

# Intra-regional heterogeneity of AI-based scoring systems: Evidence from New York, New Jersey, and Connecticut (2007–2023)

M. NAPO Tchin

December 5, 2025

## Abstract

This article extends a previous tri-state analysis of scoring systems in the New York metropolitan area by focusing on cross-state heterogeneity. Using HMDA and ACS data for New York (NY), New Jersey (NJ), and Connecticut (CT), we compare the pre-AI period (2007–2017), dominated by human expertise, to the period of widespread adoption of machine learning models (2018–2023). The study relies on logistic models, Random Forests, and XGBoost, as well as a set of fairness metrics (disparate impact, equal opportunity, calibration) computed from a common structural Random Forest model.

The results show (i) a generalized increase in approval rates in all three states after the introduction of AI, (ii) a sharp improvement in predictive performance, with AUCs rising from around 0,60 in the pre-AI period to 0,90 and above in the AI period, and (iii) the persistence of marked disparities across income groups and racial groups, with particularly pronounced gaps for Black borrowers. Connecticut appears to be both the most accurate and the least polarized state, whereas New Jersey concentrates the strongest “tensions” between performance, socio-economic segmentation, and risks of disparate impact. These findings suggest that federal-level regulatory harmonization may conceal highly heterogeneous decision regimes and risk profiles within the Tri-State region itself.

## 1 Introduction

Recent work on U.S. mortgage credit documents a twofold transition. On the one hand, there has been a progressive substitution of algorithmic scoring systems for traditional underwriting practices, in a context of intensified competition among lenders and a reconfiguration of access to credit (Agarwal, Ben-David, & Seru, 2015; Bhutta, Blair, & Dettling, 2021; Fuster, Goldsmith-Pinkham, Ramcharan, & Vickery, 2019; Consumer Financial Protection Bureau, 2019). On the other hand, concerns about non-discrimination and fairness in automated decision systems have grown significantly (Barocas & Selbst, 2016; Mullainathan & Spiess, 2017; Kearns & Roth, 2020).

In the credit domain, several studies show that the introduction of machine learning can simultaneously improve predictive accuracy and alter the distribution of credit access across groups (Fuster et al., 2019; Bartlett, Morse, Stanton, & Wallace, 2022; Hardy, Logan, & Pustejovsky, 2023). Beyond finance, similar patterns have been documented in computer vision and

facial recognition applications, where systematic performance gaps across demographic groups have been identified (Buolamwini & Gebru, 2018). Taken together, these contributions support the view that AI models operate within socio-technical systems already structured by power relations, mechanisms of segregation, and historical trajectories (Massey & Denton, 1993; Rothstein, 2018; Squires, 2017; Selbst, Boyd, Friedler, Venkatasubramanian, & Vertesi, 2019).

The objective of this article is twofold. First, we propose an *intra-regional* decomposition that examines each of the three states in the Tri-State region separately, in terms of approval rates, income gradients, and predictive performance. Second, we assess the extent to which the fairness properties of the models—in particular the disparate impact by race—vary across states and interact with the local structure of the credit market and the spatial legacy of residential segregation (Gyourko & Tracy, 2014; Rothstein, 2018).

Our decomposition shows that the transition to AI has not been homogeneous: while the three states converge toward similar levels of predictive performance, they remain differentiated in terms of socio-economic segmentation and disparate impact risks. This internal heterogeneity is crucial from a regulatory standpoint, since models that are acceptable at an aggregated tri-state level may produce contested outcomes in specific jurisdictions (Cowgill & Tucker, 2020; Kearns & Roth, 2020).

## 2 Data and empirical strategy

The data used here are identical to those in the previous tri-state study, but are exploited at a finer level of aggregation. The `hmda_tristate_core_2007_2024_v2` database provides 14 163 286 mortgage applications for the Tri-State region. After restricting to the years 2007–2017 and to the three states, we obtain a pre-AI cohort of 13 512 746 observations, distributed as follows: 6 192 532 for NY, 5 281 836 for NJ, and 2 038 378 for CT (Table 1). Approval is defined using the regulatory variable `action_taken` (1 = loan originated).

For the AI period, we use the merged HMDA+ACS dataset `hmda_acs_tristate_2018_2023_FINAL`, which combines detailed HMDA information with socio-economic characteristics from the ACS at the *census tract* level (median income, poverty rate, unemployment, racial composition, etc.). The AI cohort (2018–2023) contains 571 820 observations, of which 274 560 for NY, 207 169 for NJ, and 90 091 for CT.

Table 1: Sample sizes by state and period

	Pre-AI 2007–2017	AI 2018–2023
New York (NY)	6 192 532	274 560
New Jersey (NJ)	5 281 836	207 169
Connecticut (CT)	2 038 378	90 091

### 2.1 Income segments and racial groups

Income groups are constructed from the distribution of ACS median income at the *census tract* level, using terciles: *Low*, *Middle*, and *High*. At the tri-state level, these groups are of comparable

size (around 172 to 177 thousand observations each, excluding missing values), which ensures stable estimation in each state.

Racial groups are based on the HMDA variable `derived_race`, recoded for fairness analyses into five aggregate categories: *White*, *Black / African-American*, *Asian*, *Other* (pooled minority categories), and *Not reported*. This aggregation follows supervisory practice when evaluating disparate impact risk using HMDA data and builds on debates about residential segregation and differential access to credit (Massey & Denton, 1993; Squires, 2017).

## 2.2 Predictive modeling and fairness

To address RQ2, we estimate, separately by state and period (pre-AI vs AI), three classification models: a standard logistic regression, a Random Forest, and XGBoost. The common predictors are `loan_purpose`, `loan_type`, `hoepa_status`, and `year`, encoded via indicator variables. The data are split into training and test samples (70/30, stratified on the outcome), which allows us to compare out-of-sample performance in terms of accuracy, F1-score, and ROC AUC (see Table 4). This design follows best practices in applied economics and econometrics using machine learning (Friedman, Hastie, & Tibshirani, 2001; Mullainathan & Spiess, 2017).

For RQ4, we estimate a structural Random Forest on the entire tri-state AI cohort, with predictors `loan_purpose`, `loan_type`, `hoepa_status`, `state_code`, and `year`. Based on binary and probabilistic predictions (respectively  $\hat{y}$  and  $\hat{p}$ ), we compute, for each (state, racial group) pair, a set of fairness metrics: predicted positive rate, disparate impact vs White, true positive rate (TPR) and equal opportunity difference, false negative rate (FNR), positive predictive value (PPV) and predictive parity difference, and a calibration measure (Brier score). Groups with fewer than 150 observations are excluded to avoid unstable estimates. The main fairness results are summarized in Table 5.

Table 2: Overall approval rates by state and period

	Pre-AI (2007–2017)	AI (2018–2023)	Change (pp)
New York (NY)	46,7%	55,4%	+8,7
New Jersey (NJ)	45,6%	51,4%	+5,7
Connecticut (CT)	49,0%	57,2%	+8,2

Table 3: Odds ratios (OR) by income group in the AI period (ref. High)

State	Middle vs High	Low vs High
New York (NY)	0,887	0,80
New Jersey (NJ)	0,848	0,70
Connecticut (CT)	0,89	0,70

Table 4: Out-of-sample ROC AUC by state, period, and model

Period	State	Logit	Random Forest	XGBoost
Pre-AI	NY	0,596	0,609	0,609
	NJ	0,573	0,592	0,591
	CT	0,570	0,595	0,595
AI	NY	0,856	0,913	0,912
	NJ	0,850	0,899	0,898
	CT	0,886	0,925	0,925

Table 5: Disparate impact (DI) by race and state in the AI period (ref. White)

	Black	Asian	Other	Not reported
New York (NY)	0,81	0,91	0,94	0,80
New Jersey (NJ)	0,80	0,95	0,96	0,92
Connecticut (CT)	0,87	0,88	0,94	0,92

### 3 State-level results

#### 3.1 New York

**Approval rates and descriptive gradients.** In New York, the overall approval rate rises from about 46,7% in the pre-AI period to 55,4% in the AI period (Table 2), i.e., an increase of roughly +8,7 percentage points. In the AI period, approval rates vary markedly by race and income level, with lower rates for Black borrowers and for low-income tracts. ANOVA on the AI period confirms substantial heterogeneity in approval rates by race and income, with large  $F$ -statistics and near-zero p-values.

**Income effects in a logit model.** The logit model  $approval \sim income\_group$  (ref. High) reveals a monotonic gradient (Table 3). In the AI period, the approval odds ratio (OR) for Middle-income households is 0,887, and 0,80 for Low-income households, corresponding to a significant reduction in approval probabilities for households at the bottom of the income distribution.

**Predictive performance and ROC.** Predictive performance follows a clear trajectory (see Table 4): in the pre-AI period, the ROC AUC of the logistic regression is about 0,596, and around 0,609 for Random Forest and XGBoost. In the AI period, AUCs reach 0,856 for Logit and about 0,913 for ensemble tree models. The AI-period ROC curves (Figure 1) lie well above the diagonal, indicating a strong ability of the model to reproduce the approval decisions observed in the New York market.

**Fairness and disparate impact.** From a fairness perspective, the structural Random Forest model reveals a hierarchy of predicted positive rates (Table 5). The disparate impact (DI) for Black borrowers is about 0,81 relative to White borrowers, close to but above the regulatory 80% rule. Asian and *Other* groups display high DI values, while *Not reported* cases also cluster around

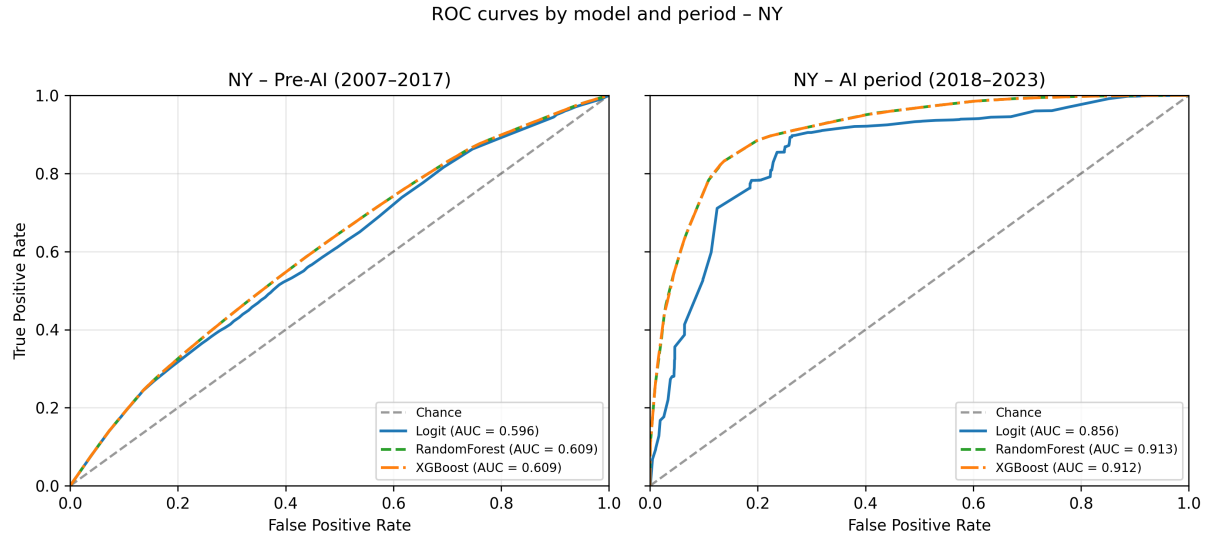


Figure 1: ROC curves by model and period for New York (pre-AI 2007–2017 and AI period 2018–2023).

0,80. AI makes the decision process more predictable but leaves a substantial disadvantage for Black borrowers and for applications with missing race information, in line with the diagnoses of indirect discrimination and proxy discrimination discussed by Barocas and Selbst (2016) and Kearns and Roth (2020).

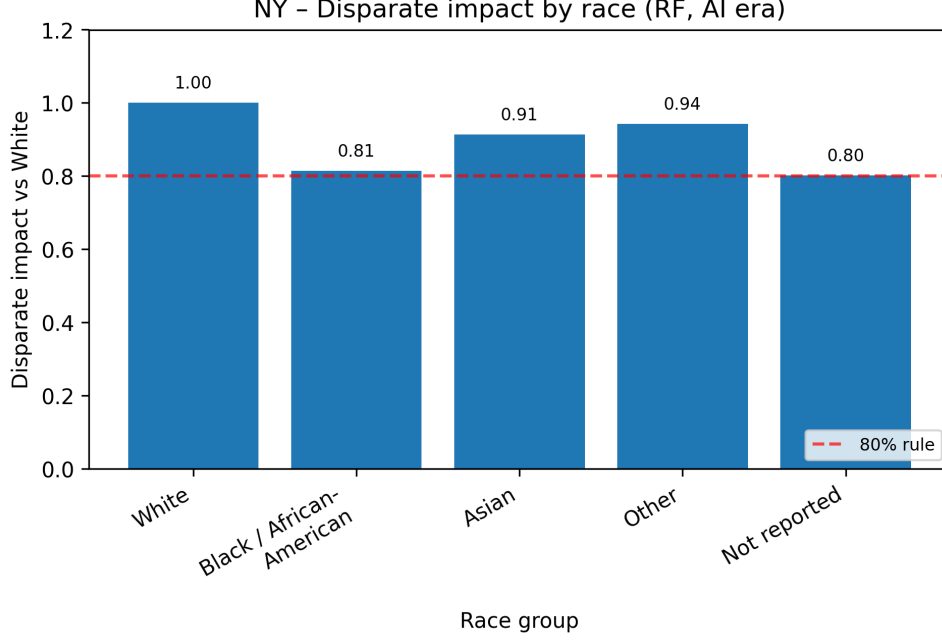


Figure 2: Disparate impact by racial group in New York in the AI period (structural Random Forest, ref. White).

### 3.2 New Jersey

**Approval rates.** In New Jersey, the overall approval rate increases from about 45,6 % in the pre-AI period to 51,4 % in the AI period (Table 2), a more modest rise than in New York and Connecticut. In the AI period, approval rates by race remain strongly differentiated, with much lower levels for Black borrowers and for *Not reported* cases. ANOVA yields the highest *F*-statistics in the region, indicating particularly pronounced racial and socio-economic segmentation, consistent with evidence on the fragmentation of credit access (Bhutta et al., 2021).

**Logit and income gradient.** In the logit model, the Middle-income group has an OR of 0,848, and the Low-income group an OR of 0,70 relative to High-income borrowers (Table 3). These are among the most adverse odds in the region for low-income households and translate into about a 30% reduction in approval probabilities for the Low group, echoing the analysis of Agarwal et al. (2015) on competition and risk selection in lending portfolios.

**Predictive performance.** Performance patterns are similar to those in NY, but at slightly lower levels (Table 4). In the AI period, the Logit AUC is around 0,85, and about 0,90 for Random Forest and XGBoost. The AI-period ROC curve (Figure 3) signals a highly systematic decision process, while still raising questions about the equity of the replicated decisions.

**Fairness: a borderline case.** The fairness metrics suggest that New Jersey is the most sensitive case in the Tri-State region (Table 5). The DI for Black borrowers is close to 0,80, i.e., exactly at the 80% rule threshold. Asian and *Other* groups show high DI values, while the *Not reported* group reaches a DI of about 0,92 but exhibits weaker calibration. The combination of a steep income gradient, strong racial disparities, and high predictive power of the AI model

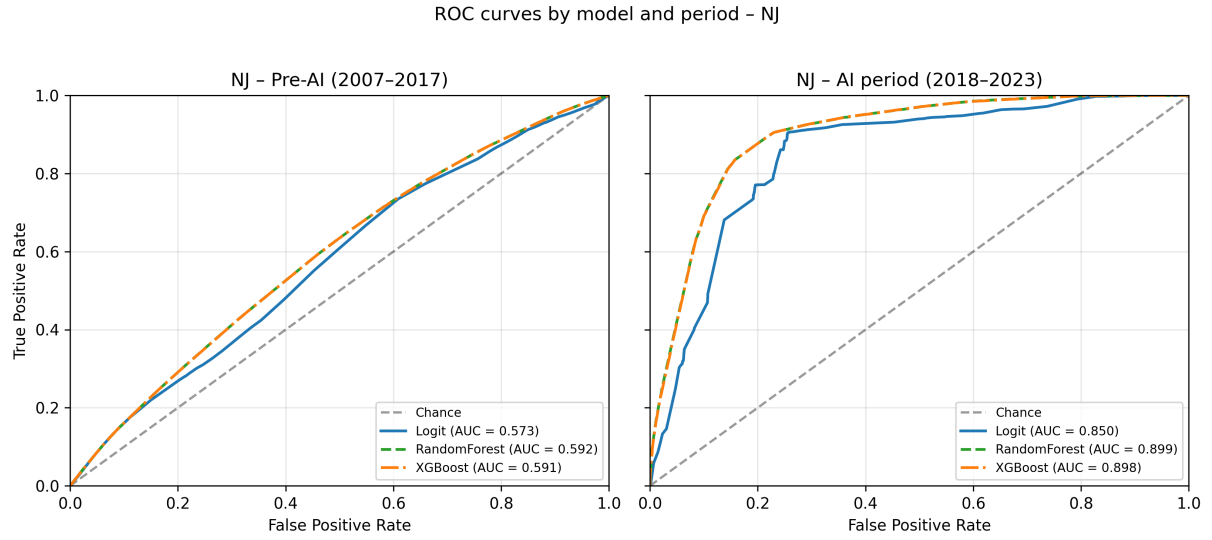


Figure 3: ROC curves by model and period for New Jersey (pre-AI 2007–2017 and AI period 2018–2023).

makes New Jersey an environment where legal risks related to disparate impact are particularly acute, resonating with recent debates on discrimination in fintech credit (Bartlett et al., 2022).

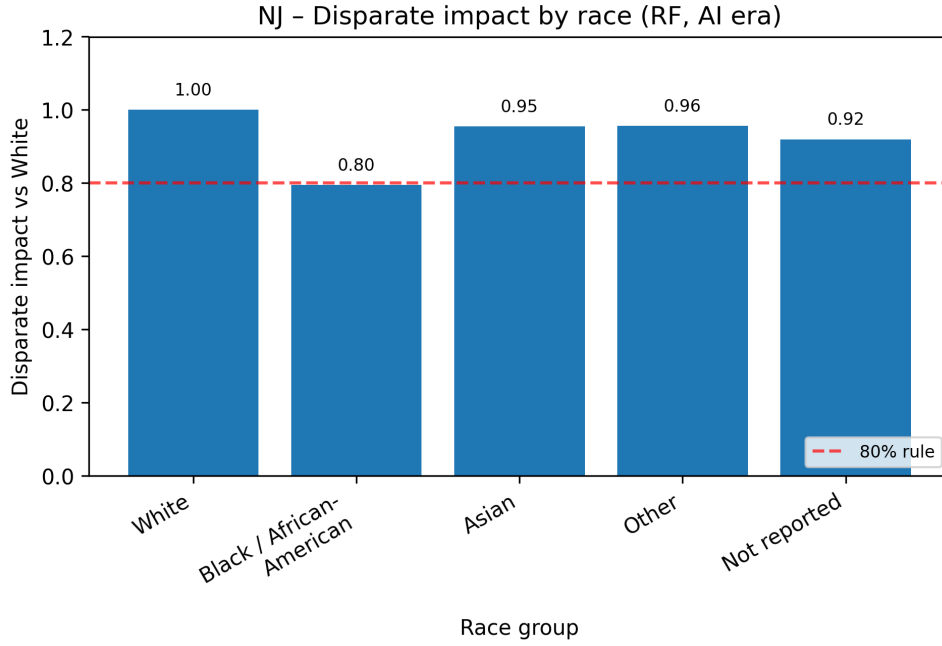


Figure 4: Disparate impact by racial group in New Jersey in the AI period (structural Random Forest, ref. White).

### 3.3 Connecticut

**Approval rates and gradients.** Connecticut stands out with the highest approval rates in the region. The transition from the pre-AI to the AI period is associated with an increase in the overall approval rate from about 49,0% to 57,2% (Table 2). In the AI period, approval rates by race exceed those observed in the other two states, but low-income tracts remain clearly disadvantaged.

ANOVA yields statistically significant  $F$ -statistics, but less extreme than in New Jersey, suggesting a more homogeneous market, despite the presence of systematic differences. This profile is consistent with the regional analyses of Gyourko and Tracy (2014) on the diversity of local mortgage markets.

**Logit and income.** The estimated ORs for Connecticut are very close to those for New Jersey in the Low-income group (Table 3): the OR is around 0,70 for Low-income households and 0,89 for Middle-income households. Belonging to the lower tail of the income distribution is thus associated with a significant reduction in approval probability, albeit in an environment with higher overall approval rates.

**Predictive performance and ROC.** Connecticut exhibits the best predictive performance in the region (Table 4). In the AI period, the Logit AUC reaches about 0,886, and Random Forest and XGBoost about 0,925. The AI-period ROC curves (Figure 5) approach the optimal frontier, indicating an almost perfect alignment between estimated probabilities and lenders' decision rules. This result is consistent with the literature on the performance gains obtained through ensemble methods (Friedman et al., 2001; Mullainathan & Spiess, 2017).



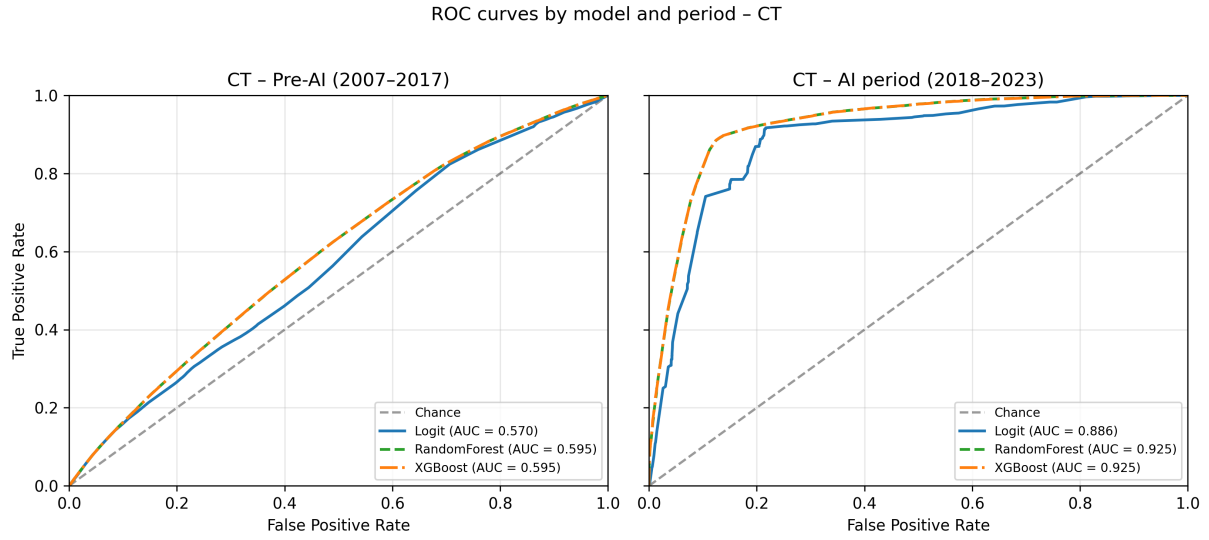


Figure 5: ROC curves by model and period for Connecticut (pre-AI 2007–2017 and AI period 2018–2023).

**Fairness.** From a fairness standpoint, Connecticut appears slightly less polarized than NY and NJ, although disparities remain meaningful. The DI for Black borrowers is about 0,87, and around 0,88 for Asian borrowers (Table 5). The *Other* and *Not reported* groups lie around 0,94 and 0,92. Equal opportunity and predictive parity metrics show relatively modest gaps compared to Whites, and calibration is generally better than in the other states, suggesting somewhat lower disparate impact risk. This configuration illustrates that high performance and relatively contained disparities can coexist, without fully eliminating inequities (Hardy et al., 2023).

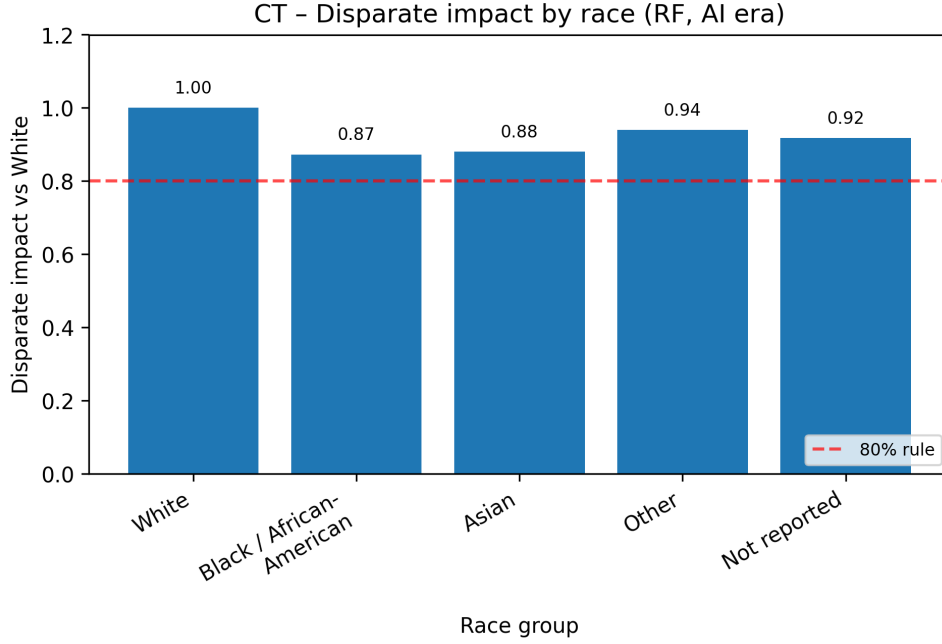


Figure 6: Disparate impact by racial group in Connecticut in the AI period (structural Random Forest, ref. White).

## 4 Cross-state comparison and implications

The state-level decomposition summarized in Tables 2 to 5 reveals a recurring combination of phenomena alongside marked specificities:

- **Predictive performance: convergence toward highly powerful AI.** In all three states, AI-period AUCs substantially exceed those observed in the pre-AI period (increases of about +0,30). New York and New Jersey converge to AUCs around 0,90, while Connecticut reaches 0,92–0,93. The shift toward highly discriminating AI systems is therefore a robust regional phenomenon, not confined to a single jurisdiction, in line with broader analyses of machine learning in economics (Mullainathan & Spiess, 2017).
- **Income gradients: New Jersey / Connecticut asymmetry.** New Jersey and Connecticut both display pronounced income gradients, especially for Low-income households ( $OR \approx 0,70$ ), whereas New York shows equally significant but slightly less steep gradients ( $OR \approx 0,80$ ). This result matters for the regulation of access-to-credit policies in areas with high income heterogeneity and resonates with evidence of declining credit access for specific household segments (Bhutta et al., 2021).
- **Racial disparities: a continuum of risks.** In all three states, Black borrowers systematically exhibit lower DI values than White borrowers, with DI ranging from 0,80 (NJ) to 0,87 (CT). New Jersey appears closest to the regulatory threshold, while Connecticut combines high performance and slightly higher DI values, though still below unity. Non-disclosure of race is associated, in all states, with low DI and weaker calibration, suggesting unobserved selection mechanisms, in line with concerns raised by Barocas and Selbst (2016) and Selbst et al. (2019).

- **Tension between performance and fairness.** The more predictable (high-AUC) the decision rule becomes, the more structural differences across groups risk being locked-in by AI models. New Jersey illustrates this tension: highly accurate models coupled with DI and income gradients that brush against regulatory tolerance thresholds. This issue lies at the core of the literature on the “ethical algorithm” and the trade-offs between predictive accuracy and fairness (Kearns & Roth, 2020; Cowgill & Tucker, 2020).

From a peer-reviewed journal perspective, these results argue for a regulatory approach that explicitly accounts for territorial heterogeneity: merely checking global AI performance or aggregate disparate-impact thresholds can obscure local regimes that are substantially different or even problematic. This echoes work in urban economics and urban sociology showing that local housing and credit markets are deeply shaped by the history of public policies on segregation and fair housing (Massey & Denton, 1993; Rothstein, 2018; Squires, 2017).

## 5 Conclusion

Drawing on a harmonized HMDA+ACS dataset covering seventeen years and three states, this article has shown that the transition to AI-based mortgage scoring systems in the Tri-State region comes with a dual movement. On the one hand, predictive capacity surges, with AI-period AUCs around 0,90 in all three states, reflecting the institutionalization of highly structured decision pipelines, consistent with advances in statistical learning (Friedman et al., 2001; Mullainathan & Spiess, 2017; Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018). On the other hand, income gradients and racial disparities remain substantial, with particularly high disparate impact risks in New Jersey and, to a lesser extent, in New York, aligning with broader diagnoses about the role of algorithmic systems in reproducing inequalities (Barocas & Selbst, 2016; Hardy et al., 2023).

Connecticut offers an interesting counterpoint: it combines the highest predictive performance with relatively less extreme fairness gaps, suggesting that efficiency and equity can, to some extent, be reconciled, albeit under specific institutional and socio-economic configurations. Two main avenues for future research emerge: (i) incorporating additional variables (credit scores, default histories, finer geographic granularity) to better disentangle fundamental risk from proxy-based bias; (ii) assessing mitigation strategies (reweighing, post-processing of scores, fairness constraints in optimization) applied at the state level rather than at a national or regional aggregate, in the spirit of Kearns and Roth (2020).

As AI systems become the “operational language” of mortgage markets, the key challenge is no longer merely to measure their performance, but to understand how it unfolds in local contexts, and how regulators can reconcile distributive justice requirements with the statistical realities of portfolios and territories (Cowgill & Tucker, 2020; Consumer Financial Protection Bureau, 2019).

## References

Agarwal, S., Ben-David, I., & Seru, A. (2015). The effects of credit competition on lenders’ risk

- taking. *Journal of Financial Economics*.
- Barocas, S., & Selbst, A. (2016). Big data’s disparate impact. *California Law Review*, 104(3), 671–732.
- Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the fintech era. *Journal of Finance*.
- Bhutta, N., Blair, J., & Dettling, L. (2021). The decline of access to credit. *Federal Reserve Board Working Paper*.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the conference on fairness, accountability, and transparency (fat\*)*.
- Consumer Financial Protection Bureau. (2019). *Mortgage market trends* (Tech. Rep.). CFPB.
- Cowgill, B., & Tucker, C. (2020). Economics of artificial intelligence. *Journal of Economic Perspectives*.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*. Springer.
- Fuster, A., Goldsmith-Pinkham, P., Ramcharan, R., & Vickery, J. (2019). Predictably unequal? the effects of machine learning on credit markets. *Review of Financial Studies*.
- Gyourko, J., & Tracy, J. (2014). A regional analysis of mortgage lending. *Regional Science and Urban Economics*.
- Hardy, B., Logan, T., & Pustejovsky, J. (2023). Ai, segregation, and financial inequality. *Economic Policy Review*.
- Kearns, M., & Roth, A. (2020). *The ethical algorithm*. Oxford University Press.
- Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics*.
- Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometrics approach. *Journal of Economic Perspectives*, 31(2), 87–106.
- Rothstein, R. (2018). *The color of law: A forgotten history of how our government segregated america*. Liveright Publishing.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*.
- Squires, G. (2017). *The fight for fair housing*. Routledge.