdens for New Yorkers Among Nation’s Highest,” Feb. 2024.
- [6] NYU Furman Center, “State of Renters and Their Homes 2024.”
- [7] NYC HPD and U.S. Census Bureau, “2023 New York City Housing and Vacancy Survey: Selected Initial Findings,” 2024.
- [8] NYU Furman Center, “Understanding Different Segments of New York City’s Rent-Stabilized Housing Stock,” 2024.
- [9] NYC Rent Guidelines Board, “Rent Laws of 2019” (HSTPA, S.6458).
- [10] Federal Reserve Board, “Survey of Consumer Finances 2022, summary extract.”
- [11] M. E. Stone, “What is housing affordability?” *Housing Policy Debate*, 17(1):151–184, 2006.
- [12] M. Schwartz and E. Wilson, “Who can afford to live in a home?” U.S. Census Bureau, 2008.