Escaping the Subprime Trap in Algorithmic Lending

Adam Bouyamourn
Princeton University, NJ, USA

Alexander Williams Tolbert
Emory University, Atlanta, GA, USA

Abstract

Disparities in lending to minority applicants persist even as algorithmic lending finds widespread adoption. We study the role of risk-management constraints, specifically Value-at-Risk and Expected Shortfall, in inducing inequality in loan approval decisions, even among applicants who are equally creditworthy. We contribute an analysis of 431,551 loan applications recorded under the Home Mortgage Disclosure Act, illustrating that disparities in data quality are associated with higher rates of loan denial and higher interest rate spreads for Black borrowers. We develop a formal model in which a mainstream bank (low-interest) is more sensitive to variance risk than a subprime bank (high-interest). If the mainstream bank has an inflated prior belief about the variance of the minority group, it may deny that group credit indefinitely, never learning the true risk of lending to that group, while the subprime lender serves this population at higher rates. We call this "The Subprime Trap": an equilibrium in which minority borrowers can borrow only from high-cost lenders, even when they are as creditworthy as majority applicants. We show that a finite subsidy can help minority groups escape the trap by covering enough of the mainstream bank's downside so that it can afford to lend to, and thereby learn the true risk of lending to, the minority group. Once the mainstream bank has observed sufficiently many loans, its beliefs converge to the true underlying risk, and competition drives down the interest rates of subprime loans.

2012 ACM Subject Classification Applied computing → Economics

Keywords and phrases Algorithmic fairness, algorithmic lending, Risk management, Value-at-Risk, Algorithmic Philosophy

Digital Object Identifier 10.4230/LIPIcs.FORC.2026.6

Related Version *Previous Version*: https://arxiv.org/abs/2502.17816

1 Introduction

Algorithmic lending has grown rapidly as scalable ML methods achieve wide adoption among both existing lenders and new market entrants, bringing with them the possibility of significantly improving the fairness of financial decisions based on observable data about loan applicants [47, 20, 15, 60]. However, inequalities persist in both loan approval rates and interest rates charged to minority applicants versus white applicants, and a switch to algorithmic lending procedures does not necessarily improve outcomes on either metric [62, 68, 59, 39, 4, 11, 37]. These inequities exist against a background of historic discrimination in US retail banking to individuals [59], businesses [16], and a long-standing "racial wealth gap": persistent differences in median household wealth by ethnic group [24, 31, 5].

Notably, disparities in interest rates and loan approvals persist even when minority applicants have comparable credit scores to majority applicants [12, 58, 29]. This combination of facts creates a puzzle: are lenders selecting on observable risk factors? And if not, why, besides explicit discrimination, might they be failing to lend to minority applicants?

Several empirical findings help us to understand this puzzle. First, ethnic groups sort across lenders, with minorities more likely to accept loans from high-interest rate banks versus conventional lenders, even conditional on credit score. Bayer et al. [12] write:

7th Symposium on Foundations of Responsible Computing (FORC 2026).
Editor: Huijia (Rachel) Lin; Article No. 6; pp. 6:1–6:27
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

> African-American and Hispanic borrowers tend to be more concentrated at high-risk lenders. Strikingly, this pattern holds for all borrowers even those with relatively unblemished credit records and low-risk loans... High-risk lenders are not only more likely to provide high cost loans overall, but are especially likely to do so for African-American and Hispanic borrowers. These lenders are largely responsible for the differential treatment of equally qualified borrowers.

Modern algorithmic lending systems operationalize risk assessment through machine learning models trained on historical data. Minority applicants disproportionately have "thin credit files" with limited payment history data, and this data sparsity directly manifests as higher prediction variance in ML models [17]. When banks face risk-management constraints, the higher predicted variance translates into higher rates of loan denial. Because denied applicants never generate repayment data, banks cannot then learn that their variance estimates were inflated, creating a self-reinforcing equilibrium of exclusion.

It is this set of stylized facts that motivates our model. How do minorities with good credit scores end up with no choice but to accept high-cost loans? To answer this question, we consider the role of risk management constraints when banks have imperfect information about their applicant pool. Lending is a problem of imperfect information: banks do not know the exact probability that a given borrower will repay a loan, and so must use observable characteristics as a proxy for repayment probability [1, 28]. Banks rely on credit scoring, which is subject to known biases, and has lower quality information about the creditworthiness of minority applicants [17]. Reliance on credit scores can therefore lead banks to hold inaccurate prior beliefs about minority applicants, which can sustain equilibria in which minority groups systematically lose out [26, 48].

We study the effect of Value-at-Risk (VaR) constraints on loan approval decisions [7, 45]. Our analysis focuses on a formal model in which a mainstream, or "low interest rate", bank exhibits heightened sensitivity to variance risk relative to a subprime, or "high interest rate", bank. We show that when the mainstream bank has inflated prior beliefs about the variance repayments from the minority group, VaR constraints bind, which means that it will then systematically refrain from extending credit to that group. This reluctance then prevents the bank from updating its beliefs regarding the true risk, thereby locking minority borrowers into higher-cost subprime lending arrangements. We describe this mechanism as "the subprime trap".

1.1 Our Contributions

First, we contribute an analysis of risk management constraints to the study of algorithmic fairness. We help address a puzzle of general interest: why might banks fail to extend loans even when they would otherwise be profitable? Second, we bring details of real-world banking practices into the study of algorithmic fairness. We introduce a two-bank equilibrium framework in which the mainstream bank enforces a more stringent VaR requirement than its subprime counterpart. Under these conditions, an erroneous initial estimate of the variance of the minority group leads the mainstream bank to permanently refrain from lending to that group. Third, we formalize the resulting equilibrium (Theorem 13) and demonstrate that the minority group is consequently confined to subprime loans, despite having the same average creditworthiness as the majority group. Fourth, we establish that the provision of a modest, finite subsidy, or partial guarantee, can resolve this equilibrium inefficiency. By adequately mitigating downside risk, the subsidy enables the main bank to lend and learn about the creditworthiness of the minority group. With an updated assessment of risk, the bank eventually meets its VaR requirements without continued external intervention, thereby extending mainstream credit at favorable rates.

This investigation contributes to the literature by highlighting how inaccurate risk metrics, in conjunction with risk-based capital constraints, can systematically exclude certain borrower groups from low-interest credit. Our results further imply that targeted subsidies can effectively rectify informational failures, yielding improved outcomes for both lenders and minority borrowers.

1.2 Related Literature

Our work contributes to several strands of literature, including algorithmic fairness in lending, the economics of discrimination, and the role of informational frictions in sustaining suboptimal market equilibria. In the domain of algorithmic fairness, recent research has examined how increasingly sophisticated data-driven methods in credit scoring may inadvertently perpetuate or exacerbate disparities [10, 9, 50]. While algorithmic approaches promise enhanced accuracy, empirical evidence indicates that minority applicants often face higher interest rates or are denied credit outright [11, 29].

Discrimination is a classical topic in economics, studied in a large theoretical and empirical literature [13, 56, 53, 6, 40, 27]. Building on this tradition, subsequent studies have explored how negative stereotypes or misperceptions of risk may translate into adverse outcomes for minority borrowers [34, 26]. In a manner analogous to these analyses, our work demonstrates that inflated beliefs about repayment variance can result in a self-reinforcing equilibrium whereby minority groups are persistently relegated to subprime lending markets.

A related literature examines the role of informational frictions and the corrective impact of policy interventions [2, 3, 63]. Cai et al. [21] studied contexts in which banks can acquire improved information about borrowers' creditworthiness through selective experimentation. Donahue and Barocas [33] study the trade-offs between solidaristic insurance policies, in which minorities are subsidized in accordance with risks that are causally related to their minority status, and actuarially-fair insurance policies, in which everyone pays premia equal to their marginal risk.

Our work also relates to work on delayed feedback with respect to implementation decisions in machine learning settings [52, 55, 22]. We describe a setting in which mainstream banks are stuck in a negative feedback loop: their failure to learn about applicant creditworthiness is self-sustaining. Our work is also related to literature on suboptimal outcomes with bandits. Honda and Takemura [42] prove that Thompson Sampling's optimality depends critically on prior specification, while Ghosh et al. [38] show that any algorithm achieving optimal performance under perfect model specification must suffer linear regret under misspecification. This mirrors our setting where banks with incorrect variance priors never explore lending to certain groups. The dueling bandits framework of Yue et al. [67] captures competition between lenders, establishing information-theoretically optimal algorithms for pairwise comparisons. Frazier et al. [36] study how to incentivize exploration when arms are pulled by self-interested agents – our approach, in which subsidies encourage banks to "learn through lending", is closely related to this setting.

Our work is also related to the literatures on selective labeling, causal fairness, and algorithmic censoring. In each case, labels are produced by a causal process, in which parameters of interest must be identified for the target, fairness-relevant quantity of policy interest to be identified [49, 66, 44, 25].

Finally, our work contributes to a broader effort to model rational but suboptimal decision-making by or with respect to disadvantaged groups. Diana et al. [32] model pessimism traps and herding behavior as individuals grapple with cycles of noisy and censored information when making decisions about potentially higher reward but riskier ends. Our argument is

Table 1 Differential Data Quality and Lending Outcomes by Race.

Panel A: Differential Missingness By Race			
	N	Missing Income	Missing DTI
Black Applicants	19,263 (4.5%)	2.3%	4.0%
White Applicants	412,288 (95.5%)	1.2%	1.3%
Difference		1.1pp	2.7pp
Ratio (Black/White)		1.9	3.1

Panel B: Mortgage Denial Rates			
	Complete Data	Missing Income	Difference
Black Applicants	26.1%	57.6%	+31.5pp
White Applicants	9.8%	23.1%	+13.3pp
Differential Effect			+18.2pp

Panel C: Interest Rate Spreads[1] on Approved Loans			
	Complete Data	Missing Income	Difference
Black Applicants	0.85%	1.12%	+0.27pp
White Applicants	0.62%	0.74%	+0.12pp
Differential Effect			+0.15pp

similar in broad outline to that of Foster and Vohra [35] and Hu and Chen [43], who argue that temporary interventions in the labor market to address discrimination may lead to long-run improvements in the fairness of observed decisions.

2 Empirical Motivation

We analyzed 2024 Home Mortgage Disclosure Act (HMDA) data, examining 431,551 conventional home purchase mortgage applications by single applicants listed as Black or White on loan applications. The HMDA requires that financial institutions that report at least 60,000 applications and loans in the previous calendar year must report data on lending applicants and loan outcomes. We examine racial disparities in two lending outcomes: loan denial decisions ($N = 431,551$) and interest rate spreads conditional on approval ($N = 247,329$).

This analysis highlights that 1) Black applicants have higher rates of missing data 2) Black applicants with missing income and personal-debt-to-income data face higher loan denial rates both marginally, and conditional on having missing data; and 3) if approved for loans, Black applicants pay higher interest rate spreads both marginally, and conditional on having missing data.

While this analysis is descriptive, it suggests a puzzle: why do differences in data quality lead to higher rejection rates and higher interest rates for Black applicants specifically? We propose an explanation for this observed phenomenon in our formal model below.

[1] Rate spread measures how much more expensive a borrower's loan is compared to the best rates available to prime borrowers at the time of origination. A spread of 1% means the borrower pays 1 percentage point more in annual interest than the benchmark rate.

3 Preliminaries and Model

We analyze a multi-period environment indexed by $t = 1, 2, \ldots, T$. At the onset of each period t, the high-interest (subprime) bank selects an interest-rate premium $\nu_t \in [0, \nu^{\max}]$. Each loan applicant belongs to one of two groups and applies simultaneously to both the mainstream (low-interest) bank and the subprime bank. The mainstream bank charges a baseline rate normalized to 1, while the subprime bank charges $1 + \nu_t$.

Each bank evaluates whether to approve or reject each applicant in that period, seeking to maximize its expected profit subject to a risk-management constraint. If an applicant receives approvals from both banks, the applicant typically accepts the offer with the lower interest rate; if only one bank approves the loan, it is accepted. After the loans are finalized, each bank privately observes its payoff. Consequently, a bank that rejects an application or whose offer is declined does not observe the repayment outcome for that applicant.

At the end of period t, both banks update their beliefs regarding the repayment distribution for each group, employing a Bayesian learning procedure described in Section 3.2.5. These updated beliefs affect the subprime bank's subsequent choice of premium ν_{t+1} and both banks' lending decisions in period $t + 1$.

3.1 Loan Applicants

3.1.1 Payoff distributions

There are two groups, denoted by $i \in \{W, B\}$. W represents the racial majority, while B represents the racial minority. Each group has a payoff distribution $\pi_{it} \sim \mathcal{N}(\mu_i, \sigma_i^2)$. The bank's return from approving a loan depends on this realized payoff.

▶ **Assumption 1** (Equal expected creditworthiness).

$$\forall t : \mathbb{E}\left[\pi_{Wt}\right] = \mu_W = \mu_B = \mathbb{E}\left[\pi_{Bt}\right] > 0$$

First, we assume that both groups have the same expected creditworthiness. We do this because we want to study the specific case where minority and majority racial groups are equally creditworthy. We allow their variances, σ_W^2 and σ_B^2, to differ.

Second, we assume that this is *common knowledge*: both banks know that each group is equally creditworthy.

We suppose that each group strictly prefers to accept a loan rather than not, and strictly prefers to accept the cheapest loan. We assume that applicants randomize if both loans have the same price. We denote each group's decision to accept the loan with $S_i \in \{0, 1\}$. We further assume that there exists an upper bound $\bar{\nu} > 0$ such that applicants accept a loan if and only if the interest rate is at most $1 + \bar{\nu}$. This reflects a reservation utility: borrowers prefer no loan to a loan at sufficiently punitive rates. In the main model, we assume $\nu_t \leq \bar{\nu}$ so that Bank H's offer is always accepted when no cheaper alternative is available.

3.1.2 Observable Applicant Characteristics

We suppose that individuals are associated with repayment-probability-relevant applicant characteristics: repayment history, assets, income, net debt, and so forth. We suppose that this information is differentially missing by group: minority borrowers have fewer complete characteristics. We formalize this simply by supposing that for each group, the number of complete applicant files, or the *credit history for group i*, C_i is binomial, so that we observe np_i complete applicant files and $n(1 - p_i)$ incomplete applicant files.

▶ **Assumption 2** (Differential Information about Applicants)**.**

$$p_B < p_W \textit{ so that } \mathbb{E}[C_B] < \mathbb{E}[C_W]$$

For simplicity, we set $p_W = 1$, and $p_B < 1$. This is WLOG, because what matters for the derivation of our results is the difference in information between the groups.

3.2 Lenders

The model features two banks: Bank L, which offers low-interest (mainstream) loans, and Bank H, which provides high-interest (subprime) loans. We refer to banks with the index j. In each period, both banks decide whether to extend a loan to each group. $A_{ijt} \in \{0, 1\}$ denotes whether bank j approves a loan to group i at time t.

3.2.1 Bank payoff functions

If the bank issues a loan to an applicant from group i in period t, it earns the payoff π_{it}. We normalize the interest rate of Bank L to 1, so that its profit Π in period t is given by

$$\Pi_{Lt} = \sum_i S_{iLt} A_{iLt} \pi_{it}$$

Bank H charges an interest rate of $1 + \nu_t$. It has profit function:

$$\Pi_{Ht} = \sum_i S_{iHt} A^{(t)}_{iHt} \left[(1 + \nu_t) \max\{\pi_{it}, 0\} + \min\{\pi_{it}, 0\}\right].$$

Here, positive returns are amplified by a factor of $1 + \nu_t$, whereas losses are incurred on a dollar-for-dollar basis. We suppose that this interest rate differential reflects its higher cost of capital.

3.2.2 Risk Management Constraints

We next suppose that banks face a *risk management* or *solvency* constraint. That is, there is some level of financial loss that is "unacceptable" to the bank. This may be due to regulatory constraints, such as Basel III or Dodd-Frank, or liquidity constraints [51]. This loss threshold is probabilistic: the bank is willing to accept some nonzero risk that their loss from making a loan falls below a certain threshold, but wants to specifically limit the probability that this occurs to less than a risk tolerance $\alpha\%$. Typically, this value is set to 1% or 5% [45].

We model this with a per-period Value-at-Risk (VaR) constraint [7]. That is, we require that the bank's anticipated profit in each period must be greater than some constant $\rho < 0$ with probability at least $1 - \alpha$. We can think of ρ as the bank's maximum acceptable loss. Equivalently, the bank is willing to accept an $\alpha\%$ risk that their profit will fall below the bank's risk management threshold ρ.

▶ **Definition 3** (Value-at-Risk (VaR))**.**

$$VaR_\alpha(X) = -\inf\{x \,|\, \mathbb{P}[X \le x] > \alpha\}$$

We adapt this by requiring that each bank j in period t faces the constraint that:

$$\mathbb{P}(\Pi_{jt} < \rho) \le \alpha$$

This simply states that the bank will accept a risk of at most $\alpha\%$ that their profit in period t falls below ρ.

We suppose that this constraint is *lexically prior* to the profit-maximization objective: the bank must satisfy the VaR constraint in order to lend at all.

Combining each bank's payoff function with their VaR constraint, we can write each bank's optimization problem as follows:

$$\arg\max_{\boldsymbol{A}_{jt}} \ \Pi_{jt}\left(\boldsymbol{A}_{jt}\right) \quad \text{subject to} \quad \mathbb{P}(\Pi_{jt}\left(\boldsymbol{A}_{jt}\right) \leq \rho) \leq \alpha$$

Each bank chooses the approval decisions that maximize profit subject to satisfying its VaR constraint. As we shall see below, the maximand depends on the expected value of the applicants' payoff function μ_i, while the VaR depends on both the mean and the variance of the applicant group's payoff function.

3.2.3 Prior Bank Beliefs

The actual profit Π_{jt} is unknown *ex ante*, however: it is only observed once a loan has been approved and accepted. Banks must therefore instead make decisions based on prior beliefs about the profitability of approving a loan.

Note that, at $t = 1$, the bank cannot use observed repayment information in the determination of the loan application. Instead, we suppose that the bank uses historical credit data to estimate the risk of lending to specific groups.

Denote by $\hat{\Pi}_{jt}$ the bank's posterior estimate of profit from lending at time t. Each bank instead must solve the *feasible* problem:

$$\arg\max_{\boldsymbol{A}_{jt}} \ \hat{\Pi}_{jt}\left(\boldsymbol{A}_{jt}\right) \quad \text{s. t.} \quad \mathbb{P}\Big(\hat{\Pi}_{jt}\left(\boldsymbol{A}_{jt}\right) \leq \rho\Big) \leq \alpha$$

Which, importantly, depends on bank beliefs about applicants.

We assume that each bank j has initial prior beliefs regarding the parameters of the distribution of returns from issuing a loan to a member of group i.

We assume, as before, that the mean of both payoff distributions is known and common knowledge. We assume that the scale of each distribution is unknown, however.

▶ **Definition 4** (Bank's prior beliefs). *We suppose that each bank has an Inverse-Gamma prior on the variance of each group, so that:*

$$\sigma^2_{ij,0} \sim \textit{Inv-}\Gamma(a_0, b_0)$$

3.2.4 Differential Credit Records Drive Prior Variances

We suppose that each bank estimates a model of repayment probability based on historical credit data to form its initial beliefs. That is, on having observed a credit file with effective sample size $C_i = np_i$, each bank forms a prior about each group's variance as follows:

$$\sigma^2_{ij}|C_i \sim \text{Inv-}\Gamma\left(a_0 + \frac{np_i}{2}, b_0 + \frac{np_i S_i^2}{2}\right)$$

Since we suppose that $p_B < p_W = 1$, we have (assuming[2] that $S_B^2 = S_W^2 = 1$):

[2] We assume that $S_B^2 = S_W^2 = 1$ for simplicity. This assumption is conservative in this model. Suppose instead that $S_B^2 > 1 > S_W^2$, which would be consistent with differential rates of missingness. Then, group B's variance, given the credit record data, would be $\text{Inv-}\Gamma\left(a_0 + \frac{np_B}{2}, b_0 + \frac{np_B S_B^2}{2}\right)$, which is strictly larger than the prior with $S_B^2 = 1$.

$$\sigma^2_{Bj,0}|C_B \sim \text{Inv-}\Gamma\left(a_0 + \frac{np_B}{2}, b_0 + \frac{np_B}{2}\right)$$
$$\sigma^2_{Wj,0}|C_W \sim \text{Inv-}\Gamma\left(a_0 + \frac{n}{2}, b_0 + \frac{n}{2}\right)$$

So that:

$$\hat{\sigma}^2_{Bj,0} = \mathbb{E}[\sigma^2_{Bj,0}|C_B] > \mathbb{E}[\sigma^2_{Wj,0}|C_W] = \hat{\sigma}^2_{Wj,0}$$

Or, in other words, based on observable credit histories, each bank has a prior belief that the variance of the minority group is larger than the variance of the majority group.

3.2.5 Learning Through Lending

In each period, if a bank extends a loan to an applicant from group i, it observes the realized repayment and updates its risk assessment for that group via Bayesian updating.

▶ **Lemma 5** (Belief Updating via Bayes Rule). *If bank j lends to group i, it observes a return π_{it} and then updates its posterior beliefs about the variance σ^2_i. Suppose that the bank observes a sequence of returns $\{\pi_1, \pi_2, \ldots, \pi_M\}$. Then, it has posterior belief about the variance of group i:*

$$\sigma^2_B|\pi_1, \ldots, \pi_M \sim$$
$$Inv\text{-}\Gamma\left(a_0 + \frac{np_B + M}{2}, b_0 + \frac{np_B + \sum_{m=1}^{M}(\pi_{Bm} - \mu_B)^2}{2}\right)$$

So that:

$$\mathbb{E}[\sigma^2_B|\pi_1, \ldots, \pi_M] = \frac{b_0 + \frac{np_B + \sum_{m=1}^{M}(\pi_{Bm} - \mu_B)^2}{2}}{a_0 + \frac{np_B + M}{2} - 1}$$

In other words, the bank's posterior belief about the variance of the minority group depends both on the depth of historical credit files (np_B) and the observed variance of repayments ($\sum_{m=1}^{M}(\pi_{Bm} - \mu_B)^2$), conditional on lending to the minority group.

A key assumption in our analysis is that a bank updates its estimate of the variance for group i only when it observes repayment outcomes from that group. Thus, if a bank never lends to group B, it will not receive data on the performance of loans to that group, and incorrect prior belief about B's variance will not be updated.

3.3 Our Modeling Choices

Inaccurate prior beliefs

Our model supposes that banks have inaccurate beliefs about the creditworthiness of applicants. This is a plausible outcome when decisions are made on the basis of credit scoring data that is either biased or noisy. Credit scores may be inherently noisier representations of underlying default risk for minority groups [17], while the components of the score may themselves be low quality indicators of repayment ability [61]. Empirical work shows that banks are more likely to avoid lending to minority neighborhoods, thereby lowering the quality of information received about lenders in those neighborhoods [18]. Noisier estimates of credit risk are likely to lead to inflated prior estimates of variance in practice.

Risk-pooling

We study three cases: where banks lend unilaterally to each group, and where banks lend to both groups. In the unilateral setting, there is no risk pooling, and banks make decisions based only on single group characteristics. However, when banks are willing to lend to both groups, risk pooling occurs, where the lower-risk group essentially subsidizes the risk of lending to the higher-risk group. This case is studied in [33]. Below, we derive variance thresholds for each of the three cases.

Constrained Exploration

We suppose that the bank learns about applicants via two channels: historical credit data, and lending. We model the case in which per-period financial risk constraints prevent banks from engaging in constrained exploration – they do not lend to a sample of loss-making applicants in order to learn information about applicants who did not receive loans. In finance, inferring the repayment behavior of applicants who were in fact denied credit is known as the reject inference problem [8, 54], and is fundamentally a causal inference problem: the counterfactual is not observed unless the bank designs its lending pipeline to causally identify group characteristics [41, 57]. Our model characterizes the equilibrium that obtains when banks do not solve this problem. Banks could recover group parameters through intentional experimentation, if per-period constraints do not bind, and the expected value of information is positive. The subsidy mechanism we describe in Theorem 18 can be understood as compensating the bank for the cost of exploration.

4 Single-Period Setting

We study the stage game to show how the VaR constraint affects lending to group B in each period. We first derive variance thresholds for each bank lending to a single group (unilateral thresholds), then derive thresholds for lending to both groups simultaneously (pooled thresholds). The pooled thresholds depend on the bank's belief about group B's variance, holding group W's variance fixed.

▶ **Lemma 6** (Unilateral Lending Thresholds)**.** *Consider bank j lending to a single group i with payoff $\pi_i \sim \mathcal{N}(\mu, \hat{\sigma}_i^2)$. The bank's VaR constraint is satisfied for a given group if and only if $\hat{\sigma}_i^2 \leq \tilde{\sigma}_j^2$, where:*

$$\tilde{\sigma}_L^2 = \left(\frac{\rho - \mu}{\Phi^{-1}(\alpha)}\right)^2 \qquad \tilde{\sigma}_H^2 = \left(\frac{\rho - (1+\nu)\mu}{\Phi^{-1}(\alpha)}\right)^2$$

These represent the thresholds under which each bank is willing to lend unilaterally to a given group.

Each bank also has a pooled variance threshold, under which it is willing to lend to both groups.

▶ **Lemma 7** (Pooled Variance Threshold for Bank L)**.** *Consider Bank L, which offers loans at a normalized interest rate of 1. Bank L lends to both groups in period t if and only if group B's variance satisfies:*

$$\hat{\sigma}_{BLt}^2 \leq \tilde{\sigma}_L^{2,pool} \equiv \left(\frac{\rho - (\mu_W + \mu_B)}{\Phi^{-1}(\alpha)}\right)^2 - \sigma_W^2$$

In other words, there is an upper bound $\tilde{\sigma}^L$ on the Bank's beliefs about the variance of group B, such that the bank is only willing to lend to group B if it believes that group has lower variance than this threshold. Otherwise, its risk of a shortfall exceeds $\alpha\%$.

A similar analysis for the subprime bank H shows that it tolerates a higher level of variance. We have:

▶ **Lemma 8** (Pooled Variance Threshold for Bank H). *Bank H lends to group B and group W in period t if and only if $\hat{\sigma}^2_{BHt} \leq \tilde{\sigma}^H_{pool}$, where:*

$$\tilde{\sigma}_H^{2,pool} = \left(\frac{\rho - (1+\nu_t)(\mu_W + \mu_B)}{\Phi^{-1}(\alpha)}\right)^2 - \sigma_W^2$$

▶ **Corollary 9** (Ordered variance thresholds). *For $\nu_t > 0$ and $\alpha < 0.5$:*

$$\tilde{\sigma}_L^2 < \tilde{\sigma}_H^2 \quad \text{and} \quad \tilde{\sigma}_L^{2,pool} < \tilde{\sigma}_H^{2,pool}$$

That is, Bank H's variance threshold exceeds Bank L's in both the unilateral and pooled settings.

Intuitively, gains are scaled by the factor $(1+\nu_t)$, providing the high-interest rate bank insulation against downside risk. When the perceived variance of group B exceeds $\tilde{\sigma}_L^{2,\text{pool}}$ but not $\tilde{\sigma}_H^{2,\text{pool}}$, Bank L declines to lend while Bank H approves the loan.

5 The Subprime Trap

We now consider the multi-period setting in which lending decisions influence the evolution of banks' beliefs about borrower creditworthiness.

5.1 Assumptions

We make several assumptions. These model stylized facts that characterize the subprime trap. First, we suppose that the variance of both groups is *in fact* below the threshold $\tilde{\sigma}_L^2$, though the variances are not necessarily identical.

▶ **Assumption 10** (Both groups are in fact creditworthy).

$$\sigma_W^2 \leq \sigma_B^2 < \tilde{\sigma}_L^2$$

This is intended to describe a situation in which the low-interest rate bank would lend to both groups under perfect information.

We characterize beliefs under imperfect information in which lending to group B does not occur. We suppose that Bank L holds a prior that the variance of repayments for group B is $\hat{\sigma}^2_{BL,0}$ such that:

▶ **Assumption 11** (L's prior variance for group B is above its risk threshold). $\hat{\sigma}^2_{BL,0} > \tilde{\sigma}_L^{2,pool}$

Where $\tilde{\sigma}_L^{2,\text{pool}}$ is the maximum variance that Bank L (the main lender) can tolerate under its constraint VaR. Based on its initial assessment, lending violates the risk limits of Bank L.

Third, Bank H, which operates at a higher interest rate and has a correspondingly higher risk tolerance. The high-rate bank has a lending threshold $\tilde{\sigma}_H^2$ such that $\tilde{\sigma}_H^2 > \tilde{\sigma}_L^{2,\text{pool}}$. Denoting bank H's prior by $\hat{\sigma}^2_{BH,0}$, we suppose that:

▶ **Assumption 12** (H's prior variance for group B is below its risk threshold). $\hat{\sigma}^2_{BH,0} \leq \tilde{\sigma}_H^2$

Consequently, Bank H is prepared to lend to group B.

5.2 Equilibrium

In each period, banks choose simultaneously whether to lend to groups W and B. Both types of bank approve group W's loan application, and group W chooses the lower-rate, mainstream bank. Based on their prior beliefs about B's variances, H approves B's loan application, but L does not. If Bank L does not lend to group B, then group B is left with the subprime option from Bank H. But since Bank L's belief update depends on observing a return from group B, if it does not lend to group B at period t, it does not have an updated posterior belief to use as its prior at period $t+1$: its beliefs do not change. The crucial observation is that if Bank L persistently withholds loans to group B, it never observes the repayment data needed to update its inflated variance estimate, and its belief remains at $\hat{\sigma}^2_{BL,0}$.

We now formalize this result.

▶ **Theorem 13** (Subprime Trap Equilibrium)**.** *Suppose Assumptions 1, 2, 10, 11, and 12 hold. Then, there exists a Bayesian subgame-perfect equilibrium in which, in every period, Bank L lends exclusively to group W, and Bank H lends to group B at rate $1+\bar{\nu}$, with high probability, permanently relegating group B to subprime loans.*

Proof. We show by induction that the low-cost bank never lends to group B.

First, in the base case, the bank does not lend to group B. We have that $A_{ij1} = 0$, since, by Lemma 7, if the bank's prior belief is $\hat{\sigma}^2_{BL0}$ and $\hat{\sigma}^2_{BL0} > \tilde{\sigma}^{2,\text{pool}}_L$, so that the VaR constraint is violated, lending does not occur in period 1.

To show the induction step, we show that if the bank does not lend to B in period t, then it does not lend to B in period $t+1$. For this, we have that $A_{ijt} = 0 \implies \hat{\sigma}^2_{ijt+1} = \hat{\sigma}^2_{ijt}$, because no updating occurs. but since $A_{ijt+1} = 1$ if and only if $\hat{\sigma}_{ijt+1} \leq \tilde{\sigma}^{2,\text{pool}}_L$, we have that $\hat{\sigma}^2_{Bt} > \tilde{\sigma}^{2,\text{pool}}_L$ and $\hat{\sigma}^2_{ijt+1} = \hat{\sigma}^2_{ijt} \implies \hat{\sigma}^2_{ijt+1} > \tilde{\sigma}^2_L$, so that $A_{ijt+1} = 0$. This verifies the induction step.

This shows by induction that the low-cost bank never lends to B. Since the bank *does* lend to W, we have that the optimal decision for Bank L is to lend exclusively to group W in every period.

Next, consider Bank H. By hypothesis, Bank H believes that $\sigma^2_{BH0} \leq \tilde{\sigma}^2_H$, and therefore lending to group B satisfies its risk constraint in the initial period. Moreover, since the true mean μ_B is positive, lending to group B is expected to yield a positive return. Bank H lends to group B with high probability, because in each stage there is a (vanishing) probability of observing a payoff realization that increases Bank H's posterior belief about group B's variance above its threshold. This probability goes to zero by the SLLN, however. Since H is the monopolist wrt group B, it charges the maximum rate it can in each period, which is $1+\bar{\nu}$.

The borrowers act accordingly. Group W, faced with a lower interest rate from Bank L, accepts that offer, while group B, having no offer from Bank L, accepts the loan from Bank H.

Finally, note that because Bank L never lends to group B, it never observes any repayment outcomes from that group. This implies that the variance estimate for group B remains at $\hat{\sigma}^2_{BL0} > \tilde{\sigma}^{2,\text{pool}}_L$ in every period. As a consequence, there is no incentive for Bank L to deviate from its strategy of lending only to group W, and similarly, neither Bank H nor the borrower groups have incentives to deviate from their prescribed actions. Hence, the described strategy profile constitutes a subgame-perfect equilibrium. ◀

▶ Remark 14 (Purely Informational Failure). Even though the true variance of group B is such that $\sigma_B^2 < \tilde{\sigma}_L^{2,\text{pool}}$, Bank L remains unaware of this fact unless it extends credit to group B. The absence of new data prevents Bank L from updating its risk assessment, leading to an inefficient outcome in which group B continues to access only subprime credit.

▶ Remark 15 (Equal Creditworthiness). It is noteworthy that the equilibrium outcome emerges solely from differences in prior variances. Since $\mu_B = \mu_W = \mu$, the two groups are identical in terms of average creditworthiness; the discrepancy stems entirely from the combination of incorrect beliefs about borrower variance, and the risk assessment constraints imposed by VaR.

6 Escaping the Subprime Trap via Subsidies

We now show that a subsidy or partial lending guarantee can help group B escape the subprime trap. The key idea is to cover sufficient downside risk so that Bank L is induced to lend to group B. When this occurs, Bank L can gather information on group B's repayment performance. Since group B is in fact creditworthy (Assumption 10), Bank L's beliefs will eventually converge to group B's actual variance, which is low enough that L's VaR constraint will be satisfied. Once enough observations have been accumulated, Bank L will update its risk assessment so that the estimated variance satisfies $\hat{\sigma}_{BLt}^2 < \tilde{\sigma}_L^2$, eliminating the need for further subsidy.

▶ **Lemma 16** (Learning through Lending). *Recall that $\hat{\sigma}_{BLm}^2$ denotes Bank L's posterior mean estimate of group B's variance after observing m repayment outcomes (Lemma 5). Then: $\hat{\sigma}_{BLm}^2 \to \sigma_B^2$ almost surely as $m \to \infty$*

By Lemma 7, this is the condition required for Bank L to lend to group B, so that the VaR constraint is satisfied. In other words, if the mainstream bank *were* to lend to group B for a sufficiently long period of time, it would learn about the variance of returns due to lending to group B. The problem is then how to induce the bank to do so.

To encourage Bank L to lend to group B, we introduce a subsidy mechanism.

First, we define our subsidy as the smallest side-payment that would allow the bank to satisfy its VaR constraint in period t. We have:

$$s_t^* = \inf \{ s \geq 0 \ : \ \mathbb{P}\left(\Pi_{jt} + s < \rho\right) \leq \alpha \}$$

This allows us to solve for the optimal subsidy.

▶ **Lemma 17** (Optimal Subsidy). *In each period t, the required subsidy is equal to:*

$$s_t^* = \max\Big\{0,\ \rho - (\mu_W + \mu_B) - \Phi^{-1}(\alpha)\sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2}\Big\}$$

If $s_t^* \leq 0$, no subsidies are needed; otherwise, the regulator covers any shortfall up to s_t^*, ensuring that Bank L meets its VaR restriction.

We formalize the adaptive subsidy mechanism in Algorithm 1.

Algorithm 1 Adaptive Subsidy to Escape the Subprime Trap.

Require: Parameters: ρ, α, μ, prior parameters (a_0, b_0), np_B, threshold $\tilde{\sigma}_L^2$, horizon T

Initialize:
  $a_t \leftarrow a_0 + \frac{np_B}{2}$ ▷ Shape with historical data
  $b_t \leftarrow b_0 + \frac{np_B}{2}$ ▷ Scale with historical data
  $M \leftarrow 0$ ▷ Count of new loans to group B
for $t \in 1, \cdots, T$ **do**
  $\hat{\sigma}_{BLt}^2 \leftarrow \frac{b_t}{a_t - 1}$ ▷ Current estimate
  if $\hat{\sigma}_{BLt}^2 > \tilde{\sigma}_L^*$ **then**
    $s_t \leftarrow s_t^*(\hat{\sigma}_{BLt}^2)$
    Bank L lends to B with subsidy s_t
    Observe return $\pi_{B,t}$
    $M \leftarrow M + 1$
    $a_t \leftarrow a_0 + \frac{np_B + M}{2}$
    $b_t \leftarrow b_0 + \frac{np_B + \sum_{m=1}^{M} (\pi_{B,m} - \mu)^2}{2}$
  else
    No subsidy needed.
  end if
end for

The subsidy incentivizes Bank L to lend to group B, generating observations that allow its risk assessment to converge to the true variance σ_B^2. Once the updated estimate satisfies $\hat{\sigma}_{BLt}^2 < \tilde{\sigma}_L^2$, it follows from Lemma 7 that Bank L meets its VaR constraint without subsidy, and no further external support is required.

We now state the main result regarding the subsidy mechanism.

▶ **Theorem 18** (Escaping the Subprime Trap). *Assume that the true distribution of repayments for group B is $\mathcal{N}(\mu, \sigma_B^2)$ with $\sigma_B^2 < \tilde{\sigma}_L^2$, but that Bank L initially believes that the variance is $\hat{\sigma}_{BL0}^2 > \tilde{\sigma}_L^2$. Under the subsidy mechanism described above (Algorithm 1), with probability one there exists a finite time τ such that for all $t \geq \tau$, $\hat{\sigma}_{BLt}^2 < \tilde{\sigma}_L^{2,pool}$ and consequently $s_t^* = 0$. Beyond time τ, Bank L continues lending to group B without further subsidy.*

In short, the subsidy induces the bank to update its information about minority group creditworthiness. It is also temporary when there exist otherwise creditworthy borrowers trapped in subprime loans: subsidizing information discovery helps banks learn the true variance of borrower groups. Further, while algorithmic audits could correct bias, subsidies can avoid privacy hurdles and scale faster [27].

▶ **Corollary 19** (Long-term benefits of subsidy). *For $t > \tau$, $\nu_t = 0$.*

In any round in which the low-interest bank would approve group B's loan, the high-interest bank will choose to set its premium to zero. If it chooses any positive premium, neither group will accept the loan. If it chooses a premium of zero, both groups will randomize over accepting the loan from bank L and accepting the loan from bank H. Since this persists for all $t > \tau$, this reduces interest rates permanently: competition reduces the prices charged by subprime lenders.

7 Extensions

VaR and CVaR/ES

Value-at-Risk is a commonly used risk management measure in practice, and is explicitly referenced in the Basel III banking regulations [51, 23]. In Appendix A.2, we also study the Conditional Value-at-Risk, or Expected Shortfall, which is related to VaR via the formula: $ES(X) = \frac{1}{\alpha}\int_0^{\alpha} VaR_{\phi}(X)\, d\phi$. The intuitions underlying our model carry through to the CVaR/ES risk metric: there is a threshold variance at which the CVaR/ES will be violated. We show that the CVaR/ES threshold is more conservative than the VaR risk metric, and hence requires a larger subsidy to escape the subprime trap.

Bank Heterogeneity

In the above model, we consider a stylized case with two banks. The action in our model is on the *extensive margin*: whether or not banks lend at all. In practice, there is a continuum of banks with heterogeneous cost structures. We study this case in Appendix A.1.1. Banks are indexed by their type ν_j, reflecting funding costs: a bank of type ν_j charges rate $1 + \nu_j$. We show that there exists a cutoff type $\bar{\nu}^*$ such that banks with $\nu_j \geq \bar{\nu}^*$ lend to minority applicants at rate $1 + \bar{\nu}^*$, while banks with $\nu_j < \bar{\nu}^*$ do not. Sorting on the extensive margin persists: minorities are served by higher-cost banks at higher rates. However, competition among these banks reduces the premium relative to the two-bank model, in which Bank H charges the monopoly rate $\bar{\nu}$.

Applicant Skimming

Our main model abstracts away from individual-level applicant data. In practice, banks also select on the *intensive margin*: *which* group members to lend to. We consider a model in Appendix A.1.2 in which banks observe whether each group B applicant has a complete or incomplete credit file, and can condition lending on this information. In this setting, Bank L lends to complete-file minority applicants (this is known as "skimming", e.g. [30]), but the population variance σ_B^2 is not identified from the complete-file sample alone: σ_{ic}^2 is unobserved. This recreates the Subprime Trap for the incomplete-file sub-population. While there is now lending to minorities, the mainstream bank does not update its beliefs about a subset of minority applicants with missing data, and these applicants remain confined to subprime loans.

8 Conclusion

Inequalities in algorithmic lending practices can persist even when minority applicants are as creditworthy as white applicants. We study one explanation for this phenomenon: the role of risk-management constraints, specifically VaR, in contributing to persistent disparities in lending. We have shown that risk management constraints can lead lenders to refuse loans even with positive Net Present Value. This forces minority applicants to accept high-interest rate loans, with negative implications for their financial prospects, a situation we describe as the "subprime trap" equilibrium. We emphasize the role of inaccurate prior beliefs about the risk of lending to minority groups, which leads to systematic exclusion and higher borrowing costs.

Our theoretical results align with documented patterns in mortgage lending markets. First, the persistence of our subprime trap equilibrium mirrors the decade-long stability of racial lending disparities documented in mortgage markets, even as algorithmic lending has expanded [11]. Second, our model's prediction that mainstream banks never update their beliefs about minority creditworthiness is consistent with [12]. Third, the equilibrium's existence despite equal creditworthiness directly explains the puzzle identified by [58] and [29]: why minorities with comparable credit scores remain concentrated at high-cost lenders.

Implementation of our subsidy mechanism maps naturally onto existing policy infrastructure. The Federal Housing Administration's mortgage insurance programs already function as partial guarantees, covering lender losses when borrowers default on FHA-insured loans [65]. Similarly, the Community Reinvestment Act provides regulatory credits to banks for community development loans and investments in low- and moderate-income neighborhoods [64, 19]. Our analysis suggests these programs could achieve greater impact if restructured to explicitly target variance reduction during a learning period; guaranteeing tail risks for a finite horizon would enable banks to update their risk assessments while avoiding solvency-affecting losses.

Our analysis focuses on one specific mechanism: variance-based discrimination that occurs when financial risk constraints disincentivize exploration. Observed lending disparities likely result from multiple interacting factors including historical discrimination, geographic segregation, and differential access to financial services. Further, banks could develop causal models that include selection dynamics when deciding which loans to approve [57, 14].

A key moral of our paper is that institutional rules developed with the intention of regulating an "offline" industry – here, financial risk constraints, can have unintended consequences when applied to online algorithmic settings [50, 46]. A per-period solvency constraint can reduce risk in a static portfolio setting, but may prevent actors from costly but profitable exploration in a dynamic setting.

Nonetheless, targeted, temporary interventions, such as subsidies or guarantees, can break this cycle by allowing banks to learn true risk profiles. These findings suggest practical avenues for addressing lending disparities and offer a framework to explain failures of fairness in algorithmic lending decisions.

References

1 William Adams, Liran Einav, and Jonathan Levin. Liquidity constraints and imperfect information in subprime lending. *American Economic Review*, 99(1):49–84, March 2009. doi:10.1257/aer.99.1.49.

2 George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488, August 1970. doi:10.2307/1879431.

3 George A Akerlof and Robert J Shiller. *Phishing for phools*. Princeton University Press, Princeton, NJ, August 2015.

4 Dionissi Aliprantis, Daniel Carroll, and Eric R Young. The dynamics of the racial wealth gap. *FRB of Cleveland Working Paper*, 2022. doi:10.26509/frbc-wp-201918r.

5 Lukas Althoff and Hugo Reichardt. Jim crow and black economic progress after slavery. *The Quarterly Journal of Economics*, 139(4):2279–2330, 2024.

6 Kenneth J Arrow. What has economics to say about racial discrimination? *Journal of Economic Perspectives*, 12(2):91–100, May 1998. doi:10.1257/jep.12.2.91.

7 Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, July 1999. doi:10.1111/1467-9965.00068.

8 John Banasik and Jonathan Crook. Reject inference, augmentation, and sample selection. *European Journal of Operational Research*, 183(3):1582–1594, 2007. doi:10.1016/j.ejor.2006.06.072.

9 Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and machine learning: Limitations and opportunities*. MIT press, 2023.

10 Solon Barocas and Andrew D. Selbst. Big data's disparate impact. *SSRN Electronic Journal*, 2016. doi:10.2139/ssrn.2477899.

11 Robert Bartlett, Adair Morse, Richard Stanton, and Nancy Wallace. Consumer-lending discrimination in the fintech era. *Journal of Financial Economics*, 143(1):30–56, January 2022. doi:10.1016/j.jfineco.2021.05.047.

12 Patrick Bayer, Fernando Ferreira, and Stephen L. Ross. The vulnerability of minority homeowners in the housing boom and bust. *American Economic Journal: Economic Policy*, 8(1):1–27, February 2016. doi:10.1257/pol.20140074.

13 Gary S Becker. *The economics of discrimination*. Phoenix Books. University of Chicago Press, Chicago, IL, 2 edition, August 1971.

14 Eli Ben-Michael, Kosuke Imai, and Zhichao Jiang. Policy learning with asymmetric counterfactual utilities. *Journal of the American Statistical Association*, 119(548):3045–3058, 2024. doi:10.1080/01621459.2023.2300507.

15 Tobias Berg, Andreas Fuster, and Manju Puri. Fintech lending. *Annual Review of Financial Economics*, 14(Volume 14, 2022):187–207, 2022. doi:10.1146/annurev-financial-101521-112042.

16 David G. Blanchflower, Phillip B. Levine, and David J. Zimmerman. Discrimination in the small-business credit market. *The Review of Economics and Statistics*, 85(4):930–943, 2003. URL: http://www.jstor.org/stable/3211816.

17 Laura Blattner and Scott Nelson. How costly is noise? data and disparities in consumer credit. arXiv preprint arXiv:2105.07554 [econ.GN], 2021.

18 Laura Blattner and Scott Nelson. How costly is noise? data and disparities in consumer credit, 2021. arXiv:2105.07554.

19 Board of Governors of the Federal Reserve System. History of the Community Reinvestment Act, 2024. URL: https://www.federalreserve.gov/consumerscommunities/cra_history.htm.

20 Teresa Bono, Karen Croxson, and Adam Giles. Algorithmic fairness in credit scoring. *Oxford Review of Economic Policy*, 37(3):585–617, September 2021. doi:10.1093/oxrep/grab020.

21 William Cai, Johann Gaebler, Nikhil Garg, and Sharad Goel. Fair allocation through selective information acquisition. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, AIES '20, pages 22–28, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3375627.3375823.

22 Allison J. B. Chaney, Brandon M. Stewart, and Barbara E. Engelhardt. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pages 224–232, New York, NY, USA, 2018. Association for Computing Machinery. doi:10.1145/3240323.3240370.

23 Chia-Lin Chang, Juan-Angel Jimenez-Martin, Esfandiar Maasoumi, Michael McAleer, and Teodosio Pérez-Amaral. Choosing expected shortfall over var in basel iii using stochastic dominance. *International Review of Economics & Finance*, 60:95–113, 2019. doi:10.1016/j.iref.2018.12.016.

24 Kerwin Kofi Charles and Erik Hurst. The transition to home ownership and the black-white wealth gap. *Review of Economics and Statistics*, 84(2):281–297, May 2002. doi:10.1162/003465302317411532.

25 Jennifer Chien, Margaret Roberts, and Berk Ustun. Algorithmic censoring in dynamic learning systems. In *ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 2023.

26 Stephen Coate and Glenn C. Loury. Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 83(5):1220–1240, 1993. URL: http://www.jstor.org/stable/2117558.

27 Bo Cowgill and Catherine E. Tucker. Economics, fairness and algorithmic bias. *SSRN Electronic Journal*, 2019. doi:10.2139/ssrn.3361280.

28 Gregory S. Crawford, Nicola Pavanini, and Fabiano Schivardi. Asymmetric information and imperfect competition in lending markets. *American Economic Review*, 108(7):1659–1701, July 2018. doi:10.1257/aer.20150487.

29 Matteo Crosignani and Hanh Le. *Stakeholders' Aversion to Inequality and Bank Lending to Minorities*. Federal Reserve Bank of New York, November 2023. doi:10.59576/sr.1079.

30 Calebe de Roure, Loriana Pelizzon, and Anjan Thakor. P2p lenders versus banks: Cream skimming or bottom fishing? *The Review of Corporate Finance Studies*, 11(2):213–262, December 2021. doi:10.1093/rcfs/cfab026.

31 Ellora Derenoncourt, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. Changes in the distribution of black and white wealth since the us civil war. *Journal of Economic Perspectives*, 37(4):71–89, November 2023. doi:10.1257/jep.37.4.71.

32 Emily Diana, Alexander Williams Tolbert, Kavya Ravichandran, and Avrim Blum. Adaptive algorithmic interventions for escaping pessimism traps in dynamic sequential decisions, 2024. To appear in FORC 2025. doi:10.48550/arXiv.2406.04462.

33 Kate Donahue and Solon Barocas. Better together? how externalities of size complicate notions of solidarity and actuarial fairness, 2021. arXiv:2103.00347.

34 Hanming Fang and Andrea Moro. *Theories of Statistical Discrimination and Affirmative Action: A Survey*, pages 133–200. Elsevier, 2011. doi:10.1016/b978-0-444-53187-2.00005-x.

35 Dean P. Foster and Rakesh V. Vohra. An economic argument for affirmative action. *Rationality and Society*, 4(2):176–188, April 1992. doi:10.1177/1043463192004002004.

36 Peter Frazier, David Kempe, Jon Kleinberg, and Robert Kleinberg. Incentivizing exploration. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pages 5–22, New York, NY, USA, 2014. Association for Computing Machinery. doi:10.1145/2600057.2602897.

37 Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47, December 2021. doi:10.1111/jofi.13090.

38 Avishek Ghosh, Sayak Ray Chowdhury, and Aditya Gopalan. Misspecified linear bandits, 2017. arXiv:1704.06880.

39 Marco Giacoletti, Rawley Heimer, and Edison G Yu. Using high-frequency evaluations to estimate discrimination: Evidence from mortgage loan officers. *SSRN Electron. J.*, 2021.

40 James J. Heckman. Detecting discrimination. *Journal of Economic Perspectives*, 12(2):101–116, June 1998. doi:10.1257/jep.12.2.101.

41 Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986. URL: http://www.jstor.org/stable/2289064.

42 Junya Honda and Akimichi Takemura. Optimality of thompson sampling for gaussian bandits depends on priors, 2013. arXiv:1311.1894.

43 Lily Hu and Yiling Chen. A short-term intervention for long-term fairness in the labor market. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, WWW '18, pages 1389–1398. ACM Press, 2018. doi:10.1145/3178876.3186044.

44 Kosuke Imai and Zhichao Jiang. Principal fairness for human and algorithmic decision-making, 2022. arXiv:2005.10400.

45 Philippe Jorion. *Value at risk*. McGraw-Hill, New York, NY, 2 edition, August 2000.

46 Stephanie Kelley, Anton Ovchinnikov, David R. Hardoon, and Adrienne Heinrich. Antidiscrimination laws, artificial intelligence, and gender bias: A case study in nonmortgage fintech lending. *Manufacturing & Service Operations Management*, 24(6):3039–3059, November 2022. doi:10.1287/msom.2022.1108.

47 Amir E. Khandani, Adlar J. Kim, and Andrew W. Lo. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking &; Finance*, 34(11):2767–2787, November 2010. doi:10.1016/j.jbankfin.2010.06.001.

48 Young-Chul Kim and Glenn C. Loury. Collective reputation and the dynamics of statistical discrimination. *International Economic Review*, 59(1):3–18, 2018. doi:10.1111/iere.12260.

49 Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Human decisions and machine predictions*. *The Quarterly Journal of Economics*, 133(1):237–293, August 2017. doi:10.1093/qje/qjx032.

50 I. Elizabeth Kumar, Keegan E. Hines, and John P. Dickerson. Equalizing credit opportunity in algorithms: Aligning algorithmic fairness research with u.s. fair lending regulation. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '22, pages 357–368. ACM, July 2022. doi:10.1145/3514094.3534154.

51 Siqiwen Li. *The Value-at-Risk Methodology of Basel II and Basel III*, pages 100–114. IGI Global, 2014. doi:10.4018/978-1-4666-5950-6.ch006.

52 L. T. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt. Delayed impact of fair machine learning. In *International Conference on Machine Learning*, pages 3150–3158, 2018.

53 Shelly J Lundberg and Richard Startz. Private Discrimination and Social Intervention in Competitive Labor Markets. *American Economic Review*, 73(3):340–347, June 1983. URL: https://ideas.repec.org/a/aea/aecrev/v73y1983i3p340-47.html.

54 Rogelio A. Mancisidor, Michael Kampffmeyer, Kjersti Aas, and Robert Jenssen. Deep generative models for reject inference in credit scoring, 2021. arXiv:1904.11376.

55 Nicolò Pagan, Joachim Baumann, Ezzat Elokda, Giulia De Pasquale, Saverio Bolognani, and Anikó Hannák. A classification of feedback loops and their relation to biases in automated decision-making systems. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23, pages 1–14. ACM, October 2023. doi:10.1145/3617694.3623227.

56 Edmund S. Phelps. The statistical theory of racism and sexism. *The American Economic Review*, 62(4):659–661, 1972. URL: http://www.jstor.org/stable/1806107.

57 Drago Plecko and Elias Bareinboim. Causal fairness analysis, 2022. doi:10.48550/arXiv.2207.11385.

58 Stephen Popick. Did minority applicants experience worse lending outcomes in the mortgage market? a study using 2020 expanded hmda data. *SSRN Electronic Journal*, 2022. doi:10.2139/ssrn.4131603.

59 Lincoln Quillian, John J. Lee, and Brandon Honoré. Racial discrimination in the u.s. housing and mortgage lending markets: A quantitative review of trends, 1976–2016. *Race and Social Problems*, 12(1):13–28, January 2020. doi:10.1007/s12552-019-09276-x.

60 Nydia Remolina. The role of financial regulators in the governance of algorithmic credit scoring. *SSRN Electronic Journal*, 2022. doi:10.2139/ssrn.4057986.

61 Lisa Rice and Deidre Swesnik. Discriminatory effects of credit scoring on communities of color. *Suffolk University Law Review*, 46:935, 2013. URL: https://api.semanticscholar.org/CorpusID:149577075.

62 Andrea Romei and Salvatore Ruggieri. A multidisciplinary survey on discrimination analysis. *The Knowledge Engineering Review*, 29(5):582–638, April 2013. doi:10.1017/s0269888913000039.

63 Richard H. Thaler. Behavioral economics: Past, present, and future. *American Economic Review*, 106(7):1577–1600, July 2016. doi:10.1257/aer.106.7.1577.

64 U.S. Congress. Community Reinvestment Act of 1977, 2013. 12 U.S.C. §§ 2901-2908. URL: https://www.govinfo.gov/content/pkg/USCODE-2013-title12/pdf/USCODE-2013-title12-chap30.pdf.

65 U.S. Department of Housing and Urban Development. Single family housing, 2024. URL: https://www.hud.gov/helping-americans/Single-Family-Housing.

66 Berk Ustun, Alexander Spangher, and Yang Liu. Actionable recourse in linear classification. In *ACM Conference on Fairness, Accountability and Transparency*, 2019.

67 Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, September 2012. doi:10.1016/j.jcss.2011.12.028.

68 Indrė Žliobaitė. Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery*, 31(4):1060–1089, March 2017. doi:10.1007/s10618-017-0506-1.

A Proofs

Proof of Lemma 5. We derive the posterior distribution for σ_B^2 after observing M loan outcomes.

Initial beliefs based on historical credit data. Before observing any data, the bank starts with an uninformative prior:

$$\sigma_B^2 \sim \text{Inv-}\Gamma(a_0, b_0)$$

The bank observes np_B historical loan outcomes from group B applicants with complete files. The likelihood of observing the historical data given σ_B^2 is:

$$p(C_B|\sigma_B^2) \propto (\sigma_B^2)^{-np_B/2} \exp\left(-\frac{np_B}{2\sigma_B^2}\right)$$

By Bayes' theorem:

$$p(\sigma_B^2|C_B) \propto p(C_B|\sigma_B^2) \cdot p(\sigma_B^2)$$

$$\propto (\sigma_B^2)^{-np_B/2} \exp\left(-\frac{np_B}{2\sigma_B^2}\right) \cdot (\sigma_B^2)^{-a_0-1} \exp\left(-\frac{b_0}{\sigma_B^2}\right)$$

$$= (\sigma_B^2)^{-(a_0+\frac{np_B}{2})-1} \exp\left(-\frac{b_0 + \frac{np_B}{2}}{\sigma_B^2}\right)$$

Therefore, the initial belief is:

$$\sigma_{B,0}^2 \sim \text{Inv-}\Gamma\left(a_0 + \frac{np_B}{2}, b_0 + \frac{np_B}{2}\right)$$

Learning through lending. Given the true variance σ_B^2 and known mean μ_B, the observed returns $\{\pi_{B,1}, \ldots, \pi_{B,M}\}$ are independently distributed:

$$\pi_{B,m}|\sigma_B^2 \sim \mathcal{N}(\mu_B, \sigma_B^2) \quad \text{for } m = 1, \ldots, M$$

The prior is:

$$p(\sigma_B^2) \propto (\sigma_B^2)^{-a_0+\frac{np_B}{2}-1} \exp\left(-\frac{b_0 + \frac{np_B}{2}}{\sigma_B^2}\right)$$

The likelihood for M observations with known mean μ_B is:

$$p(\pi_{B,1}, \ldots, \pi_{B,M}|\sigma_B^2) = \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_B^2}} \exp\left(-\frac{(\pi_{B,m} - \mu_B)^2}{2\sigma_B^2}\right)$$

$$= (2\pi\sigma_B^2)^{-M/2} \exp\left(-\frac{\sum_{m=1}^{M}(\pi_{B,m} - \mu_B)^2}{2\sigma_B^2}\right)$$

$$\propto (\sigma_B^2)^{-M/2} \exp\left(-\frac{\sum_{m=1}^{M}(\pi_{B,m} - \mu_B)^2}{2\sigma_B^2}\right)$$

By Bayes' theorem:

$$
\begin{aligned}
&p(\sigma_B^2|\text{data}) \propto p(\text{data}|\sigma_B^2) \cdot p(\sigma_B^2) \\
&\propto (\sigma_B^2)^{-M/2} \exp\left(-\frac{\sum_{m=1}^{M}(\pi_{B,m}-\mu_B)^2}{2\sigma_B^2}\right) \cdot (\sigma_B^2)^{-a_0-\frac{np_B}{2}-1} \exp\left(-\frac{b_0+\frac{np_B}{2}}{\sigma_B^2}\right) \\
&= (\sigma_B^2)^{-(a_0+\frac{np_B+M}{2})-1} \exp\left(-\frac{b_0+\frac{np_B+\sum_{m=1}^{M}(\pi_{B,m}-\mu_B)^2}{2}}{\sigma_B^2}\right)
\end{aligned}
$$

This is the kernel of an Inverse-Gamma distribution with parameters:

$$
a_{post} = a_0 + \frac{np_B + M}{2}, \quad b_{post} = b_0 + \frac{np_B + \sum_{m=1}^{M}(\pi_{B,m}-\mu_B)^2}{2}
$$

so that:

$$
\sigma_B^2|\pi_{B,1},\ldots,\pi_{B,M} \sim \text{Inv-}\Gamma\left(a_0 + \frac{np_B+M}{2}, b_0 + \frac{np_B + \sum_{m=1}^{M}(\pi_{B,m}-\mu_B)^2}{2}\right)
$$

The mean of an Inverse-Gamma(a, b) distribution is $\frac{b}{a-1}$. Therefore:

$$
\mathbb{E}[\sigma_B^2|\pi_{B,1},\ldots,\pi_{B,M}] = \frac{b_0 + \frac{np_B+\sum_{m=1}^{M}(\pi_{B,m}-\mu_B)^2}{2}}{a_0 + \frac{np_B+M}{2} - 1}
$$

Which is the formula for the bank's posterior estimate of the variance of lending to group B, given that it has observed M loans in addition to prior group-level credit history data. ◀

Proof of Lemma 6. We derive the unilateral lending thresholds: the maximum variance each bank can tolerate when lending to a single group in isolation.

Consider bank j lending only to group i, so that $\Pi_{jt} = \pi_{it}$. From the bank's perspective, $\Pi_{jt} \sim N(\mu_i, \hat{\sigma}_{it}^2)$. The VaR constraint requires:

$$
P(\Pi_{jt} < \rho) \le \alpha \quad \Longleftrightarrow \quad \Phi\left(\frac{\rho-\mu_i}{\hat{\sigma}_{it}}\right) \le \alpha \quad \Longleftrightarrow \quad \frac{\rho-\mu_i}{\hat{\sigma}_{it}} \le \Phi^{-1}(\alpha).
$$

Since $\rho - \mu_i < 0$ and $\Phi^{-1}(\alpha) < 0$ for $\alpha < 0.5$, dividing by $\Phi^{-1}(\alpha)$ reverses the inequality:

$$
\hat{\sigma}_{it} \le \frac{\rho-\mu_i}{\Phi^{-1}(\alpha)},
$$

and squaring (both sides are positive):

$$
\hat{\sigma}_{it}^2 \le \left(\frac{\rho-\mu_i}{\Phi^{-1}(\alpha)}\right)^2.
$$

Substituting the parameters for each bank and using $\mu_W = \mu_B = \mu$ (Assumption 1):

Bank L (interest rate normalised to 1):

$$
\tilde{\sigma}_L^2 = \left(\frac{\rho-\mu}{\Phi^{-1}(\alpha)}\right)^2.
$$

Bank H (interest rate $1+\nu_t$, so expected payoff $(1+\nu_t)\mu$ from a single group):

$$
\tilde{\sigma}_H^2 = \left(\frac{\rho-(1+\nu_t)\mu}{\Phi^{-1}(\alpha)}\right)^2.
$$

Since $\nu_t > 0$, we have $(1+\nu_t)\mu > \mu$, so $|\rho - (1+\nu_t)\mu| > |\rho - \mu|$, and hence $\tilde{\sigma}_H^2 > \tilde{\sigma}_L^2$. ◀

Proof of Lemma 7. We derive the pooled variance threshold for Bank L.

Suppose Bank L lends to both groups. Its profit is $\Pi_{Lt} = \pi_{Wt} + \pi_{Bt}$. Since the two payoffs are independent, the bank's belief about profit is:

$$\Pi_{Lt} \sim N\left(\mu_W + \mu_B,\ \sigma_W^2 + \hat{\sigma}_{Bt}^2\right).$$

First, note that lending to both groups is profitable in expectation, since $\mathbb{E}[\Pi_{Lt}] = \mu_W + \mu_B > 0$ by Assumption 1.
The VaR constraint requires:

$$P(\Pi_{Lt} < \rho) \leq \alpha \quad \iff \quad \Phi\left(\frac{\rho - (\mu_W + \mu_B)}{\sqrt{\sigma_W^2 + \hat{\sigma}_{Bt}^2}}\right) \leq \alpha.$$

Applying Φ^{-1} to both sides:

$$\frac{\rho - (\mu_W + \mu_B)}{\sqrt{\sigma_W^2 + \hat{\sigma}_{Bt}^2}} \leq \Phi^{-1}(\alpha).$$

Multiplying by $\sqrt{\sigma_W^2 + \hat{\sigma}_{Bt}^2} > 0$:

$$\rho - (\mu_W + \mu_B) \leq \Phi^{-1}(\alpha)\sqrt{\sigma_W^2 + \hat{\sigma}_{Bt}^2}.$$

Dividing by $\Phi^{-1}(\alpha) < 0$ (which reverses the inequality) and squaring:

$$\sigma_W^2 + \hat{\sigma}_{Bt}^2 \leq \left(\frac{\rho - (\mu_W + \mu_B)}{\Phi^{-1}(\alpha)}\right)^2.$$

Isolating $\hat{\sigma}_{Bt}^2$:

$$\hat{\sigma}_{BLt}^2 \leq \underbrace{\left(\frac{\rho - (\mu_W + \mu_B)}{\Phi^{-1}(\alpha)}\right)^2 - \sigma_W^2}_{\equiv\ \tilde{\sigma}_L^{2,\text{pool}}}.$$

Next, we verify that Bank L always lends to group W unilaterally. By Assumption 2, σ_W^2 is common knowledge. The unilateral threshold from Lemma 6 requires $\sigma_W^2 \leq \tilde{\sigma}_L^2$, which holds by Assumption 10.

Finally, since the pooled threshold $\tilde{\sigma}_L^{2,\text{pool}}$ governs whether Bank L additionally lends to group B, and Assumption 11 states that $\hat{\sigma}_{BL,0}^2 > \tilde{\sigma}_L^{2,\text{pool}}$, Bank L lends to group W alone when its prior on group B's variance exceeds this threshold. ◀

Proof of Lemma 8. The argument mirrors the proof of Lemma 7, with Bank H's payoff replacing Bank L's.

When Bank H lends to both groups, its expected payoff is $(1 + \nu_t)(\mu_W + \mu_B)$, while the variance of its payoff (from the bank's perspective) remains $\sigma_W^2 + \hat{\sigma}_{Bt}^2$. The VaR constraint is:

$$\Phi\left(\frac{\rho - (1 + \nu_t)(\mu_W + \mu_B)}{\sqrt{\sigma_W^2 + \hat{\sigma}_{Bt}^2}}\right) \leq \alpha.$$

Following the same rearrangement as in the proof of Lemma 7:

$$\hat{\sigma}_{BHt}^2 \leq \underbrace{\left(\frac{\rho - (1 + \nu_t)(\mu_W + \mu_B)}{\Phi^{-1}(\alpha)}\right)^2 - \sigma_W^2}_{\equiv\ \tilde{\sigma}_H^{2,\text{pool}}}.$$

Since $(1+\nu_t)(\mu_W+\mu_B) > \mu_W+\mu_B$ for $\nu_t > 0$, and both $\rho-(1+\nu_t)(\mu_W+\mu_B)$ and $\rho-(\mu_W+\mu_B)$ are negative, we have

$$|\rho-(1+\nu_t)(\mu_W+\mu_B)| > |\rho-(\mu_W+\mu_B)|\,,$$

and hence $\tilde{\sigma}_H^{2,\text{pool}} > \tilde{\sigma}_L^{2,\text{pool}}$: Bank H's higher interest rate insulates it against downside risk, allowing it to tolerate a larger prior variance for group B. ◀

Proof of Corollary 9. The inequality $\tilde{\sigma}_L^2 < \tilde{\sigma}_H^2$ follows from the proof of Lemma 6, and $\tilde{\sigma}_L^{2,\text{pool}} < \tilde{\sigma}_H^{2,\text{pool}}$ follows from the proof of Lemma 8. Both require $\nu_t > 0$ and $\Phi^{-1}(\alpha) < 0$. ◀

Proof of Lemma 16. This follows by application of the Law of Large Numbers. ◀

Proof of Lemma 17. In each period, Bank L has belief, $\hat{\pi}_{Bt} \sim \mathcal{N}\big(\mu, \hat{\sigma}_{BLt}^2\big)$, so that:

$$\mathbb{P}(\Pi_{jt} + s < \rho) = \Phi\left(\frac{\rho-(\mu_W+\mu_B)-s}{\sqrt{\sigma_W^2+\hat{\sigma}_{BLt}^2}}\right).$$

Thus, to guarantee $\Phi\left(\frac{\rho-(\mu_W+\mu_B)-s}{\sqrt{\sigma_W^2+\hat{\sigma}_{BLt}^2}}\right) \le \alpha$, we require

$$\frac{\rho-(\mu_W+\mu_B)-s}{\sqrt{\sigma_W^2+\hat{\sigma}_{BLt}^2}} \le \Phi^{-1}(\alpha),$$

which rearranges to

$$s \ge \rho-(\mu_W+\mu_B)-\Phi^{-1}(\alpha)\sqrt{\sigma_W^2+\hat{\sigma}_{BLt}^2}.$$

The minimal such subsidy is therefore:

$$s_t^* = \max\Big\{0,\ \rho-(\mu_W+\mu_B)-\Phi^{-1}(\alpha)\sqrt{\sigma_W^2+\hat{\sigma}_{BLt}^2}\Big\}.$$ ◀

Proof of Theorem 18. We prove the theorem in three steps.

First, by design the subsidy $s^{*(t)}$ ensures that when Bank L lends to group B, the bank's VaR constraint is satisfied, that is,

$$\mathbb{P}\Big(\Pi_{jt} + s_t^* < \rho\Big) \le \alpha.$$

Thus, the presence of s_t^* makes lending feasible even under Bank L's initial risk assessment $\hat{\sigma}_{BL0}^2 > \tilde{\sigma}_L^{2,\text{pool}}$. Consequently, if the expected net profit (inclusive of the subsidy) is positive, Bank L has an incentive to lend to group B.

Second, every time Bank L extends a loan to group B, it observes a repayment drawn from $\mathcal{N}(\mu, \sigma_B^2)$. Let m be the number of such observations. By the Strong Law of Large Numbers,

$$\hat{\sigma}_{BLm}^2 \to \sigma_B^2 \quad \text{almost surely as } m \to \infty.$$

Because $\sigma_B^2 < \tilde{\sigma}_L^{2,\text{pool}}$, there exists a (finite) index m^* (and thus a finite time τ) such that for all $m \ge m^*$ the updated estimate satisfies:

$$\sigma_B^2 \le \hat{\sigma}_{BLm}^2 < \tilde{\sigma}_L^{2,\text{pool}}.$$

Third, once the updated variance estimate satisfies $\hat{\sigma}^2_{BLt} < \tilde{\sigma}^{2,\text{pool}}_L$, we can evaluate the required subsidy as

$$s^{*(t)} = \max\left\{0, \rho - (\mu_W + \mu_B) - \Phi^{-1}(\alpha)\sqrt{\sigma^2_W + \hat{\sigma}^2_{BLt}}\right\} = 0,$$

by the same calculation as in Lemma 17. Thus, for all $t \geq \tau$, Bank L is able to satisfy its VaR constraint without any subsidy. Since the number of observations required is finite almost surely, we conclude that with probability one there exists a finite τ such that for all $t \geq \tau$, $s^*_t = 0$ and Bank L continues lending to group B without further external support.

This completes the proof. ◀

A.1 Additional Results

A.1.1 Heterogeneous Banks

We now suppose that, instead of two banks, there is a continuum of bank types $\nu_j \sim F(\nu)$ on $[0, \bar{\nu}]$. Each bank j charges rate $1 + \nu_j$, and has the same risk tolerance parameters (α, ρ). Following the logic of Lemma 7, a bank of type ν_j can lend to both groups if and only if:

$$\hat{\sigma}^2_B \leq \tilde{\sigma}^2(\nu_j) \equiv \left(\frac{\rho - (1 + \nu_j)(\mu_W + \mu_B)}{\Phi^{-1}(\alpha)}\right)^2 - \sigma^2_W$$

It is straightforward to observe that $\tilde{\sigma}^2(\nu_j)$ is increasing in ν_j. Banks with higher premia have higher variance thresholds: they can lend to groups with higher prior variances.
Given prior belief $\hat{\sigma}^2_B$, define the cutoff type $\tilde{\nu}^*$ as:

$$\tilde{\nu}^* = \inf\left\{\nu : \tilde{\sigma}^2(\nu) \geq \hat{\sigma}^2_B\right\}$$

That is, $\tilde{\nu}^*$ is the *lowest*-cost bank type that can lend to group B. We can think of this as the *marginal high-cost lender.* If $\tilde{\nu}^* > \bar{\nu}$ (the upper bound of the type distribution), no bank lends to group B. If $\bar{\nu}^* \leq 0$, all banks can lend to group B at the mainstream rate. We are interested in the intermediate case:

▶ **Theorem 20** (Sorting with Heterogeneous Banks)**.** *Suppose $\tilde{\nu}^* \in (0, \bar{\nu})$. Then: Banks with $\nu_j \geq \tilde{\nu}^*$ lend to group B. The equilibrium interest rate for group B is $1 + \tilde{\nu}^*$, set by the marginal bank. Group W receives loans from all banks at the competitive rate, since σ^2_W is commonly known and satisfies the VaR constraint for all bank types.*

Proof. The VaR constraint for each bank j is satisfied for group B if and only if $\tilde{\sigma}^2(\nu_j) \geq \hat{\sigma}^2_B$, which holds if and only if $\nu_j \geq \tilde{\nu}^*$.

Among banks that lend to group B, the lowest-type bank charges $1 + \tilde{\nu}^*$. Under price competition, this bank sets the equilibrium rate: any bank with $\nu_j > \tilde{\nu}^*$ charges a higher rate and loses group B to the marginal bank.

By Assumption 10, $\sigma^2_W < \tilde{\sigma}^2_L = \tilde{\sigma}^2(0)$. Since $\tilde{\sigma}^2$ is increasing, all bank types satisfy VaR for group W. The lowest-type bank ($\nu_j = 0$) sets the competitive rate at 1. ◀

With heterogeneous banks, the interest rates paid by minority borrowers are lower than in the two-bank model: competition reduces interest rates paid by minority applicants, as $\tilde{\nu}^* < \bar{\nu}$. However, the Subprime Trap is still in effect: the mainstream bank does not lend to minority applicants, and interest rates are above those in the full information case. Since the mainstream bank does not lend to minority applicants, it does not update its beliefs. Subsidies break the trap in a similar way: they incentivize the mainstream bank to learn the true variance of group B, which makes the mainstream bank the lowest cost bank, and prices out the higher-cost banks.

A.1.2 Applicant Skimming

In practice, banks do observe individual credit information, and can make lending decisions based on this. This is discrimination on the intensive margin: which members of a group to lend to. In this extension, we show that we can recreate the logic of the Subprime Trap in this setting: even if banks lend to individuals with high-quality signals, differential missingness means that there is a still a subgroup of minority applicants to whom the mainstream bank does not lend.

We now suppose that banks observe whether each group B applicant has a complete or incomplete credit file, and can condition their lending decisions on this information. Group B consists of a fraction p_B with complete files and a fraction $1 - p_B$ with incomplete files. We denote the sub-population parameters by (μ_c, σ_c^2) and $(\mu_{ic}, \sigma_{ic}^2)$ respectively. Parallel to the argument made above, we suppose that $\mu_{ic} = \mu_c$ (complete and incomplete file applicants are equally creditworthy).

In the main model, the bank's prior on σ_B^2 is based on np_B complete historical files. These files are drawn from the complete-file sub-population. When the bank treats group B as a single population, it uses these np_B observations to estimate a variance for the whole group, producing an inflated posterior. When the bank conditions on file completeness, the same np_B observations are used to estimate σ_c^2. For the incomplete-file sub-population, the bank has no informative observations: its prior on σ_{ic}^2 is the uninformative prior (a_0, b_0).

We suppose that the bank has learned σ_c^2 from the complete-file sample, and that $\sigma_c^2 < \tilde{\sigma}_L^2$, so that the bank lends to group W and to complete-file group B applicants. The bank then evaluates whether to also lend to incomplete-file group B applicants.
Following the logic of Lemma 7, the bank lends to incomplete-file applicants if and only if:

$$\hat{\sigma}_{ic}^2 \leq \tilde{\sigma}_L^{2,\text{pool},ic} \equiv \left(\frac{\rho - (\mu_W + \mu_c + \mu_{ic})}{\Phi^{-1}(\alpha)} \right)^2 - \sigma_W^2 - \sigma_c^2$$

This threshold is more permissive than the unilateral threshold for σ_{ic}^2 alone, since the positive expected returns from group W and complete-file group B lending provide a buffer against downside risk. The bank can tolerate more uncertainty about σ_{ic}^2 than it could in isolation. Nonetheless, the threshold is finite, and the bank's uninformative prior may exceed it.

▶ **Theorem 21** (Equilibrium in Applicant Skimming). *Suppose that (i) $\sigma_c^2 < \tilde{\sigma}_L^2$, so that the VaR constraint is satisfied for complete-file group B applicants, and (ii) $\hat{\sigma}_{ic,0}^2 = b_0/(a_0 - 1) > \tilde{\sigma}_L^{2,pool,ic}$, so that the VaR constraint is violated for incomplete-file group B applicants even under the relaxed pooled threshold. Then Bank L lends to group W and to complete-file group B applicants, but not to incomplete-file group B applicants. The Subprime Trap of Theorem 13 holds for the incomplete-file sub-population.*

Proof. Partition group B into complete-file and incomplete-file sub-populations. For complete-file applicants, the bank's prior on σ_c^2 is based on np_B observations and the true variance satisfies $\sigma_c^2 < \tilde{\sigma}_L^2$ by assumption, so Bank L's VaR constraint is satisfied and it lends. For incomplete-file applicants, the bank's prior on σ_{ic}^2 is (a_0, b_0) with posterior mean $\hat{\sigma}_{ic,0}^2 > \tilde{\sigma}_L^{2,\text{pool},ic}$. The conditions of Theorem 13 are satisfied for this sub-population: Bank L does not lend, does not observe outcomes, and does not update its beliefs. By the same inductive argument, Bank L never lends to incomplete-file group B applicants. Bank H serves this sub-population at rate $1 + \bar{\nu}$. ◀

This is the equilibrium in applicant skimming. Bank L lends to minority applicants whose files are complete, and learns σ_c^2 from their repayment outcomes. Incomplete-file applicants are served only by Bank H. The differential missingness of Section 3.2.4 sustains the trap: if the bank observed complete information for all group B applicants, it could estimate σ_B^2 directly, and the trap would not arise.

However, the population variance σ_B^2 is not identified from the complete-file sample alone. The bank can learn σ_c^2, but σ_{ic}^2 is unobserved. Since $\sigma_B^2 = p_B\sigma_c^2 + (1-p_B)\sigma_{ic}^2$, the bank cannot determine whether the full population satisfies the VaR constraint. Subsidies targeted at incomplete-file applicants break the trap for this sub-population, by the same mechanism as Theorem 18.

The moral is that, even with selection on the intensive margin, where the bank chooses which applicants to lend to, the Subprime Trap still obtains: there remains a subgroup of minority applicants to whom the mainstream bank does not lend.

A.1.3 Alternative Guarantees

We extend Theorem 18 by demonstrating that the conclusion holds under a broader class of temporary risk-sharing mechanisms. Specifically, we show that the subsidy mechanism described in the theorem is not the only way to induce Bank L to lend to group B and break the subprime trap. Any guarantee mechanism that ensures Bank L satisfies its VaR constraint during an exploration phase will suffice, provided it allows the bank to accumulate sufficient repayment data to update its risk estimate. We formalize this result as the following corollary.

▶ **Corollary 22** (Robustness to Alternative Guarantees)**.** *The conclusion of Theorem 18 holds for any temporary risk-sharing mechanism that satisfies the following condition: for each period t of the exploration phase, the mechanism ensures that*

$$\Pr(\Pi_{jt} + G(t) < \rho) \leq \alpha,$$

where $G(t)$ is the guarantee provided in period t. After sufficient data collection, the updated variance estimate $\hat{\sigma}_{BLt}^2$ will satisfy $\hat{\sigma}_{BLt}^2 < \tilde{\sigma}_L^{2,pool}$, allowing Bank L to lend without further guarantees.

Proof. For any such $G(t)$, the VaR constraint is satisfied during each period t of the exploration phase. Specifically, Bank L is guaranteed that its effective return, $\pi_B^{(t)} + G(t)$, will not fall below ρ with probability exceeding α. This ensures that the bank has an incentive to lend to group B, provided the expected return is positive. The guarantee mechanism defined in the corollary generalizes the subsidy mechanism from Theorem 18. In the original subsidy framework, the guarantee function is given explicitly by $G(t) = s_t^* = \max\{0, \rho - (\mu_W + \mu_B) - \Phi^{-1}(\alpha)\sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2}\}$, which satisfies the condition $\Pr(\hat{\Pi}_{jt} + s_t^* < \rho) \leq \alpha$. The corollary allows for any $G(t)$ that satisfies the same probabilistic constraint. Once Bank L updates its risk estimate to $\hat{\sigma}_{BLt}^2 < \tilde{\sigma}_L^{2,\text{pool}}$, the VaR constraint is naturally satisfied without external support. Specifically, the guarantee $G(t)$ is no longer required, as the bank can safely lend to group B on its own. Formally, this follows from Lemma 7.

Thus, any guarantee mechanism that ensures the bank's effective return satisfies the VaR condition during a finite exploration phase will induce Bank L to lend, learn group B's true risk, and ultimately eliminate the need for the guarantee. This completes the proof. ◀

A.2 Extension to Expected Shortfall

A.2.1 Variance Thresholds under Expected Shortfall

We now derive the variance threshold under Expected Shortfall and show it is more conservative than the VaR constraint. The logic of the argument is the same throughout, however. When Bank L lends to both groups W and B:

$$\Pi_L = \pi_W + \pi_B$$

where from the bank's perspective (using its estimate $\hat{\sigma}_B^2$):

$$\Pi_L \sim \mathcal{N}(\mu_W + \mu_B, \sigma_W^2 + \hat{\sigma}_B^2)$$

$$\text{VaR}_\phi(\Pi_L) = (\mu_W + \mu_B) + \sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2} \cdot \Phi^{-1}(\phi)$$

Using the definition $\text{ES}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \text{VaR}_\phi(X)d\phi$:

$$\begin{aligned}\text{ES}_\alpha(\Pi_L) &= \frac{1}{\alpha}\int_0^\alpha \left[(\mu_W + \mu_B) + \sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2} \cdot \Phi^{-1}(\phi)\right] d\phi \\ &= (\mu_W + \mu_B) + \frac{\sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2}}{\alpha}\int_0^\alpha \Phi^{-1}(\phi)d\phi\end{aligned}$$

Using the fact that $\int_0^\alpha \Phi^{-1}(\phi)d\phi = -\phi(\Phi^{-1}(\alpha))$:

$$\text{ES}_\alpha(\Pi_L) = (\mu_W + \mu_B) - \sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2} \cdot \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}$$

Bank L requires $\text{ES}_\alpha(\Pi_L) \geq \rho$:

$$(\mu_W + \mu_B) - \sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2} \cdot \frac{\phi(\Phi^{-1}(\alpha))}{\alpha} \geq \rho$$

Rearranging for $\hat{\sigma}_B^2$:

$$\sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2} \leq \frac{(\mu_W + \mu_B) - \rho}{\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}}$$

Squaring both sides:

$$\sigma_W^2 + \hat{\sigma}_B^2 \leq \left(\frac{(\mu_W + \mu_B) - \rho}{\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}}\right)^2$$

Therefore:

$$\hat{\sigma}_B^2 \leq \left(\frac{(\mu_W + \mu_B) - \rho}{\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}}\right)^2 - \sigma_W^2 \equiv \tilde{\sigma}_L^{2,\text{pool},ES}$$

From Lemma 7, the VaR threshold is:

$$\tilde{\sigma}_L^{2,\text{pool,VaR}} = \left(\frac{\rho - (\mu_W + \mu_B)}{\Phi^{-1}(\alpha)}\right)^2 - \sigma_W^2$$

We next show that $\tilde{\sigma}_L^{2,\text{pool},ES} < \tilde{\sigma}_L^{2,\text{pool,VaR}}$, and hence, that the ES threshold is more restrictive.

Since $\frac{\phi(\Phi^{-1}(\alpha))}{\alpha} > \Phi^{-1}(\alpha)$, the denominator in the ES threshold calculation is larger, yielding:

$$\tilde{\sigma}_L^{2,\text{pool},ES} < \tilde{\sigma}_L^{2,\text{pool},\text{VaR}}$$

This shows that Expected Shortfall creates a more conservative variance threshold than VaR, making it harder for Bank L to lend to group B initially.

We can solve for the optimal subsidy under Expected Shortfall analogously:

A.2.2 Optimal Subsidy under Expected Shortfall

In order to lend to group B, Bank L requires:

$$\text{ES}_\alpha(\Pi_L + s) \geq \rho$$

where Expected Shortfall with subsidy s is:

$$\text{ES}_\alpha(\Pi_{Lt} + s) = (\mu_W + \mu_B + s) - \sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2} \cdot \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}$$

Rearranging for s:

$$s \geq \rho - (\mu_W + \mu_B) + \sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2} \cdot \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}$$

The optimal (minimum) subsidy is:

$$s_{ES,t}^* = \max\left\{0, \rho - (\mu_W + \mu_B) + \sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2} \cdot \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}\right\}$$

The VaR-derived subsidy from Lemma 17 is:

$$s_{VaR,t}^* = \max\left\{0, \rho - (\mu_W + \mu_B) - \Phi^{-1}(\alpha)\sqrt{\sigma_W^2 + \hat{\sigma}_{BLt}^2}\right\}$$

Since $\frac{\phi(\Phi^{-1}(\alpha))}{\alpha} > \Phi^{-1}(\alpha)$ for $\alpha < .1$, (e.g. $2.063 > 1.645$ when $\alpha = 0.05$):

$$s_{ES,t}^* > s_{VaR,t}^*$$

The Expected Shortfall-based subsidy is larger, reflecting the fact that Expected Shortfall is a more conservative risk criterion than Value-at-Risk.