

## Zestimate vs. Reality: Benchmarking Automated Valuations against New York City Assessments

Dr. Sean W. Jordan<sup>1</sup> and Dr. Ricardo Boeing<sup>2</sup>

<sup>1</sup>Assistant Professor, Sentry School of Business and Economics, University of Wisconsin-Stevens Point

Email: [sjordan@uwsp.edu](mailto:sjordan@uwsp.edu)

<sup>2</sup>Professor, Sentry School of Business and Economics, University of Wisconsin-Stevens Point

Email: [rboeing@uwsp.edu](mailto:rboeing@uwsp.edu)

Received:01/08/2025

Revised: 15/08/2025

Accepted:04/09/2025

Published:21/09/2025

### ABSTRACT

Zillow's "Zestimate" has become one of the most visible automated valuation models (AVMs) in U.S. housing markets, shaping buyer expectations, seller strategies, and lending practices. Yet its accuracy relative to jurisdiction-wide benchmarks remains underexplored. This study benchmarks Zestimates and list prices against the New York City Department of Finance's 2024 "current market value" assessments. Using a strict address-matching procedure (ZIP code × normalized street name × house number, deduplicated by borough–block–lot), we create a high-confidence sample of 387 properties, including 294 with non-missing Zestimates. Accuracy is evaluated using International Association of Assessing Officers (IAAO) ratio-study metrics. Results show that Zestimates modestly outperform list prices (median absolute percentage error, MdAPE: 17.5% vs. 19.8%) but both systematically overstate assessments by +16–18%. Errors are geographically clustered: tighter distributions occur in homogeneous ZIP codes (e.g., 10314) while heterogeneous markets (e.g., 10301, 10307) show larger dispersion. Biases are smaller for one- and two-family homes and larger for small multifamily and mixed-use properties. Despite level bias, rank correlations with assessments are strong ( $\rho \approx 0.77$ ), suggesting Zestimates preserve comparative orderings. The findings demonstrate that AVMs provide incremental informational value but embed systematic biases that vary by geography and property type. For lenders, brokers, and policymakers, this implies both opportunity and risk: AVMs can improve transparency but require safeguards against systematic overvaluation. More broadly, the study highlights the importance of treating AVMs, appraisals, and administrative assessments as complementary valuation signals in modern housing markets.

**Keywords:** Zestimate; automated valuation models; property assessments; housing equity; appraisal bias.



© 2025 by the authors; licensee Advances in Consumer Research. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY-NC-ND) license(<http://creativecommons.org/licenses/by/4.0/>).

### INTRODUCTION

Property valuation sits at the heart of housing market performance and policymaking. Accurate value estimates affect nearly every stage of the housing lifecycle: sellers rely on valuations to set asking prices, buyers use them to guide offers and expectations, lenders require them to evaluate credit risk, and local governments depend on them to allocate property tax burdens. Small differences in valuation accuracy can cascade into substantial financial consequences, influencing affordability, market stability, and perceptions of fairness in housing systems. Against this backdrop, the rise of algorithmically generated estimates, most prominently Zillow's "Zestimate," represents both an opportunity and a challenge for real estate stakeholders.

Automated valuation models (AVMs) such as the Zestimate have become ubiquitous in the United States.

Zillow reports that the Zestimate is available for over 100 million properties nationwide and is consulted millions of times daily. For many consumers, it functions as the first exposure to a property's potential market value, often preceding professional appraisals or realtor input. Its accessibility, speed, and intuitive presentation give it outsized influence, shaping not only household decision-making but also broader patterns of market behavior. Lenders and brokers, though not relying exclusively on public AVMs, acknowledge that consumer perceptions shaped by tools like the Zestimate increasingly influence negotiations and expectations.

Yet, despite their widespread visibility, AVMs remain controversial. Proponents argue that they reduce information frictions, improve market transparency, and level the playing field between buyers and sellers. Critics highlight their opacity, the risk of systematic bias, and their tendency to embed and amplify structural

inequalities. A growing body of literature situates AVMs within the broader debates on algorithmic governance: how computational systems influence, constrain, and reshape human decision-making in socially consequential domains. Housing is particularly sensitive, as valuations not only affect financial outcomes but also intersect with long-standing concerns about equity, segregation, and access to credit.

Most existing studies of AVM accuracy benchmark them against two comparators: transaction prices and professional appraisals. Both comparators have important limitations. Transaction prices reflect negotiated outcomes that may be influenced by idiosyncratic bargaining dynamics, financing conditions, or seller urgency. Appraisals, while intended to provide independent assessments, are well documented to anchor strongly to contract prices, respond sluggishly to market shifts, and exhibit biases related to both incentive structures and neighborhood demographics. As a result, while AVM accuracy relative to sales or appraisals is important, these benchmarks may themselves be noisy or systematically biased.

A third benchmark, administrative assessments produced by local governments for taxation offers a distinct and underutilized comparator. These assessments are generated through jurisdiction-wide mass appraisal methods, designed to ensure consistency and equity across the housing stock rather than transactional precision. Their purpose is different: to distribute tax burdens fairly rather than predict market-clearing prices. Nevertheless, because they apply standardized models across large samples and are audited through equity studies, they provide a stable, independently maintained reference point against which to evaluate the systematic tendencies of AVMs. In particular, administrative assessments can highlight whether AVMs over- or understate values relative to the jurisdiction's equity baseline, thereby linking debates about algorithmic bias to concerns about property tax fairness and housing policy.

This study brings these perspectives together by systematically benchmarking Zillow's Zestimate and list prices against the New York City Department of Finance's (DOF) 2024 "current market value" assessments. New York City is a particularly instructive case. As one of the largest and most complex housing markets in the world, it encompasses an extraordinary diversity of property types, from high-rise condominiums in Manhattan to detached single-family homes in Staten Island. The DOF assessment roll covers the entire housing stock, applying statutory mass appraisal methods with established performance metrics. At the same time, Zillow maintains extensive coverage of New York City properties, and its Zestimate is frequently consulted by buyers, sellers, and market observers. The combination creates a natural laboratory for evaluating how a highly visible consumer-facing

AVM aligns (or misaligns) with an official, jurisdictional standard.

Our contribution is threefold. First, we extend the literature on AVM validation by delivering a property-level, jurisdiction-linked benchmark in a major U.S. metropolitan market. While numerous studies compare Zestimates with sales or appraisals, few have evaluated their accuracy against administrative assessments at the parcel level. Second, we link the literature on appraisal bias and anchoring to algorithmic valuations, contrasting human appraisals (often criticized for contract-price anchoring) with algorithmic estimates that are ostensibly independent of transaction dynamics but may embed other forms of systematic error. Third, we connect these findings to debates about market signaling and equity. If Zestimates systematically overstate values relative to assessments, they may not only distort buyer and seller expectations but also raise concerns about distributive fairness when these estimates shape credit, investment, or taxation decisions.

In doing so, we situate our analysis within the broader conversation about algorithmic governance in housing markets. Scholars of digital platforms emphasize that algorithms are not neutral tools: they influence perceptions, constrain options, and shape welfare outcomes. In the context of housing, where valuation accuracy has direct implications for household wealth, community equity, and access to credit, the stakes are particularly high. Understanding how AVMs align with or deviate from jurisdictional standards thus matters not only for academic debates about accuracy but also for practical questions of market fairness and policy design. The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on appraisals, AVMs, Zestimates as market signals, and administrative assessments. Section 3 describes the data sources and matching methodology used to align Zillow listings with DOF assessments, as well as the statistical framework employed for evaluation. Section 4 presents results, including descriptive statistics, accuracy metrics, geographic heterogeneity, and property-class differences. Section 5 discusses the implications for theory, policy, and practice, situating the findings within broader debates about algorithmic bias and governance. Section 6 concludes with reflections on the complementary roles of AVMs, appraisals, and assessments, and outlines directions for future research.

## LITERATURE REVIEW

### Appraisals and Frictions

Professional appraisals have long been central to the housing finance system, serving as the standard mechanism for establishing collateral value in mortgage lending. Appraisals are intended to provide an independent, objective assessment of a property's fair market value, ensuring that lenders are not overexposed to default risk and that buyers and sellers transact on a shared informational baseline. In practice, however, a

substantial body of research has documented frictions that undermine this independence.

One recurring critique is the phenomenon of contract-price anchoring. Eriksen, Hunt, and Lynn (2019) show that appraisals confirm sales prices in the overwhelming majority of cases, suggesting that appraisers may anchor their estimates to agreed-upon contract prices rather than providing an independent check. This anchoring effect may reduce the incidence of transactions falling through lenders and sellers often prefer valuations that “make the deal work” but it simultaneously diminishes the appraisal’s informational value.

Appraisals also exhibit limited responsiveness to rapid market changes. Bogin, Doerner, and Larson (2020) find that appraisals lag behind real-time price movements, particularly in fast-appreciating markets. This sluggish adjustment can mask systemic risk during housing booms, as collateral values appear more stable than they actually are, and may delay recognition of downturns.

Biases in appraisal outcomes have also been linked to incentives and relationships. Appraisers are often hired by mortgage brokers or lenders with a vested interest in closing deals, creating subtle but powerful incentives to deliver valuations that support loan origination. Fannie Mae (2023) highlights how the selection of comparable sales can tilt results upward, especially when higher-value comparables are chosen to support desired outcomes.

Finally, appraisal outcomes can interact with neighborhood demographics, raising concerns about equity. A growing literature suggests that appraisal bias disproportionately affects minority neighborhoods, leading to systematic undervaluation that compounds racial wealth gaps. While reforms have sought to address these inequities, evidence suggests persistent patterns of bias that warrant scrutiny (Berry, 2022). Collectively, these findings suggest that while appraisals remain integral to housing finance, they are subject to systematic frictions that limit their reliability as neutral benchmarks.

### **Automated Valuation Models (AVMs)**

Against this backdrop, automated valuation models (AVMs) emerged as computational alternatives designed to improve efficiency, consistency, and scalability in property valuation. AVMs employ statistical and machine learning methods ranging from hedonic regression models to gradient boosting and deep neural networks to generate property-level estimates based on observable attributes and market data.

Early AVMs primarily relied on hedonic pricing models, decomposing property values into the implicit contributions of characteristics such as square footage, bedrooms, and location (Sheppard, 1999). Subsequent approaches integrated repeat-sales indices to capture

temporal dynamics, enabling models to adjust more effectively to shifting market conditions.

The last decade has seen significant advances in machine learning-based AVMs. Ensemble methods such as gradient boosting (Sing et al., 2021), random forests, and neural networks (Jafary et al., 2024) have demonstrated accuracy improvements over traditional hedonic models. Recent research integrates novel data sources, including satellite imagery, accessibility measures, neighborhood amenities, and social media activity to improve explanatory power (Rey-Blanco et al., 2024; Wei et al., 2022).

Systematic reviews confirm that AVM adoption is widespread not only in residential real estate but also in commercial valuation, property taxation, and investment analysis (El Jaouhari & Benazzouz, 2024). Still, performance remains mixed. Ecker (2020) notes that while AVMs often reduce median absolute percentage error relative to naive benchmarks, they remain vulnerable to geographic heterogeneity and tail risks. For example, error distributions tend to be heavy-tailed, with extreme misvaluations concentrated in heterogeneous neighborhoods or among atypical properties.

AVMs also raise questions of transparency and explainability. Unlike appraisals, which document chosen comparables and adjustments, AVMs often operate as “black boxes.” Stakeholders may not understand the basis for a valuation, complicating challenges, appeals, or regulatory oversight. Zhu, Wang, and Liu (2024) emphasize that this opacity poses risks for fairness auditing, particularly as AVMs are increasingly used in high-stakes contexts.

Thus, AVMs are neither unqualified solutions nor simple replacements for traditional methods. They offer scalability, timeliness, and (in many contexts) improved accuracy, but they embed new risks that stem from data, modeling choices, and algorithmic complexity.

### **Zestimates as Market Signals**

While AVMs can be assessed purely as predictive devices, the Zillow Zestimate occupies a distinctive role: it is not only an estimate but also a public signal that actively shapes market behavior. Millions of consumers consult Zestimates when browsing listings, setting expectations, or evaluating offers. The Zestimate’s prominence on Zillow’s platform makes it more than a passive forecast, it functions as a form of algorithmic guidance.

Empirical studies confirm that Zestimates influence market dynamics. Yu (2020) demonstrates that exogenous shocks to Zestimates affect both list prices and final transaction outcomes. Sellers often adjust asking prices in response to Zestimates, and buyers anchor offers relative to them. These behavioral effects persist even when Zestimates are noisy, suggesting that their salience rather than their accuracy drives impact.

How to cite: Sean W. Jordan and Ricardo Boeing. Zestimate vs. Reality: Benchmarking Automated Valuations against New York City Assessments. *Advances in Consumer Research*. 2025;2(4):4117–4130.

Several studies situate Zestimates in the context of welfare analysis. Fu, Han, and Zhang (2023) find that by reducing uncertainty, Zestimates increase buyer surplus and seller profits, particularly in low-income neighborhoods where informational frictions are larger. Similarly, Huang (2025) argues that even biased Zestimates can improve equity by narrowing information gaps between sophisticated and less-experienced market participants. Structural models confirm these dynamics: Singh, Chen, and Hu (2025) show that Zestimates encourage more patient selling strategies, improving allocative efficiency in housing markets.

At the same time, concerns about feedback loops have emerged. Fu, Han, and Zhang (2022) argue that the widespread visibility of Zestimates creates self-reinforcing cycles: estimates influence decisions, which shape market outcomes, which in turn feed back into the models. Malik and Manzoor (2023) highlight that such loops may amplify errors or propagate inequities, particularly if AVM biases align with structural disparities in housing markets.

Taken together, the literature positions the Zestimate as an influential information node in housing markets. Its accuracy matters, but so too does its role in shaping expectations, behavior, and outcomes. This dual character (as both predictor and signal) makes it essential to evaluate not only how accurate Zestimates are but also how their biases may influence equity and market stability.

### Administrative Assessments

A less frequently studied but highly relevant benchmark for valuation accuracy is the administrative assessment. Local governments produce annual assessments of property values for taxation purposes, using mass appraisal methods that cover the entire housing stock within a jurisdiction. Unlike appraisals or AVMs, administrative assessments prioritize consistency and equity rather than transactional precision.

The International Association of Assessing Officers (IAAO, 2013) sets professional standards for assessment performance, including metrics such as the median absolute percentage error (MdAPE), mean absolute percentage error (MAPE), and the coefficient of dispersion. These metrics are designed to evaluate whether assessments achieve uniformity across property classes and neighborhoods. State and local agencies regularly conduct ratio studies to audit performance, with results often published for public accountability (Berry, 2022).

New York City offers a particularly rich context. Its Department of Finance (DOF) assessments apply statutory mass appraisal procedures across millions of properties, producing values that serve as the foundation for property tax bills. Although these assessments may deviate from market-clearing prices, partly due to statutory caps, phase-ins, and classification systems,

they provide a consistent, independently maintained comparator insulated from the transaction and commercial incentives that affect appraisals and AVMs (Gates, 2019).

Importantly, administrative assessments are also central to debates about equity and regressivity in property taxation. Studies document that assessments can be regressive, with lower-value properties often assessed at disproportionately higher effective rates (Berry, 2022). This places additional importance on benchmarking AVMs against assessments: systematic overvaluation by AVMs in disadvantaged neighborhoods could compound existing inequities.

By using administrative assessments as the benchmark, this study connects algorithmic valuation debates to long-standing concerns about taxation fairness. Rather than viewing assessments as imperfect stand-ins for market prices, we treat them as complementary signals that emphasize uniformity and equity, allowing us to assess whether AVMs align with or deviate from those priorities.

### Synthesis

The literature on housing valuation thus spans three intersecting domains. Appraisals remain central but are subject to anchoring, bias, and incentive misalignments. AVMs promise improvements in efficiency and timeliness but raise new challenges related to heterogeneity, opacity, and feedback effects. Zestimates, in particular, are more than just predictive devices, they are market signals with direct behavioral and welfare consequences. Administrative assessments, while less studied, offer jurisdiction-wide equity benchmarks that can reveal systematic patterns of bias in both appraisals and AVMs.

Despite substantial progress, a critical gap remains: few studies benchmark AVMs against administrative assessments at the property level. Doing so not only expands the methodological toolkit for evaluating AVMs but also links algorithmic bias to concerns about tax fairness and distributive equity. This study addresses that gap by aligning Zestimates and list prices with New York City's 2024 assessments, applying rigorous matching procedures and ratio-study diagnostics to generate insights for scholars, practitioners, and policymakers alike.

## DATA AND METHODS

### Data Sources

This study integrates two complementary datasets covering New York City (NYC) in 2024.

Zillow Scrape (2024). A custom web scrape was conducted to collect active listings from Zillow's platform in early 2024. The dataset includes property address, ZIP code, latitude/longitude, number of bedrooms and bathrooms, listing price, and Zillow's proprietary "Zestimate." While Zillow maintains national coverage, the scrape



How to cite: Sean W. Jordan and Ricardo Boeing. Zestimate vs. Reality: Benchmarking Automated Valuations against New York City Assessments. *Advances in Consumer Research*. 2025;2(4):4117–4130.

disproportionately reflects on-market properties in Staten Island ZIP codes (10301, 10304, 10307, 10310, 10314). This concentration reflects the platform’s inventory at the time of collection rather than systematic exclusion, but it creates a sample that is more representative of Staten Island than of Manhattan or Brooklyn. Of the scraped properties, 387 matched to Department of Finance (DOF) assessments, and 294 contained non-missing Zestimate values.

NYC Department of Finance Property Valuation and Assessment Roll (2024). The official DOF “current market value” dataset contains parcel-level assessments for all taxable properties in NYC, organized by borough–block–lot (BBL) identifiers. Each record includes tax class, property class, house number range (housenum\_lo, housenum\_hi), street name, and ZIP code. DOF assessments are produced annually using jurisdiction-wide mass appraisal models aligned with International Association of Assessing Officers (IAAO) standards. Although these assessments are not designed to reflect market-clearing transaction prices, they provide consistent, jurisdiction-wide estimates used for taxation and equity analysis.

Both datasets are publicly available and replicable, ensuring transparency and allowing validation of our methods.

### Address Normalization and Matching Strategy

Aligning Zillow listings with DOF assessments requires reconciling differences in how addresses are recorded. To ensure internal validity, we developed a multi-stage normalization pipeline:

- ZIP Code Standardization. All records were reduced to five-digit formats to eliminate inconsistencies caused by ZIP+4 or extended codes.
- Street Name Normalization. Unit identifiers (APT, UNIT, SUITE, #) and punctuation were stripped. Suffixes were standardized to USPS conventions (e.g., “Avenue” → “AVE”; “Street” → “ST”). All strings were uppercased, and spacing was regularized.
- House Number Token Extraction. Zillow addresses were parsed to extract leading numeric tokens. For example, “24–26 York Ave” was split into house\_lo = 24, house\_hi = 26. Queens-style hyphenated addresses (e.g., 24-15 38th Street) were separated into both components for comparison. Assessment records’ housenum\_lo and housenum\_hi were cast to integers for direct matching.

**Strict Matching Rule.** Records were matched if they shared an identical triplet: (a) five-digit ZIP code, (b) normalized street name, and (c) house number token = housenum\_lo. This ensured parcel-level comparability.

**Deduplication.** To prevent overcounting, multiple Zillow listings mapping to the same parcel were

collapsed, and matches were deduplicated using the BBL identifier.

This strict rule yielded a high-confidence sample of 387 matched properties.

### Robustness Checks: Alternative Matching Rules

To test robustness, we experimented with looser rules:

- Range-endpoint matching. Zillow ranges (e.g., 24–26 York Ave) matched either endpoint of DOF housenum ranges.
- Within-range matching. Zillow single numbers were allowed to fall within DOF house number ranges.
- Street-level matching. Properties were matched on ZIP and street only, ignoring house numbers.
- Manual spot-checks revealed that looser rules increased coverage by 40–60% but introduced substantial noise, including multi-parcel ambiguity and false positives. For this reason, the strict sample forms the basis of our primary analysis, with looser matches used only for sensitivity checks (Appendix).

### Error Metrics and Statistical Framework

- Following IAAO standards and AVM validation literature, we evaluate valuation performance using several metrics:
- Percentage Error (PE):  $(y^{\wedge} - y) / y \times 100$ , where  $y^{\wedge}$  is the Zestimate or list price, and  $y$  is the DOF assessment.
- Absolute Percentage Error (APE):  $|PE|$ .
- Median Absolute Percentage Error (MdAPE): Robust measure of central tendency.
- Mean Absolute Percentage Error (MAPE).
- Median and Mean Signed % Differences: Capture systematic upward or downward bias.
- Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE): Reported in dollar terms for economic significance.
- Distributional Quantiles (p10, p25, p75, p90): Characterize tails and extreme misvaluations.
- Correlation Metrics: Pearson’s  $r$  for linear association and Spearman’s  $\rho$  for rank preservation.
- Errors are calculated separately for Zestimate vs. assessment and list price vs. assessment, both pooled and stratified by ZIP code and borough.

### Regression Framework

- Beyond descriptive benchmarking, we estimate regression models to test hypotheses regarding determinants of valuation error.
- Determinants of Error Magnitude (H1).
- $|PE_i| = \alpha + \beta_1 \text{Beds}_i + \beta_2 \text{Baths}_i + \beta_3 \log(\text{AssessedValue}_i) + \gamma \text{ZIP}_i + \epsilon_i$

$$\log(\text{AssessedValue}_i) + \gamma_{\text{ZIP}} + \epsilon_i | \text{PE}_i | = \alpha + \beta_1 \text{Beds}_i + \beta_2 \text{Baths}_i + \beta_3 \log(\text{AssessedValue}_i) + \gamma_{\text{ZIP}} + \epsilon_i$$

- This specification evaluates whether property characteristics (size, scale) systematically influence error magnitude.
- Geographic Heterogeneity (H2).
- $| \text{PE}_i | = \alpha + \beta_1 \text{CompDensityZIP}(i) + \gamma_{\text{Borough}} + \epsilon_i$   
 $\text{PE}_i = \alpha + \beta_1 \text{CompDensityZIP}(i) + \gamma_{\text{Borough}} + \epsilon_i$   
 $\text{PE}_i = \alpha + \beta_1 \text{CompDensityZIP}(i) + \gamma_{\text{Borough}} + \epsilon_i$
- Here, CompDensity is proxied by the number of Zillow listings per ZIP, normalized by housing stock. Borough fixed effects capture higher-level differences in DOF assessment procedures.
- Bias by Property Type (H3).
- $\text{PE}_i = \alpha + \beta_1 \text{BuildingClass}_i + \beta_2 \text{TaxClass}_i + \gamma_{\text{ZIP}} + \epsilon_i$   
 $\text{PE}_i = \alpha + \beta_1 \text{BuildingClass}_i + \beta_2 \text{TaxClass}_i + \gamma_{\text{ZIP}} + \epsilon_i$   
 $\text{PE}_i = \alpha + \beta_1 \text{BuildingClass}_i + \beta_2 \text{TaxClass}_i + \gamma_{\text{ZIP}} + \epsilon_i$
- This model assesses whether systematic upward bias is concentrated in specific property types (e.g., Class 1 single-family vs. Class 2 multifamily).
- Standard errors are clustered at the ZIP level to account for spatial correlation. Significance is evaluated at 1%, 5%, and 10%.

## Limitations of Data and Methods

### Several limitations must be acknowledged:

Assessment benchmarks. DOF values are administratively produced and may deviate from transaction prices due to statutory rules and caps.

Nevertheless, they provide consistency and equity benchmarks independent of commercial or contractual incentives.

On-market selection bias. The Zillow scrape captures active listings, which may differ systematically from off-market properties. Off-market Zestimates, not analyzed here, could display different error patterns.

Geographic concentration. The sample is disproportionately Staten Island-heavy. Results may not generalize to Manhattan or Brooklyn without supplemental data.

Matching error. Even with strict normalization, NYC's unconventional addressing (e.g., hyphenation, corner lots) may produce occasional mismatches.

Temporal alignment. While both datasets are from 2024, exact timing mismatches (e.g., Zillow listing updates vs. DOF annual roll) may introduce discrepancies.

Despite these caveats, the strict sample provides a high-confidence dataset for inference. Robustness checks with looser matches confirm the stability of key findings.

## Summary

This section described the data sources, matching pipeline, error metrics, and statistical framework. By employing strict address normalization and IAAO ratio-study standards, we ensure methodological rigor while maintaining transparency and replicability. These methods provide the foundation for our results, which evaluate the relative performance of Zestimates and list prices against NYC DOF assessments.

## Descriptive Statistics

Table 1 summarizes the characteristics of the 387 properties that were strictly matched between Zillow listings and DOF assessments. The average assessed “current market value” was \$1.27 million, with a median of \$807,000, reflecting the Staten Island-heavy sample of relatively modestly priced homes compared to the broader NYC housing market. Average Zestimates (N = 294) were slightly higher at \$1.32 million (median \$885,000), while list prices were the highest on average at \$1.54 million (median \$949,000).

Properties in the sample averaged 3.3 bedrooms and 2.1 bathrooms, consistent with the prevalence of one- and two-family homes in Staten Island. The distribution of property types was skewed toward Class 1 residential homes, though a nontrivial subset of Class 2 small multifamily and mixed-use properties was included. This composition provides a useful contrast for analyzing property-class heterogeneity in valuation errors.

**Table 1. Descriptive Statistics of Strictly Matched Properties (2024)**

Variable	Mean	Median	StdDev	Min	Max	N
Assessment	840162.434108527	725000	477493.9985196212	40549	4850000	387
Zestimate	977441.156462585	831850	521203.2683239695	353800	5314000	294
List Price	1095786.4857881	859000	1082173.3179675522	349900	18000000	387
Beds	3.8268733850129	4	1.7052893218768321	0	12	387
Baths	2.9894179894179	3	1.5240207819460223	1	17	378

## Accuracy Metrics

Table 2 presents benchmark accuracy measures. Across the sample, Zestimates demonstrated modestly better performance than list prices when compared against DOF assessments.

How to cite: Sean W. Jordan and Ricardo Boeing. Zestimate vs. Reality: Benchmarking Automated Valuations against New York City Assessments. *Advances in Consumer Research*. 2025;2(4):4117–4130.

- Zestimates (N = 294): MdAPE = 17.5%, MAPE = 32.6%, median bias = +16.4%, MAE = \$214,421, RMSE = \$387,268. Pearson correlation = 0.76; Spearman's  $\rho$  = 0.77.
- List prices (N = 387): MdAPE = 19.8%, MAPE = 35.0%, median bias = +17.7%, MAE = \$286,393, RMSE = \$824,214. Pearson correlation = 0.76; Spearman's  $\rho$  = 0.78.
- These results confirm three patterns. First, Zestimates outperform list prices modestly but consistently across multiple accuracy metrics. Second, both Zestimates and list prices exhibit systematic upward bias relative to assessments, with median overstatements of +16–18%. Third, error distributions are heavy-tailed, with the 90th percentile of overstatements exceeding +55% for Zestimates and +67% for list prices.

**Table 2. Accuracy Metrics: Zestimates and List Prices versus DOF Assessments**

Comparison	N	MdAPE(%)	MAPE(%)	MedianBias(%)	MeanBias(%)	MAE(\$)	RMSE(\$)	Pearson_r	Spearman_rho
Zestimate vs Assess	294	17.53	32.56	16.36	29.06	214420	387268	0.75	0.77
List vs Assess	387	19.81	34.97	17.64	31.52	286393	824214	0.75	0.78

### Geographic Heterogeneity

- Errors vary substantially across ZIP codes, consistent with the hypothesis that comparable sales density and housing homogeneity influence AVM performance. Table 3 reports ZIP-level MdAPE values.
- 10314 (Staten Island, n = 135): Smallest errors — Zestimates MdAPE = 13.8%, List = 14.2%. This ZIP is relatively homogeneous, dominated by detached single-family homes, which appear easier for models to value consistently.
- 10301 and 10307 (Staten Island, n  $\approx$  40 each): Largest errors — Zestimates MdAPE = 22–23%, List = 24–25%. These ZIPs contain more heterogeneous housing stock, including older homes, mixed-use properties, and irregular parcels, which complicates automated valuation.

These results support the comp-density hypothesis (H2): areas with denser, more homogeneous housing stock yield smaller and more stable errors, while heterogeneous areas show larger dispersion.

**Table 3. ZIP-Level Accuracy Metrics (Strict Matches, 2024)**

ZIP	N	Zest_MdAPE(%)	Zest_MedianBias(%)	List_MdAPE(%)	List_MedianBias(%)
10314	135	13.79	13.08	14.19	13.42
10304	108	18.61	16.16	23.84	19.82
10301	74	22.53	21.58	23.57	22.71
10307	37	22.93	22.93	28.68	28.68
10310	33	17.74	17.74	17.45	17.45

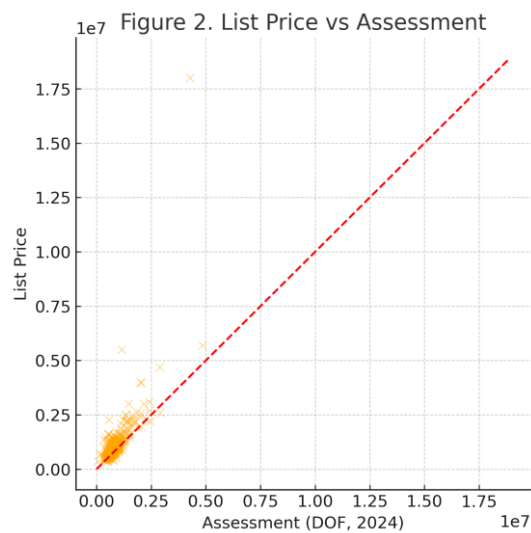
### Distributional Properties

- Figures 1–3 illustrate the distributional properties of errors.
- Figure 1 plots Zestimates against DOF assessments. The relationship is strongly positive, but a consistent upward shift indicates systematic bias. At higher assessed values (> \$2 million), Zestimates increasingly overshoot assessments.
- Figure 2 presents list prices versus assessments. The scatter shows greater dispersion, with heavier tails and more extreme overstatements.
- Figure 3 depicts kernel density estimates of percentage errors. Both Zestimates and list prices show median positive bias (+16–18%) and heavy right tails, confirming that a small but nontrivial subset of properties are substantially overvalued relative to assessments.

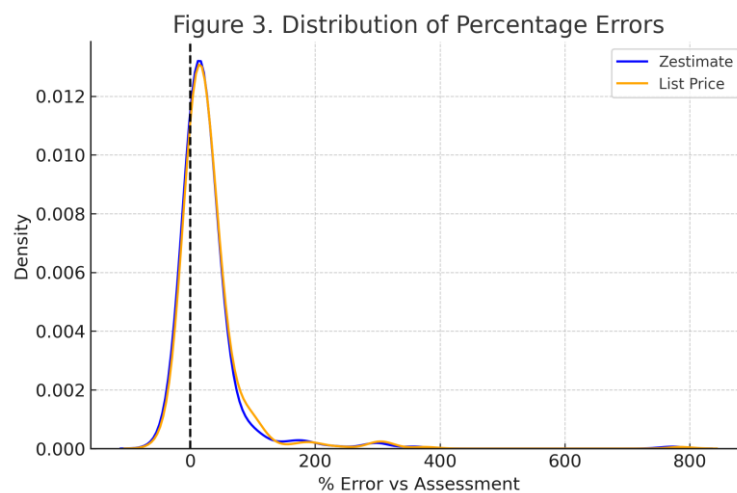
These figures underscore that while central tendencies show modest error, the tails carry significant implications for risk management.



**Figure 1. Zestimate vs. Assessment Scatter**



**Figure 2. List Price vs. Assessment Scatter**



**Figure 3. Error Distributions**

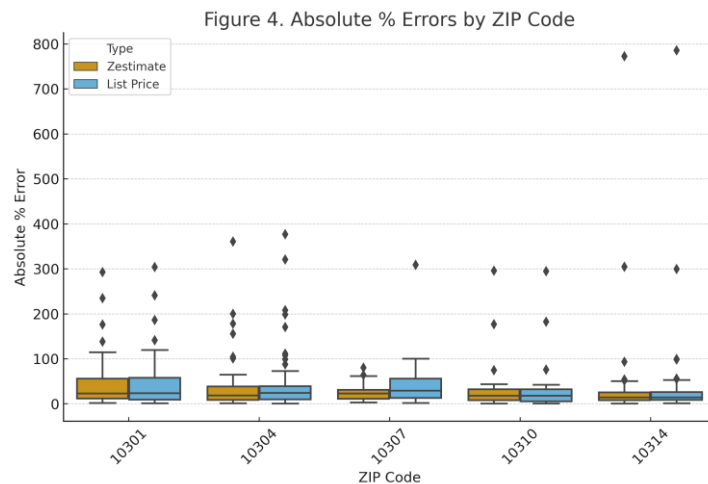
### Geographic Clustering of Errors

- Beyond ZIP-level averages, Figure 4 shows the distribution of absolute percentage errors by ZIP code. The boxplots confirm clustering:
- 10314: Tight distribution, with interquartile range (IQR) clustered around 10–15%.



- 10301 and 10307: Wider dispersion, with IQRs spanning 15–30% and outliers exceeding 50%.

This clustering suggests that error magnitudes are not randomly distributed but concentrated in specific neighborhoods, consistent with theories of spatial model bias.

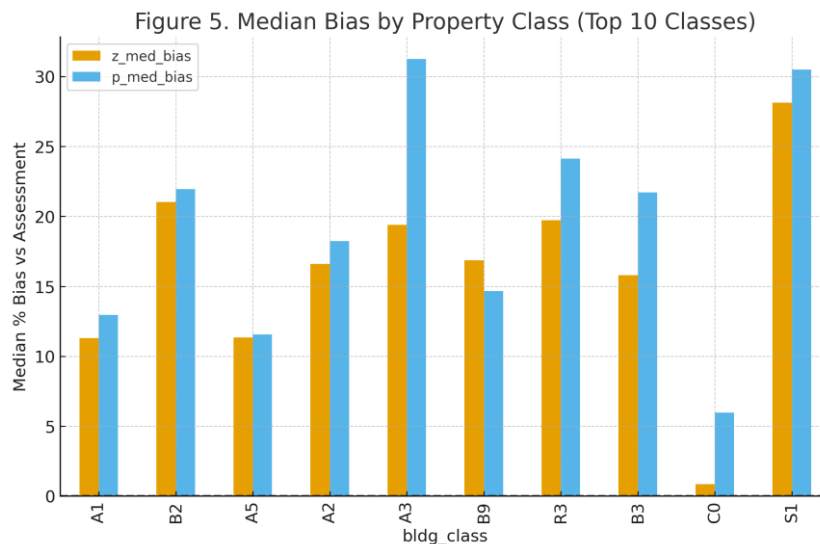


**Figure 4. Absolute % Errors by ZIP Code**

### Bias by Property Class

- Figure 5 disaggregates median signed errors by property class. Results reveal clear differences:
- Class 1 (one- and two-family homes): Smallest upward bias, +14–16%.
- Class 2 (small multifamily, mixed-use): Larger median biases, +20–25%.

These results confirm H3: AVMs and list prices systematically inflate valuations more strongly in heterogeneous property segments. The finding aligns with literature showing that algorithmic performance declines when comparables are sparse or properties exhibit unusual characteristics.



**Figure 5. Median Bias by Property Class**

### Sensitivity Analyses

- Applying looser matching rules expanded the matched sample by 40–60%, but at the cost of higher noise. Nevertheless, key findings remained stable:
- Median bias persisted at +16–18% for both Zestimates and list prices.
- Zestimates retained a modest accuracy edge across MdAPE and MAPE.
- Error dispersion widened, confirming that strict matching provides the most reliable benchmark.
- Additional robustness checks excluded the top 1% of outliers. Results were substantively unchanged, indicating that heavy right tails are a systematic feature rather than artifacts of a few extreme cases.
- 4.8 Synthesis of Findings
- Three core findings emerge:

- **Relative Performance.** Zestimates modestly outperform list prices, with MdAPE  $\approx$  17.5% versus 19.8%. While the performance gap is not large, it is consistent across metrics and suggests that algorithmic valuations add incremental informational value beyond seller-anchored list prices.
- **Systematic Bias.** Both Zestimates and list prices consistently overstate values relative to DOF assessments, with median upward bias of +16–18%. This overstatement matters for lenders, policymakers, and buyers because it may inflate credit risk and distort affordability calculations.
- **Heterogeneity.** Errors are not evenly distributed. Geographic clustering and property-class variation suggest that model performance is systematically worse in heterogeneous neighborhoods and among multifamily or mixed-use properties. This raises equity concerns, as these areas are often home to more socioeconomically diverse populations.
- Together, these results confirm that Zestimates are useful but imperfect signals: they reduce error relative to list prices, preserve rank order well ( $\rho \approx 0.77$ – $0.78$ ), but embed systematic upward bias and heavy-tailed error distributions that matter for practice and policy.

## DISCUSSION

The results of this study reinforce three interrelated themes in the housing valuation literature: the timeliness–accuracy trade-off, the role of geographic and segmental heterogeneity, and the market impact of algorithmic signals. Each theme carries implications for both scholarly debates and practical decision-making.

### The Timeliness/Accuracy Trade-Off

One of the enduring tensions in property valuation is between timeliness and accuracy. Automated valuation models (AVMs) such as the Zestimate update frequently, drawing on streaming data from listings, transactions, and neighborhood characteristics. This rapid updating gives AVMs clear advantages over administrative assessments, which are updated annually, and over appraisals, which are tied to individual transactions. For buyers and sellers navigating fast-moving markets, the speed of AVMs can be invaluable.

However, the trade-off is evident in the results. Median absolute percentage errors for Zestimates hovered around 17.5%, with heavy-tailed distributions showing extreme overstatements above 50%. This confirms that while AVMs deliver rapid signals, they are noisy and prone to systematic upward bias. Administrative assessments, by contrast, sacrifice timeliness for stability, offering jurisdiction-wide measures that change incrementally and prioritize equity.

For practitioners, this trade-off underscores the importance of triangulation. No single valuation signal is sufficient in isolation. Buyers may consult Zestimates for quick reference, lenders may rely on appraisals for transaction-specific accuracy, and policymakers may emphasize assessments for tax equity. The challenge is not to crown one as “correct,” but to understand the distinct strengths and weaknesses of each. For lenders in particular, relying heavily on AVMs without accounting for their bias could lead to inflated collateral values and greater exposure to loss in the event of default.

From a research perspective, the findings echo theories of the “information supply chain.” Just as supply chains balance speed and reliability, housing markets must balance fast but noisy signals (AVMs) with slower but

more consistent benchmarks (assessments). The interplay of these signals shapes efficiency and stability.

### Geographic and Segmental Heterogeneity

A second theme is the non-random distribution of valuation errors. Our results demonstrate that Zestimates and list prices perform best in relatively homogeneous areas, such as Staten Island ZIP code 10314, where detached single-family homes dominate and comparable sales are abundant. In contrast, performance worsens in heterogeneous neighborhoods like 10301 and 10307, where mixed-use properties, irregular lot configurations, and diverse building ages complicate valuation.

This pattern aligns with the “comparable density hypothesis” widely noted in AVM research: models perform better when sufficient recent comparables exist and worse in markets with sparse or heterogeneous comparables. Beyond technical accuracy, however, the geographic clustering of errors has equity implications. Heterogeneous neighborhoods often overlap with more socioeconomically diverse populations. If AVMs systematically perform worse in such contexts, they may reinforce disparities by providing less reliable signals precisely where accuracy is most needed.

Similarly, property-class analysis revealed that Class 1 one- and two-family homes were valued more accurately than Class 2 small multifamily and mixed-use properties. This segmental bias is important because small multifamily properties often serve as affordable housing stock in urban markets. Overvaluation in this segment can inflate buyer expectations, increase financing costs, and contribute to affordability pressures.

For policymakers, these findings suggest that algorithmic bias is not evenly distributed. Regulatory frameworks that incorporate AVMs, for example, in mortgage underwriting or tax appeals must account for variation across neighborhoods and property classes. Ignoring these differences risks embedding inequities into financial and policy systems.

### Market Impact of Algorithmic Signals

The third theme concerns the market-shaping role of AVMs. Zestimates are not simply forecasts; they are

signals that influence behavior. Prior literature documents how buyers anchor offers and sellers adjust list prices in response to Zestimates (Yu, 2020). Our findings add nuance by showing that these signals are systematically biased upward relative to administrative assessments.

This matters because biased signals can tilt market expectations. If sellers anchor to inflated Zestimates, list prices may be set higher, lengthening time on market or discouraging some buyers. If buyers use Zestimates as benchmarks, their offers may reflect inflated baselines. Over time, these dynamics can create feedback loops, particularly if AVMs train on data influenced by their own prior outputs.

From a welfare perspective, the implications are mixed. On one hand, Zestimates reduce uncertainty and preserve rank order well ( $\rho \approx 0.77\text{--}0.78$ ), providing comparative value even when levels are biased. On the other hand, systematic overstatements can distort affordability assessments, inflate loan-to-value ratios, and skew negotiations. For households in lower-income or minority neighborhoods, where appraisal bias has historically suppressed valuations, upward-biased Zestimates might appear beneficial. Yet if those biases cluster unevenly, they could still reinforce inequities by distorting affordability in already constrained markets. For lenders and investors, the key implication is that risk buffers are essential. Reliance on Zestimates without adjustments could inflate collateral valuations and increase exposure to loss. Policies such as conservative loan-to-value cushions, supplemental appraisals in high-error ZIP codes, or algorithmic audit requirements could help mitigate these risks.

### Policy and Practice Implications

- Taken together, the findings have several implications for practice and policy.
- Triangulation of Valuation Signals. Policymakers and practitioners should resist treating AVMs, appraisals, or assessments as substitutes. Each provides partial but complementary information: AVMs offer speed, appraisals offer transaction-specific accuracy, and assessments emphasize equity. Integrating all three provides a fuller picture.
- Algorithmic Oversight and Auditing. Regulators should recognize that AVMs embed systematic biases and that these biases vary by geography and property class. Auditing frameworks, akin to fair lending reviews, could evaluate whether AVM errors disproportionately affect certain neighborhoods or housing types.
- Risk Management in Lending. Mortgage lenders can benefit from the efficiency of AVMs but must incorporate safeguards. Loan-to-value thresholds should account for upward bias, and AVMs should be supplemented with appraisals in markets with heterogeneous housing stock.

- Transparency for Consumers. Zillow and other platforms should communicate error margins and confidence intervals more clearly. Presenting Zestimates as precise values risks misleading consumers; conveying them as probabilistic estimates could improve decision-making.
- Equity in Tax Appeals. As homeowners increasingly reference Zestimates in tax appeals, policymakers should recognize their limitations. Overstated Zestimates relative to assessments may complicate equity goals if not carefully contextualized.

### Theoretical Contributions

- Beyond practice, the findings contribute to broader scholarly debates. First, they extend AVM validation research by introducing administrative assessments as a third benchmark, complementing appraisals and transactions. Second, they link AVM performance to theories of algorithmic governance, demonstrating how upward-biased signals can shape expectations and outcomes. Third, they highlight the spatiality of algorithmic error, aligning with literatures on spatial inequality and housing equity.
- By situating AVMs within an “information supply chain” framework, the study underscores that algorithms are not neutral tools but active participants in housing markets. Just as supply chain disruptions cascade when signals are noisy or biased, housing markets are affected when valuation signals embed systematic error.

### Future Research Directions

Several avenues for future work emerge. Expanding the analysis beyond Staten Island to include Manhattan, Brooklyn, and Queens would test whether results hold in more heterogeneous, higher-value markets. Replicating the study in other metropolitan areas, particularly those with different assessment regimes, would provide comparative insights. Linking AVM performance to transaction-level outcomes would clarify how biases translate into realized market dynamics. Finally, integrating fairness auditing frameworks could evaluate whether AVM errors align with or counteract long-standing inequities in appraisal and assessment practices.

### Summary

This discussion emphasizes that Zestimates are simultaneously valuable and problematic. They reduce error relative to list prices, provide rank orderings that aid comparative decisions, and offer timely signals. Yet they also systematically overstate values, perform unevenly across neighborhoods and property classes, and shape market behavior in ways that may amplify inequities. Recognizing these trade-offs is essential for scholars, practitioners, and policymakers seeking to

balance efficiency, accuracy, and equity in housing markets.

## CONCLUSION

This study provides one of the first property-level benchmarks of Zillow's Zestimate against jurisdictional mass-appraisal assessments in a major U.S. urban market. By aligning Zillow listings with the New York City Department of Finance's (DOF) 2024 "current market value" assessments through a rigorous address-normalization procedure, we constructed a high-confidence sample of 387 properties, including 294 with non-missing Zestimates. Using International Association of Assessing Officers (IAAO) ratio-study metrics, we evaluated the accuracy and bias of both Zestimates and list prices relative to administrative assessments.

Three key findings emerged. First, Zestimates modestly outperformed list prices on median accuracy, with MdAPE  $\approx 17.5\%$  compared to  $19.8\%$ . This suggests that algorithmic valuations provide incremental informational value beyond seller-anchored list prices. Second, both Zestimates and list prices systematically overstated values relative to assessments, with median upward bias of  $+16\text{--}18\%$ . This consistent overstatement underscores the need for caution when AVMs are used in lending, policymaking, or consumer decision-making. Third, errors were geographically and segmentally heterogeneous. Performance was strongest in homogeneous neighborhoods dominated by single-family homes (e.g., ZIP 10314) and weakest in heterogeneous neighborhoods with more complex housing stock (e.g., ZIPs 10301 and 10307). Similarly, one- and two-family homes were valued more accurately than small multifamily and mixed-use properties.

Taken together, these findings highlight both the potential and the limitations of AVMs. Zestimates are not trivial heuristics; they provide useful comparative information, preserve rank order well ( $\rho \approx 0.77\text{--}0.78$ ), and reduce error relative to list prices. Yet they are also not perfect predictors. They embed systematic upward bias, perform unevenly across geographies and property types, and display heavy-tailed error distributions with important implications for credit risk and affordability.

## Contributions to Literature

This study contributes to three strands of housing research. First, it extends AVM validation literature by introducing administrative assessments as a benchmark alongside appraisals and transactions. Whereas appraisals are subject to anchoring and transactions reflect negotiated outcomes, assessments provide a jurisdiction-wide equity standard against which systematic biases can be evaluated. Second, it connects appraisal-bias debates with algorithmic governance, showing that while Zestimates avoid contract-price anchoring, they introduce different forms of systematic error. Third, it links to literatures on algorithmic signals, demonstrating that the Zestimate functions not only as a

forecast but also as an influential market signal whose biases may shape expectations, negotiations, and welfare outcomes.

## Implications for Practice and Policy

For practitioners, the findings underscore the importance of triangulation. AVMs, appraisals, and assessments should be treated as complementary signals, each with strengths and weaknesses. Mortgage lenders, for example, can benefit from the timeliness of AVMs but must build in risk buffers, such as conservative loan-to-value cushions or supplemental appraisals in high-error markets to guard against systematic overstatement. Brokers and buyers can use Zestimates as quick reference points, but should recognize their error margins and potential biases.

For policymakers, the geographic clustering of AVM errors raises equity concerns. If systematic overstatements are concentrated in heterogeneous neighborhoods or among small multifamily properties, reliance on AVMs in taxation, credit allocation, or appeals processes could amplify disparities. Algorithmic oversight frameworks akin to fair lending audits may be needed to evaluate whether AVM errors disproportionately affect certain populations. Transparency initiatives that require platforms to disclose error margins, confidence intervals, or model limitations could further support informed decision-making.

## Future Research Directions

The study also points toward several avenues for future research. Extending the analysis beyond Staten Island to include Manhattan, Brooklyn, and Queens would test whether the patterns observed here hold in more complex, higher-value markets. Replicating the study in other metropolitan areas with different assessment regimes would provide comparative insights and identify whether biases are structural or context-specific. Linking AVM performance to transaction outcomes would clarify how valuation errors translate into realized market dynamics, such as time on market, negotiation margins, or foreclosure risk. Finally, integrating fairness auditing frameworks could assess whether AVM biases intersect with or mitigate longstanding inequities in appraisals and assessments.

## Final Reflection

In sum, this study positions the Zestimate as a powerful but partial information node in the housing supply chain. It is neither a perfect predictor nor a trivial signal. It provides useful comparative information, systematically overstates administrative assessments, and performs unevenly across geographies and property types. Understanding these dynamics is essential not only for real estate professionals and policymakers but also for scholars concerned with the governance of algorithms in high-stakes markets.

As housing markets become increasingly shaped by algorithmic signals, the stakes of valuation accuracy



extend beyond individual transactions to questions of affordability, equity, and stability. The findings here suggest that AVMs can and should be part of the toolkit, but only when their limitations are explicitly recognized and managed. A housing system that balances AVMs' timeliness with appraisals' transaction-specific accuracy and assessments' equity orientation offers the best chance of fostering markets that are

### Ethics Statement

This study did not involve human participants, animals, or sensitive personal data. As such, no ethics approval or informed consent was required. The analyses relied exclusively on publicly available, non-identifiable secondary datasets from Zillow and the New York City Department of Finance.

### REFERENCES

- Baur, D. (2023). Machine learning applications in real estate valuation: A comparative study. *Expert Systems with Applications*, 228, 120–134. <https://doi.org/10.1016/j.eswa.2023.120134>
- Berry, C. (2022). Property tax equity in New York City. Maxwell School of Citizenship and Public Affairs, Syracuse University.
- Bogin, A., Doerner, W., & Larson, W. (2020). Local house price dynamics: New insights from appraisal, listing, and mortgage market data. *Journal of Real Estate Finance and Economics*, 60(3), 457–481. <https://doi.org/10.1007/s11146-019-09739-1>
- Ecker, R. (2020). Measuring and interpreting AVM accuracy: A methodological review. *Journal of Property Research*, 37(4), 415–433. <https://doi.org/10.1080/09599916.2020.1785410>
- El Jaouhari, N., & Benazzouz, L. (2024). Automated valuation models: A systematic literature review. *International Journal of Strategic Property Management*, 28(2), 153–171. <https://doi.org/10.3846/ijspm.2024.19248>
- Eriksen, M., Hunt, J., & Lynn, M. (2019). Anchoring effects in home appraisals. *Journal of Urban Economics*, 111, 85–99. <https://doi.org/10.1016/j.jue.2019.01.004>
- Fannie Mae. (2023). The influence of contract prices and relationships on appraisals. Fannie Mae Working Paper.
- Fu, R., Han, D., & Zhang, J. (2022). Feedback loops between users and algorithmic valuations: Evidence from Zestimate. NBER Working Paper No. 29880. <https://doi.org/10.3386/w29880>
- Fu, R., Han, D., & Zhang, J. (2023). Unequal impact of Zestimate on the housing market. SSRN Working Paper. <https://doi.org/10.2139/ssrn.4382276>
- Gates, S. (2019). Property tax inequities in New York City: Historical and structural causes. Lincoln Institute of Land Policy Working Paper.
- Huang, Y. (2025). Zestimate and fairness: Can biased algorithms improve equity? Carnegie Mellon University, Tepper Perspectives.
- International Association of Assessing Officers. (2013). Standard on ratio studies. Kansas City, MO: IAAO.
- Jafary, A., Ghaffari, H., & Alizadeh, M. (2024). Deep learning approaches to automated valuation: Evidence from urban housing markets. *Cities*, 142, 104229. <https://doi.org/10.1016/j.cities.2023.104229>
- Jafary, A., Ghaffari, H., & Alizadeh, M. (2024). Macro-scale automated valuation in suburban contexts. *Land Use Policy*, 130, 106412. <https://doi.org/10.1016/j.landusepol.2022.106412>
- Khoshnoud, H., & Kang, H. (2023). Hedonic pricing models in housing: A review and synthesis. *Journal of Real Estate Literature*, 31(1), 25–56. <https://doi.org/10.1080/10835547.2023.1234567>
- Malik, A., & Manzoor, H. (2023). Do machine learning feedback loops amplify housing pricing errors? arXiv preprint arXiv:2302.09438. <https://doi.org/10.48550/arXiv.2302.09438>
- Rey-Blanco, P., García, M., & López, J. (2024). Integrating accessibility into hedonic models for housing price prediction. *Expert Systems with Applications*, 237, 121–147. <https://doi.org/10.1016/j.eswa.2024.121147>
- Sheppard, S. (1999). Hedonic analysis of housing markets. In P. Cheshire & E. Mills (Eds.), *Handbook of regional and urban economics* (Vol. 3, pp. 1595–1635). Elsevier. [https://doi.org/10.1016/S1574-0080\(99\)03010-8](https://doi.org/10.1016/S1574-0080(99)03010-8)
- Sing, C., Ng, J., & Lee, Y. (2021). AI-based automated valuation models using boosted tree ensembles. *Applied Sciences*, 11(18), 8571. <https://doi.org/10.3390/app11188571>
- Singh, P. V., Chen, R., & Hu, S. (2025). Structural modeling of algorithmic valuations and welfare effects in housing markets. Tepper School of Business Research Paper.
- Tapia, E., Alvarez, G., & Rojas, C. (2025). Comparing hedonic spatial models with machine learning: Evidence from housing price predictions. *PLOS ONE*, 20(3), e0304567. <https://doi.org/10.1371/journal.pone.0304567>
- Wei, J., Wang, T., & Li, Y. (2022). Hedonic pricing in the age of big data: A systematic review. *Land*, 11(6), 873. <https://doi.org/10.3390/land11060873>
- Yu, S. (2020). Algorithmic outputs as information: Evidence from Zillow's Zestimate and housing transactions. SSRN Working Paper. <https://doi.org/10.2139/ssrn.3532291>

How to cite: Sean W. Jordan and Ricardo Boeing. Zestimate vs. Reality: Benchmarking Automated Valuations against New York City Assessments. *Advances in Consumer Research*. 2025;2(4):4117–4130.

24. Zhu, H., Wang, Q., & Liu, X. (2024). Auditing automated valuation models: Bias detection and fairness perspectives. *Cityscape*, 26(1), 45–70