Tradeoffs are Domain Dependent: Improving Accuracy and Fairness in Property Tax Assessments

ANONYMOUS AUTHOR(S)

Algorithmic fairness research often assumes a tradeoff between fairness and accuracy. Yet this tradeoff may not be universal. We test this assumption in the context of U.S. property tax assessment—a setting in which the output of predictive algorithms directly determines the distribution of tax obligations among homeowners. Currently, systematic assessment errors cause owners of lower-valued properties to face disproportionately high tax burdens, creating regressivity in the property tax system. Using data on 26 million property sales spanning 95% of U.S. counties, we conduct three complementary analyses. First, we find that assessment accuracy and fairness—measured using domain-relevant metrics—are strongly correlated across counties under status quo practices. Second, in simulated assessment models, we show that adding property features improves accuracy in most cases, and that when accuracy improves, fairness almost always improves as well. Third, we show that incorporating publicly available Census data into assessment models—a feasible reform in most counties—would significantly improve both accuracy and fairness relative to status quo assessments. Together, these results challenge the presumed universality of the fairness–accuracy tradeoff and demonstrate that well-designed modeling improvements can advance both fairness and accuracy in large-scale public sector systems.

CCS Concepts: • **Social and professional topics → Geographic characteristics**; • **Applied computing → Law**; **Economics**.

Additional Key Words and Phrases: vertical equity, property tax, fairness, fairness-accuracy tradeoff, regression

ACM Reference Format:

Anonymous Author(s). 2026. Tradeoffs are Domain Dependent: Improving Accuracy and Fairness in Property Tax Assessments. 1, 1 (January 2026), 24 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

For the past decade, algorithmic fairness researchers have grappled with the *fairness-accuracy tradeoff* [21, 56, 65, 71], empirically identified across domains including criminal justice [1, 18], child welfare [19], hiring [73], and advertising [57]. The widespread presence of this tradeoff has led to an implicit assumption that it is inherent: that, in most real-world settings, building fairer models means forfeiting accuracy [20]. Here, we present evidence from one domain—previously unstudied in the algorithmic fairness literature—where improvements in accuracy are accompanied by systematic and substantial gains in fairness: property taxes.

The vast majority of Americans pay property taxes. Property owners pay them directly (to the government), and renters pay them indirectly (via rents to their landlords). Property taxes are the primary source of revenue for local governments and are used to pay for public services and infrastructure like schools, roads, and fire departments [60]. To collect a property tax, taxing jurisdictions (e.g., U.S. counties) *assess* every property in their tax area by determining its fair market value, and then levy taxes in proportion to that assessed value.[1] While the property tax is intended

[1]We use *assessment* to refer to the process of determining the fair market value of a property for tax purposes. We use *appraisal* to refer to the process of determining the fair market value of a property for non-tax purposes (e.g., real estate transactions).

to be a *proportional tax* [33], multiple studies have shown that it is *regressive* in practice: individuals who own lower-valued properties pay a higher proportion of their values in taxes than individuals who own higher-valued properties [4, 5, 12, 61]. Regressivity disproportionately penalizes non-white Americans, who are more likely to own lower-valued properties due to both wealth inequality and historical segregation in the housing market [41, 66]. Because regressive property taxes are not proportional to home value, they violate the important tax policy principle of *vertical equity*, which requires that "different individuals be treated appropriately differently" [15].

Regressivity has many potential causes [12]. In some cases, it is driven by corruption or other forms of malfeasance. For example, assessors in Cook County, Illinois accepted bribes in exchange for reducing assessments of high-value properties, prompting an FBI investigation and jail time for some conspirators [35–38, 74]. In other cases, legal restrictions on assessments drive regressivity. For example, 17 U.S. states place caps on assessment increases [8, 22–24]. These caps are intended to protect individual homeowners from large hikes in tax bills, but also shift tax obligations from homeowners whose properties have appreciated quickly to those whose values have grown at slower rates [43, 75]. Regressivity has also been leveraged as a tool to suppress social movements. For example, following boycotts and pickets organized by Black residents of Edwards, Mississippi in 1966, the local assessor increased the assessed values of Black-owned homes by ten times as much as white-owned homes, with the largest increases assigned to boycott participants and organizers. [53].

However, the most significant sources of assessment regressivity today are shortcomings in the data and models used to assess properties [4, 12]. Most assessors conduct simplistic assessments based on data that can be collected at scale, without needing to inspect properties in person [14]. In contrast, the fair market value of a property is defined according to its sale price in an arms-length transaction.[2] In such transactions, buyers and sellers have access to a rich set of information, including in-person tours, inspections, and appraisals, which they use to negotiate a sale price. Because most assessors rely on comparatively sparse data, their assessments can regress to the mean market values in their jurisdiction. This results in low-value properties being, on average, over-assessed and over-taxed and high-value properties being, on average, under-assessed and under-taxed. Thus, regressivity occurs even in the absence of corruption or legal restrictions.

In this paper, we assess the hypothesis that more accurate assessments generally lead to fairer (i.e., less regressive) property taxes—in other words, that property tax assessment is not typically subject to the fairness-accuracy tradeoff. To examine this hypothesis, we take a holistic view of real-world assessment pipelines and consider whether efforts to increase assessment accuracy generally lead to increases in fairness. In doing so, we make the following key contributions:

- We describe the relationship between accuracy and regressivity in the context of property tax assessments using real-world data on 26 million property sales and assessments spanning 95% of U.S. counties over five years, finding that, across jurisdictions, assessment accuracy is positively associated with taxation fairness. This pattern suggests that methodological improvements in assessment practice—of which machine-learned approaches are a prominent example—may yield joint gains in accuracy and fairness.
- Using simulated assessment models, we find that adding additional property characteristics increases assessment accuracy in most cases, and that accuracy and fairness gains coincide more than 99% of the time.
- We propose simple computational approaches to improve assessment accuracy and fairness that can be applied to assessment processes nationwide. In particular, we estimate that the addition of publicly available Census characteristics can significantly improve assessment accuracy and fairness in hundreds of U.S. counties and may also reduce regressivity with respect to race and income.

[2]An *arms-length transaction* is one where the buyer and seller are not known to one another.

Our analysis builds on an established literature studying the distribution of assessment errors with respect to property value [e.g., 4, 12, 45, 59], income [28, 62], and race [5]. Prior work has mostly explored the effects of specific changes in assessment method on both accuracy and regressivity within a single jurisdiction [e.g., 34, 47, 59].[3]. We contribute to this literature by studying the relevance of fairness-accuracy tradeoffs in practice, using data drawn from the universe of U.S. taxing jurisdictions and variation in algorithmic methods that are realistic and feasible for jurisdictions to implement. We also identify the conditions under which gains in assessment accuracy are likely to come at the expense of regressivity, and examine heterogeneity across counties in accuracy and fairness responses to changes in assessment methodology.

We proceed as follows. In §2, we provide background on property tax assessments in the U.S. as well as on the fairness-accuracy tradeoff. We outline our methods in §3 and describe our empirical results in §4. We close by discussing the implications of our findings, avenues for future work, and limitations of our approach in §5.

2 Background and Related Work

2.1 The United States Property Tax Assessment System

In the U.S., property tax assessments are conducted at the local (e.g., county) level. Local assessors typically rely on one of three approaches to estimate the fair market value of a property. The *cost* approach values a property according to how much it would cost to reproduce it, less depreciation; the *sales* approach values a property commensurate with similar properties that have recently been sold; and the *income* approach, used for commercial or rental properties, values a property according to the income it produces [50, 58]. Local governments have discretion to conduct assessments as they see fit, subject to state-level regulation of assessment processes or outcomes. For example, as of 2025, 17 states required that local assessors cap year-over-year assessment increases; 27 states required that assessors conduct in-person property inspections as part of their assessments; and 39 states imposed limits on the time between reassessments [8, 23, 24].

While researchers have repeatedly found that regression-based assessments, which are also referred to as Automated Valuation Models (AVMs), outperform alternative methods of assessment [13, 25, 52, 72], a 2019 International Association of Assessing Officers (IAAO) survey of assessor's offices found that just 16% of survey respondents use AVMs [14].[4] The remainder rely on a combination of *valuation tables*, which provide estimates of the marginal value of property attributes (e.g. additional square footage or bedrooms) and may be either provided by the state or purchased from private vendors; individual property assessments using comparables selected by hand; or other methods. The low adoption rate of AVMs has several causes. Roughly one in five IAAO survey respondents point to capacity concerns, such as budget constraints or the lack of technical expertise necessary to properly implement an AVM. One in four cite the relative difficulty in explaining AVM predictions to taxpayers. However, the most-cited barrier to adoption is the belief that current approaches work just as well as AVMs, if not better—despite considerable empirical evidence to the contrary [14].

Furthermore, according to a 2018 survey conducted by the IAAO, only one-third to one-half of local assessors have incorporated fine-grained location or neighborhood features into their assessment analyses [3]. While traditional assessment models focus only on characteristics of a property itself (e.g., size or age), incorporating geographic data improves assessment by capturing location (e.g., proximity to schools) and neighborhood (e.g., sociodemographics of residents) information that is likely to affect a property's market value [25].

[3]An important exception is an independent working paper by Amornsiripanitch [4], which studies how incorporating the housing price index constructed by Bayer et al. [7] into property tax assessments affects regressivity and accuracy within a pooled sample of all U.S. taxing jurisdictions.

[4]The vast majority of assessor's offices do not release information about their assessment approaches to the public. Consequently, status quo assessment practices are largely opaque, with the best data available coming from surveys conducted by the International Association of Assessing Officers (IAAO). While these surveys are the best source of data for understanding the status quo, we caution that they are not representative.

2.2 The Fairness-Accuracy Tradeoff

Fair ML research often casts training fair models as a constrained optimization problem: the goal is to maximize *accuracy* (usually operationalized as predictive error) subject to a *fairness* constraint (usually operationalized as differences in model error rates between groups) [9, 18, 21, 32, 40, 54–56, 65, 82, 83]. This framing presumes the existence of a fairness-accuracy tradeoff, except under unrealistic circumstances, i.e., the presence of a perfectly accurate model [56, 71]. Because a perfectly accurate model is generally not achievable in practice, the most accurate achievable model produces incorrect predictions for at least some instances. In theory, a tradeoff is inevitable if the most accurate achievable model produces different errors across groups: if accuracy cannot be increased for the worst-off group, then the only way to achieve fairness is to reduce accuracy for the better-off groups. However, counter to the prevailing assumption in the fair ML literature that accuracy and fairness are mutually incompatible, an emerging strain of research has found that tradeoffs do not always arise in practice [20]. For example, researchers have achieved simultaneous fairness and accuracy gains across a variety of modeling tasks by collecting additional data [17] and by accounting for bias in existing data [27, 80].

2.3 Conceptual Framework

In theory, accuracy gains in property tax assessment guarantee neither fairness gains nor fairness losses. For example, consider the case where accuracy is improved by correcting errors exclusively in low-value properties that are over-valued by an assessment model: this will reduce regressivity. Alternatively, if accuracy is improved by correcting errors exclusively in low-value properties that are under-assessed, this will increase regressivity.

For at least one pair of accuracy and fairness metrics, the relationship between accuracy and fairness depends on the relative magnitudes of prediction variance and mean bias in a given assessment model. Formally, for two assessment models (A and B), if accuracy is measured using mean squared error (MSE) between ground truth sale price and predicted sale price, and fairness according to the Log Coefficient regressivity metric described in Table 1, then:

$$Fairness_B - Fairness_A = \frac{1}{2}\left[(Var_B - Var_A) + [(MeanBias_A^2 - MeanBias_B^2)] - [MSE_B - MSE_A]\right]$$

To see that the relationship between fairness and accuracy in this setting is ambiguous, assume that models A and B have the same mean prediction ($MeanBias_A = MeanBias_B$) and that model B has lower MSE ($MSE_B < MSE_A$). Then, if model B's predictions are more variable than model A's, accuracy and fairness will improve simultaneously. If model B's predictions are less variable than model A's then, depending on the magnitude of the difference in accuracy compared to the magnitude of the difference in variance, there may be a fairness-accuracy tradeoff. A formal proof of this result is provided in Appendix C.

As with other domains, a perfectly accurate model—one that sets the assessed value of each home exactly equal to its market value—is also a perfectly fair model, as it produces a proportional property tax. However, whether imperfect assessment models eventually reach a Pareto frontier wherein accuracy gains can only be achieved with corresponding decreases in fairness is an open question. *In this work, we explore empirically whether the fairness-accuracy tradeoff exists for property tax assessment, examining whether the types of accuracy improvements that are feasible in practice improve or undermine fairness.*

3 Data and Methods

3.1 Data

Our data include property and transfer records purchased from the data broker Cotality.[5] The unit of observation is a property transfer, which we filter to include only arms-length sales of single-family homes. Consistent with common

[5]Cotality was formerly called CoreLogic, and is widely used in economics research for research on property taxes and property transfers.

practice, we exclude sales of distressed or foreclosed properties under the assumption that foreclosure auctions do not reveal a property's underlying market value. For each sale, we observe transaction data (e.g., type of deed, date of sale, sale amount). We then merge sale data with property characteristics (e.g., square footage, number of bedrooms, year built), location data (e.g., county, census block group), and tax information (e.g., assessed value at time of sale). Finally, we merge in publicly available socioeconomic data from the American Community Survey (ACS) 5-year estimates. These data include block group-level neighborhood information (e.g., education levels, marital status, median age and income, unemployment rates, and share of households using SNAP). Overall, this results in a dataset of 26M sales spanning the period 2018 to 2023 and covering transactions in 2,844 out of 3,007 U.S. counties.

3.2 Counterfactual Assessment Models

Using the data described above, we build a suite of LASSO and random forest-based counterfactual AVMs. These models use property and neighborhood data to predict property sale prices, simulating the assessment process. Models are developed at the county level, using only within-county sales. We employ a chronological train-test split, such that models are trained on historical sales data and tested using the most recent year of sales, in order to align with real-world assessment practices [e.g., 68]. To preprocess the data, we first one-hot encode categorical features, encoding only categories that are present in at least 5% of observations. We then drop features whose values are missing for more than 50% of observations, and use multiple imputation by chained equations (MICE) [6] to impute any remaining missing values. Finally, we winsorize all numeric features at the 1st and 99th percentiles and subsequently normalize them. We conduct hyperparameter tuning using Bayesian optimization and five-fold cross validation, with mean absolute error between predicted and actual log sale price determining the CV score.

3.3 Evaluating Assessment Models

We evaluate each assessment model in terms of its accuracy and fairness, as measured on the held-out most recent year of sales and assessment data for each county.

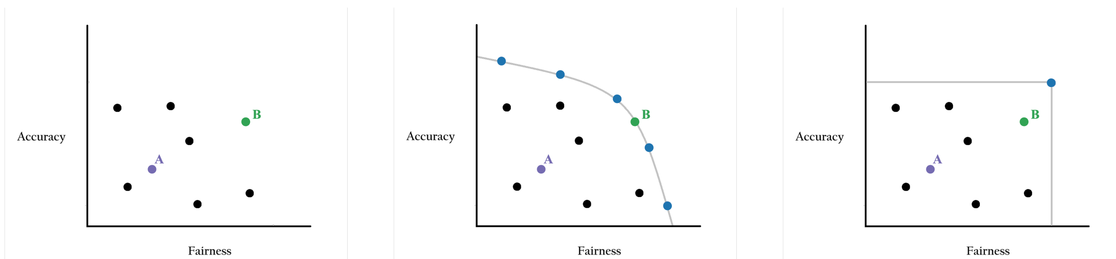
Accuracy. An accurate assessment is one that matches the true market value of a property, which is assumed to be equal to the sale price of a property in an arms-length transaction. We therefore measure accuracy as the difference between a property's assessed value in a given year (as produced by an assessment model) and its actual sale price in that year. As our primary accuracy metric, we use mean absolute percentage error (MAPE), because the percentage error between assessed and actual sale price is equivalent to the percentage difference between effective and statutory tax rates in a simple setting with no exemptions or deductions to the tax (see Appendix B). To ensure robustness, we also evaluate model accuracy using root mean squared error (RMSE) and mean absolute error (MAE).

Fairness. We define fairness as the level of regressivity produced under a particular assessment model. Regressivity occurs when a higher proportion of the tax burden falls on the owners of lower-valued properties. It is commonly used within the field of economics to measure the vertical equity of tax systems [4, 15]. Local governments, professional associations, and researchers have collectively promulgated several competing measures of assessment regressivity [61]. As we discuss in detail in §5, no metric is without shortcomings and individual metrics may not always agree with one another. Therefore, we rely on three separate regressivity metrics throughout this analysis which together cover the primary approaches currently in use. We describe each metric briefly in Table 1, and in detail in Appendix A.

Pareto Frontier Framework. To evaluate model performance, we use the framework of the *Pareto frontier*. In multi-objective optimization problems, it is often the case that objectives are in tension with one another and cannot be simultaneously maximized. The tradeoffs among the best-performing models trace out a frontier (Fig. 1b). Outcomes along the frontier are *Pareto optimal*, in the sense that no other outcome is possible which offers improvements with

Metric	Approach	Formula	Regressivity Threshold	Metric Source
Log Coefficient (LC)	Regression-Based	β_1, where $\log(R) = \beta_0 + \beta_1 \log(S)$ and $R = \frac{A}{S}$.	< 0 (significantly)	[4, 12]
Suits Index (SI)	Distribution-Based	$1 - \frac{L}{K}$, where L is the area under the curve produced by graphing the cumulative distribution of A against the cumulative distribution of S and K is the area under an analogous curve produced by a perfectly proportional tax	< 0 (significantly)	[63, 76]
Price-Related Differential (PRD)	Ratio-Based	$\frac{\bar{R}}{\bar{A}/\bar{S}}$, where $\bar{R}$ is the average $\frac{A}{S}$, $\bar{A}$ is the average A, and $\bar{S}$ is the average S	> 1.03 (strictly)	[2]

Table 1. The regressivity metrics used in our analysis. Each metric compares the sale prices (S) and assessed values (A) of properties within a given geographic area in a given year.



(a) Current State (b) Pareto Frontier (c) Pareto Point

Fig. 1. We evaluate model performance using a Pareto framework. Each point represents a model. When model accuracy and fairness are low (point **A**), as is currently the case in property tax assessments, simultaneous improvements in accuracy and fairness (point **B**) are possible. This holds regardless of whether **B** represents a point on the Pareto frontier or is a uniquely best-performing Pareto point.

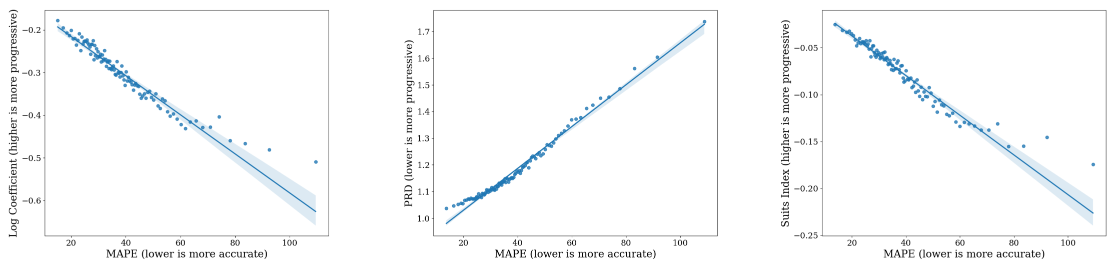
respect to both objectives. If no tradeoff exists between the two objectives, then the Pareto frontier collapses to a single point, corresponding to a unique model that strictly outperforms all other models (Fig. 1c). Such a model constitutes a *Pareto improvement* over previous models, as it offers simultaneous gains with respect to both objectives.

Empirically, we examine whether counterfactual assessment models offer Pareto improvements—simultaneous gains in fairness and accuracy—relative to baseline predictions from either status quo assessments or other counterfactual models. If these modeling adjustments yield Pareto improvements relative to these baselines, then we posit that this is evidence against the practical importance of a fairness-accuracy tradeoff in property tax assessments.[6]

4 Results

In this section, we characterize the status quo relationship between accuracy and fairness in property tax assessment, finding that even though the relationship between accuracy and fairness is theoretically ambiguous, in practice the two measures are closely related. We first show that counties with more accurate status quo assessments also tend to have more equitable assessments (§4.1). Next, we find that adding more information to hypothetical assessment models leads to accuracy gains in a majority of cases, and accuracy gains coincide with fairness gains over 99% of the time. Finally, we show that, by incorporating publicly available census data in assessment models, simultaneous accuracy and fairness gains are possible relative to status-quo assessments in 14.4% of U.S. counties without caps on annual assessment growth (§4.2).

[6] Of course, the observed compatibility of fairness and accuracy gains relative to the baselines we consider does not imply that no such tradeoffs could ever occur; for example, they might arise if the status quo baseline were substantially improved (as depicted in Fig. 1b). Our focus is on the practical importance of the fairness-accuracy trade-off in the context of the methods actually used to generate assessments in real-world settings.



(a) Log Coefficient (b) PRD (c) Suits Index

Fig. 2. Higher accuracy (as measured by MAPE) in property tax assessments is correlated with higher fairness. The figures show scatterplots of fairness and accuracy by county-year between 2018 and 2023, binned across 100 quantiles of county-level MAPE.

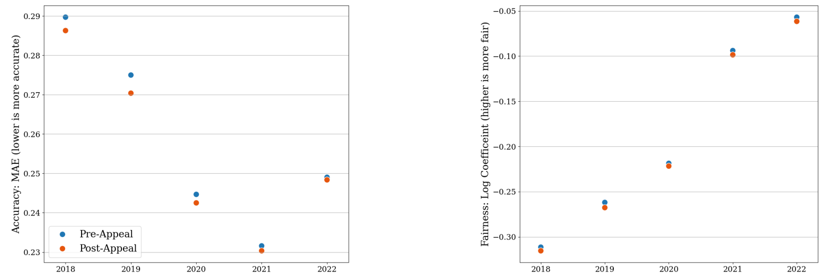
4.1 The Accuracy-Fairness Relationship

4.1.1 Assessment Accuracy and Fairness are Highly Correlated, and Both are Low for Many Counties. As shown in Fig. 2, accuracy and fairness in property tax assessment are strongly correlated, and the relationship is robust across fairness metrics. In general, counties with a higher median and average sale price ($R = -0.23$ and $R = -0.17$, respectively), sale volume ($R = -0.19$), housing stock ($R = -0.18$), and population ($R = -0.17$) have lower status quo assessment errors ($p < 0.001$ in all cases). As a result, average county-level accuracy varies across states: states with a higher proportion of rural counties such as Oklahoma, Arkansas, and Missouri tend to have higher assessment error on average. Somewhat surprisingly, county-level variance in sale price is not associated with assessment accuracy, implying that assessments are not on average higher-quality in counties with more homogenous housing markets. Almost all counties—98% between 2018 and 2023—had regressive assessments as measured by LC (i.e., LC < 0).

4.1.2 Accuracy Gains Do Not Guarantee Fairness Gains. Despite the positive correlation between accuracy and fairness in property tax assessment, more accurate assessments are not necessarily fairer. We illustrate this theoretically in Appendix C, and consider empirical tradeoffs here through the example of assessment appeals.

Improvements in accuracy can increase regressivity if accuracy gains are primarily clustered among previously over-assessed high-value properties. This commonly occurs in the real world in the context of assessment appeals [64, 79]. If a homeowner believes that their assessment is inaccurate, they can file an appeal with a county governing body. Appeals are disproportionately filed by owners of higher-valued properties, who have higher property tax bills at stake and are likely to have more resources available to navigate the appeals process. Because appeals typically reduce assessed values, the appeals system functions as an adversarial correction, increasing accuracy primarily for over-assessed, high-value properties while leaving other prediction errors untouched. The end result is that post-appeal assessments can be both more accurate and more regressive relative to pre-appeal assessments. The relationship between accuracy and fairness depends on the concentration of accuracy gains across the distribution of housing prices and on the direction of the errors that are corrected. Fig. 3 demonstrates this empirically for Cook County, IL, where between 2021 and 2023, owners of homes in the highest quintile of sale price appealed their assessments at four times the rate of those in the lowest quintile [69].

4.1.3 However, we do not observe fairness-accuracy tradeoffs in 99.5% of cases. To further examine the relationship between assessment accuracy and fairness, we develop simulated assessment models using property and owner transfer



(a) Accuracy (b) Fairness

Fig. 3. Assessment appeals can impose accuracy-fairness tradeoffs. The figures show pre- and post-appeal accuracy (as measured by MAPE) and fairness (as measured by Suits Index) among single-family home sales in Cook County from 2018-2022.

data from Cotality, examining how fairness and accuracy evolve under a good-faith attempt to improve model performance. Overall, we find that accuracy gains coincide with fairness gains in over 99% of cases. We hypothesize that models with access to more information about properties will be more accurate and equitable than models with less information. To test this, we build two LASSO models for each county according to the procedures described in §3. The first model includes three property characteristics ("sparse model"), and the second includes as many property characteristics as are available in Cotality for that county ("rich model"). Both models include six features describing the timing of each sale. Because data availability varies across counties, the property characteristics used in one county's model may differ from those in another county's model. Additionally, the difference in the number of property characteristics used in the sparse and rich model for each county ranges from 1 to 12, depending on the county. We subset to counties with at least 100 sales in our sample period (2018-2023), and to counties with a sufficient number of features available to train both sparse and rich models, leaving us with an initial sample of 2,059 counties.

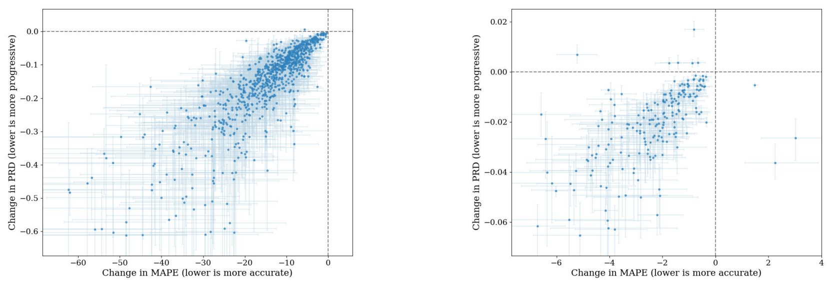
Our results are provided in Fig. 4a. We observe significant changes in both accuracy and fairness in 1,033 out of 2,059 counties (50.1%). In, 1,028 of these 1,033 counties (99.5%), we observe simultaneous gains in fairness and accuracy, while only 5 exhibit tradeoffs. Moreover, counties with larger accuracy gains also tended to have larger fairness gains. Appendix D shows the robustness of this result to alternative choices of accuracy and fairness metrics (Fig. 6).

The simple linear models developed here may be far enough below the Pareto frontier that the positive relationship we observe between fairness and accuracy is not sufficient evidence against fairness-accuracy tradeoffs in practical settings. To address this concern, in the next section, we rely on status quo assessments as our benchmark, and observe whether simultaneous gains in fairness and accuracy are possible relative to valuation methods currently in use.

4.2 Improving Property Tax Assessments Using Publicly Available Census Data

Housing prices are sensitive not only to the characteristics of homes, but also to the socioeconomic characteristics of the area surrounding a home, including local job opportunities and demographic characteristics such as age and household structure. The U.S. Census Bureau regularly disseminates estimates of these neighborhood-level characteristics, and at least one assessor's office use such data in their valuation models.[7] However, to our knowledge, this practice is not widespread, and guidance on the use of Census data in valuation models is not currently included in the technical standards promulgated by the IAAO [50, 51]. Moreover, Berry [12] identifies substantial within-county heterogeneity in assessment quality that is correlated with tract-level features such as median household income, education, and home values,

[7] Cook County, Illinois, whose assessment model documentation is publicly available: github.com/ccao-data/model-res-avm.



(a) Ablation Experiment

(b) Census Experiment

Fig. 4. Adding more information to assessment models generally leads to accuracy and fairness gains.

Notes: The figures show changes in accuracy and fairness for county-level assessment models across two experiments. The left panel shows the changes between the "sparse" and "rich" county-level LASSO-based assessment models described in Section 4. Models are trained using property and time-of-sale characteristics. The right panel shows changes in accuracy and fairness after the addition of Census block group data to county-level random forest-based assessment models. The baseline model uses only status quo predictions and time-of-sale data, while the alternative model incorporates these features in addition to Census block group characteristics. In both panels, accuracy is measured using MAPE between assessed and sale price. Fairness is measured using PRD. Changes in performance are statistically significant at a 95% confidence level after applying a Benjamini-Hochberg multiple testing correction. Confidence intervals are computed using a studentized bootstrap and displayed in light blue. Results are censored at the 2nd and 98th percentiles.

while Amornsiripanitch [4] finds that Census block-group characteristics improve assessment accuracy in aggregate across all counties. Both studies indicate that there is potential for these features to improve county-level model fit.

We examine whether Census data could improve assessment accuracy and fairness using publicly available data from the 2023 Census 5-year ACS, described in §3. We train random forest models for each county that leverage Census data alongside status-quo assessed values and time-of-sale characteristics to predict sale price. We then examine county-level changes in fairness and accuracy relative to predictions from benchmark random forest models which use only status-quo assessments and time-of-sale data. To account for statutory limitations which may impact the accuracy of status quo assessments, we drop counties located within 17 U.S. states with assessment caps [23, 24].[8] We also restrict our sample to counties with at least 100 sales in our sample period, leaving us with an initial sample of 1,459 counties.

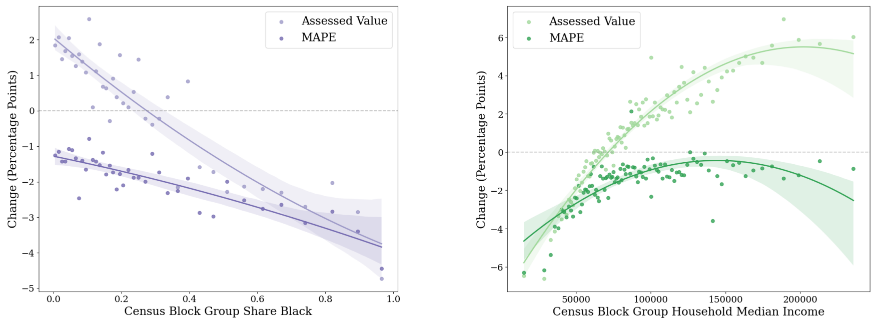
Our results are provided in Fig. 4b. We find that the addition of census features leads to statistically significant changes in fairness and accuracy for 210 out of 1,459 counties (14.4%). Among those counties, 200 (95.2%) show simultaneous fairness and accuracy gains, and only 10 (4.8%) exhibit tradeoffs. The predominance of simultaneous fairness-accuracy improvements is robust to the choice of metric, as shown in Appendix D (Fig. 7).

4.3 Census Data Also Improves Outcomes By Race and Income

While the predominant focus of this paper is regressivity with respect to sale price, there are other important and policy relevant dimensions of fairness to consider, such as race and income. We find preliminary evidence that incorporating Census characteristics in assessment models can reduce regressivity along these dimensions as well. Because these changes are reported with respect to characteristics measured at the block group level, they may be subject to aggregation bias. As such, these results are not conclusive, but are rather intended to provide a basis for future work.

4.3.1 Race. At the block group level, the addition of Census features in assessments has substantial effects on neighborhoods with relatively high shares of Black residents.[9] As shown in Fig. 5a, Census block groups with more than ∼30% Black residents see decreases in assessed value, on average, after the addition of Census features, while block groups

[8] These states include Alabama, Arizona, Arkansas, California, Florida, Georgia, Hawaii, Iowa, Louisiana, Maryland, Michigan, New Mexico, New York, Oklahoma, Oregon, South Carolina, and Texas. The annual caps range from 2 to 15%.
[9] The Census features used here and throughout the paper do not include race or ethnicity.



(a) Race

(b) Income

Fig. 5. Census features lower assessed values for predominantly Black and low-income neighborhoods and increase accuracy overall.

Notes: The figure shows changes in assessed value and MAPE relative to time-adjusted status-quo assessments after the addition of Census block group features in assessment models. Changes are calculated at the block group level, and observations are grouped by bins of the share of Census block group residents that are Black (a) and household median income (b). Lines of best fit are calculated using second-degree polynomials.

with fewer Black residents tend to see increases in assessed value. Recall that because assessments are proportional to tax obligations, these results indicate that predominantly Black neighborhoods would see reductions in tax obligations, on average, from the addition of Census features in assessment models. Census features lead to across-the-board gains in accuracy, with slightly larger gains in percentage terms for predominantly Black neighborhoods.

4.3.2 Income. With respect to income, as shown in Fig. 5b, the addition of Census features tends to reduce assessments in lower-income neighborhoods and increase assessments in higher-income neighborhoods, with adjustments roughly proportional to income levels. Accuracy improves across the board, with the largest reductions in error in percentage terms concentrated among low-income neighborhoods. These results indicate that the addition of Census features could shift assessed values and corresponding property tax obligations up the income distribution, making the tax more progressive with respect to income as well as property values.

5 Discussion

Our findings provide strong evidence against fairness-accuracy tradeoffs in property tax assessments relative to the current frontier of assessment models. While we show that tradeoffs are theoretically and empirically possible, we also find that the practical importance of these tradeoffs is limited. Rather, improvements in model accuracy tend to coincide with fairness gains, a conclusion which is robust across the vast majority of U.S. taxing jurisdictions and to various combinations of domain-relevant accuracy and fairness metrics. Moreover, we demonstrate that Pareto improvements are possible relative to the frontier of status quo models and predictions for a significant portion of counties in our data, and that these improvements can be achieved by incorporating publicly available Census data into assessment models. These findings lie in sharp contrast to other settings—predominantly, but not exclusively, involving classification—where fairness-accuracy tradeoffs have been identified. In those settings, it is often the case that enforcing notions of fairness, such as statistical parity or equalized odds, necessitates reductions in predictive accuracy, either empirically or through the joint incompatibility of fairness and accuracy as modeling objectives [1, 10, 18, 19, 57, 73].

The key distinguishing features of our setting are, first, a focus on vertical equity (treating different individuals appropriately differently) as opposed to horizontal equity (treating similar individuals from different groups in similar ways), and second, our conception of fairness, which compares the distribution of assessment errors relative to ground-truth sale prices. In this framework, tradeoffs are theoretically possible, but they are not implied by the mutual incompatibility of fairness and accuracy, and we do not observe them empirically in the overwhelming majority of cases. This implies that our findings could generalize to other forms of wealth taxation beyond the property tax, such as taxes on inherited wealth and overall assets which are prevalent in countries outside the U.S. [26]. We also find preliminary evidence that the modeling improvements we propose here could reduce regressivity with respect to race and income.

5.1 Limitations and Extensions

In this section, we discuss limitations of our analysis caused by modeling choices as well as by inherent challenges to measuring fairness and accuracy in property tax assessment at scale. Throughout, we highlight opportunities for future work.

5.1.1 Data Access and Representativeness. Most counties do not publicly release data on property sales and assessments. As a result, we relied on brokered data from Cotality to conduct a national analysis. However, this brokered data has been shown to contain a non-representative sample of SFH sales when compared to county-level property records [42]. As a result, while evidence of regressivity exists in both brokered data and county-level records, the levels of regressivity depend on the source. In addition, we chose to exclude counties with fewer than 100 sales during our sample period from our analysis (3.9% of counties). We note that our suggestions to improve assessment quality may not be as effective in counties with lower sales volumes, as those counties have a sparsity of data to train assessment models. Future work might consider the possibility of data sharing between counties, and the relative tradeoffs, if any, associated with performing assessments at higher units of government, such as the state level, which may help alleviate data sparsity at the local level.

More fundamentally, the standard approach for evaluating assessment quality (which we employ) is to compare assessed values to sale prices for the population of homes that sell in a given year. However, those homes may differ in important ways from the overall population of assessed properties, and these differences can bias estimates of assessment quality. For example, homes that sell frequently may correspond to investment properties with newer finishes than otherwise similar homes, leading to an upward bias in assessed values, particularly in neighborhoods experiencing gentrification.

5.1.2 Does Sale Price Represent the True Value of a Home? Throughout this analysis, we have assumed that sale price represents the true value of a home. However, sale prices may deviate from true values for a number of reasons. For example, there may be idiosyncratic differences in negotiating skill across buyers and sellers, which can contribute to regressivity [12]. Sale price and true home values may also differ for more systematic reasons. Racial biases on the part of buyers, sellers, and other market actors, as well as historical legacies of redlining and disinvestment from non-white neighborhoods, can introduce racial disparities in sale prices that are unrelated to the observable characteristics of a home. Indeed, Perry et al. [70] find that, in appraisals conducted at time of sale, properties in majority-Black neighborhoods are devalued by as much as 23% relative to comparable homes in majority non-Black neighborhoods, even after controlling for property and neighborhood characteristics. Race-based devaluation of homes both within and across neighborhoods is an important social problem with implications for wealth accumulation and intergenerational wealth transfer. To the extent that property tax assessments fail to account for race-based differences in home values, they can engender or worsen racial wealth gaps. Currently, this means that Black and Hispanic residents pay an estimated 10 and 13% more in property taxes, respectively, for the same bundle of public services, after controlling for taxing jurisdiction, income, and property characteristics [5]. Encouragingly, we find that the inclusion of Census block group characteristics in assessment models can identify a portion of these differences, increasing accuracy and lowering average assessments in predominantly Black neighborhoods. This suggests that assessors can achieve Pareto improvements in accuracy and fairness in aggregate with respect not only to home values, but also race, at little to no cost.

Addressing what remains of the racial gap in property tax assessments after accounting for neighborhood-level features is a more challenging task. Property tax assessors could theoretically include homeowner race as an explicit feature in assessment models, recognizing the contextual value of race in predicting home prices. However, property tax assessors typically do not know individual homeowners' race, and would therefore be forced to rely on imputed measures of race, which introduce additional complications [29, 30]. In addition, under prevailing anti-discrimination law, government entities may engage in race-conscious decisionmaking only if they can demonstrate a "strong basis in evidence" that disparities exist, that they are personally accountable for creating them, and that corrective action is necessary and justified [44]. This is a high evidentiary standard. Recent court decisions have trended towards an interpretation of the law that would bar the explicit inclusion of race in assessment models regardless of the intended purpose. A more promising approach for interventions may therefore be to establish a firmer understanding of the underlying mechanisms driving bias in sale price and intervene directly at those levels. For example, if racial biases drive wedges between market value and sale price for homes sold by minorities, then jurisdictions might consider blinded sales. While the examination of accuracy-fairness tradeoffs in property tax assessments with respect to homeowner race is not the central focus of this paper, we provide evidence at the aggregate level which suggests that improving assessment accuracy can help reduce racial assessment disparities. Directions for future work include examining the impact of individual-level race and other demographic data on model performance, and assessing the legal and procedural challenges of including such data as explicit features of assessment models.

5.1.3 No Standard Measure of Regressivity. Throughout this paper, we have estimated regressivity by comparing assessed values to sale prices. However, there is no agreed-upon 'best' method for doing this—while the IAAO recommends relying on ratio-based approaches [2], economists have argued that those approaches can inflate estimates of regressivity [63]. Different regressivity metrics can disagree with one another on whether regressivity is present, further complicating measurement [63]. Furthermore, there are several other dimensions of regressivity that may be of interest to assessors and policymakers. For example, as an alternative to assessed values, one could examine property tax bills, which differ from assessments to the extent that taxpayers claim the credits and deductions offered by their taxing jurisdictions. In particular, many jurisdictions offer homestead exemptions to taxpayers who live in the property being taxed, as well as tax credits to seniors. Such exemptions can, but do not always, offset assessment regressivity [48, 61]. One could also examine the regressivity of assessments or property tax bills relative to household income, as opposed to property sale price. This is the traditional approach that economists take when estimating the progressivity of a tax, and we provide preliminary evidence in this paper that reducing assessment regressivity with respect to sale price can also reduce statutory regressivity with respect to block-group level median incomes. Estimating tax regressivity with respect to individual-level incomes, however, is more complex, requiring the linking of administratively separated data sets and methodological decisions over the appropriate measure of income (annual versus lifetime), incidence (statutory versus economic), and conception of the property tax itself (capital tax versus fee for service) [67].

5.2 Recommendations

We close by making several recommendations for the literature on accuracy-fairness tradeoffs and for tax assessors and policymakers seeking to improve the fairness and accuracy of property tax assessments.

5.2.1 Recommendations for Studies of Fairness-Accuracy Tradeoffs. Our paper engages with two important critiques of the existing literature on fairness–accuracy tradeoffs: first, that the data and metrics employed in this work are too often disconnected from real-world applications and domain-specific conceptions of equity [77]; and second, that the bias mitigation strategies found to impose tradeoffs are often ad hoc and intervene at just a single point in the modeling pipeline [16]. We find that, in a concrete ML application of practical import, and with a holistic approach that spans the entire modeling pipeline and applies domain-relevant fairness metrics, tradeoffs are the exception rather than the rule, and that there exist substantial opportunities for Pareto improvements relative to the status quo.

Our findings do not negate the statistical and theoretical validity of fairness–accuracy tradeoffs identified in previous work. Indeed, tradeoffs may arise in assessments under alternative modeling choices, baselines, or fairness definitions that abstract away from how domain experts have historically conceptualized equity. However, our results suggest that the practical relevance of these tradeoffs depends critically on examining the full modeling pipeline, using domain-appropriate metrics, and evaluating performance relative to models currently in use. Greater attention to domain-relevant notions of fairness can help the fair ML literature more accurately identify when tradeoffs have real-world significance.

5.2.2 Recommendations for Assessment Methodology

Collect More High-Quality Data on Properties. One of the fundamental challenges of property tax assessment is that assessors often have less information about properties than buyers and sellers. Assessors typically cannot see the interiors of properties and instead rely only on information they can gather from properties' exteriors and from permitting records. We find that adding more information to assessment models generally leads to Pareto improvements in model performance. While there are concrete limits to the amount of information assessors can collect, there are likely also opportunities for many local governments to improve the coverage and quality of the data at assessors' disposal. Our findings indicate that more frequent and comprehensive field surveys, alongside investments in data validation, would in most cases lead to to better and more equitable assessments. However, not every county may be in a position to invest additional resources into collecting and maintaining more current and comprehensive data.

Incorporate Census Block Group Characteristics In Assessments. We find that roughly 14.4% of counties in states without assessment caps could improve the accuracy and uniformity of their assessments by adding Census block group features into their assessment models. This data is high quality, publicly available, and can be incorporated into assessment pipelines at little to no cost. While counties should independently evaluate the impact of Census data on model performance, in many cases this data is likely low-hanging fruit for counties looking to improve assessments.

Use Tree-Based Automated Valuation Models As discussed in Section 2, only a small minority of assessor's offices have adopted AVMs [14], despite a wealth of evidence indicating that AVMs outperform simpler assessment methods [13, 25, 52, 72]. Moreover, among offices that have adopted AVMs, linear regressions remain the model of choice, even though tree-based models generally outperform linear models in predicting housing prices [11, 31, 46, 81]. In the real world, counties that have adopted tree-based AVMs have seen noteworthy accuracy and fairness improvements. For example, Cook County, Illinois, which publicly overhauled its assessment process after a high-profile investigation documented widespread property tax regressivity [35–38, 74], has used a LightGBM-based assessment model since 2019. In that time, Cook County has reduced property tax obligations by over one billion dollars for homeowners in the bottom 70% of the housing price distribution, and has also come into compliance with industry standards regarding assessment uniformity [11]. We recommend that assessor's offices consider adopting AVMs, and tree-based AVMs in particular.

Release Open Data on Sales and Assessments. Publicly available data on sales and assessments, alongside transparent documentation of assessment methodology, would provide all stakeholders—policymakers, academic researchers, and ordinary citizens alike—the opportunity to understand how property tax bills are determined and how assessments might be improved. More transparency surrounding county-level assessment practices could also help paint a more thorough picture of the current frontier of assessment practices and assist counties in learning from one another about new approaches to assessment. Moreover, open data would enable researchers to validate the coverage and accuracy of brokered datasets, which form the basis for much of the existing research on property tax assessments. Cook County, which publicly releases its assessment data and modeling details, can serve as a template here.[10]

5.2.3 Recommendations for Assessment Policy

Consider Implementing Book-Tax Conformity for Assessments. "Book-tax conformity" refers to the concept that corporations should report the same income for tax purposes that they report on their financial statements, with the idea that doing so will encourage firms to reduce tax avoidance [39]. The analog in the property tax setting would be conformity between assessed values and appraised values/sale prices. Homeowners would prefer that their assessed values remain as low as possible to minimize their property tax bills, but may prefer to have higher appraised values and sale prices to maximize their proceeds from the sale of a home, or to secure home equity loans. Aligning assessed values with appraisals and sale prices would exploit this tension and eliminate regressivity for homes that sell. It could also reduce racial disparities in regressivity: Black-owned homes have been shown to be consistently under-appraised at time of sale [70] even as they are over-assessed in the property tax system [5], as discussed in detail in §5.1.2. Of course, these advantages apply only to homes that are appraised or that sell. Moreover, book-tax conformity could cause discrete jumps in tax obligations upon sale that introduce frictions in the housing market. Ultimately, the potential benefits of book-tax conformity in assessments are context dependent.

Reconsider Statutory Limits on Assessments A majority of states regulate the way assessments are conducted and the impact of assessments on property tax bills. In particular, 17 U.S. states impose caps on the year-over-year growth in assessed values [8, 23, 24, 78]. These caps limit the impact that improved assessment methodology can have on property tax regressivity. While more accurate predictions can lower assessed values and tax bills for over-assessed (and disproportionately low-value) properties, assessment caps constrain the extent to which these tax obligations can be shifted to under-assessed (and disproportionately high-value) properties, undermining the impact of assessment methodology on vertical equity. Additionally, 15 U.S. states only require counties to re-assess properties every five to six years, while an additional 10 states enact no requirements at all on the frequency of re-assessments [23]. Even if initial assessments are equitable, differences in market dynamics across neighborhoods and property types can drive differential wedges between assessed and market value over time [12]. As such, whether an assessment model is accurate or not matters less and less as the period between assessments grows. While improvements in assessment methodology are a promising avenue for improving the vertical equity of the property tax in many jurisdictions, state-level regulations on assessments will inevitably mediate how much modeling improvements can reduce assessment regressivity.

6 Conclusion

This paper examines whether property tax assessments are subject to fairness-accuracy tradeoffs, and finds that good-faith efforts to increase assessment accuracy generally lead to fairer, more uniform assessments, even though tradeoffs are both theoretically and empirically possible. In particular, adding Census block group characteristics to assessment models could significantly increase both assessment accuracy and fairness, with accuracy and fairness gains coinciding in 95% of cases. This finding stands in stark contrast to many other settings where fairness-accuracy tradeoffs have been examined, and contributes to a small but growing literature examining tradeoffs in regression settings. We provide recommendations for policymakers and assessors and suggest directions for future work, including examining other dimensions of fairness and addressing data and measurement challenges in the assessment space. Fundamentally, our work challenges the assumption that fairness and accuracy are necessarily in tension, and shows that machine learning can jointly advance both objectives, revealing a wider range of domains where fairness comes for free.

[10] github.com/ccao-data/model-res-avm

Generative AI Usage Statement

We used ChatGPT (GPT-5) for minor editing (specifically, suggesting words to cut to reduce the length of the abstract) and to assist in the generation of figures by modifying existing code to improve formatting. We did not use generative AI for any other purpose.

References

[1] 2016. Machine Bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2] 2017. STANDARD ON Mass Appraisal of Real Property: A criterion for measuring fairness quality, equity and accuracy. International Association of Assessing Officers. https://www.iaao.org/wp-content/uploads/StandardOnMassAppraisal.pdf

[3] 2018. *Summary Findings of IAAO's GIS Technology Survey for Assessment Professionals.* Technical Report. International Association of Assessing Officers, Kansas City, MO. https://www.iaao.org/wp-content/uploads/GIS_SURVEY2018.pdf

[4] Natee Amornsiripanitch. 2022. *Why Are Residential Property Tax Rates Regressive?* Working paper (Federal Reserve Bank of Philadelphia) 22-02. Federal Reserve Bank of Philadelphia. 22–02 pages. doi:10.21799/frbp.wp.2022.02

[5] Carlos F Avenancio-León and Troup Howard. 2022. The Assessment Gap: Racial Inequalities in Property Taxation. *The Quarterly Journal of Economics* 137, 3 (July 2022), 1383–1434. doi:10.1093/qje/qjac009

[6] Melissa J. Azur, Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf. 2011. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research* 20, 1 (2011), 40–49. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mpr.329 doi:10.1002/mpr.329

[7] Patrick Bayer, Marcus Casey, Fernando Ferreira, and Robert McMillan. 2017. Racial and ethnic price differentials in the housing market. *Journal of Urban Economics* 102 (2017), 91–105.

[8] Joyce Beebe, John W. Diamond, and Jennifer Rabb. 2025. The Surprising Impact of the 20% Appraisal Cap in Texas. *Center for Tax and Budget Policy Research Paper* (2025). https://www.bakerinstitute.org/research/surprising-impact-20-appraisal-cap-texas

[9] Richard Berk, Hoda Heidari, Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. 2017. A Convex Framework for Fair Regression. In *FAT ML (FATML)*. FAT ML, Halifax, Nova Scotia, Canada. doi:10.48550/arXiv.1706.02409

[10] Richard Berk, Hoda Heidari, Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. 2017. A convex framework for fair regression. *arXiv preprint arXiv:1706.02409* (2017).

[11] Christopher Berry. 2025. *An Evaluation of Progress on Residential Assessment Fairness in Cook County.* Technical Report. Mansueto Institute for Urban Innovation, University of Chicago, Chicago, IL. https://bpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/6/2330/files/2025/09/Kaegi-Evaluation-9_04.pdf

[12] Christopher R. Berry. 2021. Reassessing the Property Tax. *SSRN Electronic Journal* (March 2021). doi:10.2139/ssrn.3800536

[13] Paul E Bidanset, Michael McCord, John R Lombard, Peadar Davis, and William McCluskey. 2017. Accounting for locational, temporal, and physical similarity of residential sales in mass appraisal modeling: the development and application of geographically, temporally, and characteristically weighted regression. *Journal of Property Tax Assessment & Administration* 14, 2 (2017), 5–13.

[14] P. E. Bidanset and Ron Rakow. 2022. Survey on the use of automated valuation models (AVMs) in government assessment offices: An analysis of AVM use, acceptance, and barriers to more widespread implementation. *Journal of Property Tax Assessment & Administration* 19, 2 (Nov. 2022). doi:10.63642/1357-1419.1250

[15] Emily Black, Hadi Elzayn, Alexandra Chouldechova, Jacob Goldin, and Daniel Ho. 2022. Algorithmic Fairness and Vertical Equity: Income Fairness with IRS Tax Audit Models. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. ACM, Seoul Republic of Korea, 1479–1503. doi:10.1145/3531146.3533204

[16] Emily Black, Rakshit Naidu, Rayid Ghani, Kit Rodolfa, Daniel Ho, and Hoda Heidari. 2023. Toward operationalizing pipeline-aware ML fairness: A research agenda for developing practical guidelines and tools. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–11.

[17] Irene Chen, Fredrik D Johansson, and David Sontag. 2018. Why Is My Classifier Discriminatory?. In *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.), Vol. 31. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2018/file/1f1baa5b8edac74eb4eaa329f14a0361-Paper.pdf

[18] Alexandra Chouldechova. 2016. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. In *Fairness, Accountability, and Transparency in Machine Learning (FAT ML) 2016 (FATML)*. arXiv. doi:10.48550/arXiv.1610.07524

[19] Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAT*)*. PMLR, 134–148. https://proceedings.mlr.press/v81/chouldechova18a.html

[20] A. Feder Cooper and Ellen Abrams. 2021. Emergent Unfairness in Algorithmic Fairness-Accuracy Trade-Off Research. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES)*. Association for Computing Machinery, New York, NY, USA, 46–54. doi:10.1145/3461702.3462519

[21] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. Algorithmic Decision Making and the Cost of Fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. Association for Computing Machinery,

New York, NY, USA, 797–806. doi:10.1145/3097983.3098095

[22] Alan S Dornfest, Jennifer Rearich, T Douglas Brydon III, and Richard Almy. 2019. State and Provincial Property Tax Policies and Administrative Practices (PTAPP): 2017 Findings and Report. *Journal of Property Tax Assessment & Administration* 16, 1 (2019). https://doi.org/10.63642/1357-1419.1211

[23] Alan S Dornfest, Jennifer Rearich, T Douglas Brydon III, and Richard Almy. 2019. State and Provincial property tax policies and administrative practices (PTAPP): 2017 findings and report. *Journal of Property Tax Assessment & Administration* 16, 1 (2019), 43–130.

[24] Alan S Dornfest, Steve Van Sant, and Rick Anderson. 2014. State and provincial property tax policies and administrative practices (PTAPP): 2012 update of 2009 compilation and report. *Journal of Property Tax Assessment & Administration* 11, 3 (2014), 15–116.

[25] Gabriela Droj, Anita Kwartnik-Pruc, and Laurențiu Droj. 2024. A comprehensive overview regarding the impact of GIS on property valuation. *ISPRS International Journal of Geo-Information* 13, 6 (2024), 175.

[26] Marcus Drometer, Marco Frank, Maria Hofbauer Pérez, Carla Rhode, Sebastian Schworm, and Tanja Stitteneder. 2018. Wealth and inheritance taxation: An overview and country comparison. *ifo DICE Report* 16, 2 (2018), 45–54.

[27] Sanghamitra Dutta, Dennis Wei, Hazar Yueksel, Pin-Yu Chen, Sijia Liu, and Kush Varshney. 2020. Is There a Trade-Off Between Fairness and Accuracy? A Perspective Using Mismatched Hypothesis Testing. In *Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 119)*, Hal Daumé III and Aarti Singh (Eds.). PMLR, 2803–2813. https://proceedings.mlr.press/v119/dutta20a.html

[28] Robert H Edelstein. 1979. An appraisal of residential property tax regressivity. *Journal of Financial and Quantitative Analysis* 14, 4 (1979), 753–768.

[29] Hadi Elzayn, Emily Black, Patrick Vossler, Nathanael Jo, Jacob Goldin, and Daniel E. Ho. 2024. Estimating and Implementing Conventional Fairness Metrics With Probabilistic Protected Features. In *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. 161–193. doi:10.1109/SaTML59370.2024.00016

[30] Hadi Elzayn, Evelyn Smith, Thomas Hertz, Cameron Guage, Arun Ramesh, Robin Fisher, Daniel E Ho, and Jacob Goldin. 2024. Measuring and Mitigating Racial Disparities in Tax Audits*. *The Quarterly Journal of Economics* 140, 1 (09 2024), 113–163. arXiv:https://academic.oup.com/qje/article-pdf/140/1/113/60122543/qjae027.pdf doi:10.1093/qje/qjae027

[31] Gang-Zhi Fan, Seow Eng Ong, and Hian Chye Koh. 2006. Determinants of house price: A decision tree approach. *Urban Studies* 43, 12 (2006), 2301–2315.

[32] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and Removing Disparate Impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. Association for Computing Machinery, New York, NY, USA, 259–268. doi:10.1145/2783258.2783311

[33] Glenn Fisher. 2002. History of Property Taxes in the United States. EH.Net Encyclopedia. https://eh.net/encyclopedia/history-of-property-taxes-in-the-united-states/

[34] Vincent J Geraci. 1977. Measuring the benefits from property tax assessment reform. *National Tax Journal* 30, 2 (1977), 195–205.

[35] Jason Grotto. 2017. The Tax Divide | An era of errors. Chicago Tribune. https://apps.chicagotribune.com/news/watchdog/cook-county-property-tax-divide/houlihan.html

[36] Jason Grotto. 2017. The Tax Divide | An unfair burden. Chicago Tribune. https://apps.chicagotribune.com/news/watchdog/cook-county-property-tax-divide/assessments.html

[37] Jason Grotto. 2017. The Tax Divide | The problem with appeals. Chicago Tribune. https://apps.chicagotribune.com/news/watchdog/cook-county-property-tax-divide/appeals.html

[38] Jason Grotto and Sandhya Kambhampati. 2017. The Tax Divide | Commercial breakdown. Chicago Tribune. https://apps.chicagotribune.com/news/watchdog/cook-county-property-tax-divide/index.html

[39] Michelle Hanlon and Terry Shevlin. 2005. Book-tax conformity for corporate income: An introduction to the issues. *Tax policy and the economy* 19 (2005), 101–134.

[40] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems (NIPS, Vol. 29)*. Curran Associates, Inc. https://papers.nips.cc/paper/2016/hash/9d2682367c3935defcb1f9e247a97c0d-Abstract.html

[41] Lee Harris. 2004. 'Assessing' Discrimination: The Influence of Race in Residential Property Tax Assessments. *Journal of Land Use & Environmental Law* 20, 1 (2004). https://law.fsu.edu/sites/g/files/upcbnu1581/files/JLUEL/jluel-v20n1.pdf

[42] Emma Harvey and Evelyn Smith. 2025. Hidden Errors in Big Data: Brokered Data Obfuscates Measures of Vertical Equity. *Working paper* (2025).

[43] Andrew T. Hayashi. 2014. Property Taxes and Their Limits: Evidence from New York City. *Stanford Law & Policy Review* 25 (01 2014). https://law.stanford.edu/publications/property-taxes-limits-evidence-new-york-city/

[44] Daniel E Ho and Alice Xiang. 2020. Affirmative algorithms: The legal grounds for fairness as awareness. *U. Chi. L. Rev. Online* (2020), 134.

[45] Timothy R Hodge, Daniel P McMillen, Gary Sands, and Mark Skidmore. 2017. Assessment inequity in a declining housing market: The case of Detroit. *Real Estate Economics* 45, 2 (2017), 237–258.

[46] Jengei Hong, Heeyoul Choi, and Woo-sung Kim. 2020. A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea. *International Journal of Strategic Property Management* 24, 3 (2020), 140–152.

[47] Yilin Hou, Lei Ding, David J Schwegman, and Alaina G Barca. 2025. Assessment frequency and equity of the property tax: Latest evidence from Philadelphia. *Journal of Policy Analysis and Management* 44, 2 (2025), 483–507.

[48] Keith Ihlanfeldt and Luke P Rodgers. 2022. Homestead exemptions, heterogeneous assessment, and property tax progressivity. *National Tax Journal* 75, 1 (2022), 7–31.

[49] International Association of Assessing Officers (IAAO). 2013. *Standard on Ratio Studies.* Technical Standard. International Association of Assessing Officers, Kansas City, MO. https://www.iaao.org/wp-content/uploads/Standard_on_Ratio_Studies.pdf

[50] International Association of Assessing Officers (IAAO). 2018. *Standard on Automated Valuation Models (AVMs).* Technical Standard (2nd edition). International Association of Assessing Officers, Kansas City, MO. https://www.iaao.org/wp-content/uploads/Standard_on_Automated_Valuation_Models.pdf

[51] International Association of Assessing Officers (IAAO). 2021. *Standard on Data Quality.* Technical Standard. International Association of Assessing Officers, Kansas City, MO. https://www.iaao.org/wp-content/uploads/Standard_on_Data_Quality.pdf

[52] MAS Jennifer, Rearich RES, et al. 2021. Garbage in, garbage out: Implications of data quality for valuation models. *Journal of Property Tax Assessment & Administration* 18, 1 (2021), 1.

[53] Andrew W Kahrl. 2024. Citizens and taxpayers. In *The Black tax: 150 years of theft, exploitation, and dispossession in America*. University of Chicago Press.

[54] Faisal Kamiran and Toon Calders. 2009. Classifying without discriminating. In *2nd International Conference on Computer Control and Communication 2009 (IC4).* 1–6. doi:10.1109/IC4.2009.4909197

[55] Faisal Kamiran and Toon Calders. 2012. Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems* 33, 1 (Oct. 2012), 1–33. doi:10.1007/s10115-011-0463-8

[56] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2017. Inherent Trade-Offs in the Fair Determination of Risk Scores. In *Innovations in Theoretical Computer Science (ITCS) 2017 (ITCS).* arXiv. doi:10.48550/arXiv.1609.05807

[57] Allison Koenecke, Eric Giannella, Robb Willer, and Sharad Goel. 2023. Popular Support for Balancing Equity and Efficiency in Resource Allocation: A Case Study in Online Advertising to Increase Welfare Program Awareness. *Proceedings of the International AAAI Conference on Web and Social Media* 17 (June 2023), 494–506. doi:10.1609/icwsm.v17i1.22163

[58] Carol Kokinis-Graves. 2005. Use of the cost, income and sales-comparison approaches in the valuation of real estate. *J. St. Tax'n* 24 (2005), 23.

[59] Tzu-Chin Lin. 2010. Property tax inequity resulting from inaccurate assessment—The Taiwan experience. *Land Use Policy* 27, 2 (2010), 511–517.

[60] Katherine Loughead, Jared Walczak, and Eddie Koranyi. 2022. Unpacking the State and Local Tax Toolkit: Sources of State and Local Tax Collections (FY 2020). Tax Foundation. https://taxfoundation.org/data/all/state/state-local-tax-collections/

[61] Daniel McMillen and Ruchi Singh. 2020. Assessment Regressivity and Property Taxation. *The Journal of Real Estate Finance and Economics* 60, 1-2 (Feb. 2020), 155–169. doi:10.1007/s11146-019-09715-x

[62] Daniel McMillen and Ruchi Singh. 2020. Assessment regressivity and property taxation. *The Journal of Real Estate Finance and Economics* 60, 1 (2020), 155–169.

[63] Daniel McMillen and Ruchi Singh. 2023. Measures of vertical inequality in assessments. *Journal of Housing Economics* 61 (Sept. 2023), 101950. doi:10.1016/j.jhe.2023.101950

[64] Daniel P McMillen. 2013. The effect of appeals on assessment ratio distributions: Some nonparametric approaches. *Real Estate Economics* 41, 1 (2013), 165–191.

[65] Aditya Krishna Menon and Robert C. Williamson. 2018. The cost of fairness in binary classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAT*).* PMLR, 107–118. https://proceedings.mlr.press/v81/menon18a.html

[66] Mark Nichols, Tonya Simpson, Maia Rosenfeld, Jared Kofsky, and Tommy Brooksbank. 2025. 'Highballed': How disproportionate property taxes are forcing some Americans out of their homes. ABC News. https://abcnews.go.com/US/highballed-disproportionate-property-taxes-forcing-americans-homes/story?id=124312846

[67] Wallace E Oates and William A Fischel. 2016. Are local property taxes regressive, progressive, or what? *National Tax Journal* 69, 2 (2016), 415–433.

[68] Cook County Assessor's Office. 2025. About Automated valuation model for all class 200 residential properties in Cook County. https://github.com/ccao-data/model-res-avm

[69] Cook County Treasurer's Office. 2025. A Broken Property Tax Appeals System. https://www.cookcountytreasurer.com/pdfs/appealsreportanalysis/AppealsReportAnalysisEnglish.pdf

[70] Andre Perry, Jonathan Rothwell, and David Harshbarger. 2018. *The Devaluation of Assets in Black Neighborhoods.* Technical Report. The Brookings Institution. https://www.brookings.edu/articles/devaluation-of-assets-in-black-neighborhoods/

[71] Geoff Pleiss, Manish Raghavan, Felix Wu, Jon Kleinberg, and Kilian Q Weinberger. 2017. On Fairness and Calibration. In *Advances in Neural Information Processing Systems (NIPS, Vol. 30).* Curran Associates, Inc., Long Beach, CA. https://papers.nips.cc/paper/2017/hash/b8b9c74ac526fffbeb2d39ab038d1cd7-Abstract.html

[72] Carmela Quintos. 2014. Improving assessment equity in mass appraisal models. *Journal of Property Tax Assessment & Administration* 11, 4 (2014), 53–64.

[73] Joaquin Quiñonero Candela, Yuwen Wu, Brian Hsu, Sakshi Jain, Jennifer Ramos, Jon Adams, Robert Hallman, and Kinjal Basu. 2023. Disentangling and Operationalizing AI Fairness at LinkedIn. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT).* Association for Computing Machinery, New York, NY, USA, 1213–1228. doi:10.1145/3593013.3594075

[74] Jon Seidel. 2025. Former Cook County assessor official gets probation after helping feds. Chicago Sun-Times. https://chicago.suntimes.com/2025/07/29/former-cook-county-assessors-official-gets-probation-after-helping-with-feds-lengthy-probe

[75] Marc V. Shaw. 2021. THE ROAD TO REFORM: A Blueprint for Modernizing and Simplifying New York City's Property Tax System. https://www.nyc.gov/assets/propertytaxreform/downloads/pdf/final-report.pdf

[76] Daniel B Suits. 1977. Measurement of tax progressivity. *The American Economic Review* 67, 4 (1977), 747–752.

[77] Kiri Wagstaff. 2012. Machine learning that matters. *arXiv preprint arXiv:1206.4656* (2012).

[78] Nada Wasi and Michelle J White. 2005. Property tax limitations and mobility: the lock-in effect of California's Proposition 13.

[79] Rachel N Weber and Daniel P McMillen. 2010. Ask and ye shall receive? Predicting the successful appeal of property tax assessments. *Public Finance Review* 38, 1 (2010), 74–101.

[80] Michael Wick, swetasudha panda, and Jean-Baptiste Tristan. 2019. Unlocking Fairness: a Trade-off Revisited. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/373e4c5d8edfa8b74fd4b6791d0cf6dc-Paper.pdf

[81] Seckin Yilmazer and Sultan Kocaman. 2020. A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land use policy* 99 (2020), 104889.

[82] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P. Gummadi. 2017. Fairness Constraints: Mechanisms for Fair Classification. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, 962–970. https://proceedings.mlr.press/v54/zafar17a.html

[83] Han Zhao and Geoffrey J. Gordon. 2022. Inherent Tradeoffs in Learning Fair Representations. *Journal of Machine Learning Research* 23, 57 (2022), 1–26. http://jmlr.org/papers/v23/21-1427.html

A Regressivity Metrics

A.1 Suits Index

The Suits Index, introduced by [76] as a tool for evaluating tax progressivity, is closely related to the Gini coefficient in its reliance on Lorenz-style curves. Unlike the Gini, however, it is derived from a single cumulative curve. The steps for computing the index are:

(1) Rank homes in ascending order by sale price.
(2) For this ordered list, compute cumulative shares of both sale prices and assessed values, ranging from 0 to 100 percent.
(3) Plot the cumulative share of assessed values (y-axis) against the cumulative share of sale prices (x-axis).
(4) Define the index as

$$S = \frac{(K - L)}{K} \tag{1}$$

which represents the area beneath the 45-degree line less the area under the cumulative assessment curve. Because both axes span 0–100, $K = 0.5 * 100 * 100 = 5,000$. Thus, the Suits Index simplifies to $S = 1 - \frac{1}{5000} * L$.

We approximate L using the discrete method outlined in [63]. An index value of 0 reflects perfectly proportional assessments ($K = L$). Negative values arise when $L > K$, meaning that, for example, the bottom 10% of homes by sale price represent more than 10% of total assessed value, which is taken as evidence of regressivity. Positive values, in contrast, indicate progressivity.

A.2 Price-Related Differential (PRD)

The price-related differential, or PRD, is a measure which compares the average assessment ratio to the sale price weighted average ratio among the population of asssessed homes. Formally, the PRD is given by:

$$PRD = \frac{\bar{R}}{\sum_i (\frac{A_i}{S_i})(\frac{S_i}{\sum_j S_j})} = \frac{\bar{R}}{\bar{A}/\bar{S}} \tag{2}$$

Where $\bar{R}$ is the average assessment ratio, $\bar{A}$ the average assessed value, and $\bar{S}$ the average sale price. PRD values greater than 1 indicate that high-priced homes have lower assessment ratios than low-priced homes, which is an indication of regressivity. Per IAAO standards, the acceptable range for the PRD is between 0.98 and 1.03 [49].

A.3 Regression-Based Measures

Regression-based measures of assessment regressivity capture linear associations between sale price and assessed value. Numerous specifications have been adopted in the assessment literature, including regressions of sale price on assessed value, of assessed value on sale price, and of sales ratios on sale price, alongside logged versions of these specifications and more complicated formulations such as the coefficient of price-related bias [49].

The measure used in the proof provided in Appendix C is derived from a regression of log sales ratios on log sale prices. More formally, the measure is estimated as the coefficient β_1 from the following expression:

$$log(A/S) = \beta_0 + \beta_1 log(S) + \varepsilon \tag{3}$$

Where A is assessed value and S is sale price. Intuitively, if assessments are regressive, then sales ratios will generally decrease as sale prices increase, and β_1 will be negative.

[63] uses simulated data to establish that regression-based measures of assessment inequality are generally biased downwards, towards finding regressivity. This bias stems from a number of sources, including measurement error in

sale price, endogeneity bias, and the functional form of assessment models. However, [63] also establish that regression-based measures are sensitive to changes in simulated regressivity even if the levels are biased. The majority of the applications of this metric in this paper, in particular in Section 4 and Appendix C, consider changes in regressivity between assessment models rather than the overall level of regressivity. If the extent of bias is uncorrelated with the method of assessment, as is the case for measurement error in sale price, then the bias will difference out, and the resulting estimate of the change in regressivity will be valid. In cases where we examine the level of regressivity, we show robustness of these results to other metrics.

B Assessment Accuracy and Effective Tax Rates

In this section, we illustrate the relationship between our primary assessment accuracy metric (MAPE) and property tax rates. Specifically, we show that the percentage difference between a property's *statutory* and *effective* tax rates is equivalent to the mean absolute percentage error between assessed value and sale price in a setting with no exemptions or deductions to the property tax.

First, let the statutory property tax rate—the nominal rate of tax established by state or local law—be denoted τ. Similarly, let $\hat{\tau}$ denote the effective tax rate—the actual rate of tax paid relative to a property's underlying market value. Let A denote assessed value, and S denote market value. If there are no exemptions or deductions to the tax, we have that:

$$\tau A = \hat{\tau} S \tag{4}$$

That is, the statutory rate applied to the assessed value of the home is equal to the effective tax rate applied to the home's underlying market value.

Re-arranging and subtracting 1 from both sides yields:

$$\frac{A}{S} = \frac{\hat{\tau}}{\tau} \iff \frac{A}{S} - 1 = \frac{\hat{\tau}}{\tau} - 1 \iff \frac{A - S}{S} = \frac{\hat{\tau} - \tau}{\tau} \tag{5}$$

The left-hand side of the final term is the percentage error between assessed value and sale price, while the right-hand side is the percentage error between effective and statutory property tax rates. One can therefore evaluate mean absolute percentage error across the population using either term and arrive at the same result.

C Relationship Between Fairness and Accuracy: Log Coefficient and Mean Squared Error

This section provides a theoretical analysis the relationship between fairness and accuracy using the regression-based definition of model fairness described in Appendix A and mean squared error as the measure of model accuracy.

Consider two assessment models, denoted 1 and 2. Each model predicts ground-truth price, denoted S. The predictions of each model are given by $\hat{S}_i$, $i \in \{1, 2\}$.

Let fairness be measured using the coefficient β_1 from the following regression:

$$log(\frac{\hat{S}_i}{S}) = \beta_0 + \beta_1 log(S) + \varepsilon \tag{6}$$

If β_1 is negative, then the log ratio of assessed to actual sale price is decreasing in log sale price, indicating that assessments are regressive. Therefore, increases in β_1 signal that fairness is improving between one assessment model and the next.

The formula for $\hat{\beta}_1$ is given by:

$$\hat{\beta}_1 = \frac{cov(log(\hat{S}_i/S), log(S))}{var(log(S))} = \frac{cov(log(\hat{S}_i), log(S))}{var(log(S))} - 1 \tag{7}$$

Because $var(log(S))$, the variance of log ground-truth price, is unchanging between models, the key determinant of model fairness is the term in the numerator, $cov(log(\hat{S}_i), log(S))$. More specifically, if:

$$cov(log(\hat{S}_2), log(S)) - cov(log(\hat{S}_1), log(S)) > 0 \tag{8}$$

Then the value of β_1 will be higher under model 2 than under model 1, and consequently model 2 will be fairer according to the regression coefficient metric than model 1.

Let s denote $log(S)$ and $\hat{s}_i$ denote $log(\hat{S}_i)$. We can express the covariance between predicted and actual log prices for model i as:

$$cov(\hat{s}_i, s) = \frac{1}{2}(\sigma^2_{\hat{s}_i} + (\mu_{\hat{s}_i} - \mu_s)^2 - MSE(\hat{s}_i, s) + \sigma^2_s) \tag{9}$$

To see this, note that the mean squared error between $\hat{s}_i$ and s is given by:

$$\begin{aligned} MSE(\hat{s}_i, s) &= E[\hat{s}_i^2] - 2E[\hat{s}_i s] + E[s^2] \\ &= \sigma^2_{\hat{s}_i} + \mu^2_{\hat{s}_i} - 2(cov(\hat{s}_i, s) + \mu_{\hat{s}_i}\mu_s) + \sigma^2_s + \mu^2_s \\ &= \sigma^2_{\hat{s}_i} + (\mu_{\hat{s}_i} - \mu_s)^2 - 2cov(\hat{s}_i, s) + \sigma^2_s \end{aligned} \tag{10}$$

Which yields the expression in Equation 9 after some re-arranging. Given this, we can express the difference in the covariance between predicted and actual values between model 2 and model 1 as:

$$cov(\hat{s}_2, s) - cov(\hat{s}_1, s) = \frac{1}{2}\left[(\sigma^2_{\hat{s}_2} - \sigma^2_{\hat{s}_1}) + [(\mu_{\hat{s}_1} - \mu_s)^2 - (\mu_{\hat{s}_2} - \mu_s)^2] - [MSE(\hat{s}_2, s) - MSE(\hat{s}_1, s)]\right] \tag{11}$$

It follows that the relationship between fairness and accuracy under these metrics depends on the relative size of the following three terms, which together determine the change in model fairness $cov(\hat{s}_2, s) - cov(\hat{s}_1, s)$:

(1) The change in the variance of model predictions, $(\sigma^2_{\hat{s}_2} - \sigma^2_{\hat{s}_1})$,

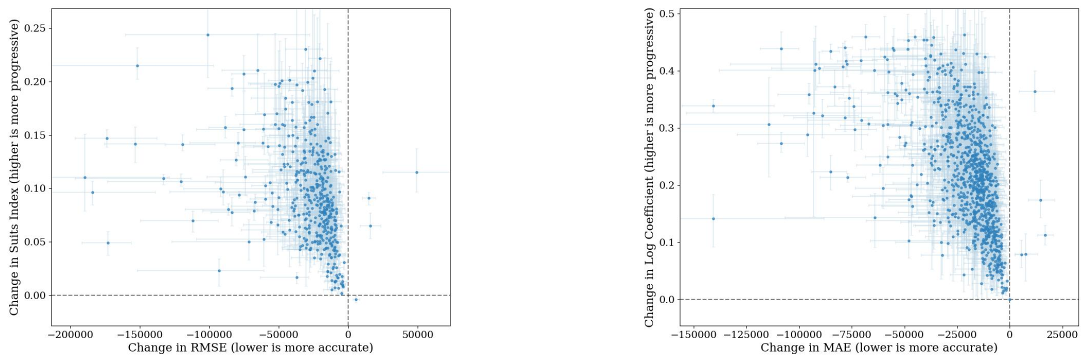
(2) The change in the squared difference between mean predicted and mean ground-truth log price, $(\mu_{\hat{s}_1} - \mu_s)^2 - (\mu_{\hat{s}_2} - \mu_s)^2$,

(3) And the change in model accuracy, $-[MSE(\hat{s}_2, s) - MSE(\hat{s}_1, s)]$

The relationship between fairness and accuracy in this setting is therefore ambiguous. To see this, assume that models 1 and 2 have the same mean prediction, $\mu_{\hat{s}_1} = \mu_{\hat{s}_2} = \mu_{\hat{s}}$. This sets the value of term (2) to zero. Further, let model 2 have lower mean squared error than model 1, so term (3) is positive. Whether fairness increases alongside accuracy now depends entirely on the difference in the variance of the predictions between models 1 and 2, given by term (1). If model 2's predictions are significantly less variable than model 1, and the magnitude of the difference exceeds the magnitude of model 2's accuracy gains, then model 2 will be less fair than model 1, despite being more accurate. If, however, model 2's predictions are more variable than model 1's predictions, or exhibit only slight decreases in variance relative to model 1, then accuracy and fairness will improve simultaneously.

In summary, simultaneous improvements in fairness and accuracy in the assessment setting are possible, as are tradeoffs, and the realized outcome depends on changes in the distribution of predictions relative to ground-truth prices.

D Robustness Checks

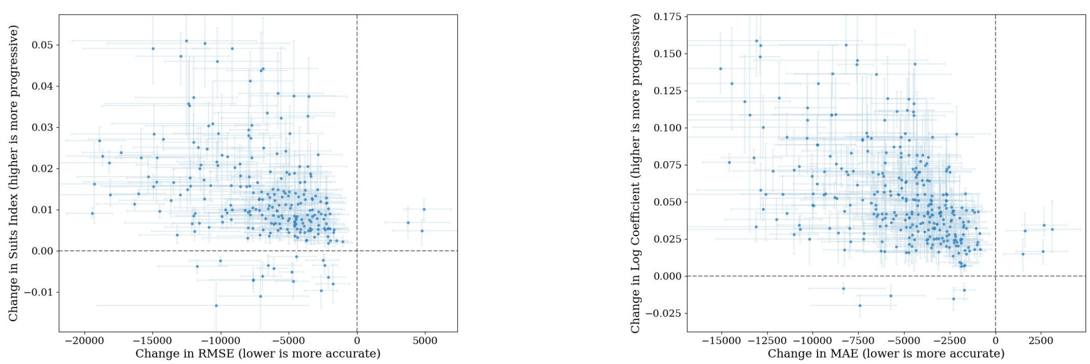


(a) RMSE, Suits Index

(b) MAE, LC

Fig. 6. Pareto Improvements in Model Performance after Adding More Property Characteristics to Assessment Models are Robust Across Metrics

Notes: The figures show changes in accuracy and fairness between the “sparse" and “rich" county-level LASSO-based assessment models described in Section 4. Models are trained using property and time-of-sale characteristics. Accuracy is measured using root mean squared error (left panel) and mean absolute error (right panel) between assessed and sale price. Fairness is measured using the Suits Index (left panel) and LC (right panel) measures described in A. Changes in performance are statistically significant at a 95% confidence level after applying a Benjamini-Hochberg multiple testing correction. Confidence intervals are computed using a studentized bootstrap and displayed in light blue. Results are censored at the 2nd and 98th percentiles.



(a) RMSE, Suits Index

(b) MAE, LC

Fig. 7. Pareto Improvements in Model Performance after Incorporating Census Data are Robust Across Metrics

Notes: The figures show changes in accuracy and fairness for county-level random forest assessment models following the addition of Census characteristics. The baseline model for each county uses only status quo predictions and time-of-sale data. The alternative model incorporates these features in addition to Census block group characteristics. Accuracy is measured using the change in root mean squared error (left panel) and mean absolute error (right panel) between predicted and actual sale price as we move from the baseline model to the alternative model. Fairness is measured using the Suits Index and LC measures described in Appendix A. Changes in performance are statistically significant at a 95% confidence level after applying a Benjamini-Hochberg multiple testing correction. Confidence intervals are computed using a studentized bootstrap and displayed in light blue. Results are censored at the 2nd and 98th percentiles.