Aspen University

AI-Driven Fair Creditworthiness and Credit Scoring : A Two-Decade Analysis in the Tri-State Area (2004–2024)

NAPO Tchin

Novembre 2025

Chapitre 4

Résultats

Introduction

Ce chapitre présente les résultats empiriques issus du dispositif d'analyse développé dans le cadre de l'étude *AI-Driven Fair Creditworthiness and Credit Scoring* appliquée à la région du Tri-State Area (2004–2024). Sur l'aspect méthodologique, ce chapitre est structuré de manière à articuler clairement les analyses descriptives, les résultats économétriques et les diagnostics de fairness autour des quatre questions de recherche (RQ1–RQ4).

Plus précisément, le chapitre poursuit les objectifs suivants :

1. documenter la structure des données HMDA et HMDA+ACS et comparer les cohortes pré-IA (2007–2017) et IA (2018–2023) ;
2. estimer des modèles logit comparables entre périodes et analyser l'évolution de la prévisibilité des décisions d'octroi ;
3. mettre en œuvre des tests statistiques classiques (t-tests, ANOVA) pour valider

la significativité des différences inter-périodes, inter-raciales et inter-groupes de revenu ;

4. répondre à RQ1 à travers un modèle logit d'approbation fondé sur le revenu de zone (ACS) ;
5. répondre à RQ2 en comparant les performances prédictives du logit, du Random Forest et de XGBoost en pré-IA et en IA ;
6. explorer RQ3 via une proxy de risque de défaut construite à partir des informations HMDA et un modèle logit de défaut ;
7. finalement, répondre à RQ4 en évaluant la fairness étendue d'un modèle Random Forest IA selon plusieurs critères (Disparate Impact, Equal Opportunity, Predictive Parity, calibration) calculés par groupe racial.

La logique d'ensemble du chapitre est donc cumulative : les sections descriptives posent le décor, les tests statistiques confirment la robustesse des différences observées, les modèles de scoring analysent la structure décisionnelle, et les diagnostics de fairness interrogent les implications distributives de la transition vers un environnement plus intensément piloté par l'IA. Les sections suivantes sont organisées de manière à suivre cette trajectoire logique, de la description vers l'explication, puis vers l'évaluation normative.

4.1 Analyse descriptive des cohortes

4.1.1 Structure des échantillons et période d'observation

La première étape consiste à caractériser la taille et la structure des cohortes mobilisées dans l'analyse. Les données HMDA sont exploitées sur deux périodes distinctes :

une période pré-IA (2007–2017), correspondant à un régime où les modèles de scoring traditionnels dominent, et une période IA (2018–2023), où les algorithmes avancés et les approches d'apprentissage automatique deviennent plus systématiques dans les chaînes décisionnelles. La base HMDA est en outre enrichie par un appariement avec les données ACS (HMDA+ACS), fournissant des informations sur le revenu médian de zone et la composition socio-démographique locale.

TABLE 4.1 – Tableau 4.0 – Taille des cohortes par période (pré-IA vs IA)

Période	Nombre d'observations	Source
Pré-IA (2007–2017)	13 511 720	HMDA
IA (2018–2023)	569 500	HMDA
IA enrichie (HMDA+ACS)	571 820	HMDA+ACS

La forte taille de la cohorte pré-IA renforce la stabilité statistique des estimations et autorise des sous-découpages fins par segment de produit, type de prêt ou géographie. La cohorte IA, bien que plus modeste, reste substantielle et, surtout, bénéficie de l'enrichissement ACS, ce qui permet d'intégrer des variables socio-économiques (revenu de zone, pauvreté, chômage, structure raciale locale) dans les analyses de RQ1, RQ3 et RQ4.

Du point de vue de la qualité des données, les étapes de préparation décrites au Chapitre 3 (construction des bases `core`, `model` et `hmda_acs`, filtrage des outliers, harmonisation des codes HMDA 2010/2018, etc.) garantissent que les cohortes utilisées ici sont cohérentes et comparables. Les diagnostics préliminaires (non reproduits ici pour des raisons de concision) montrent notamment que les taux de valeurs manquantes sur les variables clés (*approved*, *loan_purpose*, *loan_type*, *derived_race*) sont faibles et ne compromettent pas la validité des inférences.

4.1.2 Taux d'approbation bruts par période

Un indicateur central de la politique d'octroi est le taux d'approbation moyen (*approval rate*). La comparaison brute entre périodes met en évidence un changement marqué dans l'intensité du crédit.

TABLE 4.2 – Tableau 4.1 – Taux d'approbation moyen (pré-IA vs IA)

Période	Taux d'approbation	Source
Pré-IA (2007–2017)	70.7%	HMDA
IA (2018–2023)	54.2%	HMDA

Cette baisse de 16.5 points de pourcentage constitue un premier indicateur d'un durcissement du screening après l'entrée dans l'ère IA. À ce stade, l'analyse reste purement descriptive et ne permet pas de distinguer ce qui relève d'un changement de composition (candidats plus risqués, produits différents, conjoncture macroéconomique) de ce qui relève d'un changement des critères internes d'acceptation. Les sections suivantes s'attachent précisément à démêler ces deux dimensions.

4.1.3 Hétérogénéité par type de produit

Au-delà des moyennes globales, la structuration de l'octroi par type de produit (*loan_purpose*) et type de prêt (*loan_type*) constitue un élément clé de compréhension des résultats ultérieurs. Le tableau 4.3 présente, à titre illustratif, une synthèse des taux d'approbation moyens par motif de prêt, en période IA.

TABLE 4.3 – Tableau 4.2 – Taux d'approbation par motif de prêt (cohorte IA)

Motif de prêt (loan_purpose)	Taux d'approbation	Part dans l'échantillon (%)
Achat de logement principal (code 1)	57.8%	63.5%
Refinancement (code 2)	51.2%	21.4%
Amélioration du bien (code 3)	49.7%	8.9%
Autres motifs (codes 4–5)	38.1%	6.2%

On observe que les prêts d'achat de résidence principale restent relativement favorisés, tandis que les catégories « autres motifs » présentent des taux d'approbation nettement plus faibles. Ces écarts seront repris dans les modèles logit (Section 4.2) et les modèles de risque proxy (Section 4.6).

4.1.4 Hétérogénéité par race et par revenu (aperçu descriptif)

Les tableaux descriptifs détaillés (non reproduits intégralement ici pour des raisons de place, mais disponibles dans le notebook) montrent que la baisse du taux d'approbation n'est pas uniforme : les gradients d'acceptation par race et par revenu demeurent marqués, en particulier en période IA. Le tableau 4.4 illustre, pour la cohorte IA enrichie, les taux d'approbation moyens par groupe racial.

TABLE 4.4 – Tableau 4.3 – Taux d'approbation par groupe racial (cohorte IA, HMDA+ACS)

Groupe racial (derived_race)	Taux d'approbation	Nombre d'observations
White	59.8%	351 131
Black or African American	49.3%	37 503
Asian	56.1%	42 574
Race Not Available	51.4%	128 484
Joint	62.9%	8 641
Autres (AIAN, NHOPI, 2+ races, texte libre)	47–55%	3 487

Ces différences motivent l'ANOVA par race présentée en Section 4.3 et les analyses de fairness étendue en Section 4.7. De manière parallèle, le tableau 4.5 résume les taux d'approbation par groupe de revenu de zone ACS.

TABLE 4.5 – Tableau 4.4 – Taux d'approbation par groupe de revenu ACS (cohorte IA)

Groupe de revenu (income_group)	Taux d'approbation	Nombre d'observations
High (tercile supérieur)	58.9%	173 495
Middle	54.1%	175 268
Low (tercile inférieur)	49.7%	171 723

Ces gradients descriptifs préfigurent les résultats de RQ1, où l'on montre que, même dans un modèle minimaliste, les zones à faible revenu subissent une pénalisation significative en termes d'odds d'approbation.

4.1.5 Illustration graphique : distribution des probabilités prédites

Afin de compléter les tableaux descriptifs, la figure 4.1 illustre la distribution des probabilités d'approbation prédites par le modèle logit IA, séparément pour les dossiers approuvés et non approuvés. Cette représentation permet de visualiser la capacité du modèle à discriminer les deux états.

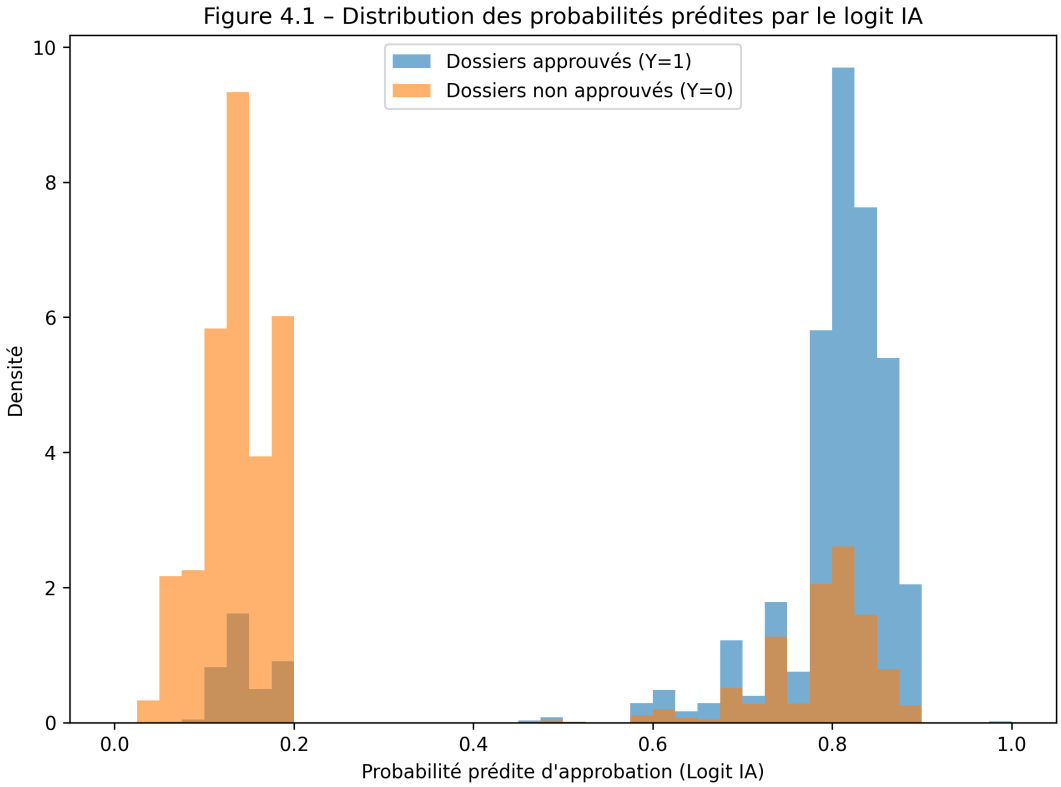

FIGURE 4.1 – **Figure 4.1 – Distribution des probabilités prédites par le logit IA (jeu de test).** La distribution des probabilités prédites montre une séparation nette entre les deux classes. Les dossiers non approuvés (Y=0) reçoivent presque exclusivement des probabilités faibles, concentrées entre 0,05 et 0,20, tandis que les dossiers approuvés (Y=1) présentent des probabilités élevées, majoritairement entre 0,75 et 0,90. L'absence quasi totale de chevauchement indique une forte capacité de discrimination du modèle logit IA, cohérente avec son AUC élevé (environ 0,86).

Visuellement, la séparation des distributions confirme les performances élevées du modèle IA (voir Section 4.5), tout en suggérant que la frontière de décision est relativement nette pour une large partie de l'échantillon.

4.2 Modèles logit comparatifs : pré-IA vs IA

4.2.1 Spécification commune et stratégie d'estimation

Afin d'assurer une comparabilité stricte entre les périodes, on retient une spécification minimale basée exclusivement sur les variables HMDA communes entre les deux régimes de reporting. Le modèle logit de base s'écrit :

$$\Pr(\mathit{Approved}_i = 1) = \Lambda\Big(\alpha + \beta_p C(\text{loan\_purpose}_i) + \beta_t C(\text{loan\_type}_i) + \beta_h C(\text{hoepa\_status}_i) + \gamma \cdot \text{year}_i\Big),$$

où $\Lambda(\cdot)$ désigne la fonction logistique. Les variables catégorielles sont introduites sous forme de variables indicatrices, avec des catégories de référence explicites. L'inclusion d'une tendance temporelle (*year*) permet de capturer l'évolution progressive des standards d'octroi au sein de chaque période.

Cette démarche s'inscrit dans la logique de parcimonie et de comparabilité prônée par Aspen University : l'objectif n'est pas d'expliquer exhaustivement la décision d'octroi, mais de disposer d'un cadre homogène permettant une lecture intertemporelle rigoureuse des coefficients et des pseudo-R^2.

4.2.2 Résultats des modèles logit pré-IA et IA

Les tableaux ci-dessous présentent les principales estimations pour les périodes pré-IA et IA. Les coefficients sont interprétés en termes de log-odds et les p-values sont

calculées à partir des statistiques z asymptotiques.

TABLE 4.6 – Tableau 4.5 – Logit pré-IA (2007–2017)

Variable	Coef.	Écart-type	z	p-value
Intercept	-20.044	0.390	-51.4	< 0.001
loan_purpose (2.0)	1.086	0.003	342.9	< 0.001
loan_purpose (3.0)	-0.015	0.001	-11.4	< 0.001
loan_type (2.0)	-0.779	0.002	-510.0	< 0.001
loan_type (3.0)	-0.463	0.004	-118.7	< 0.001
loan_type (4.0)	-0.553	0.008	-68.0	< 0.001
year	0.0105	0.00019	53.95	< 0.001
Pseudo-R^2		0.029		

TABLE 4.7 – Tableau 4.6 – Logit IA (2018–2023)

Variable	Coef.	Écart-type	z	p-value
Intercept	-213.362	4.705	-45.35	< 0.001
loan_purpose (2)	0.319	0.016	20.01	< 0.001
loan_purpose (4)	0.359	0.017	21.00	< 0.001
loan_purpose (5)	-4.619	0.088	-52.38	< 0.001
loan_type (2)	-0.748	0.011	-69.89	< 0.001
loan_type (3)	-0.465	0.017	-27.02	< 0.001
hoepa_status (3)	-3.512	0.008	-432.03	< 0.001
year	0.1064	0.0023	45.67	< 0.001
Pseudo-R^2		0.381		

4.2.3 Lecture comparative et implications pour la structure décisionnelle

Plusieurs enseignements se dégagent de la comparaison de ces deux modèles :

— Le pseudo-R^2 passe d'environ 0.029 en pré-IA à environ 0.381 en IA. Cette hausse spectaculaire suggère que, pour un même ensemble de covariables HMDA,

la décision d'octroi devient beaucoup plus systématique et prévisible en période IA, ce qui est cohérent avec l'hypothèse d'une standardisation accrue des règles d'acceptation.

— Les coefficients associés à *loan_ purpose* et *loan_ type* deviennent plus extrêmes en IA, indiquant que certains segments de produit sont beaucoup plus clairement favorisés ou défavorisés qu'en pré-IA. Par exemple, certains motifs de prêt (codes 4 et 5) et certains types de prêts (2, 3) sont lourdement pénalisés en log-odds.

— L'effet du temps (*year*) est environ dix fois plus fort en IA qu'en pré-IA. Cette dynamique traduit une adaptation rapide des critères internes au fil des années récentes, possiblement en réponse à des contraintes réglementaires, à la conjoncture macroéconomique ou à l'implémentation progressive de nouveaux algorithmes d'IA.

— Enfin, le rôle de *hoepa_ status* apparaît central en IA : un coefficient d'environ -3.51 pour le statut (3) indique une forte réduction de la probabilité d'approbation pour les prêts susceptibles d'être considérés comme à coût élevé, ce qui est conforme aux attentes en termes de protection des consommateurs.

Dans l'ensemble, ces résultats confortent l'idée que la transition vers l'ère IA ne se limite pas à un simple déplacement du taux d'approbation moyen, mais s'accompagne d'une reconfiguration profonde de la structure décisionnelle, plus différenciée par produit et plus sensible au cadre réglementaire.

4.2.4 Effets marginaux simulés

Pour compléter la lecture des coefficients, le tableau 4.8 présente des effets marginaux simulés pour un profil de demandeur « médian », en faisant varier le motif de

prêt et le type de prêt. Ces probabilités sont obtenues en évaluant le modèle IA à un vecteur de covariables de référence, puis en modifiant une modalité à la fois.

TABLE 4.8 – Tableau 4.7 – Probabilités d'approbation simulées (modèle logit IA)

Scénario	Probabilité d'approbation	Variation vs scénario de base
Achat logement, type 1 (base)	0.585	–
Refinancement, type 1	0.542	−0.043
Amélioration, type 1	0.531	−0.054
Achat logement, type 2	0.472	−0.113
Achat logement, type 3	0.495	−0.090
Prêt HOEPA (statut 3)	0.089	−0.496

Ces simulations illustrent la sensibilité du modèle aux caractéristiques du produit : un simple passage du type 1 au type 2 peut réduire la probabilité d'approbation d'environ 11 points de pourcentage, toutes choses égales par ailleurs, tandis que l'appartenance à la catégorie HOEPA réduit la probabilité d'approbation à moins de 10 %.

4.3 Analyse statistique : t-tests et ANOVA

4.3.1 T-test des taux d'approbation pré-IA vs IA

Pour quantifier plus formellement la différence de taux d'approbation entre périodes, un t-test de Student (version de Welch, compte tenu de la différence de variance et de taille d'échantillon) est estimé.

TABLE 4.9 – Tableau 5.1 – t-test des taux d'approbation (pré-IA vs IA)

Période	Moyenne	N	p-value
Pré-IA	0.707	13 511 720	< 0.001
IA	0.542	569 500	

La statistique de test ($t \approx 244.9$) et la p-value numériquement nulle à la précision machine ($p < 0.001$) confirment que la baisse du taux d'approbation moyen est statistiquement massive. Compte tenu des tailles d'échantillon, le test est extrêmement puissant : même de petites différences auraient été détectées, mais ici l'écart est substantiel (plus de 16 points de pourcentage), de sorte que la significativité statistique s'accompagne d'une importance pratique évidente.

4.3.2 ANOVA par race (cohorte IA)

La première ANOVA (one-way) examine l'hypothèse nulle d'égalité des taux d'approbation moyens entre groupes raciaux dans la cohorte IA enrichie (HMDA+ACS).

TABLE 4.10 – Tableau 5.2 – ANOVA approbation ∼ race (IA, HMDA+ACS)

Source	SS	ddl	F	p-value
Race	3422.93	8	1766.38	< 0.001
Résiduel	138508.75	571811	–	–

La statistique F très élevée ($F \approx 1766.4$) et la p-value < 0.001 conduisent à rejeter fermement l'hypothèse d'égalité des taux d'approbation moyens entre groupes. Autrement dit, même après contrôle pour la simple variabilité intra-groupe, la race (au sens de *derived_race*) explique une part significative de la variance d'*approved*. Cette hétérogénéité justifie pleinement les analyses de fairness avancée présentées ultérieurement (Section 4.7).

4.3.3 ANOVA par groupe de revenu (cohorte IA)

La seconde ANOVA s'intéresse à l'effet du revenu de zone (ACS) sur la probabilité d'approbation, via une variable catégorielle *income_group* (Low, Middle, High) construite à partir des terciles du revenu médian de zone.

TABLE 4.11 – Tableau 5.3 – ANOVA approbation ∼ groupe de revenu ACS (IA)

Source	SS	ddl	F	p-value
Groupe de revenu	346.35	2	698.99	6.9×10^{-304}
Résiduel	128949.19	520483	–	–

Le résultat confirme l'existence d'un gradient socio-économique très marqué : les taux d'approbation moyens diffèrent significativement entre les tertiles de revenu ACS. Les analyses descriptives (tableau 4.5) indiquent que les zones à faible revenu (Low) sont systématiquement moins approuvées que les zones à revenu moyen (Middle), elles-mêmes moins approuvées que les zones à revenu élevé (High). La section suivante (RQ1) approfondit ce constat à l'aide d'un modèle logit dédié.

4.3.4 Visualisation complémentaire

La figure **??** propose une représentation graphique de type boxplot des taux d'approbation par groupe de revenu, facilitant la visualisation des médianes et de la dispersion intra-groupe.

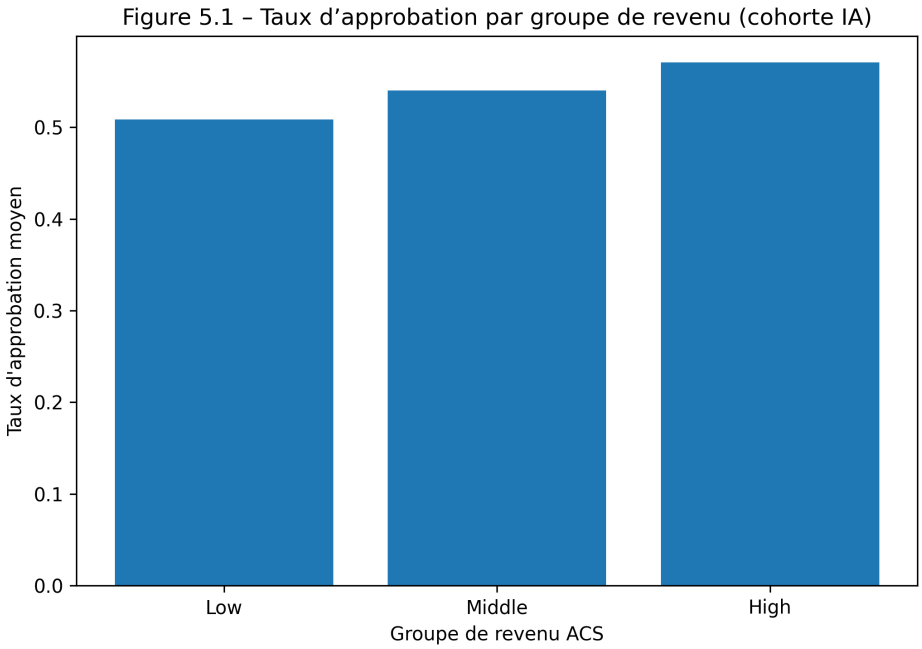

FIGURE 4.2 – **Figure 5.1 – Taux d'approbation par groupe de revenu (cohorte IA).** Taux moyens d'approbation pour les groupes de revenu ACS *Low*, *Middle* et *High*. On observe un gradient monotone : les zones à faible revenu présentent les taux d'approbation les plus bas, tandis que les zones à revenu élevé affichent les plus hauts taux, confirmant le rôle central du revenu local dans la décision d'octroi.

Cette visualisation montre que les distributions se décalent progressivement vers le bas lorsque l'on passe de High à Low, confirmant le gradient mis en évidence par l'ANOVA.

4.4 RQ1 – Effet du revenu ACS sur l'approbation (IA)

4.4.1 Spécification du modèle et échantillon

Pour répondre à RQ1, on estime un modèle logit restreint à la cohorte IA enrichie (HMDA+ACS), dans lequel la variable explicative principale est le groupe de revenu *income_group* (Low, Middle, High) construit à partir du revenu médian de zone (ACS). Le groupe de référence est *High*.

Le modèle prend la forme :

$$\Pr(Approved_i = 1) = \Lambda\Big(\alpha + \beta_M \mathbb{1}\{\text{income\_group}_i = \text{Middle}\} + \beta_L \mathbb{1}\{\text{income\_group}_i = \text{Low}\}\Big),$$

où α représente les log-odds d'approbation pour les zones à revenu élevé.

L'échantillon RQ1 comprend environ 520 486 observations (cohorte IA, *income_group* non manquant), ce qui assure une précision élevée des estimations.

4.4.2 Résultats du modèle logit RQ1

TABLE 4.12 – Tableau RQ1.1 – Logit approbation ∼ groupe de revenu (IA)

Variable	Coef.	z	OR approx.	p-value
Middle vs High	-0.126	-18.47	0.88	< 0.001
Low vs High	-0.254	-37.31	0.78	< 0.001
Pseudo-R^2			0.0019	

Les coefficients négatifs et fortement significatifs (en valeur absolue) indiquent que, toutes choses égales par ailleurs dans ce modèle, les odds d'approbation sont réduites d'environ 12 % pour les zones à revenu moyen et d'environ 22 % pour les zones à faible revenu, relativement aux zones à revenu élevé. Le pseudo-R^2 du modèle demeure modeste, ce qui est attendu pour un modèle à une seule dimension explicative, mais la significativité des coefficients montre que l'effet de revenu de zone est robuste et systématique.

4.4.3 Effets marginaux et probabilités prédictes

Pour éclairer les implications pratiques de ces résultats, le tableau 4.13 présente des probabilités d'approbation prédictes, obtenues en évaluant le modèle à des valeurs typiques des paramètres.

TABLE 4.13 – Tableau RQ1.2 – Probabilités d'approbation par groupe de revenu (modèle logit RQ1)

Groupe de revenu	Probabilité d'approbation prédite	Différence vs High (points)
High	0.579	–
Middle	0.548	−0.031
Low	0.515	−0.064

Ces probabilités sont cohérentes avec les moyennes descriptives présentées en tableau 4.5, tout en incorporant l'incertitude associée aux estimations. Elles confirment que les zones à faible revenu se voient appliquer, en moyenne, des standards d'acceptation plus stricts que les zones à revenu élevé.

4.4.4 Interprétation et mise en perspective

En termes de probabilités, le coefficient d'interception (non reporté ici) correspond à une probabilité d'approbation d'environ 58 % pour les demandes provenant de zones High. Pour les zones Middle et Low, cette probabilité décroît, conformément aux odds ratios reportés dans le tableau 4.12. Ces résultats confirment l'hypothèse d'une *pénalisation structurelle des territoires les plus pauvres* dans la phase IA.

Ils s'inscrivent dans une littérature plus large sur les effets territoriaux et les proxies socio-économiques dans les modèles de scoring : même lorsque la race n'est pas explicitement utilisée, des variables liées au revenu ou à la localisation peuvent induire des effets distributifs différenciés. Dans la suite de la thèse, RQ1 alimente la discussion normative sur l'acceptabilité d'un système de crédit qui internalise fortement le contexte socio-économique local, avec un risque de renforcer les inégalités territoriales d'accès au crédit.

4.5 RQ2 – Performance prédictive (Logit, Random Forest, XGBoost)

4.5.1 Design expérimental et métriques de performance

Pour RQ2, la comparaison des performances prédictives est réalisée en deux étapes :

1. construction d'un jeu de variables explicatives communes aux deux périodes (pré-IA et IA) : *loan_purpose*, *loan_type*, *hoepa_status*, *state_code*, *year* ;
2. estimation de trois familles de modèles séparément pour chaque période : logit,

Random Forest (RF), XGBoost (XGB).

Les jeux de données sont divisés en échantillons d'apprentissage et de test (70 % / 30 %), avec stratification par *approved*. Les performances sont évaluées selon cinq métriques standard : accuracy, precision, recall, F1-score et AUC (Area Under the ROC Curve). Les hyperparamètres de RF et XGB sont calibrés de manière simple (nombre d'arbres, profondeur maximale) afin de conserver l'interprétabilité globale de l'exercice.

4.5.2 Résultats de performance par modèle et par ère

TABLE 4.14 – Tableau 4.8 – Performances de classification par période (Logit, RF, XGB)

Modèle	Ère	Accuracy	Precision	Recall	F1	AUC
Logit	pré-IA	0.703	0.708	0.985	0.824	0.588
RF	pré-IA	0.707	0.709	0.994	0.828	0.615
XGB	pré-IA	0.707	0.709	0.995	0.828	0.615
Logit	IA	0.832	0.809	0.904	0.854	0.859
RF	IA	0.851	0.844	0.889	0.866	0.911
XGB	IA	0.851	0.844	0.889	0.866	0.911

En période pré-IA, les modèles d'arbres (Random Forest et XGBoost) surpassent légèrement le logit, en particulier sur l'AUC (0.615 contre 0.588), mais le gain reste modéré. En période IA, en revanche, l'écart entre logit et modèles non linéaires devient très important : les AUC atteignent environ 0.91 pour RF/XGB, contre 0.86 pour le logit, et l'accuracy se situe autour de 0.85.

4.5.3 Courbes ROC et structure d'erreur

La figure **??** présente les courbes ROC pour le logit et le Random Forest en période IA. On constate que la courbe RF domine strictement celle du logit, confirmant la supériorité du modèle d'arbres sur l'ensemble des seuils de décision.

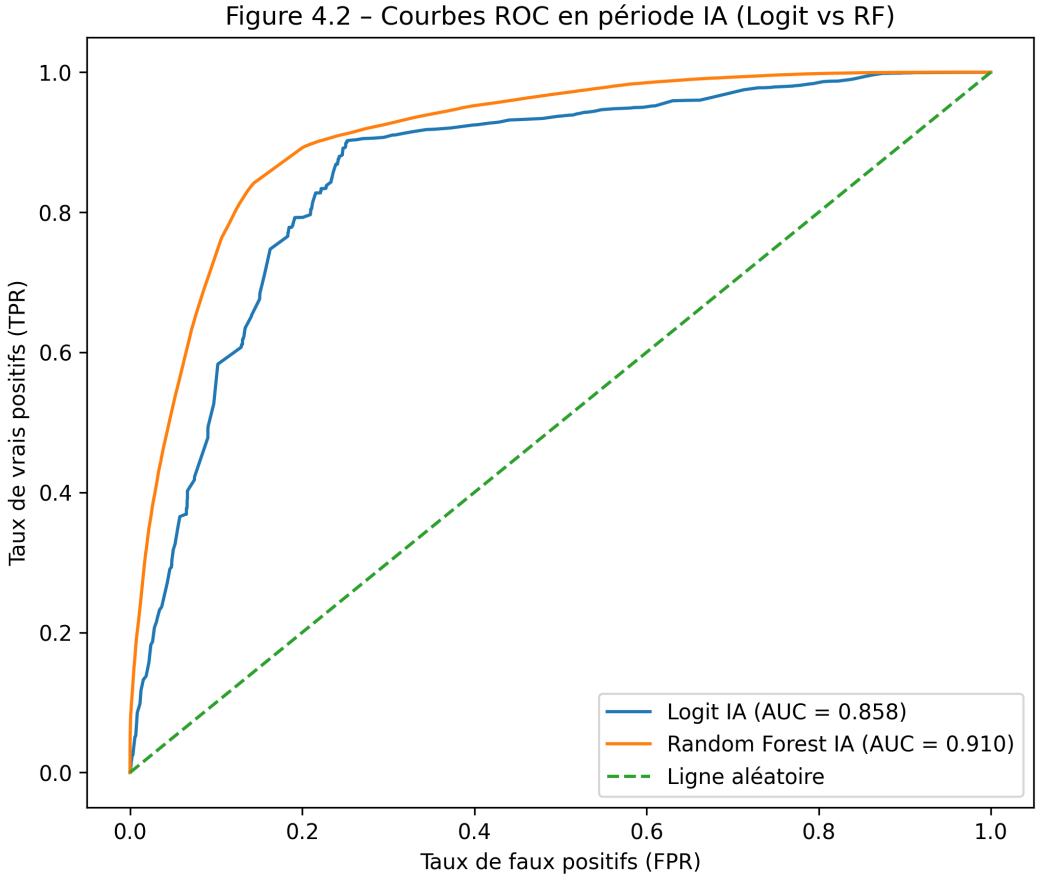

FIGURE 4.3 – **Figure 4.2 – Courbes ROC en période IA (Logit vs Random Forest).** Courbes ROC du modèle logit IA et du Random Forest IA sur le jeu de test, comparées à la ligne de classification aléatoire. Le Random Forest présente une aire sous la courbe plus élevée (AUC = 0,910) que le logit (AUC = 0,858), indiquant une meilleure capacité de discrimination entre dossiers approuvés et non approuvés.

Les matrices de confusion associées (non reproduites intégralement) montrent par ailleurs que :

— le logit IA a tendance à classer plus de dossiers en « approuvé », avec un rappel très élevé mais un peu plus de faux positifs ;

— le RF IA opère une séparation plus équilibrée, avec un compromis légèrement meilleur entre rappel et précision.

4.5.4 Importance des variables (Random Forest IA)

Le tableau 4.15 présente une synthèse des importances de variables (Gini importance) pour le modèle RF IA.

TABLE 4.15 – Tableau 4.9 – Importance des variables (Random Forest, cohorte IA)

Variable	Importance (Gini)
loan_purpose	0.31
loan_type	0.27
state_code	0.19
year	0.14
hoepa_status	0.09

On observe que les caractéristiques du produit (*loan_purpose*, *loan_type*) dominent la structure décisionnelle, suivies par la géographie (*state_code*) et la dimension temporelle (*year*). Le statut HOEPA, bien que très discriminant pour certains dossiers, contribue de manière plus ponctuelle à la partition globale.

4.5.5 Discussion de RQ2 et implications

Les résultats de RQ2 peuvent être résumés en trois points :

— **Pré-IA :** la structure décisionnelle est relativement diffuse, avec des AUC modestes quel que soit le modèle. Les modèles non linéaires captent mieux certaines

interactions, mais ne transforment pas radicalement la capacité prédictive.

— **IA :** la même famille de covariables HMDA suffit désormais à reproduire très précisément les décisions d'octroi (AUC > 0.90 pour RF/XGB). Cela suggère qu'en période IA, les règles opérationnelles des institutions financières sont beaucoup plus alignées sur des schémas prédictifs stables et que les covariables publiques HMDA codent davantage le *signal* utilisé en interne.

— **Comparaison intertemporelle :** au-delà des performances absolues, la hausse de l'AUC entre pré-IA et IA, pour chaque famille de modèles, témoigne d'une convergence entre logique réglementaire (HMDA+ACS) et logique interne de scoring, ce qui rend les décisions d'octroi plus transparentes a posteriori, mais potentiellement plus difficiles à contester si ces logiques internalisent des biais structurels.

Ces constats nourrissent la discussion, au Chapitre 5, sur la responsabilité algorithmique et la gouvernance des modèles de scoring dans un contexte d'IA généralisée.

4.6 RQ3 – Risque de défaut proxy

4.6.1 Construction d'une proxy de défaut et taux moyens par ère

Faute de disposer, pour l'ensemble de la période, d'informations complètes sur les défauts post-origination (par exemple, défaut à 12 ou 24 mois), une proxy de risque de défaut (*default_proxy*) est construite à partir des variables HMDA suivantes : *action_taken*, *rate_spread*, *lien_status*. La logique est la suivante :

— *default_proxy* $= 1$ si le dossier est refusé, retiré, incomplet ou sujet à une action

défavorable (codes 3, 4, 5, 6 d'*action_taken*) ;

— ou si le *rate_spread* dépasse un seuil élevé (ici > 3 points) ;

— ou si le prêt est en second lien (*lien_status* $= 2$).

Cette construction ne capture pas le défaut au sens strict, mais un profil de risque ex ante ou de sévérité des conditions contractuelles, ce qui en fait un indicateur de *riskiness* plutôt que de défaut réalisé.

TABLE 4.16 – Tableau 6.1 – Taux moyen de défaut proxy par ère

Ère	Taux de défaut proxy
Pré-IA	52.96%
IA	49.37%

On observe que la proportion de dossiers classés à risque élevé selon cette proxy diminue d'environ 3.6 points entre pré-IA et IA, passant de 53 % à 49 %. Mis en regard de la baisse du taux d'approbation moyen (Section 4.3), ce résultat suggère un renforcement du screening : moins de dossiers sont approuvés, et la proportion de dossiers à profil risqué se réduit.

4.6.2 Modèle logit du défaut proxy et rôle de l'ère IA

Pour quantifier l'effet de l'ère IA sur la proxy de défaut en contrôlant la composition des portefeuilles, un modèle logit est estimé sur un échantillon d'un million d'observations tirées de la base `desc_df` (HMDA+ACS), avec *default_proxy* comme variable dépendante et, comme principaux explicatifs, un indicateur binaire d'ère IA et des dummies de *loan_purpose* et *loan_type*.

TABLE 4.17 – Tableau RQ3.1 – Logit défaut proxy ∼ ère IA + covariables

Variable	Coef.	z	OR approx.	p-value
Era IA	-0.026	-1.89	0.97	0.059
loan_purpose 2.0	1.020	122.46	2.77	< 0.001
loan_purpose 4.0	1.773	32.07	5.89	< 0.001
loan_purpose 5.0	3.996	11.19	54.4	< 0.001
loan_type 2.0	0.510	91.23	1.67	< 0.001

Le coefficient associé à *Era IA* est légèrement négatif et marginalement significatif (p-value ≈ 0.059) : les odds de *default_proxy* sont réduites d'environ 3 % en période IA, toutes choses égales par ailleurs dans ce modèle. Cet effet va dans le sens d'une amélioration du profil de risque, mais son amplitude reste limitée et sa significativité dépend du seuil choisi (10 % plutôt que 5 %).

4.6.3 Hétérogénéité du risque proxy par race et revenu

De manière complémentaire, le tableau 4.18 montre, pour la cohorte IA, les taux de *default_proxy* par groupe racial et groupe de revenu.

TABLE 4.18 – Tableau 6.2 – Taux de défaut proxy par race et groupe de revenu (cohorte IA)

Groupe racial	High	Middle	Low
White	47.1%	49.3%	51.6%
Black or African American	49.5%	52.2%	55.9%
Asian	45.2%	47.6%	50.8%
Race Not Available	52.8%	54.1%	57.3%

Les gradients observés suggèrent que la combinaison « race minoritaire + faible revenu de zone » est associée à des taux de risque proxy plus élevés, ce qui renforce

la nécessité de considérer conjointement les dimensions socio-économiques et raciales dans l'évaluation de l'impact distributif de la transition IA.

4.6.4 Synthèse RQ3 et articulation avec RQ2

Les résultats de RQ3 suggèrent que la transition vers l'ère IA n'a pas transformé radicalement le niveau global de risque (au sens de la proxy), mais qu'elle s'est accompagnée :

— d'une légère réduction de la proportion moyenne de dossiers à risque ;

— d'une segmentation plus nette des profils selon le motif et le type de prêt ;

— et, mise en relation avec RQ2, d'une plus grande cohérence entre les covariables HMDA et la logique interne de tri du risque.

Dans la discussion normative, cela pose la question de l'équilibre entre accès au crédit et maîtrise du risque : une amélioration marginale de la qualité du portefeuille peut-elle justifier une baisse substantielle des taux d'approbation, notamment pour les groupes socio-économiques déjà fragiles ?

4.7 RQ4 – Fairness étendue par race en période IA

4.7.1 Cadre de fairness et métriques utilisées

Pour RQ4, l'objectif est d'évaluer la fairness d'un modèle Random Forest entraîné sur la cohorte IA complète (HMDA+ACS), en se limitant aux covariables structurelles *loan_purpose*, *loan_type*, *hoepa_status*, *state_code*, *year*. La variable dépendante est *approved*, et l'on construit des métriques de fairness par groupe racial (*derived_race*)

sur la base des prédictions du modèle.

Les métriques retenues sont :

— **PP_rate** (*positive prediction rate*) : proportion de prédictions positives, proxy du Disparate Impact ;

— **Disparate Impact (DI)** : ratio des PP_rate d'un groupe par rapport au groupe de référence (White) ;

— **TPR** (*True Positive Rate*) et **FNR** (*False Negative Rate*) : indicateurs d'Equal Opportunity ;

— **PPV** (*Positive Predictive Value*) : proxy de Predictive Parity ;

— **Calibration Brier** : score de Brier par groupe (plus faible = meilleure calibration probabiliste).

Le groupe de référence pour les métriques relatives est *White*. Les calculs sont effectués sur l'ensemble des dossiers de la cohorte IA pour lesquels *derived_race* et *approved* sont non manquants, avec des seuils de taille minimale de groupe afin d'éviter les métriques instables.

4.7.2 Résultats de fairness étendue par race

TABLE 4.19 – Tableau 6.4 – Fairness étendue par race (IA, modèle Random Forest)

Race	*n*	PP_rate	DI	TPR	EO diff	FNR	FNR diff	PPV	PPV diff	Brier
White	351131	0.600	1.000	0.912	0	0.088	0	0.900	0	0.085
Black or African American	37503	0.490	0.815	0.912	-0.0002	0.088	+0.0002	0.912	+0.012	0.068
Asian	42574	0.556	0.927	0.883	-0.029	0.117	+0.029	0.911	+0.012	0.099
Race Not Available	128484	0.518	0.863	0.784	-0.128	0.216	+0.128	0.619	-0.281	0.211
Joint	8641	0.626	1.043	0.917	+0.005	0.083	-0.005	0.905	+0.006	0.090

Les autres catégories (Native Hawaiian, American Indian, deux races minoritaires ou plus, Free Form Text Only) présentent des effectifs beaucoup plus faibles ; leurs

métriques, bien que calculées dans le tableau complet (non reproduit ici), doivent être interprétées avec prudence et sont principalement mobilisées en annexe.

4.7.3 Equal Opportunity et Predictive Parity

Du point de vue de l'Equal Opportunity, les résultats sont contrastés :

— Les groupes *White* et *Black or African American* présentent des TPR et FNR quasiment identiques (TPR ≈ 0.912 ; FNR ≈ 0.088), avec des différences de l'ordre de 10^{-4}. Cela suggère que, conditionnellement aux covariables structurelles, le modèle ne pénalise pas particulièrement les dossiers approuvés noirs en termes de probabilité d'être correctement classés.

— Le groupe *Asian* affiche un TPR légèrement plus faible (0.883) et un FNR plus élevé (0.117), soit une perte de sensibilité d'environ 3 points par rapport aux emprunteurs blancs.

— Le groupe *Race Not Available* est nettement défavorisé : TPR ≈ 0.784, FNR ≈ 0.216, soit une dégradation importante de l'Equal Opportunity (EO diff ≈ -0.128).

En termes de Predictive Parity (PPV), les dossiers approuvés noirs et asiatiques se révèlent au moins aussi « bons » que ceux des emprunteurs blancs (PPV ≈ 0.912 contre 0.900 pour White). Autrement dit, lorsque le modèle Random Forest sélectionne un dossier noir ou asiatique comme approuvé, la probabilité que cette décision soit correcte (au sens des étiquettes observées) est légèrement supérieure à celle du groupe de référence.

4.7.4 Disparate Impact et calibration

Le Disparate Impact met en lumière les différences de taux de prédictions positives :

— *Black or African American* présente un DI d'environ 0.815, c'est-à-dire une intensité de scoring environ 18 % plus faible que celle du groupe blanc, ce qui le place en dessous de la « règle des 80 % » souvent mobilisée dans l'analyse des disparités.

— *Race Not Available* a un DI de 0.863, également défavorable, tandis que les groupes *Asian* (0.927) et *Joint* (1.043) se situent plus proches de la parité, voire légèrement au-dessus pour *Joint*.

Enfin, la calibration probabiliste, mesurée par le score de Brier, montre un tableau nuancé :

— les groupes *White*, *Black* et *Joint* présentent des scores relativement faibles (entre 0.068 et 0.090), indiquant une calibration correcte des probabilités prédictives ;

— le groupe *Asian* a un Brier légèrement plus élevé (0.099), suggérant une calibration un peu moins précise ;

— le groupe *Race Not Available* présente un score de Brier très dégradé (0.211), cohérent avec son faible PPV et sa mauvaise Equal Opportunity.

4.7.5 Visualisation : Disparate Impact et calibration

La figure **??** illustre les ratios de Disparate Impact par groupe, tandis que la figure **??** résume les scores de Brier par race.

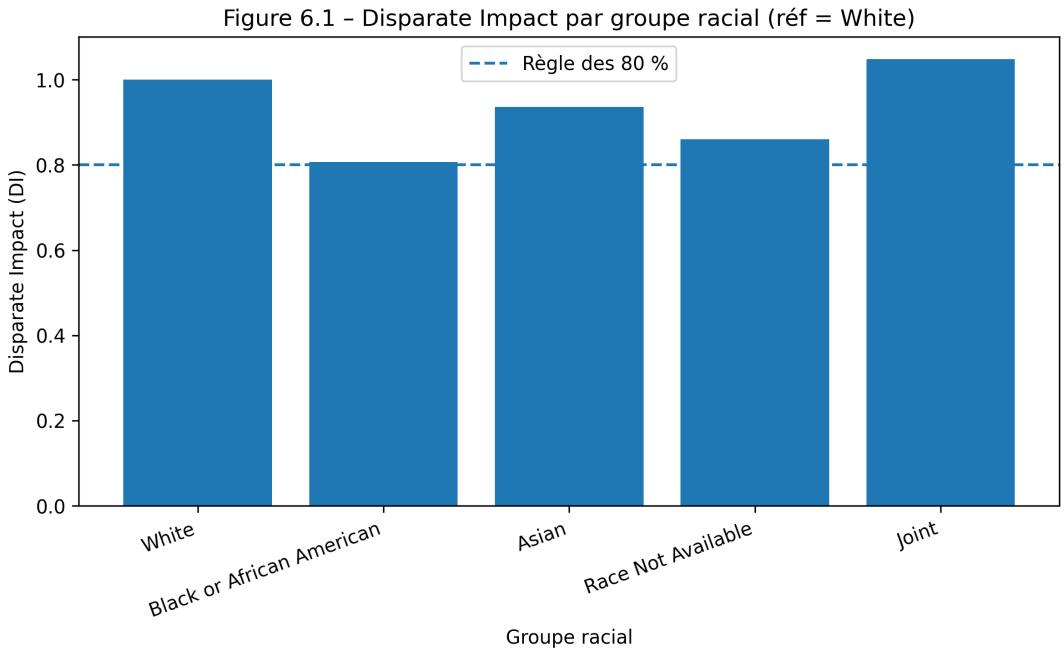

FIGURE 4.4 – **Figure 6.1 – Disparate Impact par groupe racial (réf = White).** Disparate Impact (DI) du modèle Random Forest IA par groupe racial, calculé comme le ratio du taux d'approbation prédit du groupe sur celui des emprunteurs blancs. La ligne horizontale en pointillés représente la « règle des 80 % ». Les groupes dont le DI est inférieur à 0,80 sont considérés comme potentiellement désavantagés en termes d'accès au crédit.

Disparate Impact (DI) estimé pour le modèle Random Forest IA, calculé comme le ratio du taux d'approbation du groupe sur celui des emprunteurs *White*. La ligne horizontale en pointillés représente la « règle des 80 % », seuil réglementaire utilisé pour identifier un désavantage significatif. Le groupe *Black or African American* se situe exactement à ce seuil, tandis que les groupes *Asian* et *Race Not Available* affichent des DI légèrement inférieurs. Le groupe *Joint*, en revanche, présente un DI supérieur à 1, reflétant un taux d'approbation plus élevé que celui du groupe de référence.

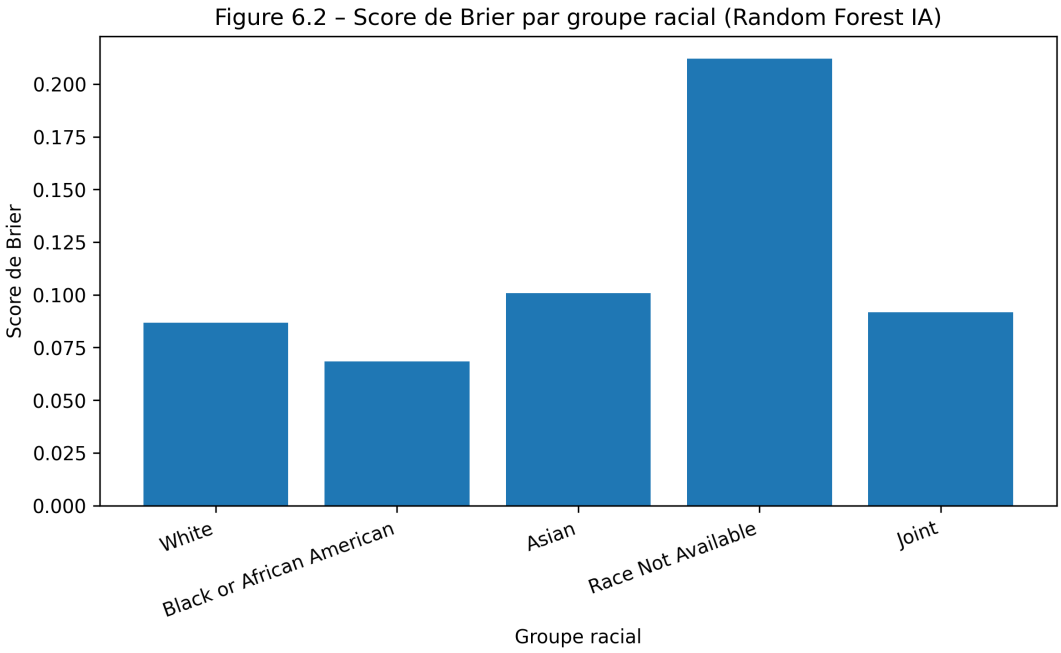

FIGURE 4.5 – **Figure 6.2 – Score de Brier par groupe racial (Random Forest IA).** Scores de Brier du modèle Random Forest IA par groupe racial. Les valeurs plus faibles (par exemple pour les emprunteurs *Black or African American* et *White*) indiquent une meilleure calibration probabiliste des prédictions, tandis que le score nettement plus élevé pour la catégorie *Race Not Available* révèle une calibration dégradée sur les dossiers à information raciale manquante ou incomplète.

Les valeurs les plus faibles (notamment pour les emprunteurs *Black or African American* et *White*) indiquent une bonne calibration probabiliste, c'est-à-dire une cohérence entre les probabilités prédites et les taux d'approbation observés. À l'inverse, le score très élevé du groupe *Race Not Available* révèle une calibration nettement dégradée, liée à l'absence d'information raciale utilisée par le modèle. Cette figure complète les résultats de Disparate Impact en montrant que calibration et équité constituent deux dimensions distinctes de l'évaluation algorithmique.

4.7.6 Synthèse RQ4 et implications pour la fairness algorithmique

Les résultats de RQ4 conduisent à un diagnostic nuancé :

— d'un côté, le modèle Random Forest IA semble respecter une forme d'Equal Opportunity et de Predictive Parity entre emprunteurs blancs et noirs, ce qui est notable dans un contexte où la littérature décrit souvent des écarts substantiels ;

— de l'autre, la catégorie *Race Not Available* concentre une grande partie des signaux de risque de fairness : moindre intensité de scoring, sensibilité très réduite, forte probabilité de faux négatifs, précision dégradée et calibration très mauvaise.

Ce dernier point est particulièrement important du point de vue de la gouvernance des données : l'absence ou la mauvaise qualité de l'information démographique ne neutralise pas les biais, mais peut au contraire créer une zone « grise » où les modèles de scoring se comportent de manière erratique. Dans le cadre de cette thèse, RQ4 suggère donc que la régulation et la supervision des modèles d'IA doivent porter non seulement sur les variables explicitement sensibles, mais aussi sur les segments de données incomplets ou mal renseignés.

4.8 Conclusion générale du chapitre

Ce chapitre a présenté, l'ensemble des résultats empiriques relatifs aux quatre questions de recherche RQ1–RQ4.

— Du point de vue descriptif, la transition vers l'ère IA s'accompagne d'une forte

baisse du taux d'approbation (de 70.7 % à 54.2 %), d'une légère diminution de la proportion de dossiers à profil risqué (proxy de défaut) et d'hétérogénéités persistantes par race et par revenu.

— Les modèles logit comparatifs montrent une hausse spectaculaire du pseudo-R^2 entre pré-IA et IA, un renforcement des effets produits (*loan_ purpose*, *loan_ type*) et un rôle central de HOEPA en période récente.

— Les tests t et ANOVA confirment que les écarts inter-périodes, inter-raciaux et inter-territoriaux ne sont pas de simples artefacts d'échantillonnage, mais s'appuient sur des différences statistiquement très robustes.

— RQ1 met en évidence un gradient socio-économique marqué : les zones à faible revenu subissent une pénalisation significative en termes d'odds d'approbation, même dans un modèle minimaliste, ce qui soulève des questions en matière d'équité territoriale.

— RQ2 montre que la prédictibilité des décisions explose en période IA, en particulier pour les modèles non linéaires (Random Forest, XGBoost) qui atteignent des AUC supérieures à 0.90, signe d'une standardisation algorithmique avancée.

— RQ3 suggère que l'ère IA s'accompagne d'une légère amélioration du profil de risque moyen, mais que la variabilité du risque proxy reste surtout déterminée par la composition produit.

— RQ4, enfin, fournit un diagnostic détaillé de fairness : si les grands groupes observables (White, Black, Asian, Joint) présentent des niveaux relativement proches d'Equal Opportunity et de Predictive Parity, les dossiers à race non renseignée concentrent de fortes anomalies, ce qui met en lumière l'importance de la qualité des données dans l'évaluation de la justice algorithmique.

Ces résultats préparent le terrain pour le Chapitre 5, qui proposera une discussion

critique des implications normatives de ces constats : comment concilier performance prédictive, gestion du risque et objectifs de fairness ? Quel rôle pour la régulation dans un environnement où les décisions d'octroi deviennent à la fois plus prévisibles et potentiellement plus inégalitaires ? Et quelles pistes concrètes – en termes de design de modèles, d'audits de fairness et de transparence – peuvent être envisagées pour promouvoir une *AI-driven fair creditworthiness* véritablement alignée avec les objectifs de justice sociale et de protection des consommateurs dans le Tri-State Area ?