

MASTER ÉCONOMISTE D'ENTREPRISE



RECHERCHE, RÉALISATION, RESTITUTION

Les déterminants de l'approbation d'un prêt

Étudiants

BENICIA EKUBA KABUIKU
LUCAS PIRES

Promotion M2 2019-2020

Professeurs

SILVIA CONCETTINI
YANN KOSSI

Université de Tours

Table des matières

1	L'introduction générale	4
1.1	Une exclusion bancaire	4
1.2	Un dispositif anti-discrimination	5
2	Le cadre théorique	6
2.1	L'asymétrie d'information	6
2.2	Des acteurs actifs	7
3	Les travaux empiriques	8
3.1	Les variables principales	8
3.2	Une discrimination avérée	9
4	La base de données	10
4.1	La présentation de l'échantillon	10
4.2	Le biais sélection	10
5	L'analyse descriptive	11
5.1	La répartition des demandes de crédits	11
5.2	Les caractéristiques intrinsèques	11
5.3	Les disparités dans les exigences liées au prêt	13
5.4	Les matrices de corrélations	14
5.5	Les variables importantes	14
6	L'approche théorique des différents modèles	16
6.1	La méthodologie de la régression logistique	16
6.2	Les principes de la décomposition non linéaire	17
7	Les résultats empiriques	18
7.1	Les régressions logistiques	18
7.2	La décomposition de Fairlie	20
7.3	Les régressions selon l'ethnie	21

8	La conclusion finale	22
	Bibliographie	24
A	Annexes : Le sommaire des variables	25
B	Annexes : La programmation	26

1 L'introduction générale

1.1 Une exclusion bancaire

Si, tendanciellement, la demande de prêt hypothécaire a crû très fortement, les conditions d'accès à ce produit sont dans le même temps devenues plus restrictives, excluant une partie des ménages de ce système. Ces formes d'exclusion, très différentes d'un pays à l'autre, sont partout identifiées par les pouvoirs publics, qui tentent d'y apporter des réponses variables.

L'accessibilité des particuliers aux crédits est, en effet, une préoccupation assez récente dans les pays développés. La prise de conscience des conséquences économiques et sociales du problème d'exclusion bancaire a contraint progressivement la plupart des gouvernements à intervenir. Le système bancaire américain, par exemple, compte aujourd'hui dix-mille institutions réparties sur l'ensemble du territoire. Malgré une densité correcte du réseau, les disparités géographiques de localisation dégagent une cruelle vérité : les banques refusent non seulement à s'installer mais aussi à consentir des prêts dans les quartiers pauvres et minoritaires. Les États-Unis excluent donc près de 20% de sa population de son système bancaire. Géographiquement, la stratégie des banques américaines est celle de l'encerclement (redlining) autour des zones pauvres qui se retrouvent donc totalement exclues du système bancaire. Cette politique consiste en effet à refuser des prêts aux habitants sur la base de leur domiciliation, accentuant d'autant plus le processus de paupérisation et de dévalorisation de ces quartiers. En 1989, la FED de Boston¹ montre que les douze plus grandes institutions de la ville gèrent cinq fois plus d'agences sur les zones WASP² que sur celles des hispaniques ou afro-américaines. Entre 1978 et 1988, les banques ferment 40% de leurs agences dans les quartiers pauvres, tandis que parallèlement, la densité du réseau WASP s'accroît de 30%. La France n'est pas en reste puisque des tests ont été réalisés en 2017 dans l'agglomération de Lyon et révèlent des inégalités de traitement significatives, essentiellement par rapport à l'origine migratoire des candidats.

1. Federal Reserve Bank of Boston : couvre la Nouvelle Angleterre et l'ensemble du Connecticut

2. White Anglo-Saxon Protestant : désigne l'archétype de l'Anglo-Saxon issu d'Europe de l'Ouest

1.2 Un dispositif anti-discrimination

Dans ce concert des discriminations relatives à la plupart des pays riches, les systèmes judiciaires rétorquent par l'écriture de lois pour en réprimer les pratiques. Toutefois, ces textes diffèrent entre pays sur les mesures et sanctions pratiquées en cas de discrimination avérée.

Le cadre légal américain comporte quatre lois spécifiques visant à combattre l'exclusion systématique de populations, en particulier les afro-américains et les hispaniques. Parmi celles-ci tout d'abord la loi sur le logement équitable (FHA)³ qui, dès 1968, interdisait toute pratique discriminatoire en matière de prêt immobilier. En 1974, la loi sur l'égalité des chances en matière de crédit (ECOA)⁴ vient renforcer ce dispositif en ne portant plus uniquement sur les crédits immobiliers mais sur tous les types de crédit. Par ces lois, sont rendues illégales les pratiques bancaires qui excluent sur la base de l'ethnie, la religion, la nationalité, le genre ou encore l'âge. En 1975, une étape supplémentaire est franchie par le vote de la loi sur la divulgation des hypothèques résidentielles (HMDA)⁵ qui exige des institutions de dépôts réglementées au niveau fédéral ainsi que de certaines sociétés de crédit spécialisées de fournir aux régulateurs des données statistiques annuelles sur leur activité de prêt immobilier. L'objectif de cette loi était alors de pouvoir apprécier la répartition des prêts octroyés afin de cerner la réalité et de mieux quantifier l'ampleur du phénomène de l'accès différencié au crédit de certains quartiers, en particulier de ceux à revenus faibles à modérés. Une quatrième loi, portant sur le réinvestissement communautaire (CRA)⁶ est promulguée en 1977 et vise à encadrer les pratiques bancaires auprès des populations en difficultés. Ces quatre lois visant à assurer l'égalité s'inscrivent donc dans un ensemble de règles plus vaste, qui dépasse la sphère des activités financières. Elles font en effet suite à un contexte politique, économique et social singulier, celui des années soixante qui a vu naître d'importants mouvements contestataires.

3. Fair Housing Act : interdit à quiconque de pratiquer la discrimination à la location

4. Equal Credit Opportunity Act : encadre les actions financières des populations marginalisées

5. Home Mortgage Disclosure Act : impose les banques à publier des données hypothécaires

6. Community Reinvestment Act : lutte contre la discrimination communautaire de certains quartiers

2 Le cadre théorique

2.1 L'asymétrie d'information

L'introduction de l'hypothèse d'asymétrie d'information par Akerlof (1970)[1] dans la théorie économique est un élément fondamental pour appréhender le processus d'octroi de prêt et la discrimination qui en découle. Cette notion centrale suppose en effet que dans une transaction économique, une partie au contrat est plus informée qu'une autre.

Bien que ce déséquilibre soit davantage rencontré dans le secteur des assurances, le marché bancaire n'est pas écarté puisque les emprunteurs sont plus informés que les prêteurs sur les divers paramètres de leurs projets. Contredisant par extension l'hypothèse de transparence du modèle standard de concurrence pure et parfaite, on peut distinguer deux situations d'information asymétrique : l'antisélection qui se déclare au moment de la signature du contrat (ex-ante) et l'aléa moral qui intervient après l'approbation du prêt (ex-post). Plus en détail, la sélection adverse, peut conduire à une allocation inefficace du crédit et notamment à un phénomène de rationnement. En effet, le partage inéquitable du savoir concernant le risque de défaillance attaché aux crédits rend problématique l'identification des bons emprunteurs. Ainsi, selon Stiglitz et Weiss (1981)[25], à défaut de pouvoir fixer un taux d'intérêt qui correspond au risque effectif du projet à financer, la banque applique un taux reflétant la qualité moyenne des emprunteurs. Une telle pratique conduit alors à pénaliser les individus dont le projet est peu risqué en leur faisant payer une prime de risque plus élevée que leur risque effectif, et à avantager inversement les agents détenant des projets risqués. S'en suit le risque moral qui est identifié par l'incomplétude de l'information qui provient du comportement non observable susceptible d'être entrepris suite à la signature du contrat. En effet, lors de l'octroi de crédit, le prêteur se positionne dans un cadre d'insuffisance d'information à propos des actions prises par l'emprunteur dont il doit contrôler l'activité pour vérifier qu'il ne cherche pas à dissimuler les profits réels découlant de son projet d'investissement. Pour conclure, d'après les travaux de Leland et Pyle (1977)[19], l'intermédiation financière reste donc insuffisante pour qu'elle puisse résoudre le problème d'antisélection et d'aléa moral. En effet, le savoir bancaire reste d'une part incomplète, tant que le rendement du projet d'investissement est lié à l'évolution de l'environnement, et d'autre part asymétrique dans le sens où l'information est inégale entre prêteur et emprunteur.

2.2 Des acteurs actifs

Malgré cette distribution informationnelle inéquitable entre les demandeurs et les établissements de crédit, ces acteurs ne sont toutefois pas passifs comme ont pu le suggérer les précédents postulats. En effet, si nous adoptons une perspective dynamique (temporelle), ces derniers vont adopter des comportements révélateurs d'information afin de démêler le vrai du faux et ceux à travers deux méthodes.

En jouant sur les différents termes des contrats de crédit (taux d'intérêt, garanties, clauses contractuelles, maturité), les prêteurs peuvent en effet obtenir de l'information sur la qualité réelle des emprunteurs. Une banque peut en effet proposer une gamme de contrats multicritères, structurés de telle manière que le choix d'un contrat par le client soit révélateur de l'information privée détenue par celui-ci. D'après les travaux de Bester (1985)[7], ce comportement actif des banques nécessite de considérer, au-delà du seul taux d'intérêt, un niveau de garantie externe. Ainsi, les emprunteurs les plus risqués auront tendance à choisir les contrats les plus onéreux et requérant le moins de sûretés puisqu'ils ont une probabilité plus importante de faire défaut et, en conséquence, de perdre la valeur de la sûreté engagée. Ces mécanismes incitatifs conduisent les emprunteurs à révéler leurs véritables natures tout en limitant l'incitation de ces derniers à adopter un comportement opportuniste une fois le prêt accordé. Enfin, concernant les « bons emprunteurs », il est capital pour ces derniers de prendre des mesures pour se signaler aux prêteurs face au phénomène d'antisélection précédemment évoqué. En effet, Arrow (1973)[2] et Spence (1974)[24] stipulent qu'à travers une variable financière ou « signal », les bons demandeurs peuvent annoncer publiquement les informations qui permettent aux prêteurs de distinguer leurs projets de ceux de moins bonne qualité. Dans cette logique, les banques ont élaboré le crédit scoring, un outil statistique qui permet de mesurer la fiabilité et la solvabilité d'un emprunteur à partir d'un certain nombre de variables que nous allons aborder dans la partie suivante.

3 Les travaux empiriques

3.1 Les variables principales

Les études empiriques s'accordent sur un ensemble de variables à prendre en compte dans l'approbation ou non d'un prêt. Les travaux de Blanchard *et al.* (2007)[8] retiennent, par exemple, les caractéristiques individuelles (l'âge, l'éducation, la profession, etc.) et financières (la rentabilité, la solvabilité, le revenu, les liquidités et les antécédents de crédit du demandeur).

Tel le noyau dur d'un atome, cette liste non-exhaustive de variables explicatives est en effet similaire à un grand nombre d'études comme celle de Turner *et al.* (2003)[26] ou encore Pager et Shepherd (2008)[21]. Tout comme les employeurs peuvent associer le diplôme des candidats à la productivité d'un travailleur, les prêteurs peuvent associer ces attributs à la solvabilité. De plus, les recherches récentes apportées par Wamalwa (2016)[27] renforcent également cette idée puisque l'éducation et les richesses nettes des emprunteurs sont, dans une large mesure, associées au remboursement du prêt. Ainsi, ces variables ont pour but premier de déterminer la capacité du demandeur à soutenir l'emprunt ainsi que d'établir l'historique de crédit de celui-ci afin de déceler les risques et les défaillances potentielles de ce dernier. De plus, Haubrich (1989)[15] montre que les institutions financières ont toujours reconnu l'importance des relations de long terme, mais que la reconnaissance explicite du rapport prêteur-emprunteur parmi les économistes fut pendant trop longtemps négligée. Les travaux de Siles *et al.* (1994)[23] démontrent en effet, que les relations sociales entre une banque et ses clients affectent considérablement la probabilité d'approbation du prêt. De plus, dès lors que les établissements de crédits ont pour objectif de maximiser leur profit, les relations de clientèle conduisent ces derniers à favoriser leurs anciens clients (Bédué et Lévy, 1997[4]). Ainsi, conformément aux travaux de Perterson et Rojan (1994)[22], ce lien « quasi-précieux » permet d'améliorer l'accès au crédit mais aussi de réduire le déficit d'information entre offreurs et demandeurs (Berger et Udell, 1995[6]). Bien entendu, l'aspect dynamique de la relation de long terme nécessite, au plan méthodologique, l'utilisation de modèles multipériodiques.

3.2 Une discrimination avérée

Bien que le lien tissé entre une banque et ses clients soit important, une hypothèse avancée par Calomiris *et al.* (1994)[11] suggère que la discrimination à l'égard des minorités sur le marché du crédit pourrait résulter de l'absence d'affinités culturelles entre les agents de prêts blancs et les candidats minoritaires.

En introduisant dans sa réflexion une variable de préférence pour la discrimination, Becker (1957)[3] développe une nouvelle facette de l'économie et devient le premier à mesurer l'ampleur et l'origine de ces comportements illégaux. Bien que ses travaux soient axés sur les inégalités salariales, Crawford et Rosenblatt (1999)[13] confirment néanmoins l'existence d'une discrimination substantielle dans l'octroi des prêts contre les entreprises appartenant à la communauté afro-américaine. Ces affirmations concordent avec les résultats antérieurs de Glennon et Stengel (1994)[14] dont l'analyse appuie la conclusion suivante : la moitié des écarts des taux de refus des prêts peut être attribuée à des différences de traitement par communauté. L'article de Blanchflower *et al.* (1998)[9] vient également étayer ce même postulat en appuyant le fait que les afro-américains sont deux fois plus susceptibles de se voir refuser un crédit. Ainsi, les conclusions au sujet de la discrimination ethnique sont remarquablement robustes puisque des preuves similaires ont été confirmées tout au long des dernières décennies à l'encontre des minorités. La seconde forme de discrimination débattue dans la littérature socio-économique est géographique. Selon Hutchinson *et al.* (1977)[18] ainsi que Bradbury *et al.* (1989)[10] les banques américaines refusent d'accorder des crédits dans des zones caractérisées par une forte concentration de minorités ethniques, quelle que soit la solvabilité des demandeurs de prêt. Toutefois, comme l'a noté Benston (1981)[5], la plupart de ces études sont inadéquates car elles ne parviennent pas à contrôler suffisamment les caractéristiques de l'emprunteur. En outre, comme l'ont souligné Holmes et Horvitz (1994)[16], ces études ne contrôlent pas suffisamment les différences de risque entre différentes zones géographiques. Des études plus récentes sur la réduction des effectifs ont donné des résultats quelque peu mitigés. Holmes et Horvitz (1994)[16], dans leur étude de la redlining à Houston, au Texas, ne parviennent pas à trouver des preuves claires de la pratique, alors que Canner *et al.* (1991)[12] examinant des données nationales révèlent davantage de preuves de cette pratique. Pour conclure, la prédominance de la couleur de peau dans la probabilité d'obtenir un prêt, indépendamment des caractéristiques de solvabilité semble témoigner d'un certain échec des lois évoquées précédemment. Celles-ci n'auront pas réussi à contrecarrer la segmentation sociale et ethnique de la bancarisation américaine.

4 La base de données

4.1 La présentation de l'échantillon

La base de données utilisée pour notre étude est initialement issue des données HMDA de 1990 recueillies par la banque fédérale de Boston sur les emprunteurs et les demandes de prêt figurant dans la région statistique métropolitaine (MSA)¹ de cette institution.

Composé à l'origine de 4022 observations, cet échantillon fut rendu célèbre pour avoir été utilisé par Munnell *et al.* (1992)[20] dans l'élaboration d'un modèle des déterminants des décisions de prêts hypothécaires dans l'État du Massachusetts. Toutefois, cette base de données fut remodelée par la suite par Hunter et Walker (1996)[17] en raison des observations jugées atypiques et aberrantes. C'est donc des travaux de ces derniers qu'est directement tiré la base de données sur laquelle nous avons travaillé. Composé de 1989 demandes de prêts conventionnels, celle-ci comprend en tout cinquante variables explicatives mesurant la capacité de l'emprunteur à soutenir le prêt (ex : salaires mensuels), les antécédents en matière de crédit (ex : déclarations de faillite), les pertes potentielles par défaut (ex : ratios prêt-valeur) et les caractéristiques personnelles du demandeur (ex : statut marital). S'ajoutent à cela quelques variables qualitatives liées au recensement comme le taux d'inoccupation des parcelles.

4.2 Le biais sélection

Toute méthode de recueil d'information comporte des biais pouvant être définis comme une déformation ou encore un travers. Dans notre base de données, la répartition des différentes ethnies n'est pas alignée à la réalité de la population de Boston dans les années 90.

En effet les blancs représentent 84,5% des individus contre 9,9% pour les afro-américains et 5,6% pour les hispaniques. Ce constat apporte une problématique ainsi qu'une réalité qui sont dans l'ordre le risque d'hétérogénéité et le fait que les minorités, partant du principe que leurs demandes seront refusées, ne font aucune démarche quantifiable par la base de données. Ce biais de sélection risque de se traduire par la sur-représentation de certaines variables mesurées dans la base de données.

1. Metropolitan Statistical Area : définit la région métropolitaine de Boston-Cambridge-Newton

5 L'analyse descriptive

5.1 La répartition des demandes de crédits

Notre premier réflexe, lors de l'analyse descriptive de la base de données, fut d'identifier la répartition d'octroi de crédit selon l'ethnie du demandeur dont les résultats sont présentés dans le tableau de contingence suivant :

Tableau 5.1 : Répartition des demandes de crédits

	Total	Blanc	Minorité
Approuvé	87,73%	90,84%	70,78%
Rejeté	12,17%	9,16%	29,22%

Bien que les afro-américains et les hispaniques soient sous-représentés dans notre échantillon, nous constatons des différences significatives sur la répartition des demandes de crédits entre ces derniers et les blancs. En effet 90,84% des prêts hypothécaires contractés par la population blanche ont été acceptés contre 70,78% pour les minorités. Récemment, ce constat a quelque peu évolué d'après les données HMDA collectées en 2018 puisque 85,50% des crédits contractés par des minorités ont été approuvés contre 93,20% pour les blancs. Le fait d'appartenir à un groupe ethnique semble donc jouer un rôle dans le processus d'octroi de prêts. Toutefois, la conclusion selon laquelle les prêteurs hypothécaires discriminent activement les minorités n'est pas appropriée à cette étape de l'étude.

5.2 Les caractéristiques intrinsèques

L'écart constaté précédemment peut s'expliquer, en partie, par les caractéristiques intrinsèques de nos deux sous-populations. En effet les blancs et les minorités présentent des apports, des ratios et des historiques de crédit bien différents comme en témoigne le tableau ci-dessus et les tests de Student et du Chi2 effectués.

Tableau 5.2 : Caractéristiques des deux populations

Variables	Total	Blanc (1)	Minorité (2)	Différence (1-2)
Apports				
Revenu mensuel total	5195,55	5326,91	4478,62	848,29 *
Richesse nette	266,57	291,19	132,19	159,00 *
Actifs liquides	4618,03	5453,65	57,36	5396,29 *
Ratios				
Ratio prêt/prix	77,06%	75,72%	84,40%	-8,68% *
Ratio dépenses/revenus	24,79%	24,69%	25,37%	-0,68%
Ratio obligations/revenus	32,39%	32,02%	34,41%	-2,39% *
Historiques du crédit				
Absence d'histo. hypothécaire	33,18%	35,16%	22,40%	12,76% *
Absence de retard de paiement	63,85%	62,05%	73,70%	-11,65% *
Retard de paiement	1,91%	1,78%	2,60%	-0,82% *
Caractéristiques individuelles				
Marié	65,86%	65,99%	65,15%	0,84%
Grandes études	77,17%	79,06%	66,88%	12,18% *
Auto-entrepreneur	12,92%	13,33%	10,71%	2,62% *

En 1980, près de seize ans après la fin de la ségrégation raciale aux États-Unis, le fossé entre les blancs et les minorités est toujours présent. En effet, les apports moyens des hispaniques ou des afro-américains sont bien moins conséquents que les blancs, en particulier au niveau des actifs liquides (5396,29) et du revenu mensuel (848,29). Témoinnant des inégalités de l'époque, ces chiffres ont un rôle important dans l'estimation du montant du prêt hypothécaire, du taux d'intérêt et du délai de remboursement par les banques. Pour ce qui est des ratios, les minorités détiennent des quotients moins favorables puisqu'ils présentent des coefficients budgétaires plus élevés. Par conséquent, les dépenses allouées à l'alimentation, le logement et les obligations représentent une part importante de leur revenu disponible, c'est pourquoi ils ont une propension à épargner plus faible vis-à-vis des blancs. Ces inégalités se reflètent également dans le système éducatif américain des années 80 puisque les afro-américains et les hispaniques sont significativement moins nombreux à poursuivre leurs études après le lycée (high school). Les minorités occupent ainsi des métiers moins qualifiés, moins rémunérés et plus sensibles au chômage. Malgré cela, ces derniers présentent de meilleurs historiques de crédit (absence de retard de paiement et d'historique hypothécaire) que les blancs. Pour ce qui est du statut d'auto-entrepreneur, nous comptons plus de travailleurs indépendants parmi les demandeurs blancs (13,33%) que parmi les minoritaires (10,71%).

5.3 Les disparités dans les exigences liées au prêt

Ce tableau illustre, en moyenne, les exigences requises de voir son prêt hypothécaire approuvé ou non selon l'ethnie, dans notre cas les blancs et les minorités, à Boston dans les années 90.

Tableau 5.3 : Différentes exigences selon l'ethnie

Variables	Blanc		Minorité	
	Approuvé	Rejeté	Approuvé	Rejeté
Capacité à soutenir un prêt				
Ratio dépenses/revenus	24,59%	25,67%	24,77%	26,80%
Ratio obligations/revenus	31,66%	35,54%	33,14%	37,4%
Richesse nette	282,26	379,71	145,17	100,74
Risque d'insolvabilité				
Retard de paiement	0,9%	1,9%	0,5%	3,3%
Probabilité de chômage	3,85%	4,38%	3,78%	3,86%
Auto-entrepreneur	12,57%	20,78%	11,47%	8,89%
Ratio prêt/prix	74,87%	84,09%	83,46%	86,69%
Caractéristiques du prêt				
Taux ajustable	30,26%	32,47%	34,86%	27,78%
Prêt supérieur à 30 ans	85,50%	84,97%	91,28%	95,45%

Nous constatons que les ratios exigés en moyenne pour nos deux sous-populations ne diffèrent que très légèrement. En effet, il semble que pour qu'un prêt hypothécaire soit approuvé, les dépenses du demandeur ne doivent pas dépasser 1/4 de son revenu et 1/3 de ses obligations. Néanmoins, les richesses nettes exigées en moyenne diffèrent selon l'ethnie du demandeur puisque les minorités détiennent moins d'apports financiers que les blancs. Ce constat peut apporter un élément de réponse au fait que le ratio prêt/prix est en moyenne plus proche de 100% pour les hispaniques et les afro-américains que pour les blancs. En effet, le manque de ressources et d'apports financiers contraint les minorités à demander des prêts couvrant plus de 8/10 du prix du bien souhaité, soit environ 10% de plus que les blancs. De plus, nos deux sous-populations ne jouissent en moyenne pas des mêmes modalités dès lors que le prêt est accordé. En effet, par rapport aux blancs, les minorités disposent de prêt hypothécaire généralement plus long et à taux flexible. Un délai de remboursement allongé conjoint à des taux plus sujet à la conjoncture font exploser le coût du prêt pour les minorités.

5.4 Les matrices de corrélations

Dans l'optique d'évaluer la dépendance entre nos variables, nous avons représenté graphiquement deux corrélogrammes distincts qui se différencient par leurs méthodes de constructions et de lectures. En effet, la première mesure l'intensité des liens qualitatifs via le test de Cramer, tandis que la seconde évalue la dépendance entre nos variables quantitatives grâce au test de Pearson.

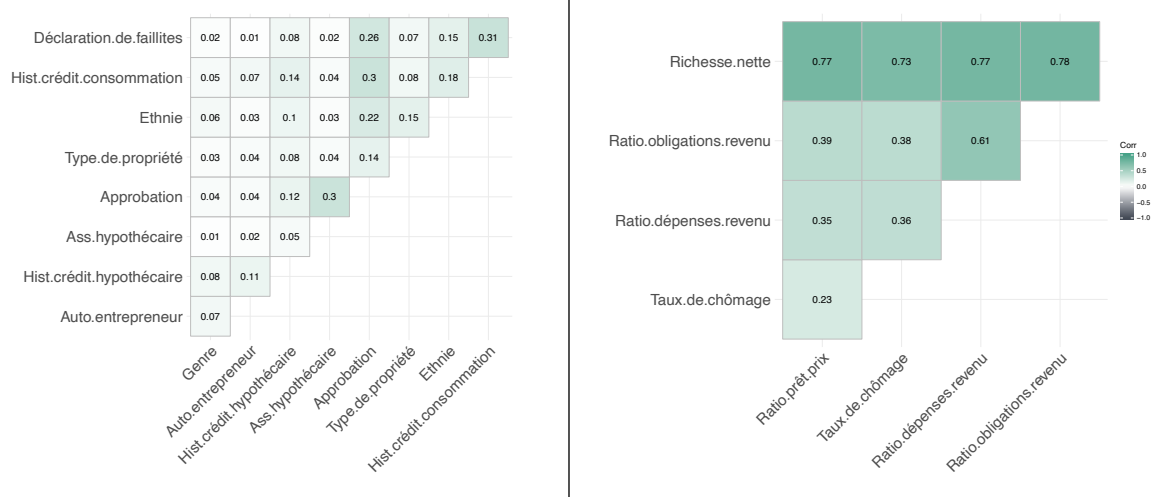


Tableau 5.4 : Matrices de corrélations

La première matrice de corrélation indique de forte dépendance entre l'approbation d'un prêt et les variables suivantes : l'historique de crédit à la consommation (0,30), l'obtention d'une assurance hypothécaire (0,30), les déclarations de faillites (0,26) et l'ethnie du demandeur (0,22). Ce constat coïncide avec la littérature socio-économique évoquée précédemment puisque sont retenus principalement les antécédents de crédit du demandeur mais également, dans une moindre mesure, la couleur de peau de l'individu. Au-delà de cette conclusion, nous ne constatons aucunes corrélations significatives entre nos variables qualitatives si ce n'est entre la déclaration de faillites et l'historique de crédit à la consommation (0,31). Le second corrélogramme présente des coefficients élevés entre nos variables quantitatives impliquant un risque de multicollinéarité dans la construction de notre modèle de régression. Cela s'avère problématique, car il peut augmenter la variance des coefficients de régression et les rendre instables et difficiles à interpréter. Dans notre cas, les richesses nettes d'un individu (variable récurrente dans les études empiriques) sont fortement liées ($>0,70$) à l'ensemble des ratios mesurés dans la base de données.

5.5 Les variables importantes

En raison du grand nombre de variables que contient notre base de données, il fut nécessaire de choisir certaines d'entre-elles pour la construction de notre modèle de régression. Pour ce faire, nous nous sommes basés sur la littérature et sur la forêt aléatoire.

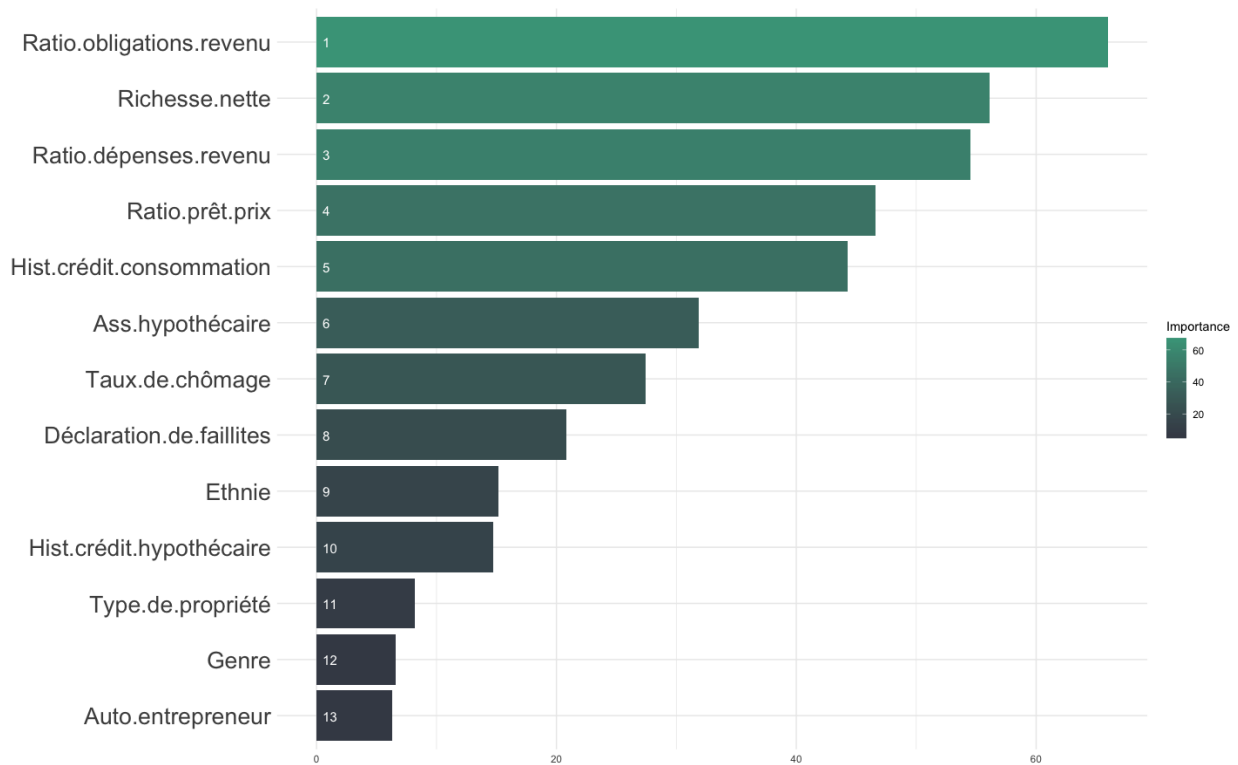


Tableau 5.5 : Variables importantes

Cet algorithme puissant d'apprentissage automatique permet de faire des prédictions en se basant sur l'agrégation de plusieurs arbres de décision. Le graphique ci-dessus présente le nombre d'occurrence des variables explicatives pour déterminer si le prêt hypothécaire fut accepté ou rejeté. Nous constatons que les premières places du podium sont détenues par des variables financières (ratio dette/revenu, richesse nette, ...) mais aussi celles liées à l'historique de crédit du demandeur (historique de crédit à la consommation et hypothécaire). L'ethnie du demandeur est en neuvième position témoignant ainsi d'une plausible discrimination dans le processus d'octroi de prêt hypothécaire à Boston en 1980. Nous constatons également que le genre du candidat joue son importance ainsi que son statut de travail. Toutefois seul notre modèle de régression, à travers la significativité des coefficients associés aux variables, nous apportera des conclusions robustes. De plus nous conserverons pour la suite de cette étude 11 variables explicatives, jugées essentielle selon la littérature et les statistiques descriptives réalisées précédemment, pour identifier les déterminants de l'approbation d'un prêt hypothécaire.

6 L'approche théorique des différents modèles

6.1 La méthodologie de la régression logistique

Nous proposons d'analyser en trois étapes les inégalités existantes dans le processus d'approbation des prêts hypothécaires entre la population blanche et minoritaire. La première partie est menée à l'aide de régression logistique dans le but de mettre en évidence cette différence identifiée à plusieurs reprises par les institutions américaines.

Notre analyse économétrique des effets d'affinités culturelles sur les marchés du crédit hypothécaire est effectuée dans le cadre des modèles de choix dichotomiques. La décision du prêteur quant aux candidats aux prêts à approuver peut être considérée comme n'acceptant que les candidats dont la solvabilité dépasse un niveau critique. L'analyse repose sur l'estimation des déterminants de l'approbation d'un prêt hypothécaire à l'aide d'un modèle logit :

$$S_i = \alpha + X_i\beta + \epsilon_i$$

Nous supposons que la solvabilité, S_i du i -ème demandeur telle que perçue par le prêteur peut être exprimée comme une combinaison linéaire des caractéristiques de l'emprunteur. Où β est un vecteur de paramètres et X_i est le vecteur de variables informationnelles pour le i -ème demandeur. Le terme d'erreur ϵ_i représente des facteurs connus du prêteur mais non observables par le banquier. Nous supposons que ce bruit aléatoire a une moyenne nulle et une variance σ^2 . Bien sûr, cet indice continu de solvabilité, S_i , n'est pas observé (variable latente). Nous distinguons seulement si une demande de prêt hypothécaire est acceptée ou refusée. Soit :

$$\begin{aligned} Y_i &= 1 & \text{si } S_i > 0 \\ Y_i &= 0 & \text{si } S_i \leq 0 \end{aligned}$$

La probabilité que le i -ème demandeur se voit accorder un prêt par la banque peut donc s'écrire : $\mathbb{P}(Y_i = 1) = F(X_i\beta)$ où $F(\cdot)$ est la fonction de distribution cumulative de ϵ , en supposant que ce dernier a une distribution symétrique. Étant donné que les paramètres β ne sont identifiés que jusqu'à un scalaire positif, nos estimations sont en fait des estimations de β/σ .

6.2 Les principes de la décomposition non linéaire

Nous proposons, ensuite, de décomposer les inégalités d'accès aux crédits entre les deux populations en utilisant la méthodologie proposée par Fairlie (1999). Il s'agira, en outre, d'étudier la contribution relative des déterminants socio-économiques susceptibles d'expliquer la différence de prêts approuvés entre nos deux populations.

La méthode de décomposition proposée par Fairlie est similaire à celle initialement proposée par Blinder (1973) et Oaxaca (1973), permettant d'étudier les discriminations existantes sur le marché du travail entre deux groupes de population. Fairlie (1999) propose une analyse équivalente qui s'applique à des modèles non linéaires où la variable d'intérêt est binaire, comme notre variable approuvé. Il s'agit de décomposer en deux parties la différence de probabilités moyennes de voir sa demande de prêt hypothécaire acceptée entre la population blanche et minoritaire. La première partie de la décomposition fait référence à une différence attribuable à la distribution des caractéristiques observables alors que la seconde partie représente une différence attribuable aux effets de ces caractéristiques. Selon Fairlie, la décomposition dans le cas des modèles non-linéaires peut s'exprimer de la manière suivante :

$$\underbrace{\bar{Y}^B - \bar{Y}^M}_{\hat{\Delta}_b \text{ (brut)}} = \underbrace{\left[\sum_{i=1}^{N^B} \frac{F(X_i^B \hat{\beta}^B)}{N^B} - \sum_{i=1}^{N^M} \frac{F(X_i^M \hat{\beta}^B)}{N^M} \right]}_{\hat{\Delta}_e \text{ (expliqué)}} + \underbrace{\left[\sum_{i=1}^{N^M} \frac{F(X_i^M \hat{\beta}^B)}{N^M} - \sum_{i=1}^{N^M} \frac{F(X_i^M \hat{\beta}^M)}{N^M} \right]}_{\hat{\Delta}_i \text{ (inexpliqué)}}$$

Comme pour la décomposition d'Oaxaca-Blinder, nous obtenons ainsi un contrefactuel répondant à la question suivante : quel serait le taux d'approbation des individus minoritaires si leurs caractéristiques étaient valorisées de la même manière que pour les blancs ? Concernant la notation, \bar{Y}^B est la probabilité moyenne que la demande de crédit soit acceptée dans la population j (avec $j = B, M$ pour la population blanche et minoritaires respectivement), X^j correspond à la distribution des caractéristiques observables au sein de la population considérée, $\hat{\beta}^B$ représente les coefficients estimés affectés aux caractéristiques observées, N^j fait référence à la taille de l'échantillon de chaque sous-population et $F(\cdot)$ représente la fonction de distribution cumulative qui suit une loi logistique. Le choix de la population blanche comme population de référence suggère ainsi des discriminations en défaveur de la population minoritaire. Toutefois, la méthodologie proposée par Fairlie ne permet pas de mesurer la contribution relative des coefficients estimés, propre à chaque caractéristique observable. C'est pourquoi, nous proposons en dernière partie de comparer les odd-ratios associés aux caractéristiques observables entre la population blanche et minoritaire à l'aide de deux régressions logistiques menées séparément au sein des deux populations.

7 Les résultats empiriques

7.1 Les régressions logistiques

Le tableau ci-dessous présente les résultats des analyses logit ayant pour objet d'étudier l'influence des caractéristiques individuelles et financières du demandeur sur la probabilité d'obtenir un prêt. Le premier modèle contrôle l'ensemble des variables retenues sur la base de la littérature économique et des statistiques descriptives tandis que le second représente la régression retenue par l'algorithme de sélection de modèles (critères AIC).

Tableau 7.1 : Les régressions par parcimonie

	Modèle 1			Modèle 2		
	Coefficients	Écart-types	Odds-ratio	Coefficients	Écart-types	Odds-ratio
Constante	2,400 ***	0,664		2,918 ***	0,480	
Caractéristiques individuelles						
Blanc	1,187 ***	0,173	3,043	1,180 ***	0,169	3,253
Marié	0,500 ***	0,179	1,577	0,418 ***	0,156	1,519
Homme	-0,165	0,204	0,848			
Personnes à charges	-0,054	0,076	0,947			
Diplômé	0,071	0,179	1,074			
Caractéristiques financières						
Ratio dépenses/revenus	-2,009 *	1,076	0,134	-2,051 **	1,063	0,129
Ratio prêt/prix	-1,537 ***	0,440	0,215	-1,561 ***	0,458	0,210
Durée du prêt	0,001	0,001	1,001			
Assurance hypothécaire	-4,724 ***	0,750	0,009	-4,721 ***	0,748	0,009
Déclaration de faillite	-1,813 ***	0,207	0,163	-1,809 ***	0,205	0,164
Retard de paiement	0,154	0,176	1,166			

Les seules caractéristiques personnelles qui semblent avoir une influence significative dans la décision d'approbation du prêt hypothécaire sont le statut marital et l'ethnie du demandeur. S'il est évident qu'être marié est fortement lié à un revenu total plus important, la couleur de peau ne l'est pas forcément. En effet, le coefficient positif et statistiquement significatif suggère qu'après avoir pris en compte le statut marital, l'assurance hypothécaire, la déclaration de faillite ainsi que les ratios dépenses/revenus et prêt/prix, l'ethnie du demandeur joue toujours un rôle dans la décision du prêteur d'approuver ou non le prêt. Selon notre régression logistique, une personne de couleur a donc trois fois moins de chance de voir sa demande de prêt approuvée vis-à-vis d'un blanc, toutes choses

égales par ailleurs. Ce résultat apporte donc un élément de réponse à notre problématique – les déterminants de l’approbation d’un prêt hypothécaire – mais témoigne également d’un certain échec de la loi sur la lutte contre la discrimination ethnique. Néanmoins, la piste d’une discrimination de genre fut écartée puisqu’après contrôle de la variable homme, il semble que celle-ci n’est aucune influence significative sur la probabilité d’approbation d’un prêt. En effet, par souci de cannibalisation, nous avons testé la variable de genre sans le statut marital afin d’identifier le moindre effet d’être une femme dans le processus d’octroi d’un prêt. Cette seconde forme de discrimination n’a donc certes pas été décelée dans cette analyse, possiblement dû à la sur-représentation des hommes (81,45%), mais est une réalité selon Blanchard *et al.* (2007)[8]. De plus, bien que leurs signes associés soient cohérents, il semble que le nombre de personnes à charges ainsi que le niveau d’éducation n’ont aucun effet significatif dans l’approbation d’un prêt. Le manque de granularité dans cette dernière variable et la très faible variance du nombre de personnes à charges ont sûrement eu leur importance.

Concernant les variables financières significatives, il semble surprenant que le fait de détenir une assurance hypothécaire ait une impacte négative sur la décision finale du banquier. Effectivement, bien qu’elle protège le prêteur des éventuels défauts de remboursement de son prêt immobilier, cette garantie est très onéreuse et représente donc des charges supplémentaires pour le demandeur. Par ailleurs et sans surprise, la déclaration de faillite et le ratio dépenses/revenus ont également des impacts négatives sur la probabilité d’approbation d’un prêt hypothécaire. En effet, le premier réflexe du banquier est d’identifier les dépenses du demandeur par rapport à son revenu puisqu’il permet de donner une idée précise du niveau d’endettement du ménage. D’autre part, la déclaration de faillite, synonyme d’incident de remboursement, a également une répercussion négative sur la probabilité d’approbation d’un prêt hypothécaire puisqu’elle témoigne d’une situation de surendettement qu’a traversée l’emprunteur. Pour terminer, le ratio prêt/prix (LTV)¹ a lui aussi un effet négatif sur la décision final du banquier puisqu’il mesure le niveau d’endettement d’un emprunteur au regard de la valeur vénale de ses actifs immobiliers. Par conséquent, plus ce ratio est élevé, plus le risque de l’emprunt est important.

Pour conclure, nous nous sommes concentrés sur six caractéristiques objectives de l’emprunteur qui sont généralement considérées comme des facteurs importants pour mesurer la qualité de ce dernier. Il s’agit de l’ethnie, du statut marital, de la présence ou non d’une assurance hypothécaire et d’une déclaration de faillite ainsi que des ratios prêt/prix et dépenses/revenus. Pour tenir compte de la possibilité que ces facteurs soient perçus différemment par les prêteurs en fonction de l’ethnie du demandeur, chacune de ces six variables a été mise en interaction avec l’ethnie du demandeur par la suite.

1. Loan To Value : indicateur utilisé lors de l’octroi du prêt immobilier

7.2 La décomposition de Fairlie

Étant donné que les modèles confirment l'hypothèse d'un effet de sélection sur la base de l'ethnie, nous allons compléter cette analyse via la décomposition proposée par Fairlie. Il s'agit d'examiner la part de la différence d'approbation d'un prêt entre les blancs et les minorités attribuable à une différence de distribution des caractéristiques observables.

Tableau 7.2 : Les écarts expliqué et inexpliqué

	Estimations	Part en %
Écart brut ($\hat{\Delta}_b$)	0,20	100%
Écart expliqué ($\hat{\Delta}_e$)	0,05	25,83%
Écart inexpliqué ($\hat{\Delta}_i$)	0,15	74,17%

Sachant que 90,84% des demandes de prêts contractées par des blancs ont été approuvées contre 70,78% pour les minorités, la différence de probabilité moyenne de voir sa requête acceptée entre les deux populations s'élève à 20,06 points de pourcentage. Les résultats montrent que 25,83% de l'écart brut sont expliqués par la différence de distribution des caractéristiques observables entre les deux populations. Ainsi, cette différence d'approbation de prêt diminuerait d'environ 26% si les minorités avaient les mêmes caractéristiques que les blancs. La décomposition des inégalités dans l'approbation des prêts montre que les caractéristiques observables permettent d'expliquer plus de 1/4 de la différence totale, ce qui signifie qu'environ 3/4 de $\hat{\Delta}_b$ est attribuable à une différence dans les coefficients estimés. Cette part est considérée comme une approximation des discriminations car elle suggère que l'effet des caractéristiques observées sur l'approbation d'un prêt est différent selon la population considérée.

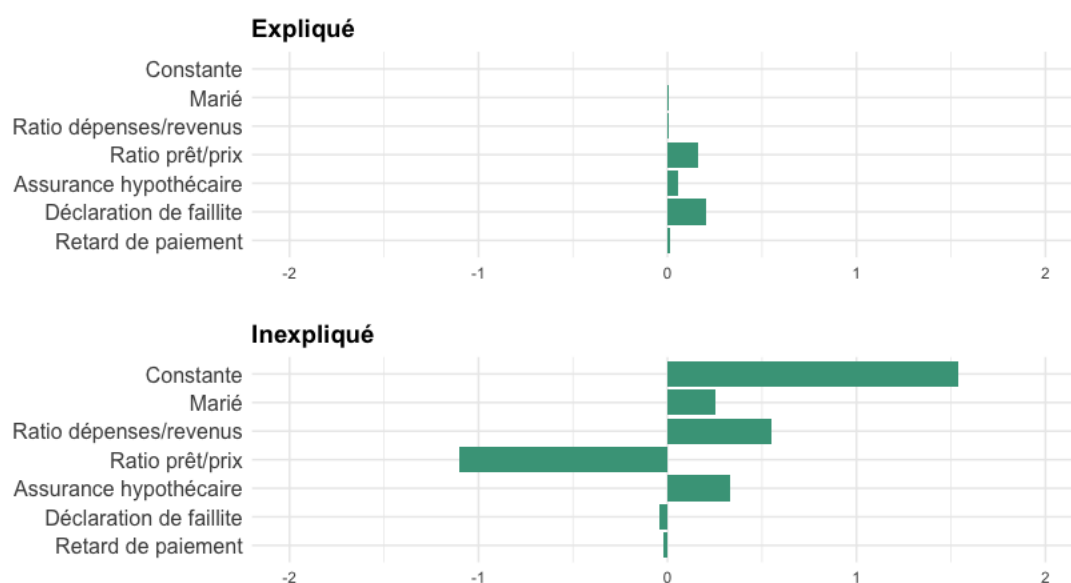


Tableau 7.3 : Contribution des variables aux écarts expliqué et inexpliqué

La partie supérieure du graphique ci-dessus présente la contribution de chaque variable à l'écart expliqué. Les variables qui contribuent le plus positivement à l'écart expliqué sont dans l'ordre la déclaration de faillite, le ratio prêt/prix ainsi que l'acquisition d'une assurance hypothécaire. Autrement dit, presque l'intégralité de l'écart brut mesuré entre blancs et minorités tient au fait que les afro-américains et les hispaniques déclarent davantage de faillite personnelle (15,91% vs 5,23%), détiennent des ratios prêt/prix plus élevés (0,84 vs 0,76) et contractent plus souvent une assurance hypothécaire (2,61% vs 1,31%). La partie inférieure du graphique ventile l'écart inexpliqué par variable : de même qu'il est possible de détailler les contributions de chaque variable à l'effet de composition, nous pouvons aussi obtenir le détail de l'écart inexpliqué. Cependant, des hypothèses supplémentaires et des précautions particulières sont nécessaires pour interpréter ces résultats détaillés. Nous pouvons toutefois remarquer que la part de la constante dans l'écart inexpliqué est très importante, ce qui signifie que les variables sélectionnées dans notre analyse économétrique ne suffisent pas à capter la totalité de la discrimination sur le marché bancaire américain.

7.3 Les régressions selon l'ethnie

Afin de compléter cette analyse, nous avons par la suite comparé dans une dernière étape les signes, les coefficients et les odd-ratios associés aux différents déterminants de l'approbation d'un prêt séparément dans la population blanche et minoritaire.

Tableau 7.4 : Les régressions selon l'ethnie

	Modèle : Blancs			Modèle : Minorités		
	Coefficients	Écart-types	Odds-ratio	Coefficients	Écart-types	Odds-ratio
Constante	4.074 ***	0.552		2.538 ***	0.887	
Marié	0.511 ***	0.190	1.667	0.123	0.287	1.131
Ratio dépenses/revenus	-1.327	1.293	0.265	-3.509 *	2.015	0.030
Ratio prêt/prix	-1.849 ***	0.520	0.157	-0.547	0.825	0.579
Assurance hypothécaire	-4.597 ***	0.759	0.010	-17.469	796.964	0.000
Déclaration de faillite	-1.909 ***	0.256	0.148	-1.629 ***	0.335	0.196
Retard de paiement	0.095	0.206	1.099	0.168	0.327	1.182

Nous constatons dans un premier temps que les signes associés aux différentes variables ne varient pas d'une population à une autre comparés à leurs significativités et leurs odd-ratios. En effet, les quatre variables financières retenues jusqu'ici ont un effet négatif sur la probabilité d'obtenir un prêt mais ne sont pas significatives selon la population étudiée. Par exemple, le ratio dépenses/revenus des blancs n'a pas d'influence sur la décision finale du banquier tandis qu'il l'a pour les afro-américains et les hispaniques. Cela signifie donc qu'à l'instar des blancs, les minorités sont davantage réprimandées sur ce « reste à vivre », prouvant encore une fois une différence de traitement lors du processus d'octroi d'un prêt. L'effet du statut marital est également différent dans les deux sous-populations puisque seuls les blancs jouissent de l'influence positive du fait d'être marié sur la probabilité d'approbation d'un prêt.

8 La conclusion finale

Cette étude a examiné « une voie » par laquelle un traitement différentiel pourrait affecter l'accès des minorités au crédit et aux possibilités d'accession à la propriété. Elle a révélé que les demandeurs de prêts hypothécaires afro-américains et hispaniques de la région de Boston dans les années 90 étaient plus susceptibles d'être refusés que les demandeurs blancs présentant des caractéristiques similaires.

Bien que les demandeurs issus de minorités ont, en moyenne, un endettement plus important, un ratio prêt/prix plus élevé et déclarent plus souvent des faillites que les demandeurs blancs, ces facteurs n'éliminent pas totalement la disparité entre les taux de refus. En effet, ces désavantages qui n'expliquent qu'environ 1/4 de ces différences selon la décomposition proposée par Fairlie, soulève la question de la discrimination sur le marché bancaire. Ainsi, en fin de compte, il subsiste un écart statistiquement significatif, qui est associé à la couleur de peau du demandeur puisque, comme en témoigne les deux régressions logistiques basées sur chaque ethnie, les coefficients sont significativement différents au profit de la population blanche. Il est important de clarifier l'objectif limité de cette analyse ; elle fait abstraction de la discrimination qui peut se produire ailleurs dans l'économie. Par exemple, si les minorités sont victimes de discrimination dans l'éducation ou sur le marché du travail, elles auront des revenus plus faibles et leurs demandes peuvent refléter des ratios dépenses/revenus et un rapport prêt/prix plus élevés ou encore des antécédents de crédit plus médiocres. Même dans le cadre de l'attention particulière accordée par les prêteurs conventionnels, la mesure des obstacles rencontrés par les minorités doit être mise en perspective ; un traitement différent peut se produire à de nombreuses étapes du processus de prêt. Par exemple, les minorités peuvent être découragées de demander un prêt hypothécaire à la suite d'un processus de présélection. De même, si les demandeurs blancs sont plus susceptibles que les demandeurs issus de minorités d'être « coachés » au moment de remplir la demande, ils auront des demandes plus fortes que les minorités se trouvant dans une situation similaire. Dans ce cas, les ratios et autres informations financières figurant dans la demande finale, qui ont fait l'objet de cette analyse, peuvent eux-mêmes être le produit d'un traitement différentiel. Cette étude n'explore donc pas la mesure dans laquelle le processus d'octroi a lieu, mais se concentre plutôt sur l'impact de l'ethnie sur les décisions des prêteurs concernant les demandes finales reçues des emprunteurs potentiels.

Bibliographie

- [1] G. Akerlof. The market for lemons : quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 1970.
- [2] K. Arrow. What has economics to say about racial discrimination ? *The Journal of Economic Perspectives*, 1973.
- [3] G. Becker. The economics of discrimination : economics research studies. *University of Chicago Press*, 1957.
- [4] A. Bedue and N. Levy. Bank-business relationship and cost of credit. *Financial Economics Review*, 1997.
- [5] G. Benston. Mortgage redlining research : a review and critical analysis. *Journal of Bank Research*, 1981.
- [6] A. Berger and G. Udell. Relationship lending and lines of credit in small firm finance. *The Journal of Business*, 1995.
- [7] H. Bester. The role of collateral in credit markets with imperfect information. *European Economic Review*, 1985.
- [8] L. Blanchard, B. Zhao, and J. Yinger. Do lenders discriminate against minority and woman entrepreneurs ? *Urban Economics*, 2007.
- [9] D. Blanchflower, P. Levine, and D. Zimmerman. Discrimination in the small-business credit market. *The Review of Economics and Statistics*, 1998.
- [10] K. Bradbury, K. Case, and D. Constance. Geographic patterns of mortgage lending in boston. *New England Economic Review*, 1989.
- [11] C. Calomiris, C. Kahn, and S. Longhofer. Housing-finance intervention and private incentives : helping minorities and the poor. *Journal of Money and Banking*, 1994.
- [12] G. Canner, S. Gabriel, and M. Woolley. Understanding mortgage lending patterns. *Southern Economic Journal*, 1991.
- [13] G. Crawford and E. Rosenblatt. Differences in the cost of mortgage credit implications for discrimination. *The Journal Of Real Estate Finance and Economics*, 1999.
- [14] D. Glennon and M. Stengel. An evaluation of the federal reserve bank of boston's study of racial discrimination in mortgage lending. *Economic Policy Analysis*, 1994.
- [15] J. Haubrich. Financial intermediation : Delegated monitoring and long-term relationships. *Journal of Banking Finance*, 1989.

- [16] A. Holmes and P. Horvitz. Mortgage redlining : race, risk, and demand. *Journal of Finance*, 1994.
- [17] W. Hunter and M. Walker. The cultural affinity hypothesis and mortgage lending decisions. *The Journal of Real Estate Finance and Economics*, 1996.
- [18] P. Hutchinson, J. Ostas, and D. Reed. A survey and comparison of redlining influences in urban mortgage lending markets. *Real Estate Economics*, 1977.
- [19] H. Leland and D. Pyle. Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance*, 1977.
- [20] A. Munnell, G. Tootell, L. Browne, and J. McEneaney. Mortgage lending in boston interpreting hmda data. *The American Economic Review*, 1992.
- [21] D. Pager and H. Shepherd. Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, 2008.
- [22] M. Petersen and R. Rajan. The benefits of lending relationships : evidence from small business data. *The Journal Of Finance*, 1994.
- [23] M. Siles, S. Hanson, and L. Robison. Socio-economics and the probability of loan approval. *Review of Agricultural Economics*, 1994.
- [24] M. Spence. Market signaling : Informational transfer in hiring and related screening processes. *Harvard University Press*, 1974.
- [25] J. Stiglitz and A. Weiss. Credit rationing in markets with imperfect information. *The American Economic Review*, 1981.
- [26] M. Turner, E. Godfrey, S. Ross, and R. Smith. A paired testing study of discrimination in mortgage lending. *Economics Working Papers*, 2003.
- [27] W. Wamalwa. Determinants of loan repayment by borrowers from micro-financial institutions in nakuru county kenya. *The Journal Of Investment Management*, 2016.

A Annexes : Le sommaire des variables

Tableau A.1 : Statistiques des variables quantitatives retenues dans le modèle

Variables quantitatives				
Variables	Moyenne	Ecart-type	Min.	Max.
Ratio dépenses/revenus	0,2479	0.071	0,010	0,720
Ratio prêt/prix	0,7706	0.189	0,02105	2,5714
Personnes à charges	0.7712	1.104	0,0	8,0
Durée du prêt	341,2	64,962	6,0	480,0

Tableau A.2 : Répartition des modalités des variables qualitatives retenues dans le modèle

Variables qualitatives	
Variables	Répartitions
Approuvé	Oui : 87,7%
	Non : 12,3%
Blanc	Oui : 84,5%
	Non : 15,5%
Homme	Oui : 80,9%
	Non : 19,1%
Diplômé	Oui : 77,2%
	Non : 22,8%
Marié	Oui : 65,8%
	Non : 34,2%
Assurance hypothécaire	Oui : 1,5%
	Non : 98,5%
Déclaration de faillite	Oui : 6,9%
	Non : 93,1%
Retard de paiement	Oui : 36,1%
	Non : 63,9%

B Annexes : La programmation

```
library(ggcorrplot)
library(ggplot2)
library(naniar)
library(missMDA)
library(skimr)
library(FactoMineR)
library(GGally)
library(dplyr)
library(tidyr)
library(labelled)
library(randomForest)
library(stargazer)
library(gridExtra)
library(cowplot)
library(GGally)
library(questionr)
library(ggthemes)
library(MASS)
```

:

```
data<-read.csv("loanapp.csv",sep=";",dec="," );
names(data);
\head(data)
str(data);
summary(data);
```

:

```
data$Blanc<-as.factor(data$Blanc);
data$Marié<-as.factor(data$Marié);
data$Retard.paiement<-as.factor(data$Retard.paiement);
data$Approuvé<-as.factor(data$Approuvé);
data$Homme<-as.factor(data$Homme);
data$Déclaration.de.faillite<-as.factor(data$Déclaration.de.faillite);
data$Assurance.hypothécaire<-as.factor(data$Assurance.hypothécaire);
data$Diplômé<-as.factor(data$Diplômé);
```

:

```

FiltreVar<-NULL
for(i in 1:ncol(data)){FiltreVar<-c(FiltreVar,is.numeric(data[,i]))}
names(data)[FiltreVar]
data_quantitative<-data[,FiltreVar]
data_qualitative<-data[,!FiltreVar]
cv<-function(x,y){
  table<-table(x,y)
  chisq<-suppressWarnings(chisq.test(table))$statistic
  cramer<-sqrt(chisq/(NROW(x)*(min(dim(table))-1)))
  cramer
}
cramer.matrix<-function(y,fill=TRUE){
  col.y<-ncol(y)
  V<-matrix(ncol=col.y,nrow=col.y)
  for(i in 1:(col.y-1)){
    for(j in (i+1):col.y){
      V[i,j]<-cv(y[,i],y[,j])
    }
  }
  diag(V)<-1
  if (fill){
    for (i in 1:ncol(V)){
      V[,i]<-V[i,]
    }
  }
  colnames(V)<-names(y)
  rownames(V)<-names(y)
  V
};

```

:

```

ggcorrplot(cramer.matrix(data_qualitative),hc.order = TRUE, type = "upper",
  colors = c("#424955", "white", "#46A288"),lab = TRUE, lab_size = 4.5,
  digits = 2,insig = c("pch", "blank"), pch = 1, pch.col = "black",pch.cex
=1,tl.cex = 20));
ggcorrplot(cramer.matrix(data_quantitative),hc.order = TRUE, type = "upper"
,colors = c("#424955", "white", "#46A288"),lab = TRUE, lab_size = 4.5,
  digits = 2,insig = c("pch", "blank"), pch = 1, pch.col = "black",pch.cex
=1,tl.cex = 20));

```

:

```

forest<-randomForest(as.factor('Approuvé')~.,data=data,method="class",parms
=list(split="gini"),ntree = 500);
varImportance <- data.frame(Variables = row.names(importance(forest)),
  Importance = round(importance(forest)[, 'MeanDecreaseGini'],2));
rankImportance <- varImportance %>% mutate(Rank = paste0(dense_rank(desc(
  Importance))));

```

:

```
ggplot(rankImportance, aes(x = reorder(Variables, Importance), y =
  Importance, fill = Importance)) +
  geom_bar(stat='identity') + labs(x = '', y = '') + coord_flip() + theme_
  minimal() + scale_fill_gradient(low="#424955", high="#46A288") +
  geom_text(aes(x = Variables, y = 0.5, label = Rank), hjust=0, vjust=0.55,
    size = 4, colour = "white") +
  theme(axis.text.y = element_text(size = 20)) + theme(legend.position='
    none');
```

:

```
modele1<-glm(as.factor(Approuvé) ~ as.numeric(Ratio.dépense.revenu) +
  as.numeric(Ratio.prêt.prix) +
  as.integer(Durée.du.prêt) +
  as.factor(Assurance.hypothécaire) +
  as.factor(Déclaration.de.faillite) +
  as.factor(Retard.paiement) +
  as.factor(Homme) +
  as.integer(Personnnes.à.charges) +
  as.factor(Diplômé) +
  as.factor(Blanc) +
  as.factor(Marié),
  data = data, family=binomial (link='logit'));
modele2 <- stepAIC(modele1);
stargazer(modele1, modele2, type="text", align=TRUE);
round(odds.ratio(modele1), 3);
round(odds.ratio(modele2), 3);
Approuvé.pred <- predict(modele2, type = "response", newdata = data)
table(Approuvé.pred > 0.5, data$Approuvé);
```

:

```
prop.B<-mean(data$Approuvé[data$Blanc=="1"]);
prop.M<-mean(data$Approuvé[data$Blanc=="0"]);
ecart=prop.B-prop.M;
logitB<-glm(Approuvé ~ as.factor(Marié) +
  Ratio.dépense.revenu +
  Ratio.prêt.prix +
  as.factor(Assurance.hypothécaire) +
  as.factor(Déclaration.de.faillite) +
  as.factor(Retard.paiement),
  data = data[data$Blanc==1,], family=binomial (link='logit'));
logitM<-glm(Approuvé ~ as.factor(Marié) +
  Ratio.dépense.revenu +
  Ratio.prêt.prix +
  as.factor(Assurance.hypothécaire) +
  as.factor(Déclaration.de.faillite) +
  as.factor(Retard.paiement),
  data = data[data$Blanc==0,], family=binomial (link='logit'));
```

:

```
stargazer(logitB, logitM, type="text", column.labels=c("Blanc", "Minorité"),
  align=TRUE);
round(odds.ratio(logitB), 3);
round(odds.ratio(logitM), 3);
```

:

```
data$pB<-predict(logitB, data, type='response');
prop.cM<-mean(data$pB[data$Blanc==0], na.rm=TRUE);
prop.cM;
expl<-prop.B - prop.cM;
inexpl<-prop.cM - prop.M;
expl.p<-round((expl/ecart)*100, 2);
inexpl.p<-round((inexpl/ecart)*100, 2);
coeffs.B<-logitB$coefficients;
coeffs.M<-logitM$coefficients;
```

:

```
X.B <- model.matrix(~ as.factor(Marié) +
  Ratio.dépense.revenu +
  Ratio.prêt.prix +
  as.factor(Assurance.hypothécaire) +
  as.factor(Déclaration.de.faillite) +
  as.factor(Retard.paiement),
  data = data[data$Blanc==1,]);
X.moy.B<-apply(X.B, 2, mean);
X.M <- model.matrix(~ as.factor(Marié) +
  Ratio.dépense.revenu +
  Ratio.prêt.prix +
  as.factor(Assurance.hypothécaire) +
  as.factor(Déclaration.de.faillite) +
  as.factor(Retard.paiement),
  data = data[data$Blanc==0,]);
X.moy.M<-apply(X.M, 2, mean);
delta.X.beta<-(X.moy.B- X.moy.M)*coeffs.B;
delta.beta.X<-(coeffs.B- coeffs.M)*X.moy.M;
```

:

```
resultats <- data.frame(Variables = c("Constante",
  "Marié",
  "Ratio dépenses/revenus",
  "Ratio prêt/prix",
  "Assurance hypothécaire",
  "Déclaration de faillite",
  "Retard de paiement"),
  Explique = delta.X.beta,
  Inexplique = delta.beta.X);
```

:

```

graph.expl<-ggplot(data=resultats, aes(x=Variables, y=Explique)) +
  geom_bar(stat="identity", fill="#46A288") +
  scale_x_discrete(limits=c("Retard de paiement", "Déclaration de faillite",
    "Assurance hypothécaire", "Ratio prêt/prix", "Ratio dépenses/revenus",
    "Marié", "Constante")) +
  coord_flip() + theme_minimal() + theme(axis.title.y = element_blank()) +
  scale_y_continuous(name="", limits=c(-2, 2)) + ggtitle("Expliqué") +
  theme(plot.title = element_text(size=14, face="bold"), axis.text.y =
    element_text(size=12));
graph.inexpl<-ggplot(data=resultats, aes(x=Variables, y=Inexplique)) +
  geom_bar(stat="identity", fill="#46A288") +
  scale_x_discrete(limits=c("Retard de paiement", "Déclaration de faillite",
    "Assurance hypothécaire", "Ratio prêt/prix", "Ratio dépenses/revenus",
    "Marié", "Constante")) +
  coord_flip() + theme_minimal() + theme(axis.title.y = element_blank()) +
  scale_y_continuous(name="", limits=c(-2, 2)) + ggtitle("Inexpliqué") +
  theme(plot.title = element_text(size=14, face="bold"), axis.text.y =
    element_text(size=12));
plot_grid(graph.expl, graph.inexpl, labels=c("", ""), ncol = 1, nrow = 2);

```

: