OVERRIDING THE ALGORITHM: GENDER, COMPLIANCE, AND RISK MANAGEMENT*

Jose G. Montalvo, UPF-BarcelonaGSE-IPEG
Marta Reynal-Querol, ICREA-UPF-BarcelonaGSE-IPEG

Abstract

We examine gender differences in compliance with AI-based lending recommendations and propose a novel mechanism that can explain rational gender differences in algorithm-informed decisions. Loan screening increasingly relies on objective, algorithmic measures. Using detailed bank data, we analyze loan officers' compliance by gender. We find that women comply more with algorithmic recommendations than men, which leads to substantially lower default rates on their portfolios. We show that female loan officers have nearly a 15% lower delinquency rate than male officers. We identify a novel mechanism that can explain managerial gender differences in the context of algorithm-informed decisions. This pattern is consistent with gender bias operating through a mistake–punishment tradeoff, which may partly explain women's higher compliance with AI recommendations.

Keywords: *Gender, compliance, delinquency, credit risk*

JEL Classification numbers: G21, J16

*Corresponding author: Jose G. Montalvo, Universitat Pompeu Fabra, C/ Ramon Trias Fargas 25-27, Barcelona, 08005, Spain (email: jose.garcia-montalvo@upf.edu, phone: (34) 687869541). Marta Reynal-Querol, ICREA - Universitat Pompeu Fabra, C/ Ramon Trias Fargas 25-27, Barcelona, 08005, Spain (email: marta.reynal@upf.edu). We acknowledge the excellent research assistance provided by Ece Yagman and Daniele Alimonti. We are especially grateful to the financial companies that shared with us the data for this study, and to its members who kindly answered our questions about the data and provided qualitative insights. Financial support from the Spanish Department of Science and Innovation project PID2023-150768NB-I00 and the Government of Catalonia (ICREA) is gratefully acknowledged. We are also thankful for the financial support provided by the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2022).

I Introduction

The integration of artificial intelligence and machine learning algorithms into organizational decision-making represents one of the most consequential economic transformations of the early twenty-first century. From medical diagnostics to warehouse operations, from judicial sentencing to financial lending, algorithmic systems increasingly augment, replace, or constrain human judgment in contexts where decisions carry substantial economic and social consequences. The speed and breadth of AI adoption are extraordinary. By late 2024, nearly 40% of the US population aged 18–64 will use generative AI, with 23% of employed respondents having used it for work at least once in the previous week and 9% using it every workday (Bick et al., 2025). Relative to each technology's first mass-market product launch, work adoption of generative AI has been as fast as the personal computer and faster than the internet overall. These tools promise substantial productivity gains: randomized control trials document significant productivity improvements (Dell'Acqua et al., 2023; Noy & Zhang, 2023), and recent macroeconomic modeling suggests that current levels of adoption of generative AI could increase TFP by 0.5–0.7% over a decade (Bick et al., 2025; Acemoglu, 2025).

Yet, this rapid AI transformation reveals fundamental tensions inherent in human–machine collaboration. Research on human–algorithm interaction has documented systematic patterns of deviation from algorithmic prescriptions, even in contexts where algorithms demonstrably outperform human judgment. Dietvorst et al. (2015) shows experimentally that participants continue to override algorithmic recommendations even after observing that algorithms outperform human judgment. This phenomenon, termed "algorithm aversion," has been observed across multiple contexts. In Alibaba Group warehouse operations, Sun et al. (2021) find that packers deviate from algorithmic prescriptions for 5.8% of packages. Interestingly, these deviations increase the packing time by 48.3% and substantially reduce operational efficiency. Similarly, Lebovitz et al. (2022) document how radiologists in US hospitals abandoned machine learning algorithms after using them for several months, suggesting that algorithm aversion persists even among trained professionals in high-stakes medical decision-making.

However, the phenomenon of human overriding algorithmic recommendations does not have to be uniformly detrimental to organizational performance. Sun et al. (2021) propose a fundamental distinction: deviations from algorithmic prescriptions can be beneficial or harmful depending on their source. Information deviations occur when humans may possess objective information that is not available to the algorithm and use this information to execute more feasible or optimal actions than the algorithm recommends. For example, warehouse workers may observe that certain items can be folded or compressed and therefore fit in smaller boxes than algorithms predict, or they may know that items require protective wrapping that the algorithm failed to account for. These information deviations are typically beneficial to operational performance. In contrast, complexity deviations stem from workers' aversion, inability, or discretion to precisely implement algorithmic prescriptions that are actually optimal. When algorithmic recommendations become too complex to understand or execute, workers resort to simpler but suboptimal decisions. Sun et al. (2021) find that complexity deviations typically affect operational performance.

The organizational context fundamentally shapes whether deviations improve or harm outcomes. Nikpayam et al. (2024) shows that delegation itself creates systematic incentives for over-reliance on algorithms. In their laboratory experiments, when a manager delegates AI-assisted decisions to an employee, information asymmetries create conditions where the manager cannot directly observe the employee's private information. The manager must therefore infer decision quality from observable information: the machine's recommendation and the ultimate outcome. Because the manager cannot verify whether an employee's decision was optimal given that employee's private information, the manager systematically over-weights observable machine recommendations when evaluating the employee's performance. This creates rational incentives for employees to adhere closely to algorithmic recommendations even when their own private judgment contradicts the algorithm. The experimental evidence is striking: delegation increases algorithmic over-reliance, and this over-reliance persists and intensifies over time as employees learn that managers apply asymmetric outcome bias: penalizing employees whose algorithmic overrides lead to bad outcomes while failing to reward employees whose overrides lead to good outcomes (Nikpayam et al., 2024). This organizational dynamic implies that human–algorithm interaction cannot be understood as a purely technical problem of algorithm design or as an individual-level problem of algorithm aversion. Rather, it is fundamentally an organizational and institutional problem shaped by how managers evaluate and reward employee decisions.

An emerging yet remarkably underdeveloped dimension of human–algorithm interaction concerns systematic gender differences in AI adoption, algorithm trust, and the propensity to override algorithmic recommendations. As AI systems proliferate, gender gaps in adoption and compliance threaten to generate new forms of economic inequality and perpetuate existing biases.

Research documents substantial gender differences in how individuals perceive and interact with algorithms. Women express significantly lower trust in AI systems and report higher levels of concern about fairness and bias in algorithmic decision-making (Otis et al., 2024). Experimental evidence shows that, in general, women perceive algorithms as more discriminatory than men, even when algorithms are demonstrably unbiased (Utz, 2024), are less willing to delegate decisions to algorithms and are more likely to override algorithmic recommendations, particularly in high-stakes contexts (Dhar et al., 2025). Critically, Dhar et al. (2025) show that women react more strongly to algorithmic mistakes, with greater reductions in trust following algorithm failures compared to men. These gender differences in algorithm trust and overriding propensity raise a paradox. On one hand, if algorithms systematically encode biases—particularly in domains like finance and hiring where historical discrimination is well-documented—then women's greater algorithmic caution may represent superior judgment and protection against discrimination. On the other hand, if women's lower algorithm compliance reflects algorithm aversion or systematic distrust that exceeds the algorithmic error rate, it could leave women missing out on productivity gains and better decisions that algorithms provide. Tang et al. (2025) find that after the emergence of ChatGPT, the increase in productivity of male researchers is 6.4% higher than that of female researchers. Understanding the sources of gendered algorithmic compliance is therefore essential both for equity (ensuring women are not disadvantaged by biased algorithms) and for efficiency (ensuring organizations realize productivity gains from AI).

However, a critical but underexplored mechanism linking gender to algorithmic compliance involves differential punishment for errors depending on gender. Although algorithm aversion and algorithm trust are typically understood as individual-level psychological phenomena, research on organizational behavior reveals that gender shapes how mistakes are evaluated and punished, which in turn creates differential incentives for algorithmic compliance.

Sarsons (2022) shows that primary care physicians update their beliefs about surgeon quality differently according to surgeon gender. When a patient dies after surgery, referrals to female surgeons drop significantly more than referrals to male surgeons, even controlling for the frequency of adverse outcomes. This suggests that female physicians have "fewer chances to make mistakes" than male surgeons—a single mistake can substantially damage a woman's career more than an equivalent mistake damages a man's career. In the context of financial services, Egan et al. (2022) find that, following misconduct incidents, female financial advisors are more likely than male advisors to lose their jobs and spend substantially more time unemployed looking for new positions. The punishment differential persists even after controlling for the severity of misconduct, suggesting that gender shapes how organizational actors evaluate and penalize employee mistakes.

These findings suggest that the observed gender differences in algorithmic compliance may not reflect fundamental differences in risk aversion, overconfidence, or algorithm trust. Rather, they may reflect rational responses to asymmetric accountability structures. If women face harsher consequences for mistakes than men do – for a given level of performance – then women have stronger incentives to follow conservative decision-making strategies that minimize the risk of bad outcomes. Algorithmic recommendations represent such a conservative strategy: following an algorithm's recommendation shifts responsibility for the decision to the algorithm, which can serve as a defense if outcomes are poor ("I followed the algorithm's recommendation"). Overriding the algorithm, by contrast, places full responsibility on the decision-maker's judgment, exposing that decision-maker to accountability if outcomes disappoint.

In organizational contexts where women face differential punishment for mistakes, following algorithmic recommendations becomes a rational strategy for risk management and career preservation. Women's higher compliance with algorithmic recommendations thus reflects not inferior judgment but rather optimization given the accountability structure they face. Conversely, if men face more forgiving accountability structures –where mistakes are more readily excused or attributed to external factors—they have greater freedom to exercise discretion and override algorithmic recommendations when they believe their judgment is superior.

We investigate these questions empirically using an exceptionally detailed dataset from a large Spanish bank's credit screening operations that used AI procedures to screen loans. The financial sector is a particularly well suited industry to analyze the interaction between AI and humans since it adopted machine learning algorithms long ago. The institutional setting provides several critical advantages for studying human–algorithm interaction and gender differences in algorithmic compliance. First, the loan screening process relied on an algorithmic system that provided both a quantitative risk score and an explicit categorical recommendation for each loan application. Loan officers could "very exceptionally" override these recommendations by providing written justifications, creating a natural setting to study discretionary deviations from algorithmic prescriptions. Second, we observe not only whether loan officers followed algorithmic recommendations, but also actual ex post loan performance, allowing us to evaluate whether overriding recommendations improved or harmed outcomes. Third, loan officers operated under intense short-term incentives to generate loan volume (bonuses tied 60% to loan production) while facing long-term career consequences from high delinquency rates, creating exactly the tension between short-term compliance incentives and long-term accountability that organizational theory and the delegation framework of Nikpayam et al. (2024) predict will shape algorithmic compliance.

We document three main findings. First, female loan officers demonstrate significantly higher compliance with algorithmic recommendations than male officers. Since, as we argued before, the literature shows that women show significantly lower trust in AI systems, this result should be related to a strong incentive to follow the algorithmic recommendations in this particular case. For applications where the algorithm recommends rejection (negative or very negative recommendations), male loan officers approve 13 percentage points more frequently than female loan officers. This substantial differential in override propensity cannot be explained by differences in the risk characteristics of applications screened by male versus female officers – the distribution of risk scores and algorithmic recommendations is virtually identical across genders. The difference in compliance reflects behavioral choices, not selection differences.

Second, this gender difference in algorithmic compliance translates into economically significant differences in loan performance. Loans screened by female officers exhibit delinquency rates 1.5 to 2.5 percentage points lower than loans screened by male officers, even after controlling for risk scores and detailed applicant characteristics. Given that the sample average delinquency rate is 12%, this represents a 12–20% relative improvement in loan performance. This performance advantage is robust to multiple specifications, alternative outcome measures (delinquency at 30, 60, and 90 days), alternative loan types (mortgages vs. consumer loans), and inclusion of various controls and fixed effects.

Third, we provide evidence that gender bias in punishment could explain the differential algorithmic compliance. Using survival analysis, we find that loan officers who accumulate high proportions of non-performing loans face differential career consequences by gender. At 4% accumulated bad loans, men have a substantially lower probability of being demoted or dismissed compared to women with identical performance records. This asymmetric punishment creates rational incentives for women to comply more strictly with algorithmic recommendations –particularly when overriding the system to approve risky loans that might later default.

These findings have implications beyond the credit-risk context. As organizations increasingly embed AI recommendation systems into management workflows, the aggregate benefit from those systems depends not only on their predictive accuracy but also on the degree to which managers comply with their output. Our results suggest that compliance is not gender-neutral: institutional structures that penalize women more severely for errors may, as a byproduct, induce a higher degree of AI-alignment among female managers. Conversely, in settings where the penalty gradient is equalized—as in our data after 2008, when the bank prohibited recommendation overrides entirely—gender differences in performance disappear. This "compliance channel" constitutes a novel mechanism linking gender, institutional incentives, and the organizational returns to algorithmic decision-support tools.

Our paper contributes to several strands of the literature. *First*, we contribute to the rapidly growing literature on human–algorithm interaction. Previous work has established that managers who deviate more from algorithmic recommendations typically produce worse outcomes (Hoffman & Li, 2018), but has not examined whether this tendency varies by gender. We provide the first field evidence of a gender gap in AI-recommendation compliance using a large-scale administrative dataset with explicit recommendation information. Although previous research has established that humans systematically deviate from algorithmic recommendations (Sun et al., 2021; Angelova et al., 2022; Dietvorst et al., 2015), we document gender as a previously unexamined dimension of algorithmic compliance behavior. Our findings complement recent evidence on gender gaps in AI adoption (Humlum & Vestergaard, 2024; Otis et al., 2024; Bick et al., 2025) by demonstrating that gender differences extend beyond the decision to adopt AI tools to encompass how individuals actually use those tools, specifically their propensity to follow or override algorithmic recommendations.

Second, we contribute to the literature on gender differences in financial decision-making and credit risk management, moving beyond differences in the treatment of risk to a new margin: the willingness to override algorithmic guidance. Previous studies have documented gender differences in the way loan officers evaluate soft information (Beck et al., 2013; Bellucci et al., 2010; Huang et al., 2024) and in risk preferences in investment decisions (Barber & Odean, 2001). However, these studies do not observe explicit algorithmic recommendations or risk scores. Our setting, where we observe both risk scores and algorithm recommendations, enables for a precise analysis of the compliance channel. We show that male and female officers receive applications with identical risk characteristics, but make systematically different decisions conditional on algorithmic recommendations. This allows us to isolate the compliance mechanism from alternative explanations. In this vein, our article relates to the literature on gender and rule compliance (Tittle, 1980; Rosenbloom, 2009), extending evidence from traffic regulations and pedestrian behavior into a high-stakes economic environment with explicit financial incentives.

Third, we extend the literature on discrimination in punishment and differential career consequences by gender. Sarsons (2022) demonstrate an asymmetric career update after mistakes in medicine, and Egan et al. (2022) document harsher punishment of female financial advisors for misconduct. We add a novel behavioral channel: differential punishment creates rational incentives for differential algorithmic compliance. We propose and document a rational explanation for this compliance differential: women face a harsher career penalty when the loans they approve go delinquent, creating stronger incentives to follow the system's guidance even when short-run volume incentives push in the opposite direction. This "mistake-punishment asymmetry" operates as an implicit tax on female discretion, shaping compliance behavior without any change in the formal rules applied to men and women. This provides a concrete mechanism through which gender discrimination in organizational accountability structures translates into observable differences in operational decisions and ultimately organizational performance.

Fourth, we contribute to the literature on the productivity effects of algorithms and artificial intelligence. Recent work by Hoffman & Li (2018) and Paravisini & Schoar (2015) analyzes how algorithms affect decision quality and organizational performance. However, this literature typically treats algorithm adoption as binary (on or off) or examines aggregate compliance rates without investigating heterogeneity in compliance by employee characteristics. We show that achieving productivity gains from algorithms depends critically on compliance patterns, which vary systematically by gender. This suggests that to optimize algorithmic systems for organizational performance, implementation strategies must account for systematic heterogeneity in how different groups of decision-makers interact with algorithms.

Fifth, our paper contributes methodologically by showing the value of combining algorithmic prescriptions with ex post performance outcomes to evaluate the economic consequences of algorithmic overrides. Unlike settings where the true state of the world remains partially unobserved (as in verification-bias settings), we observe actual loan performance. This enables a direct assessment of whether the overriding algorithmic recommendations improve or harm outcomes. We can answer the question: when male loan officers override algorithmic rejections and approve risky loans, do these overrides reflect superior human judgment (information deviations) or erroneous decisions based on bias or overconfidence (complexity deviations)? By examining subsequent loan delinquency, we find that male overriding of algorithmic rejections leads to substantially worse outcomes, suggesting that these deviations are complexity deviations that reduce organizational performance. This approach could be profitably applied to other contexts where algorithms provide recommendations and the outcomes are observable.

Sixth, our findings contribute to policy debates on AI implementation and equity. As organizations increasingly deploy AI systems, design choices about how humans interact with algorithms become critical. If women face differential accountability for mistakes, creating organizational incentives for greater algorithmic compliance among women, this could simultaneously harm productivity (by preventing valuable human overrides) and perpetuate biases embedded in algorithms (as women's compliance ensures biased algorithms' recommendations are followed). Our results suggest that realizing the productivity potential of AI while simultaneously reducing algorithmic bias requires attention not only to algorithm design but to the organizational structures and incentive systems within which humans interact with algorithms.

The remainder of this paper is organized as follows. II provides institutional background on the bank's loan screening procedures and the algorithmic recommendation system. III describes the data sources, measurement of key variables, and summary statistics. IV presents our main results on gender differences in compliance and their relationship to delinquency. V provides extensive robustness checks, including alternative outcome measures, alternative specifications, and analyses by loan type. VI explores mechanisms, including detailed analysis of the mistake–punishment trade-off by gender. VIII concludes with implications for algorithmic system design, organizational implementation strategies, and policy.

II Institutional setup

We analyze the decisions of loan officers using information from a large bank in Spain. Our database includes more than 400,000 applications for loans during the period 2000-2012.[1] We consider the portfolio of mortgages and consumer loans, which jointly represented the largest share of the portfolio of loans of the bank. Our choice is based on the fact that an AI procedure provided the risk score and a recommendation for these two categories of loans. Mortgages represented 34% of the loans in our dataset.

Depending on the size of the requested loan, the decision was either made at the branch or was elevated to a specialized committee in the bank's central servicing department.We only consider applications that were decided in the branches. During the analyzed period, most households' applications for loans were initiated at a branch of the bank. One basic operating principle was the delegation of the ability to authorize different types of loans. The internal circulars issued by central servicing to the branches[2] confirm that during the period of study, the loan officers at the branches could approve mortgages of up to 350,000 euros.[3] We do not consider a small number of applications that requested amounts above the concession limits at the branches, which were sent to the bank-wide committee.[4]

The banks' branches were small and handled all types of operations. Each branch had only one loan officer, who made the decision on each loan application that met the delegation limits. There were no online or phone applications. During the analysis period, there were 1,507 loan officers and 22% were women. Since we do not have a randomized experiment, it is important to explain the assignment of loans to officers. As we already noted, there is only one loan officer per office, and therefore there is no choice at a particular branch. In the data section, we show that the distribution of the risk score of applications to male and female loan officers is practically identical, as is the distribution of applications by recommendation grade. Therefore, the risk profile of the clients of the male and female loan officers is identical.[5]

The procedure for screening the applications was as follows. The loan officer met with the clients and gathered all the information required by the system, typing it directly into the computer. After entering all the data of the applicant[6] the loan officer's screen showed information on the risk score of the application

1. We do not have repeated applications.

2. We had access to all internal communications between central servicing and the branches. They precisely describe all the procedures of the bank with respect to the functioning of branches.

3. Note that the average price of a typical house in Spain, at that time, was approximately 155,000 euros, meaning that loan officers in the branches could authorize most mortgages, and for personal loans, the limit before delegation to central servicing was 110,000 euros.

4. Paravisini & Schoar (2015) analyzes the role of credit scores in the productivity of credit committees.

5. We also show that the results are robust to the use of fixed effects in the branch.

6. we have all the information on scoring trials (any change in the original typing of the information in the application). In

and the recommendation derived from the policies of the bank.[7] In the process of screening loans, the risk score and recommendation from the system were very salient.[8]

Therefore, the loan officers knew the risk score and the recommendation on each application before making their decision, although they could "exceptionally" overrule this recommendation. Internal Circular A2-038/00 states that when the recommendation system provided a favorable recommendation (positive or very positive), the loan could automatically be granted. If the recommendation was unfavorable (negative or very negative), then the operation should be denied. However, the internal circulars add that "in exceptional cases the officer can ultimately approve the loan, explaining why she disagrees with the recommendation provided by the system."[9]

This option was frequently used by loan officers who, during the period 2002-2008, granted approximately 80% of the loans that the system recommended rejecting.[10] At the end of 2008, a new internal circular eliminated the possibility of using the exceptional option for granting loans with a rejection grade from the recommendation system. This implied a substantial reduction in the production of mortgages.

The loan officers only made discretionary judgments about whether to approve or reject an application. They had no discretion about the interest rate, the loan amount, the maturity, or the amount of each installment. In general, loan officers have a very limited role in setting prices when mortgage approvals are made by underwriting algorithms. In the Spanish banking system, the interest rate was set independently of the financial situation of the clients (risk scoring, type of contract, debt situation, etc.) (Mayordomo et al., 2019) and the banks ration credit through quantities (approval of loans) instead of prices (Bentolila et al., 2017). In the case of the bank that provided the data, the internal circulars precisely explained rate setting. The common standard rate for a mortgage, for instance, could be reduced by 0.1 points for subscribing to life insurance; 0.05 additional points for buying home insurance; 0.1 points for getting a credit card; and 0.1 points for direct payment of paychecks to the bank account. No reference is made to any influence of the risk score or any other variable on the interest rate. This is, moreover, a general feature of Spanish banks:

fact, in our case, the system saved all the keys typed by the loan officer. Loan officers were aware of the fact that their actions on the keyboard were saved for further inspection. For these reasons, we do not observe a significant amount of scoring trials.

7. Examiners did not have access to the algorithm of the recommendation system.

8. Managers and loan officers reported in personal interviews that the recommendation produced by the AI procedure was very salient.

9. However, loan officers were aware of the ability of the recommendation system to predict delinquency. Internal Circular A2-111/02 reports that after two years of application, the recommendation system had shown very good predictive power. Since then, many internal circulars insist on the excellent predictive power of the recommendation system.

10. In this sense, the Spanish banking crisis was a classical banking crisis derived from excessive risk taking, as was the case in many other countries.

interest rates were insensitive to loan characteristics (risk scores, LTV, etc.) and clients' financial situation (type of contract, employment situation, etc.) in the segment of retail banking clients.

Incentives are an important part of the story. The compensation of the loan officers included a fixed salary plus a bonus. The bonus scheme provided a strong incentive to approve loans. The variable pay was a function of several indicators and could reach 20% of the fixed pay if the objective was reached 100% in all indicators. The indicators related to the number of loans approved by an officer amounted to 60% of the variable pay.[11] There was no variable pay if the indicator did not reach at least 50% of the objective. The important fact is that male and female loan officers had the same set of incentives. Therefore, there was a strong incentive in the short run to generate volume, and this was the case for male and female loan officers. In fact, women had less than perfect compliance with the rule, but in any case complied more often than men.

Therefore, loan officers were compensated on the basis of short-term performance, mostly the yearly production of loans. This is similar to the situation in the mortgage companies described in Huang et al. (2024), although in that industry only the loan volume and no other factor were part of the compensation scheme. However, there is a relevant difference between Spanish banks and US mortgage companies in terms of the measurement of the performance of loan officers. Mortgage companies sell all the loans they originate in the secondary market.[12] Spanish banks could not deconsolidate lending risk by securitizing mortgages in the US style, and therefore risk could not be spun off the balance sheets of banks by creating special purpose vehicles. The securitization process in the Spanish case was mainly based on covered bonds.[13] This means that most of the risk did not exit the balance sheets of Spanish banks despite the securitization process. Therefore, while in mortgage companies that immediately sell risk, there is little incentive to closely monitor default rates, this was not the case in the Spanish context. Therefore, in the long run, a high proportion of defaulted loans could imply important financial problems for the bank, and therefore it is reasonable that the rate of defaulted loans is a critical variable for the long-term evaluation of loan officers.

The strong incentives to generate new business in the banking industry had important consequences in the Spanish financial sector. The large increase in the default rate and the worsening of the capital ratios were mostly concentrated in savings and loan institutions. Before the beginning of the financial crisis, there were 45 savings and loan firms. At the beginning of the crisis, the Bank of Spain promoted the merger of savings and loan institutions in an attempt to improve their financial situation. Finally, with the exception

11. Other variables included in the bonus formula were the generation of commissions, the number of insurance policies, etc.

12. Huang et al. (2024) reports that 93.4% of mortgages originated by mortgage companies were sold at the end of the calendar year in the sample period analyzed.

13. Banco de España (2017)

of two very small institutions, those mergers of savings and loan firms were transformed into banks and sold to other Spanish banks at very low prices since, in most cases, their financial situation was close to bankruptcy. Our data come from one of those mergers of savings and loan firms.[14]

III Data

IIIA. Characteristics of the dataset

It is thus important to assess the relevance of compliance with the rules relative to the differential results in the lending decisions of male versus female loan officers. We had access to several datasets. The first, the validation database, contained all the variables needed for the construction of the internal scoring model. It includes many financial variables[15] considered at the time of the original screening of the operation and all the characteristics of the applicants that were recognized as potentially relevant, or predictive, in the scoring model.

The risk scoring model was quite sophisticated and most probably[16] included not only demographic, financial, and personal characteristics of the applicants (age, marital status, occupation, type of contract, indebtedness, etc.) but also variables related to the relationship between the client and the bank, transactionality (number of transactions per month, length of the commercial relationship, average amount held in the account during the previous year, etc.) and the type of product (loan-to-value ratio of mortgage, etc.).

The second dataset includes a variety of performance measures used by the bank to validate the risk scoring model. Obviously, the validation model must be confronted with the performance of loans. The performance database is, by its structure, quite different from the validation dataset. The latter captures a still picture at the time of approval of each loan, while the performance database includes the accumulated performance since the approval. For instance, it contains, among many other variables, indicators describing whether there were any late payments[17] of more than 30, 60, or 90 days since the origination of the loan.[18] In most of the exercises, we measure the performance of a loan using the regulatory definition of delinquency: loans having a late payment of more than 90 days at any point until the last period of observation.

14. We signed an NDA that prevents us from identifying those institutions.

15. This information came from the main dataset of the bank, which included all financial information on the accounts and products of the bank used to produce financial statements, regulatory reports, etc.

16. The bank did not share the exact model that was used to calculate the risk scores due to the confidentiality of the commercial algorithm.

17. Including the number of late payments during the life of the loan.

18. We had also access to datasets with the temporal evolution of these performance indicators. In the last section of the paper, we use this information to construct the known evolution of the performance of each loan officer at each point in time.

A third dataset provided information on the loan officers' employment history while in that position. It included not only demographic characteristics of the officer (e.g., gender, age) but also the duration of their tenure in the position and the branches at which he or she had been working. The merger of this information and the previous dataset provided information on the performance of loan officers on different dimensions.

A fourth dataset covered all the characteristics of the loans that were approved: maturity, amount, type, purpose, etc. This information also came from the bank's main financial dataset.

The database resulting from merging these four datasets is not only very detailed but also includes information that makes it unique. First, it contains data not only on loans granted but also applications denied. While the information comes from several financial institutions, the high number of loans and the length of time over which the data are available provide confidence in the external validity of the results.[19] Second, the database includes the risk score and the recommendation generated by the AI algorithm, one of the main novelties of the analysis presented in this paper. Generally, researchers working with administrative data on individual loans rarely have access to the internal scoring of loans or the recommendations.

IIIB. Characteristics of scoring and the recommendation algorithm

An important aspect of our paper is the use of the scoring as a conditioning variable to control for the risk and quality of loans when analyzing the relationship between delinquency and gender of the loan officer. In fact, credit score models are generally not publicly available since they are a very sensitive element of the credit risk management of financial institutions. Rajan et al. (2015) argue that the scoring models used in the US during the period 1997-2006 were unstable because securitization changed the incentives of lenders. The securitization process that took place in the US during the period 2000-2006 did not happen in the financial system of many other countries. As noted above, banking regulation in the Spanish case did not allow for the deconsolidation of SPVs created with securitized mortgages, and therefore, banks could not improve their capital ratios by securitizing mortgages in the US style since they could not take the risk off of their balance sheets. In addition, as we show later in this section, the internal risk models of the banks were validated every year and updated if there was any significant loss of predictive power. Internal documents of the bank show that the AUC[20] of the scoring model was systematically over 80% during the period under

19. The use of detailed banking information from one, or a few, loan providers instead of the whole sector is not uncommon in the recent literature. See, for instance, Campbell & Cocco (2015) and Rajan et al. (2015).

20. The AUC, or area under the ROC curve, is the usual measure to check the discrimination ability of a binary classifier. It compares the sensitivity of the procedure (true positive rate) with the false positive rate (one minus the specificity). The integral of that area, normalized, is the AUC. It basically measures the probability of correctly identifying a good loan if faced with one random good and one random bad loan.

study.

The bank provided two scores: a behavioral score[21] and a proprietary score (BKS). The former was used to offer small amount, preapproved loans, while the latter was used when the client did not have enough data to construct the behavioral score or when the amount applied for was over the limit of the preapproved loan. All the loans in our database were screened using the proprietary score. We use the behavioral score as an additional measure of the quality of the applicants and for robustness purposes.

The bank did not share with us the specification of their risk scoring models but only the variables they had available for this purpose. To check the accuracy of their claims relative to the quality of their scoring model, we constructed our own model using the variables included in the validation dataset. In particular, we considered a variety of demographic and financial characteristics of the borrowers: age, marital status, type of job contract, number of years in the current job, loan type, destination of the loan, debt-to-income ratio, debt over wealth, loan-to-value ratio (in the case of mortgages), monthly mortgage payment over 6-month average bank account balance, nationality of the client, average bank balance over 6 months, 6/12 month bank balance ratio, an indicator for whether the individual is a bank client, and number of years as a bank client.[22] Using this specification, we derived the AUC for consumer loans (Figure I) and mortgages (Figure II). Our specification covers the full period and, therefore, is not strictly comparable with the results of the internal documents of the bank. The area under the ROC curve was 77.7 in the case of mortgages and 74.5 in the case of consumer loans. These results confirm the good quality of the data supporting the risk scoring model[23].

Financial institutions use diverse scoring models for different clients and products, and this is true of our data on proprietary scoring. It is, for example, common to have one model for clients and another for non-clients, as the respective availability of data is very different. It is also common to use diverse models to score applications for distinct products (mortgages, consumer loans, etc.). In fact, tables may also change over time when the models are updated. Furthermore, these scores generate different tables by product and/or client that evolve over time, with diverse ranges of variation. For this reason, the risk scores of the different models are frequently aligned into one adjusted score that synthesizes all the tables and allows one to check the goodness of fit of the risk management system as a whole. While the bank provided the aligned behavioral score, it did not provide an adjusted proprietary score.

We consequently generated a standardized proprietary score (BKS), that we name "adjusted score" to

21. This score was produce by TRIAD, a standardized product for credit risk score sold by FICO.

22. The bank did not have the gender of the client among the variables available for the construction of the risk score, and therefore, it could not have used it to generate the model.

23. An AUC of 70% or greater is the goal in information-rich environments such as that discussed in this paper.

distinguish it from the aligned score produced by the standardization of the behavioral scoring, constructed directly by the bank. We use the following procedure. Denote $F(.)$ as the distribution function of the scores of each table. The reference score function is table 0 corresponding to product 0. Therefore, for any table i, we can calculate the aligned score using the following algorithm. In step one, we run the probability of default (PD) of the table of reference (0). Then, we run the probability model for all the scoring models (i). Using the predicted probabilities derived from that model and the parameters estimated in the reference model, we can obtain the adjusted scores. The empirical findings check the robustness of the results alternatively using the aligned behavioral score of the bank and our adjusted proprietary score.

$$
\begin{aligned}
PD_0 &= F(\beta_0 * Score_0) \\
PD_i &= F(\beta_i * Score_i) \\
AdjScore &= F^{-1}(\hat{\beta}_0 * \hat{PD}_i)
\end{aligned} \tag{1}
$$

IIIC. Characteristics of the sample

The advantage of including only loans to households is that the internal risk assessment produces risk scores for all cases. Therefore, each of the loan officers had the same summary information about the quality of the loan based on the observable quantitative indicators used by the scoring system.

By contrast, loans to SMEs and micro-companies are considerably more difficult to score appropriately, and consequently, no risk scoring is usually available.[24] This is also the case for the bank that provided the data. We eliminate from the population of household loans those that were authorized by a risk committee in central servicing due to the size of the request exceeding the authorization of the loan officer at the branch. In summary, we start with 422,302 applications for mortgages and personal loans. When we analyze the determinants of delinquency, we obviously only consider approved applications. This sample consists of 362,898 observations.

Since this sample is not the result of a randomized experiment, in this section, we show that the loans screened by male and females officers are comparable. Obviously, in the context of credit risk, this means that the distribution of the loans screened by male and females loan officers have similar ex ante risk, which in terms of the information available, means that credit scores and recommendations should be similar.

Table I presents the basic statistics of the credit risk of loan applications by the gender of the loan officer who screened them. The average of the adjusted proprietary score (BKS)[25] is practically identical for

24. Loans for large corporations are mostly scored using the ratings produced by rating agencies.

25. This is described in the previous section.

applications managed by male versus female loan officers. We can also examine differences in the distribution of the score of applicants depending on the gender of the loan officer who managed the application. Figure III shows that the distributions of the standardized score of the applications adjusted by the cohort of the loan[26] are virtually identical for male and female loan officers.[27] In addition, the distributions of the recommendations on the applications submitted to male and female loan officers are identical, as shown by the second panel of Table I. Therefore, whether we consider the risk scores or the recommendations, the distributions of the applications received by male and female loan officers are very much alike.

Despite the similarity of applications received by male and female loan officers, their approval and overruling rates are quite different, as shown by Table II. The approval rate of loans is higher among male than female loan officers. The difference is four percentage points, although Table II shows that this is mostly concentrated among the loan applications with a rejection recommendation. This implies that the overruling rate, or the approval of loans notwithstanding a negative recommendation, is considerably higher for male than for female loan officers. Specifically, we observe in Table II that the overruling rate for men is 13 percentage points greater than that for women. Table II also shows that the approval rates of men and women are almost identical for applications with an acceptance recommendation. The approval rate for applications with a rejection recommendation are substantially higher for men than for women.

IV Basic results

The basic regression analyzes the relationship between the gender of the loan officer and the delinquency rate of the loans conditional on the quality of the applicant, as determined by the internal scoring rate. The basic specification is a logit model[28]

$$logit(Delinq_{ijt}) = \alpha\, male_{ijt} + \beta\, Score_{ijt} + \sum \gamma_k X_{ijkt} + \mu_t + \mu_j \tag{2}$$

where *Delinq* is a dummy variables that takes value 1 if the loan has missed any payment for more than 90 days since its origination until the end of 2012, which is the standard definition of delinquency; *male* is a dummy variable that takes value 1 if the loan officer was a man; *score* corresponds to the different versions of the score; X includes other explanatory variables; μ_t is a time dummy; and μ_j is a geographical dummy. The time variable is relevant since it is important to control for the cohort of a loan.

26. Controlling for the origination year of the loans is important to account for the cyclical situation of the economy.

27. The Kolmogorov-Smirnov test cannot reject the null that the samples of males and females were drawn from the same distribution (p=0.12).

28. A linear probability model delivers almost identical results.

Column 1 of Table III shows that loans approved by male loan officers have a delinquency rate that is 1.7 points higher than that of female loan officers.[29] This difference increases to 2.5 percentage points if we consider the cohort of the loan (Column 2). This figure is both highly statistically significant and economically important since the average delinquency rate of the loans in the sample is 12%. Conditional on the aligned behavioral score provided by the bank (Column 3), the loans approved by male loan officers have a delinquency rate that is 2.4 pp (percentage points) higher than that of female loan officers. The score is highly statistically significant in the explanation of the delinquency rate. In particular, an increase of 100 points in the score decreases the probability of delinquency 1.5 pp. The result remains basically unaffected when we add experience or demographic characteristics of the loan officer (age). Older loan officers have a higher probability of granting loans that will be delinquent.[30] However, it is not very relevant in economic terms: 10 more years of age implied an increase in the delinquency rate of 0.2 pp. Experience as a loan officer reduces the probability of delinquency of loans. These results are unaffected by the inclusion of geographical dummies.

As discussed above, the banks use different scoring models for different products, types of clients, and periods. Each of these models defines a particular scoring table. For example, Score Table 3 (SC3) was used to obtain the scoring for mortgages for non-clients during the period 2003-09. The banks worked with a proprietary scoring divided into 13 scoring tables, with different models for clients and non-clients[31] and for personal loans and mortgages. The updating of the different models over time also generated new scoring tables since the specification of the models changed.

Table IV analyzes differences in delinquency rates by loan officer gender, considering the proprietary score before adjustment. This approach avoids the need to adjust the scores to make them comparable across tables and periods. Table IV reports the baseline probability and the increase in the probability of delinquency for males (interaction effect). The basic results of Table III are supported by the use of the

29. To facilitate the interpretation of the parameters, they are expressed as average marginal effects in all tables. In addition, the variables score, age and tenure enter into the estimation divided by 100 to avoid presenting too many decimal places. Finally, the tables present the standard errors clustered at the geographical level, which are substantially larger than the robust standard errors. Abadie et al. (2023) argue that cluster adjustment is not always necessary. In particular, their procedure, by focusing on finite populations (which could be entirely or substantially sampled in the data) produce substantially smaller standard errors than the clustered alternatives, although larger than robust estimates. Although our data are similar to those in their setup, we present the clustered version of the standard errors as the worst-case scenario.

30. This result is in line with the career concerns model of Agarwal & Ben-David (2018) but in contrast to the results of Beck et al. (2013).

31. A client who opened an account less than 6 months before the calculation of the score is considered, from a scoring perspective, as a non-client given that some of the relevant variables used for the scoring of clients (e.g., average account balance over the last 6 months) cannot be calculated.

proprietary score by each scoring table. In general, female loan officers have a lower delinquency rate for loans they approved than do male loan officers. Scoring Tables 8 to 13, which correspond to the scoring models used after 2009, represent an exception to this general finding. As argued above and based on the analysis of the internal circulars, after 2008, there is a clear change in the management of credit risk. Internal Circular A2-088/08 states that branches could not approve new loans to any applicant, either holder or guarantor of a previous loan, who had had any loan delinquent for more than 30 days, a refinancing operation, or any incidence in the risk information service. It also prohibited, with no exception, the approval of applications for which the system had recommended rejection (negative and very negative recommendation). The fact that after 2008 there is not a significant effect of gender on the delinquency rate implies that when the controls on recommendations are tightened, for instance by eliminating the possibility of overriding a negative recommendation, male and female loan officers perform similarly.

That said, most of the loans in the sample belong to Scoring Tables 2 to 7, corresponding to the period prior to 2009, which show a significantly higher delinquency rate for loans granted by male loan officers. In particular, Scoring Tables 3, 5, and 6, which correspond to non-client applicants, show the largest difference in the delinquency rate between loans granted by male and female loan officers.

Table V replicates the estimation of the basic specification in Table III using our adjusted score, calculated as described in Section III. The results in Tables III and IV are confirmed. Female loan officers have a lower delinquency rate than male loan officers, ranging from 2.0 to 2.5 pp when there is a control for the cohort of the loan. As argued above, from the analysis of the internal circulars and the results in Table IV, we know that the period before 2009 (prior to the banking crisis) was quite different from that after the beginning of the crisis. We also include a final Column (7), which considers only those loans produced before 2009. The results show a difference of 2.1 pp, very similar to the findings using the full sample.

This outcome is reasonable given that after 2008, the number of loans originated is very low compared to the pre-2009 period. Interestingly, the explanatory power of the specification using our version of the adjusted score is almost double the pseudo R^2 obtained using the aligned behavioral score provided by the bank. In the following sections, we consequently check the robustness of the results to the pre-2009 sample and include our adjusted score as an indicator of the risk quality of the loan.[32]

32. The results remain basically unchanged if we use the aligned behavioral score provided by the bank.

V Robustness

In the previous section, we showed, using alternative measures of the quality of the loans, that male loan officers have a higher delinquency rate than female officers. This section investigates the robustness of this finding to the inclusion of additional explanatory variables and specifications.

VA. Adding characteristics of the loan officers

Table VI includes some robustness checks. The basic results shared above are robust to these changes. Adding as an explanatory variable the interaction between tenure and male officer shows that improvement in the ability to screen applicants increases much faster for female than for male officers. In other words, an enhanced ability to screen bad loans, understood as a reduction in the delinquency rate of the loans approved as function of the years spent as a loan officer, occurs more quickly among female than among male officers.

It is also interesting that having experienced a previous crisis as a banking employee does not immunize the loan officer from granting bad loans. Measuring exposure to a previous crisis represents a challenge. To gauge the possible influence of this experience, we employ specific time periods as reference points. The first and main threshold is defined as being hired initially before 1995, which corresponds to the previous banking crisis in the Spanish economy. The crisis, starting in 1992, involved the failure of Banesto, a major Spanish bank, and a rapid increase in the proportion of NPLs, peaking in 1995. We find that the experience of a previous financial crisis did not prevent loan officers from approving bad loans. In fact, the opposite is true: loan officers who were already working in the banking sector during the previous financial crisis present a significantly higher delinquency rate than other loan officers, although the effect is economically small. The average number of loans approved by loan officers also increases the delinquency rate.

VB. Adding the determinants of the score

In the previous tables, we used two alternative scores (the aligned behavioral score and the adjusted proprietary score) as indicators of the quality of loans. As argued in Section III, the banks' validation reports of the scoring show AUCs above 80%, implying a good level of accuracy. Nonetheless, we investigate the robustness of the results to the use of variables that are known to be determinants of the quality of loans. Specifically, rather than using the scores provided by the bank, which were calculated using a confidential model that the bank did not share with us, we generated our own scoring model to assess whether the results are robust to the direct use as explanatory variables of those factors commonly included in the calculation of scoring models.[33] The estimates are obtained using the same variables that we used to calculate the accuracy of the scoring model in Section III.

The results in Table VII, where we substitute for score with the abovementioned variables, are consistent with the previous results: loans granted by female officers present a delinquency rate between 1.6 and 1.8 percentage points lower their male counterparts. The effect of age and tenure have the same sign as before but are not significant. If we also include the adjusted score (Columns 5 and 6), we still find additional explanatory power, although the coefficient is substantially reduced compared with previous results. Table VIII shows that the basic results using the components of the score are robust to using additional controls for the demographic characteristics of the loan officer.

VC. Alternative definitions of delinquency

The performance database contains several indicators of delinquency depending on how many times, or for how long, the client missed a payment. These are codified for the whole life of the loan. Performance is measured as the number of missing payments over 30 days, 60 days, and 90 days. In previous sections, we defined a delinquent loan as having missed at least one payment for a period of 90 days. The 90-day threshold is the standard used in many countries to define an NPL.

It is hence of interest to check the robustness of our findings to changes in the measurement of the performance of the loan. Table IX considers 60 days as the threshold to classify a loan as nonperforming, while Table X considers as NPLs those missing at least one payment over 30 days. The basic results are robust to these new definitions of the performance of the loans. In fact, for periods below 90 days, the performance of male loan officers relative to female loan officers worsens with respect to the 90-day threshold. Tables XI and XII show that the results of Tables IX and X are robust to including additional controls for the demographic characteristics of the loan officers.

VD. Panel data analysis with branch fixed effects

In the previous exercises, we controlled for the location of the branches using geographical dummy variables. We could also run the exercises using a branch fixed effect. Note that the gender of the loan officer in each branch can change over time. Table XIII shows the estimation of the basic specification using branch fixed effects.[34] The results show that male loan officers have between a 1.1 and 1.5 point higher delinquency rate than female loan officers, consistent with the results of previous exercises. If we consider

33. Figures I and II already show the AUC of these scoring models for consumer loans and mortgages respectively.

34. We resume using as the definition of NPLs those missing at least one payment over 90 days.

the period prior to 2009, the estimate is 1.4 percentage points. The estimator is significant in all columns and economically important: loan officers show a 11.8% lower delinquency rate than their male counterparts. As in previous tables, the delinquency rate is higher for older loan officers, although in this case it is not significant. However, here there is no effect of tenure on the proportion of NPLs.

VE. Type of loan

As mentioned above, our sample includes two basic types of loans: mortgages and consumer/personal loans. Table XIV analyzes whether there is a differential influence of the gender of the loan officer depending on the type of loan. The results in Table XIV show that men have a worse performance than women in both types of loans.

VI Explaining the findings

The previous sections have shown that there is a significant difference between the delinquency rate of male and female loan officers conditional on the risk score of each loan. In this section, we explore alternative explanations for the findings reported in the previous section.

VIA. High delinquency with high return?

Is this difference in the delinquency rate between men and women relevant to the bank? Could it be that, despite having a higher delinquency rate, men generate a higher return-adjusted by risk than women? This is unlikely since, as explained above, the interest rate charged for loans in Spain during the analysis period was quite insensitive to the risk score of the client. A precise calculation of the risk-adjusted return on capital (RAROC) confirms the previous hypothesis. Table XV discusses the case of mortgages.[35] The interest rates on mortgages approved by male and female loan officers are almost identical (4.70% and 4.78%, respectively),[36] as are the size of the average mortgage (112,265 euros versus 112,685 euros) and the maturity (10 years in both cases). We calculate the RAROC as the ratio of revenue plus return on economic capital minus expenses minus expected loss under realistic assumptions. We assume the same loss given default (LGD) for male and female loan officers since this parameter is not available, although we assume different LGDs for mortgages (15%) and consumption loans (45%). Obviously, given the higher default rate of loans

35. The case of consumer loans returns the same conclusion.

36. Note that the small difference in the interest rate and the mechanisms used to determine it indicate that gender differences in cross selling were not relevant.

approved by men, the largest difference between men and women in the calculation is the expected loss. The operational cost is supposed to be the same regardless of the sex of the officer who approved the loan. This calculation leads to a higher RAROC for mortgages approved by women (14.39%) than men (11.92%).[37]

Could the additional profitability derived from cross selling be a good explanation for not considering the difference in delinquency rate between men and women to be relevant? As reported in previous sections, during the study period, Spanish banks rationed credit using quantities, not prices. Interest rates are basically the same irrespective of the score of the client, the LTV of the operation, etc. In fact, interest rates were a function of cross selling. The internal circulars of the bank precisely explain the rate setting. The common standard rate could be reduced by 0.1 points for subscribing to life insurance; 0.05 additional points for buying home insurance; 0.1 points for getting a credit card; and 0.1 points for direct deposit of paychecks into the bank account. No reference is made to any influence of the scoring on the interest rate. This is, moreover, a general feature of Spanish banks: interest rates are insensitive to mortgage characteristics (risk scores, LTV, etc.) in the segment of retail banking clients.[38] Since we show that the interest rate on loans managed by male and female loan officers are virtually identical, the cross-selling explanation does not seem to be particularly relevant.

Could securitization explain the difference in delinquency between male and female loan officers? As argued above, the securitization process that took place in the US during the period 2000-2006 did not occur in the Spanish financial system. Banks could not deconsolidate lending risk by securitizing mortgages in the US style, and therefore risk could not be spun off the balance sheets of banks. There is no reason to believe that financing through traditional mortgage-covered bonds had a differential impact on male and female loan officers.

VIB. Gender differences in risk attitudes

Can differences in risk attitudes between men and women explain the results? If women are more risk averse than men, as risk increases, female loan officers should reject more loans than their male counterparts. [39] Furthermore, as mentioned above, the objectives of the variable part of the salary are absolute values and not relative to other members of the organization, further reducing the role of gender associated with differences in competitive behavior.

37. The cost of capital at that time for the banking sector was approximately 12%.

38. This fact simplifies the calculation of the RAROC of each individual loan.

39. For example, Adams et al. (2016) show that female in the CFA are less tradition and conformity-oriented and more achievement-oriented than women in the general population;

In our case, when examining very positive, positive, and neutral recommendations, there is not a substantial difference in the rejection rate of female and male loan officers, although the average risk clearly increases, we move toward the neutral recommendation. The difference between the rejection rate of female and male loan officers jumps to 15 pp when the recommendation is negative even when controlling for risk as measured by the credit score.

An alternative approach is to check whether the typical difference in risk aversion between men and women can explain the results in the previous sections. There is a large literature on gender differences in risk aversion, although there is less agreement on the extent of this difference[40] We consider a type of lottery in which with probability 0.87, the loan officer receives an application with a positive recommendation, and with probability 0.13, the officer receives an application with a negative recommendation. These are the probabilities observed in our data. Using the probabilities of rejection, default conditional on rejection, etc. of male and female officers, as discussed in the following section, we can calculate the difference in risk aversion needed for female officers to choose the rejection rates observed in the data instead of the rates exhibited by male loan officers. We also estimated, using our data, the probability of losing the loan officer conditions as a function of the accumulated default rate. For reasons that will become clear in the next section, we use the same rate for men and women. Using a constant relative risk aversion function, we find that the rate of risk aversion required to justify the rejection rate of negatively recommended loans by female loan officers is 10 times the highest value obtained by Filippin & Crosetto (2016). Obviously, this is merely a back of the envelope calculation since it is unclear that constant relative risk aversion is a good representation of the utility function and there is controversy over the actual value of the gender difference in risk aversion. Nevertheless, these calculations indicate that it is difficult to claim that the gender differences observed in the data can be fully explained by differences in risk aversion.

VIC. Gender differences in rule compliance

In previous sections, we analyzed the effect of the credit score. Here, we instead consider the influence of the system's recommendations on the loan officer decisions, conditional on the risk score. The recommendation system reflected specific bank credit policies and took the form of a categorical variable with five levels. The five recommendation categories are very positive (A1), positive (A2), neutral (A3), negative (D1) and very negative (D2). Categories D1 and D2 implied a recommendation to reject the application.[41] . The

40. We consider gender differences in risk calculated in Filippin & Crosetto (2016) using the procedure Holt & Layry (2002) and the risk-elicitation procedure of Eckel & Grossman (2002).

41. The bank did not share the AI algorithms that generated the recommendations.

correlation between risk score and recommendation is far from perfect.[42]

Loan officers could, however, override the recommendation before 2008 in "exceptional cases." After 2008, as noted above, Internal Circular A2-088/08 states that the approval of applications for which the system had recommended rejection (negative and very negative recommendations) is prohibited, with no exception.

The reasons for overriding the recommendations and forcing an approval were explained using several standard sentences, codified into 13 relevant categories. Most of the reasons were very subjective justifications, such as "the applicant has good prospects of generating future business with the bank" (23%) and "the client has a positive credit history with the bank" (27.6%). Other comments downplayed important components of the scoring model. For instance, some operations with a negative recommendation were approved arguing that "the applicant has a temporary contract, but has been working continuously in recent years." There were also cases where the loan officer chose to overrule the recommendation even when the client was included on a list of known delinquent debtors[43] or had experienced issues in the payment of previous loans at the bank. In such cases, the reason given for overriding the recommendation is that "the incidence has been regularized." The reasons exhibited by men and women had similar proportions with a few exceptions. Men have a higher probability than women of claiming that the applicants had good prospects for generating future business (3 pp) and that the client had a positive credit history with the bank (2 pp). Women argued more often than men that the client was associated with trustable third parties (5 pp higher proportion than men) and that she had been a bank's client for a long time (3 pp higher than men).

As a first approximation, to understand the observed difference in the delinquency rates of loans handled by men and women, we can decompose, from a purely accounting perspective, the default rate for each gender into the delinquency rate of positive and negative recommendation loans.[44]

$$\begin{aligned} P(D|G) &= P(PR|G)P(A|PR,G)P(D|A,PR,G) \\ &\quad + P(NR|G)P(A|NR,G)P(D|A,NR,G) \end{aligned} \tag{3}$$

where D equals 1 if the loan is delinquent, PR is a positive recommendation, NR is a negative recommendation, A is approval, and G is either a man or a woman.

The difference in the delinquency rates of men and women reflects three components: differences in the likelihood of handling positively and negatively labeled applications; differences in the approval rates of positive and negative recommendation applications; and differences in the delinquency rates conditional

42. Later in this section, we show that the recommendation has explanatory power for the delinquency rates ever after controlling for risk scores.

43. For example, the applicant was listed in the ASNEF registry, the "black list" of defaulters managed by EQUIFAX.

44. To simplify the decomposition, we also include the neutral level in the positive recommendation category.

on the recommendation. The result of the decomposition shows that 60% of the difference is caused by negative recommendation loans. There is no difference in the types of application received by gender: the proportion of negative recommendations and positive recommendations are identical, as we already showed. However, there is a large difference in the probability of overriding the decision for loans with a negative recommendation. In fact, the differences in the likelihood of overriding the decision of the system explain 74% of the discrepancy in delinquency rates between men and women for the group of loans with a negative recommendation. Another 25% is explained by the difference in the component of the delinquency rate.

To analyze men and women's compliance with the rules, we ran several empirical exercises. Table XVIII presents some logit specifications to explain the differences in the approval rates of men versus women by type of recommendation. The set of explanatory variables is the same as in previous regressions. Table XVIII shows, as expected, that the approval rate decreases with the worsening of the recommendation. More interesting, the difference in the approval rates of male and female loan officers increases monotonically with the worsening of the loan classification. For very positive and positive recommendations, there is basically no difference in rejection rates. However, for neutral recommendation loans, the difference is 1.1 percentage points, which jumps to 3.3 for negative recommendation loans and 5.8 for very negative recommendation loans.

An analysis of the overruling behavior of men and women offers another perspective on the results presented in the previous paragraph. Taking into account only loans that received a negative recommendation label, Table XIX shows that men overrule the system's recommendation significantly more often than women. Consistent with the results of Table XVIII, the difference in the overruling proportion increases with the worsening of the category assigned by the recommendation system. The coefficient in the combination of negative recommendation loans by men implies that their overruling rate is 5.5 percentage points higher than that of women, which means that female loan officers comply with the rules more often than men.

In previous sections, we showed that risk scoring, in different versions, is a significant determinant of the delinquency rate. Does following the rules also provide an advantage in terms of lower delinquency rates? Table XX analyzes the determinants of the rate of delinquency by recommendation category while controlling for risk scores. As expected, the delinquency rate increases monotonically with the worsening of the loan recommendation. However, as in the previous table, conditional on recommendation, male loan officers are associated with a higher rate of delinquency, especially for loans in the negative or very negative categories.

These results indicate that female loan officers reject more loans with negative recommendations than men. Since women are less likely to override the recommendations of the AI algorithm than men, they can choose a better pool of clients even in the worst category. This explains why the loans produced by female loan officers show a lower delinquency rate than those approved by male loan officers.

Therefore, even after conditioning on risk score, the recommendation should contain relevant information. This can be seen in Tables XVIII and XX, where both both risk score and the recommendation are highly significant. We can thus interpret this result as meaning that conditional on the risk score, female loan officers follow the rules more often than male loan officers and that this provides an advantage in terms of avoiding loan delinquency.

In the first section of this article, we discussed several theories that could explain why gender may have an effect on the lower delinquency rate of loans monitored by women. Our study is unique in that, unlike previous research, the officer observes the risk score and the system recommendation before making a decision on a loan. Therefore, the initial decision is based on following or overruling the recommendation of the system.

Research in social psychology finds that women are more compliant than men.[45] There is also extensive evidence that dangerous behavior and participation in car accidents among adults are more often due to rule-breaking among men than among women. Women adhere to road signs more often than men; they also less frequently violate pedestrian rules.[46] This section has similarly shown that female loan officers tend to follow rules more often than their male counterparts, which consequently means that they generate lower levels of delinquency.

VII Gender bias in the mistake-punishment tradeoff

Why do women follow the system recommendation more often than men? High-level managers in the bank's risk department described a phenomenon they define as gender bias in the "mistake-punishment tradeoff." Specifically, one reason why female loan officers were afraid to deviate from the recommendation of the system was that if they approved a negatively recommended loan that then became an NPL, their careers would be damaged more so than those of men [47]. They therefore had a strong incentive to follow the recommendation of the system that counterbalanced, at least partially, the variable payment incentive to produce more loans.[48]

45. See the classical reference of Tittle (1980).

46. Rosenbloom (2009).

47. Conversations with managers and many loan officers indicated that female loan officers were aware of this gender bias in terms of a mistake-punishment tradeoff.

48. For loan officers, the incentive for the volume of loans produced could amount to 15% of the fixed salary if they met 100% of the objective.

The literature has recently discussed the possibility of double standards in terms of punishment for breaking the rules. Egan et al. (2022), for example, analyze gender discrimination in the financial services industry. Using panel data on misconduct reported to FINRA for 1.2 million financial advisors in the US between 2005 and 2015, they find that women face harsher punishment for misconduct. In fact, they are 20% more likely than men to lose their job after a misconduct incident. Sarsons (2022) analyzes primary care physicians (PCPs) referrals to surgeons and describes an asymmetric update in terms of gender. PCPs reject referrals to female surgeons more strongly than to male surgeons after a patient dies. She concludes that women have fewer chances to make mistakes, which in turn could mean lower promotion rates than men.

As we can link loan officers and the performance of the loans they approved, we can investigate potential gender bias in the mistake-punishment tradeoff. In particular, we are interested in whether gender is a determinant of the duration of the position as a loan officer conditional on the accumulation of NPLs, which increases the probability of being demoted or, eventually, dismissed.

To verify the performance of each loan officer, we constructed the BadL variable, which corresponds to the accumulation of bad loans by an officer. Specifically, this variable represents the proportion of loans generated by each loan officer that are delinquent for more than 90 days, according to the definition of delinquency used in the previous sections of the paper. Figure IV shows the Kaplan-Meier nonparametric estimator of the survival function for loan officers who have accumulated at least 4% of bad loans versus officers who have less than 4% of delinquent loans[49]. Figure IV shows that the probability of being demoted as a loan officer, or dismissed, increases drastically when bad loans accumulate in the portfolio of a particular loan officer. This result confirms the usefulness of this indicator as a measure for evaluating the long-term performance of loan officers. Figure V shows the estimation of the survival functions for male and female loan officers. We observe that, unconditionally, men remain in the loan officer position longer than women, although the difference in survival functions is less striking than that depicted in Figure IV. However, as these unconditional figures are not evidence of the hypothesis of gender bias relative to the mistake-punishment tradeoff, we run several duration models.

Using the bad loans variable, we can investigate the effect of the accumulation of bad loans on the careers of men and women and determine whether, conditional on the accumulation of such loans, there is a differential tenure in the position of loan officer by gender. Survival analysis modeling usually assumes

49. The results of the figures and the analysis in this section are unaffected if we used the proportion of delinquent loans conditional on a negative recommendation instead of the unconditional delinquency rate, which is not surprising given the high correlation between delinquency and recommendations.

that the values of all covariates were determined at time 0. However, in this case, it is necessary to consider at least one time-varying covariate, as the proportion of bad loans accumulated by each loan officer could change over time. The basic variables in our specification are gender, age, and the proportion of bad loans. Table XXI shows the results using different specifications for the tenure hazard rate as loan officer. The first column includes the estimation of a proportional hazard model.

$$\lambda(t,x,\beta) = \lambda_0(t)\exp(\beta_1 * male_i + \beta_2 * age_i + \beta_3 * BadL_{it}) \tag{4}$$

Our results show that, conditional on the age of the loan officer and the proportion of bad loans accumulated in the past, men have a 34% lower rate of being demoted from their position as a loan officer than women. The hazard rate increases by 3% for each point increase in the accumulation of bad loans. As expected in Figure IV, reaching a high proportion of bad loans implies a high hazard of being sacked. In particular, the hazard rate increases 12.3% for a four-point increase in the proportion of bad loans.

Columns 2 to 4 include the estimation of parametric models of increasing flexibility in terms of the shape of the hazard function. The Weibull and Gompertz models in the following two columns generate similar results, and their respective ancillary parameters, p and γ, signal that the hazard rate is increasing monotonically. Unlike the previous parametric models, the Gamma regression coefficients in Column 5 can only be reported in accelerated failure-time metric. The results here support previous findings that men have a lower probability of being promoted than women.

Table XXII includes as an additional explanatory variable the cross-product of gender by bad loans. The term captures gender bias with respect to the mistake-punishment tradeoff. In this specification, the dummy for male is not significant, as observed for the results in Table XXI. However, the interaction effect of male and the accumulated proportion of bad loans is significant. Taking the results of the proportional hazard model,[50] we observe that while the estimated log hazard function with respect to the proportion of bad loans has the same origin (since the male dummy is not significant), the slope of the female log hazard function is higher than that of men. For example, at 2% of accumulated bad loans, men have a 9% lower probability of being demoted or dismissed than women. At 4% of bad loans, men have a 16% lower probability of being demoted or dismissed than women. The results remain basically unchanged using the same basic specification for alternative parametric models.

Table XXIII adds to the specification the average number of loans produced by each loan officer every month. This variable considers the revenue side of the production of loans. The basic results of Table

50. Note that in this case, describing the model as proportional hazard is not, strictly speaking, appropriate, and we use this terminology since it is generally adopted in the literature.

XXII are fundamentally unchanged. The male dummy is not significant. The effect of bad loans and the interaction of bad loans by gender are similar to those discussed in Table XXIII. Finally, the number of loans produced by loan officers is not significant in most of the specifications. In fact, it appears that the more loans that are approved, the higher the hazard of dismissal, which may be due to a higher proportion of bad loans associated with a fast rate of loan production. This conclusion would be consistent with the results of Table VI.

Recent literature on gender bias notices that promotion gaps are not a final proof of bias, since there may be some characteristics of workers that researchers cannot observe. To try to overcome this problem, there have been several proposals around the so-called Becker outcome test for biases (gender, racial, etc.) which implies comparing managerial performance at the margin of promotion. This test has been applied to promotion decisions in the context of loan officers in mortgage companies (Huang et al. (2024))[51] that uses instrumental variables based on the firm-level average of promotion rate in each years leaving out the focal worker and her teammates as in Benson et al. (2019). Since we only have one company, this type of approach is not feasible. In any case, it is interesting that Huang et al. (2024) concludes that mortgage companies require women to have 5% greater managerial ability than men to be promoted to managerial positions, which implies that male loan officers are 15% more likely to be promoted than their female counterparts with the same sales performance and experience.

In addition, our final set of tests can be viewed as a Becker style outcome test applied to the retention margin, rather than to individual loan approvals. Becker's framework predicts that in the absence of bias, companies equate marginal performance thresholds between groups. In our setting, the relevant margin is whether a loan officer is retained or demoted/dismissed as her portfolio deteriorates. We estimate hazard models where the risk of demotion depends on the accumulated default rate of the officer and its interaction with gender. The results show that, for a given default rate and controlling for other observables, female officers have a significantly higher hazard of demotion/dismissal than male officers. This implies that the bank applies a stricter retention standard to women, in accordance with Becker's prediction that biased firms require higher performance from the disadvantaged group. Combined with our earlier evidence that women generate portfolios with lower delinquency conditional on the same scores and recommendations, this suggests that gender bias operates both at loan approval and at the employment margin, although a strict Becker outcome test at the 'marginal loan' cannot be implemented because we lack an exogenous instrument that shifts officer approval thresholds, so a strict Becker test based on LATE on individual loans is not feasible. However, we show at the loan margin that women follow the AI recommendation more often and generate

51. See also Dobbie et al. (2021) for consumer lending or Arnold et al. (2018) for bail decisions.

lower delinquency, particularly for high risk (negatively recommended) loans. At the employment margin, we show that for any given delinquency rate, women face a higher probability of demotion/dismissal. This "double asymmetry" is very much in line with Becker's prediction that biased practices (i) distort decision rules and (ii) are costly to the firm.

VIII Conclusions

In this paper, we analyze gender differences in compliance with AI recommendation systems. In particular, we study the behavior of loan officers who, when making their decisions, have at their disposal an estimate of the risk of the application and a recommendation obtained from an AI system. Since the procedures to generate the risk score and the recommendation are quite different, there is a less than perfect correlation between the two measures. Corporate guidance indicates that regardless of the risk measure, applications with a bad or very bad recommendation should not be approved. The loan officers were aware, through internal circulars, that the AI algorithm performed very well in the screening of loan applications.

Using information from a large Spanish bank, we show that female loan officers follow the recommendations of the AI system more often than men and produce loans that have a lower nonperforming rate than those screened by male loan officers. This result indicates that following corporate guidance reduces the rate of non performing loans. In fact, in our sample of close to half a million loans, female loan officers have a delinquency rate that is 1.5 to 2.5 percentage points lower than their male counterparts. This result is economically significant, as it amounts to between 12.5% and 20% of the average delinquency rate.

The fact that after 2008 there is no longer a significant effect of gender on the delinquency rate implies that when controls are tightened, for instance, eliminating the possibility of overriding the recommendation to reject a loan, male and female loan officers perform similarly. One reason for the better performance of the loans screened by women is that, even conditional on risk, female loan officers followed the recommendation of the system more often than men.

Could men be overriding the recommendation because they possess superior soft information about borrowers that the algorithm does not capture, and the additional loans they approve generate profits that offset the higher delinquency? Several pieces of evidence argue against this interpretation. First, if overrides reflect genuinely profitable private information, the risk-adjusted return of male-approved loans should equal or exceed that of female-approved loans. Our RAROC calculations show the opposite: female-approved mortgages yield a risk-adjusted return of 14.4% versus 11.9% for male-approved mortgages. Second, we show that overriding the recommendation increases the probability of delinquency even after conditioning on the risk score; under the superior-information hypothesis, the exception rate should predict lower, not higher, delinquency among overridden loans. Third, the institutional structure rules out a revenue-through-pricing mechanism: interest rates in Spanish retail banking were insensitive to borrower risk characteristics, so overriding a negatively-recommended loan cannot generate higher revenues through loan pricing. The only potential revenue channel is cross-selling, yet the near-identical interest rates on loan approved by male and female (4.70 versus 4.78%) indicate that the differences in cross-selling between officers were negligible. Finally, the logic of Hoffman & Li (2018) applies directly: if discretion generates superior results, managers who exercise more discretion should outperform those who follow the rule. They do not, suggesting that the exceptions reflect bias or mistaken beliefs rather than genuine information (Freankel, 2021).

We explore alternative explanations for this difference. One reason for analyzing the case of loan officers is that women working in the financial industry, especially in relevant positions, have chosen commission-based jobs in a male-dominated industry. Therefore, gender differences in risk attitudes, overconfidence, and competitive behavior play a less important role than in other jobs and industries. In any case, we check for indications of differences in risk attitudes and the possibility that high delinquency is also associated with high returns, without finding much support for these alternative explanations.

This higher compliance with the rules is potentially explained by a differential punishment of women versus men in the case of bad outcomes/performance. Using our data, we find that women who accumulate a high proportion of NPLs have a greater probability of being punished than men, conditional on the same level of performance. These results support a recent finding in the literature on gender double standards in punishment for misconduct.

References

Abadie, A., Athey, S., Imbens, G., & Wooldridge, J. (2023). When should you adjust standard erros for clustering? *Quarterly Journal of Economics*, *138*(1), 1-35.

Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy*, *40*(121), 13–58.

Adams, R., Barber, B., & Odean, T. (2016). Family, values and women in finance. *mimeo*.

Agarwal, S., & Ben-David, I. (2018). Loan prospecting and the loss of soft information. *Journal of Financial Economics*, *129*, 608-628.

Angelova, V., Dobbie, W., & Yang, C. S. (2022). Algorithmic recommendations and human discretion. *Working Paper*. (Available at https://law.yale.edu/sites/default/files/area/workshop/leo/judge_discretion.pdf)

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, *133*, 1885-1932.

Banco de España, B. (2017). *Informe sobre la crisis financiera y bancaria en españa, 2008-2014*.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence and common stock investment. *Quarterly Journal of Economics*, *116*, 261-292.

Beck, T., Behr, P., & Guettler, A. (2013). Gender and banking: Are women better loan officers? *Review of Finance*, *17*, 1279-1321.

Bellucci, A., Borisov, A., & Zazzaro, A. (2010). Does gender matter in bank—firm relationships? Evidence from small business lending. *Journal of Banking & Finance*, *34*, 2968-2984.

Benson, A., Li, D., & Shue, K. (2019). Promotions and the peter principle. *The Quarterly Journal of Economics*, *134*, 2085-2134.

Bentolila, S., Jansen, M., & Jimenez, G. (2017). When credit dries up: job losses in the great recession. *Journal of the European Economic Association*, *16*, 650-695.

Bick, A., Blandin, A., & Deming, D. (2025). The state of generative AI adoption in 2025. *Working Paper*. (Federal Reserve Bank of St. Louis and Harvard Kennedy School)

Campbell, J. Y., & Cocco, J. F. (2015). A model of mortgage default. *The Journal of Finance*, *70*, 1495-1554.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., ... Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Working Paper*(24-013).

Dhar, R., et al. (2025). *Overconfidence, gender, and authority attribution in algorithmic versus human advice: A cognitive perspective on compliance.* (SSRN Working Paper 5275797. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5275797)

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126.

Dobbie, W., Liberman, A., Paravisini, D., & Pathania, V. (2021). Measuring bias in consumer lending. *Review of Economic Studies*, *88*, 2799-2832.

Eckel, C. C., & Grossman, P. (2002). Sex differences and statistical stereotyping in attitudes towards financial risk. *Evolution and Human Behavior*, *23 (4)*, 281-295.

Egan, M. L., Matvos, G., & Seru, A. (2022). When Harry fired Sally: the double standard in punishing misconduct. *Journal of Political Economy*, *130*(5).

Filippin, A., & Crosetto, P. (2016). A reconsideration on gender differences in risk attitudes. *Management Science*, *62 (11)*, 3085-3105.

Freankel, A. (2021). Selecting applicants. *Econometrica*, *89*(2), 615-645.

Hoffman, K. L., M., & Li, D. (2018). Discretion in hiring. *Quarterly Journal of Economics*, 765-800.

Holt, C., & Layry, S. (2002). Risk aversion and incentive effects. *American Economic Review*, *92 (5)*, 1644-1655.

Huang, R., Mayer, E. J., & Miller, D. P. (2024). Gender bias in promotions: Evidence from financial institutions. *Review of Financial Studies*, *37*(5), 1685–1728.

Humlum, A., & Vestergaard, E. (2024). The unequal adoption of ChatGPT exacerbates existing inequalities among workers. *Proceedings of the National Academy of Sciences*, *121*(52), e2414972121.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, *33*(1). doi: 10.1287/orsc.2021

Mayordomo, S., Rachedi, O., & Rodriguez, M. (2019). Bank regulatory capital arbitrage: evidence from housing overappraisals. *WP Bank of Spain*.

Nikpayam, H., Kremer, M., & de Véricourt, F. (2024). When delegating AI-assisted decisions drives AI over-reliance. *Working Paper*. (SSRN Working Paper 4966186)

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192.

Otis, N. G., Delecourt, S., Cranney, K., & Koning, R. (2024). Global evidence on gender gaps and generative AI. *Harvard Business School Working Paper*(25-023).

Paravisini, D., & Schoar, A. (2015). The incentive effect of scores: randomized evidence from credit committees. *mimeo*.

Rajan, U., Seru, A., & Vig, V. (2015). The failure of models to predict failure: Distance, incentives and defaults. *Journal of Financial Economics*, *115*, 237–260.

Rosenbloom, T. (2009). Crossing at a red light: Behaviour of individuals and groups. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*, 383–394.

Sarsons, H. (2022). Interpreting signals in the labor market: Evidence from medical referrals. *mimeo*.

Sun, Z., Zhang, J., & Bassamboo, A. (2021). Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations. *Management Science*, *67*(9), 5811–5826.

Tang, C., Li, S. K., Hu, S., Zeng, F., & Du, Q. (2025). Gender disparities in the impact of generative artificial intelligence: Evidence from academia. *PNAS Nexus*, *4*(2), e591.

Tittle, C. R. (1980). *Sanctions and social deviance: The question of deterrence*. New York: Praeger.

Utz, S. (2024). How gender and type of algorithmic group discrimination influence ratings of algorithmic decision making. *International Journal of Communication*, *18*, 571–581.

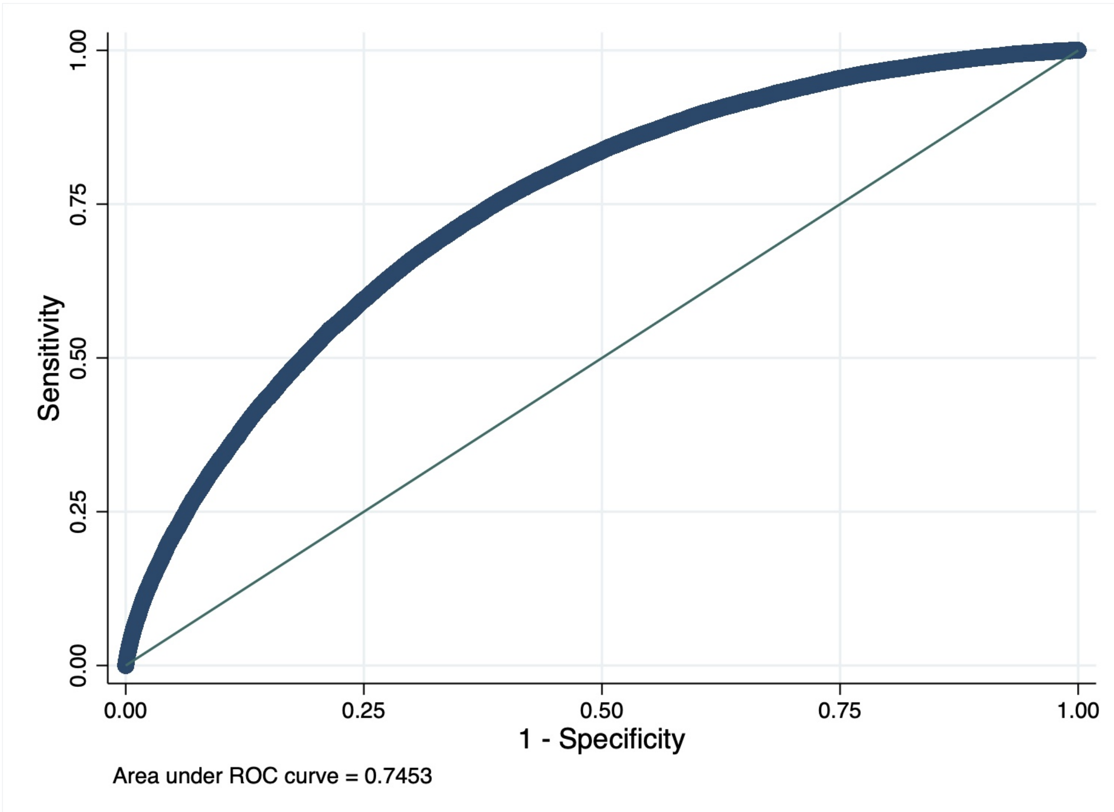

Figure I: AUC-consumer loans

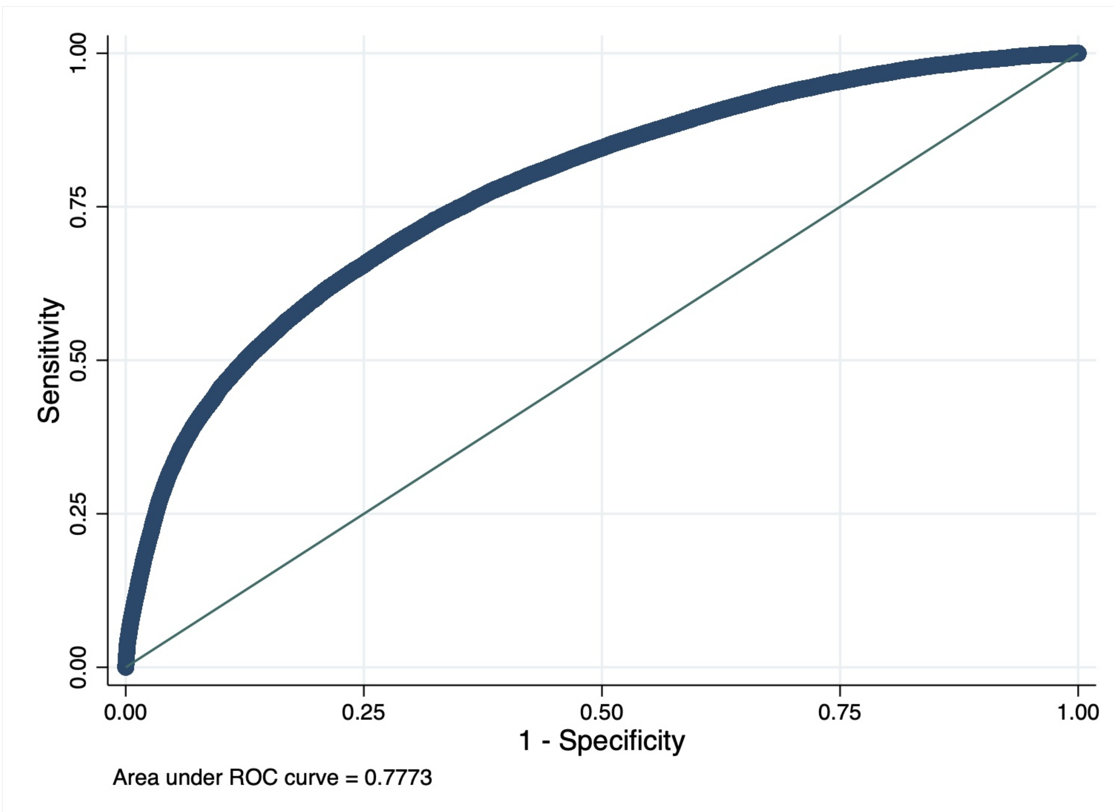

Figure II: AUC-mortgage loans

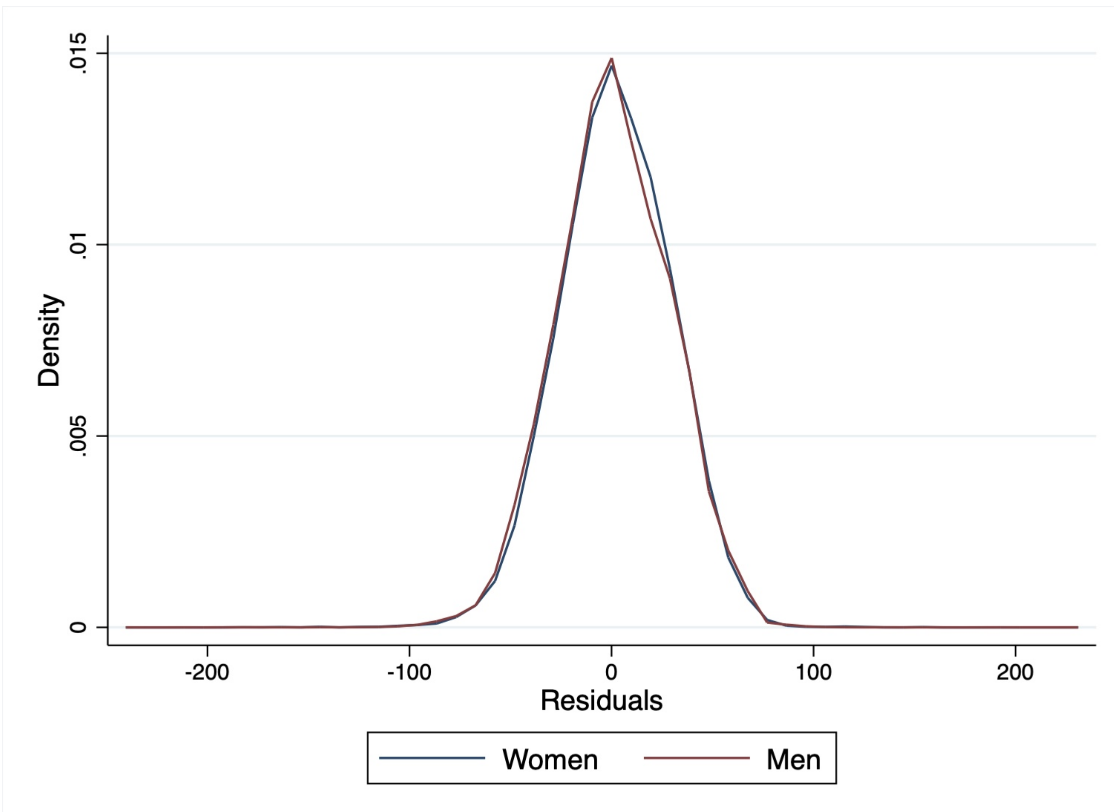

Figure III: Standardized score of applications by gender

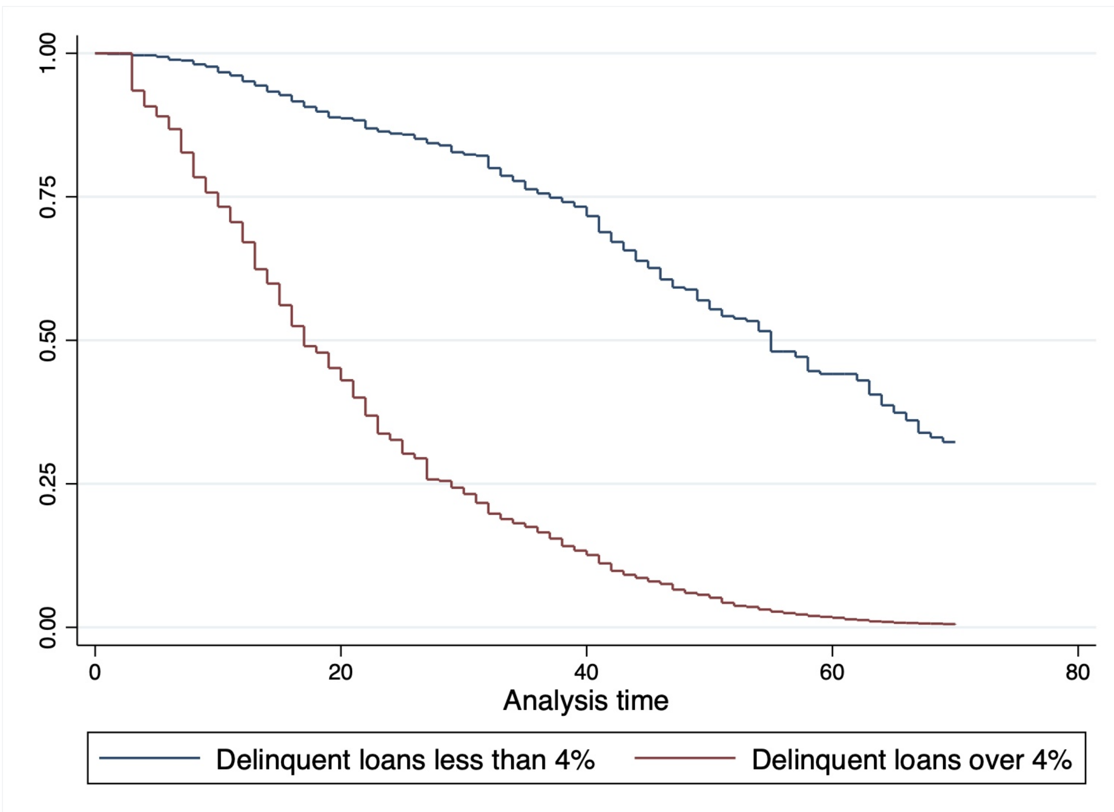

Figure IV: K-M survival estimates by proportion of delinquent loans

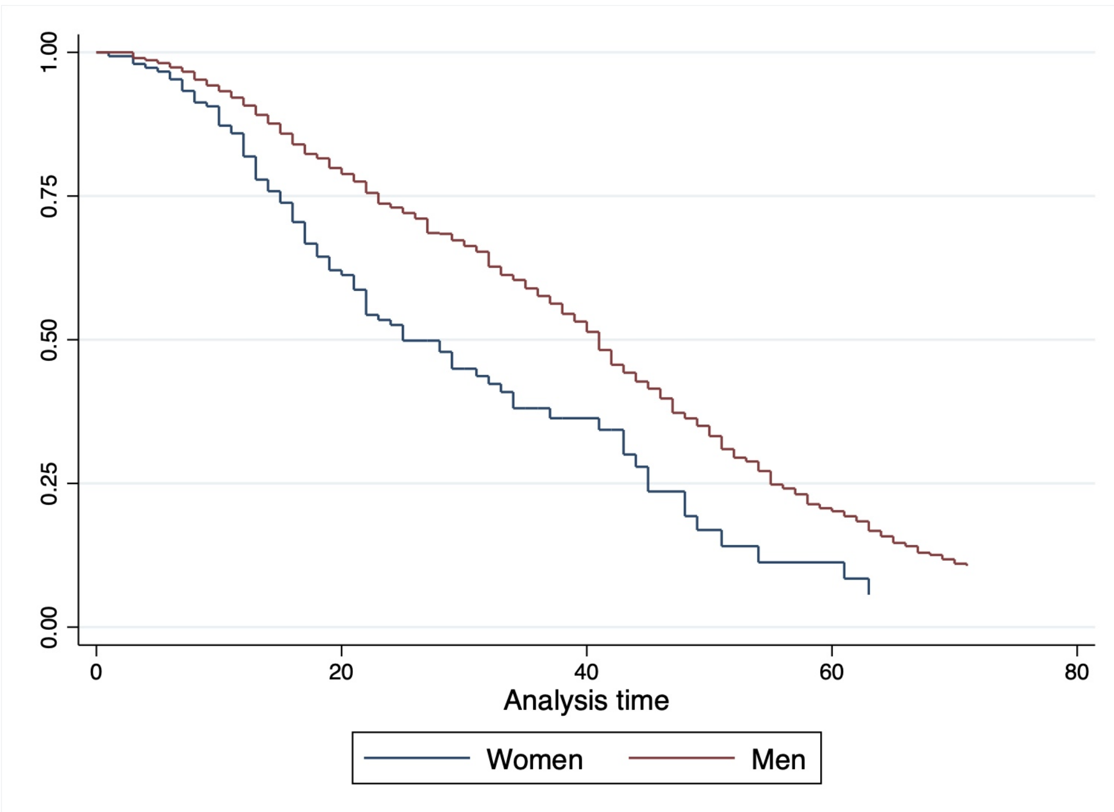

Figure V: K-M survival estimates by gender

Table I: **Descriptive Statistics: applications**

The table shows the basic descriptive statistics of the loan applications. The adjusted score is the standardized proprietary score that we describe in section IIIB. The recommendations of the applications are classified into 5 categories using decision trees.

	Total	Women	Men
Adjusted score	475	475	476
Applications by recommendation			
Very positive (A1)	0.30	0.30	0.30
Positive (A2)	0.44	0.44	0.44
Neutral (A3)	0.14	0.14	0.14
Negative (D1)	0.08	0.08	0.08
Very negative (D2)	0.04	0.04	0.04

Table II: **Descriptive statistics: decisions**

The table shows the basic descriptive statistics on the decisions taken by loan officers. For each recommendation it shows the approval rate by gender of the loan officer. It also shows the overall approval rate, the overruling share (proportion of negative recommendation loans that have been approved), and the delinquency rate, defined as a situation with a missed payment for more than 90 days by gender.

	Total	Women	Men
Approval rate by recommendation			
Very positive (A1)	0.99	0.99	0.99
Positive (A2)	0.98	0.98	0.98
Neutral (A3)	0.96	0.94	0.96
Negative (D1)	0.82	0.70	0.85
Very negative (D2)	0.66	0.59	0.67
Overall approval rate	0.95	0.92	0.96
Overruling share	0.77	0.66	0.79
Delinquency rate	0.12	0.10	0.12

Table III: **Basic regression analysis with the aligned score**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). We control for the aligned score, and age and tenure of the loan officer. For interpretation of the coefficients, we should notice that these three variables have been divided by 100 to facilitate the reading of the parameters. We also control for the cohort of the loans using a time dummy in all the regressions except in (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (6) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2012	(6) 2000-2012
Male	0.017***	0.025***	0.024***	0.022***	0.023***	0.022***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Aligned score			−0.015***	−0.015***	−0.015***	−0.015***
			(0.001)	(0.001)	(0.001)	(0.001)
Age				0.026**	0.048***	0.035***
				(0.012)	(0.014)	(0.006)
Tenure					−0.087*	−0.047*
					(0.050)	(0.026)
Time effects	No	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	No	Yes
Observations	362,898	362,898	362,898	362,898	362,898	362,871
Pseudo R^2	0.00	0.03	0.05	0.05	0.05	0.06

Table IV: **Basic regression analysis with score tables**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan across tables and periods, considering the proprietary score before adjustment. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). The table reports the baseline probability and the increase in the probability of delinquency for males. We control for the aligned score, age and tenure of the loan officer. These variables are divided by 100 We also control for the cohort of the loans using a time dummy in all the regressions except in (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (6) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2012
Table 1: Personal loans	0.148***	0.217***	0.232***	0.233***	0.230***
Non-client (2000-04)	(0.013)	(0.018)	(0.019)	(0.019)	(0.019)
Table 1 × Male	0.008	0.013	0.008	0.007	0.010
	(0.014)	(0.019)	(0.019)	(0.019)	(0.019)
Table 2: Personal loans	0.051***	0.077***	0.086***	0.086***	0.086***
Client (2000-04)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
Table 2 × Male	0.011***	0.018***	0.016***	0.016***	0.015***
	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
Table 3: Mortgage loans	0.145***	0.116***	0.128***	0.128***	0.133***
Non-client (2003-09)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Table 3 × Male	0.067***	0.070***	0.067***	0.068***	0.066***
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
Table 4: Mortgage loans	0.093***	0.074***	0.083***	0.083***	0.084***
Client (2003-09)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Table 4 × Male	0.017***	0.020***	0.018***	0.018***	0.017***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Table 5: Personal loans	0.181***	0.123***	0.135***	0.134***	0.135***
Non-client (2005-2006)	(0.015)	(0.011)	(0.012)	(0.011)	(0.012)
Table 5 × Male	0.104***	0.080***	0.077***	0.078***	0.077***
	(0.016)	(0.012)	(0.012)	(0.012)	(0.012)
Table 6: Personal loans	0.255***	0.171***	0.186***	0.185***	0.185***
Non-client (2006-09)	(0.016)	(0.012)	(0.013)	(0.013)	(0.013)
Table 6 × Male	0.066***	0.044***	0.040***	0.041***	0.042***
	(0.017)	(0.013)	(0.013)	(0.013)	(0.013)
Table 7: Personal loans	0.104***	0.073***	0.082***	0.082***	0.082***
Client (2005-2011)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
Table 7 × Male	0.023***	0.013***	0.011***	0.012***	0.011***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Table 8: Mortgage loans	0.372***	0.383***	0.393***	0.394***	0.394***
Foreigners (2009-13)	(0.044)	(0.043)	(0.044)	(0.044)	(0.043)
Table 8 × Male	0.050	0.054	0.049	0.049	0.048
	(0.049)	(0.050)	(0.050)	(0.050)	(0.050)
Table 9: Personal loans	0.130***	0.146***	0.158***	0.159***	0.159***
Foreigners (2009-13)	(0.021)	(0.023)	(0.024)	(0.024)	(0.024)
Table 9 × Male	0.047	0.058	0.056	0.055	0.054
	(0.024)	(0.026)	(0.026)	(0.026)	(0.027)
Table 10: Personal loans	0.117***	0.135***	0.148***	0.148***	0.147***

Continued on next page

Table IV: **(Continued)**

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2012
Non-client (2010-13)	(0.028)	(0.032)	(0.034)	(0.034)	(0.034)
Table 10 × Male	−0.002	−0.002	−0.004	−0.005	0.002
	(0.032)	(0.036)	(0.037)	(0.037)	(0.037)
Table 11: Mortgage loans	0.122***	0.122***	0.134***	0.134***	0.135***
Client (2010-13)	(0.010)	(0.011)	(0.012)	(0.012)	(0.012)
Table 11 × Male	0.012	0.013	0.011	0.011	0.010
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Table 12: Mortgage loans	0.038***	0.038**	0.043**	0.043**	0.045**
Non-client (2010-13)	(0.019)	(0.019)	(0.021)	(0.021)	(0.022)
Table 12 × Male	0.004	0.004	0.004	0.004	0.002
	(0.023)	(0.023)	(0.023)	(0.024)	(0.024)
Table 13: Personal loans	0.034***	0.033***	0.037***	0.037***	0.037***
Client (2011-13)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Table 13 × Male	0.005	0.005	0.004	0.004	0.005
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Age			0.029***	0.053***	0.038***
			(0.008)	(0.009)	(0.009)
Tenure				−0.092***	−0.042***
				(0.014)	(0.014)
Time effects	No	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	Yes
Observations	362,861	362,861	362,861	362,861	362,834
Pseudo R^2	0.04	0.05	0.05	0.05	0.06

Table V: **Basic regression analysis with adjusted score**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations, in column (7) we consider the period of loans originated before 2009. We control for the adjusted score, and the age and tenure of the loan officer. These variables are divided by 100. We also control for the cohort of the loans using a time dummy in all the regressions except in (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (6) and (7) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2012	(6) 2000-2012	(7) 2000-2008
Male	0.017***	0.025***	0.023***	0.020***	0.021***	0.020***	0.021***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Adjusted Score			−0.066***	−0.067***	−0.067***	−0.067***	−0.068***
			(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Age				0.030***	0.050***	0.034***	0.035***
				(0.010)	(0.016)	(0.008)	(0.008)
Tenure					−0.081	−0.042	−0.044
					(0.049)	(0.026)	(0.029)
Time effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	No	Yes	Yes
Observations	362,898	362,898	362,898	362,898	362,898	362,871	338,978
Pseudo R^2	0.00	0.03	0.10	0.10	0.10	0.10	0.10

Table VI: **Robustness checks**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan as robustness checks of the previous regression table. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations, in column (5) we consider the period of loans originated before 2009. We control for the adjusted score, the age and tenure of the loan officer. Additionally, we control for the interaction between tenure and gender of the loan officer, a dummy indicating weather the loan officer was hired before 1995, and the number of loans. We also control for the cohort of the loans using a time dummy in all the regressions. These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (4) and (5) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2008
Male	0.017***	0.017***	0.015***	0.014***	0.015***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Adjusted Score	−0.067***	−0.067***	−0.066***	−0.067***	−0.068***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Age	0.051***	0.048***	0.026	0.007	0.006
	(0.016)	(0.018)	(0.019)	(0.008)	(0.008)
Tenure	−0.277	−0.278	−0.338*	−0.292*	−0.322
	(0.186)	(0.187)	(0.192)	(0.169)	(0.202)
Tenure × Male	0.200	0.192	0.219	0.207	0.231
	(0.207)	(0.204)	(0.211)	(0.179)	(0.213)
Hired before 1995		0.003	0.005**	0.005***	0.005***
		(0.002)	(0.002)	(0.001)	(0.001)
Number of loans			0.018***	0.021***	0.023***
			(0.004)	(0.002)	(0.002)
Time effects	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	Yes	Yes
Observations	362,898	362,898	362,898	362,871	338,978
Pseudo R^2	0.10	0.10	0.10	0.10	0.10

Table VII: **Regression analysis: Including the Determinants of the Score**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations in column (7) we consider the period of loans originated before 2009. Instead of using the score, in the regressions of columns (1) to (5) we include the set of variables that determine the score used for constructing Figures II and III. These variables are age, marital status, type of contract, loan type, destination of the loan, debt-to-income ratio, loan to value ratio, total debt over wealth, average balance over six months, 6/12 months bank balance ratio, monthly mortgage payment over 6-months average bank account balance, nationality, number of years in the current job, an indicator of client and the number of years as client of the bank. We also control for the cohort of the loans using a time dummy in all the regressions except in (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (5), (6) and (7) we also include geographical controls. In columns (6) and (7) we also include the adjusted score. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2012	(6) 2000-2012	(7) 2000-2008
Male	0.013***	0.018***	0.016***	0.017***	0.016***	0.015***	0.015***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Age			0.014	0.030	0.019*	0.019*	0.022*
			(0.012)	(0.019)	(0.011)	(0.011)	(0.012)
Tenure				−0.061	−0.026	−0.026	−0.029
				(0.055)	(0.034)	(0.033)	(0.036)
Adjusted Score						−0.031***	−0.031***
						(0.001)	(0.001)
Determinants of the Score	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	Yes	Yes	Yes
Observations	330,813	330,813	330,813	330,813	330,796	330,796	317,604
Pseudo R^2	0.13	0.14	0.14	0.14	0.15	0.15	0.15

Table VIII: **Robustness Checks: Including the Determinants of the Score**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan as robustness checks of the previous regression table. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations in column (5) we consider the period of loans originated before 2009. Instead of using the score, in the regressions of columns (1) to (3) we include the set of variables that determine the score used for constructing Figures II and III. These variables are age, marital status, type of contract, loan type, destination of the loan, debt-to-income ratio, loan to value ratio, total debt over wealth, average balance over six months, 6/12 months bank balance ratio, monthly mortgage payment over 6-months average bank account balance, nationality, number of years in the current job, an indicator of client and the number of years as client of the bank. Additionally, we control for the interaction between tenure and gender of the loan officer and a dummy indicating weather the loan officer was hired before 1995. We also control for the cohort of the loans using a time dummy in all the regressions. These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (3), (4) and (5) we also include geographical controls. In columns (4) and (5) we also include the adjusted score. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2008
Male	0.016***	0.016***	0.015***	0.015***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age	0.030	0.030	0.018*	0.020*	0.022*
	(0.018)	(0.019)	(0.011)	(0.011)	(0.011)
Tenure	−0.080	−0.080	−0.061	−0.060	−0.092
	(0.113)	(0.113)	(0.115)	(0.121)	(0.137)
Tenure × Male	0.021	0.022	0.035	0.036	0.064
	(0.141)	(0.138)	(0.126)	(0.128)	(0.146)
Hired before 1995		−0.000	0.001	0.000	0.000
		(0.002)	(0.001)	(0.001)	(0.001)
Adjusted Score				−0.031***	−0.031***
				(0.001)	(0.001)
Determinants of the Score	Yes	Yes	Yes	Yes	Yes
Time effects	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	Yes	Yes	Yes
Observations	330,351	330,351	330,334	330,334	317,460
Pseudo R^2	0.14	0.14	0.15	0.15	0.15

Table IX: **Basic regression analysis with delinquency after 60 days**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan. A loan is considered delinquent if it has missed any payment for more than 60 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations in column (7) we consider the period of loans originated before 2009. We control for the adjusted score, and the age and tenure of the loan officer. We also control for the cohort of the loans using a time dummy in all the regressions except in (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (6) and (7) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2000-2012	2000-2012	2000-2012	2000-2012	2000-2012	2000-2012	2000-2008
Male	0.019***	0.029***	0.026***	0.023***	0.024***	0.022***	0.023***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Adjusted Score			−0.080***	−0.081***	−0.081***	−0.081***	−0.081***
			(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Age				0.036***	0.057***	0.043***	0.048***
				(0.009)	(0.014)	(0.005)	(0.005)
Tenure					−0.081	−0.042	−0.047
					(0.055)	(0.031)	(0.033)
Time effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	No	Yes	Yes
Observations	362,898	362,898	362,898	362,898	362,898	362,871	338,978
Pseudo R^2	0.00	0.03	0.09	0.09	0.09	0.10	0.09

Table X: **Basic regression analysis with delinquency after 30 days**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan. A loan is considered delinquent if it has missed any payment for more than 30 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations in column (7) we consider the period of loans originated before 2009. We control for the adjusted score, and the age and tenure of the loan officer. We also control for the cohort of the loans using a time dummy in all the regressions except in (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (6) and (7) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2000-2012	2000-2012	2000-2012	2000-2012	2000-2012	2000-2012	2000-2008
Male	0.018***	0.031***	0.027***	0.024***	0.024***	0.023***	0.025***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Adjusted Score			−0.106***	−0.106***	−0.106***	−0.106***	−0.106***
			(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Age				0.040***	0.057***	0.051***	0.060***
				(0.011)	(0.016)	(0.013)	(0.012)
Tenure					−0.064	−0.034	−0.049
					(0.051)	(0.036)	(0.037)
Time effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	No	Yes	Yes
Observations	362,898	362,898	362,898	362,898	362,898	362,898	338,996
Pseudo R^2	0.00	0.02	0.09	0.09	0.09	0.09	0.09

Table XI: **Robustness check with delinquency after 60 days**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan as robustness checks of the previous regression table. A loan is considered delinquent if it has missed any payment for more than 60 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations in column (5) we consider the period of loans originated before 2009. We control for the adjusted score, and the age and tenure of the loan officer. Additionally, we control for the interaction between tenure and gender of the loan officer, a dummy indicating weather the loan officer was hired before 1995, and the number of loans. We also control for the cohort of the loans using a time dummy in all the regressions. These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (4) and (5) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1)	(2)	(3)	(4)	(5)
	2000-2012	2000-2012	2000-2012	2000-2012	2000-2008
Male	0.017^{***}	0.017^{***}	0.015^{***}	0.014^{***}	0.014^{***}
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Adjusted Score	-0.081^{***}	-0.081^{***}	-0.080^{***}	-0.081^{***}	-0.081^{***}
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Age	0.058^{***}	0.052^{***}	0.026	0.008	0.012^{**}
	(0.014)	(0.016)	(0.018)	(0.006)	(0.005)
Tenure	-0.410^{*}	-0.413^{*}	-0.482^{**}	-0.452^{**}	-0.466^{**}
	(0.214)	(0.215)	(0.225)	(0.178)	(0.203)
Tenure × Male	0.336	0.317	0.350	0.353^{*}	0.355
	(0.241)	(0.239)	(0.250)	(0.194)	(0.218)
Hired before 1995		0.006^{**}	0.008^{***}	0.008^{***}	0.009^{***}
		(0.003)	(0.002)	(0.001)	(0.001)
Number of loans			0.021^{***}	0.026^{***}	0.029^{***}
			(0.005)	(0.003)	(0.003)
Time effects	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	Yes	Yes
Observations	362,898	362,898	362,898	362,871	338,978
Pseudo R^2	0.09	0.09	0.09	0.10	0.10

Table XII: **Robustness check with delinquency after 30 days**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan as robustness checks of the previous regression table. A loan is considered delinquent if it has missed any payment for more than 30 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations in column (5) we consider the period of loans originated before 2009. We control for the adjusted score, and the age and tenure of the loan officer. Additionally, we control for the interaction between tenure and gender of the loan officer, a dummy indicating weather the loan officer was hired before 1995, and the number of loans. We also control for the cohort of the loans using a time dummy in all the regressions. These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (4) and (5) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2008
Male	0.014^{***}	0.014^{***}	0.012^{**}	0.011^{**}	0.013^{***}
	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)
Adjusted Score	-0.106^{***}	-0.106^{***}	-0.106^{***}	-0.106^{***}	-0.106^{***}
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Age	0.058^{***}	0.050^{***}	0.024	0.013	0.021^{*}
	(0.015)	(0.017)	(0.021)	(0.013)	(0.012)
Tenure	-0.504^{*}	-0.508^{*}	-0.581^{**}	-0.586^{***}	-0.589^{**}
	(0.266)	(0.268)	(0.282)	(0.222)	(0.250)
Tenure $\times$ Male	0.449	0.424	0.458	0.487^{**}	0.468^{*}
	(0.287)	(0.289)	(0.300)	(0.234)	(0.259)
Hired before 1995		0.008^{***}	0.010^{***}	0.009^{***}	0.010^{***}
		(0.003)	(0.003)	(0.002)	(0.002)
Number of loans				0.030^{***}	0.033^{***}
				(0.003)	(0.003)
Time effects	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	Yes	Yes
Observations	362,898	362,898	362,898	362,898	338,996
Pseudo R^2	0.09	0.09	0.09	0.09	0.09

Table XIII: **Regressions using branch fixed effects**

The table shows estimates of a panel data specification including branch fixed effects. Notice that the gender of the loan officer in each branch can change over time. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations in column (5) we consider the period of loans originated before 2009. We control for the adjusted score, and the age and tenure of the loan officer. Additionally, we control for the interaction between tenure and gender of the loan officer and include a constant. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1)	(2)	(3)	(4)	(5)
	2000-2012	2000-2012	2000-2012	2000-2012	2000-2008
Male	0.013***	0.011***	0.011***	0.012***	0.013***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Adjusted Score	−0.058***	−0.058***	−0.058***	−0.058***	−0.057***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age		0.024**	0.030*	0.029*	0.032
		(0.011)	(0.017)	(0.016)	(0.022)
Tenure			−0.018	0.080	0.069
			(0.027)	(0.105)	(0.138)
Tenure × Male				−0.101	−0.094
				(0.118)	(0.144)
Observations	362,898	362,898	362,898	362,898	338,996
R^2	0.04	0.04	0.04	0.04	0.04
Adjusted R^2	0.04	0.04	0.04	0.04	0.04

Table XIV: **Regression Analysis with Loan Type**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). Since in 2009 the bank tight up the control on the approval of loans and prohibited exceptions of the recommendations in column (9) we consider the period of loans originated before 2009. We control for the adjusted score, and the age and tenure of the loan officer. Additionally, we control for the interaction between the gender of the loan officer and the type of loan, and a dummy indicating weather the loan officer was hired before 1995. Also, we control for the cohort of the loans using a time dummy in all the regressions except (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (8) and (9) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2012	(6) 2000-2012	(7) 2000-2012	(8) 2000-2012	(9) 2000-2008
Male	0.017***	0.025***	0.025***	0.023***	0.021***	0.021***	0.021***	0.020***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Type of loan: Mortgage	0.024***	0.010***	0.010***	−0.012*	−0.012*	−0.012*	−0.012	−0.011	−0.011
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male × Mortgage			0.029***	0.024***	0.021***	0.021***	0.022***	0.020***	0.022***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Adjusted Score				−0.067***	−0.068***	−0.068***	−0.067***	−0.068***	−0.069***
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age					0.029***	0.030***	0.046***	0.030***	0.029***
					(0.008)	(0.009)	(0.009)	(0.009)	(0.010)
Hired before 1995						−0.001	0.002	0.003*	0.003**
						(0.001)	(0.002)	(0.002)	(0.002)
Tenure							−0.083***	−0.046***	−0.051***
							(0.014)	(0.015)	(0.015)
Time effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	No	No	No	Yes	Yes
Observations	362,410	362,410	362,410	362,410	362,410	362,410	362,410	362,383	338,819
Pseudo R^2	0.00	0.03	0.03	0.10	0.10	0.10	0.10	0.10	0.10

Table XV: **Calculation of the RAROC: mortgages**

	Male	Female
Financial Income	3.605	3.671
Financial Cost	-2.559	-2.562
Financial Margin	**1.047**	**1.109**
Expected Loss	-324	-284
Operational cost	-246	-246
Return on economic capital	7	7
Net Profit (before taxes)	**485**	**587**
Taxes	-145	-176
Net Profit (after taxes)	339	411
Economic Capital	2.842	2.854
RAROC	**11,94%**	**14,39%**

Table XVI: **Basic regression analysis with the aligned score and exception rate**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). We control for the exception rate, aligned score, and age and tenure of the loan officer. For interpretation of the coefficients, we should notice that these three variables have been divided by 100 to facilitate the reading of the parameters. We also control for the cohort of the loans using a time dummy in all the regressions except in (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (6) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2012	(6) 2000-2012
Male	0.017^{***}	0.024^{***}	0.024^{***}	0.023^{***}	0.023^{***}	0.022^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Exception	-0.014^{**}	0.049^{***}	0.034^{***}	0.032^{***}	0.038^{***}	0.041^{***}
	(0.006)	(0.010)	(0.011)	(0.011)	(0.009)	(0.008)
Aligned score			-0.014^{***}	-0.014^{***}	-0.014^{***}	-0.015^{***}
			(0.001)	(0.001)	(0.001)	(0.001)
Age				0.013	0.038^{**}	0.022^{***}
				(0.010)	(0.015)	(0.006)
Tenure					-0.106^{**}	-0.067^{***}
					(0.045)	(0.021)
Time effects	No	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	No	Yes
Observations	362,660	362,660	362,660	362,660	362,660	362,633
Pseudo R^2	0.00	0.03	0.05	0.05	0.05	0.06

Table XVII: **Basic regression analysis with the aligned score and exception rate interacted with gender**

The table shows estimates of a logit specification of the impact of the gender of the loan officer on the delinquent status of a loan. A loan is considered delinquent if it has missed any payment for more than 90 days since its origination up to the final observation date (December 2012). We control for the interaction between exception rate and male, aligned score, and age and tenure of the loan officer. For interpretation of the coefficients, we should notice that these three variables have been divided by 100 to facilitate the reading of the parameters. We also control for the cohort of the loans using a time dummy in all the regressions except in (1). These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. In specification (6) we also include geographical controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis and are clustered at the geographical level.

	(1) 2000-2012	(2) 2000-2012	(3) 2000-2012	(4) 2000-2012	(5) 2000-2012	(6) 2000-2012
Male	0.022***	−0.019**	−0.006	−0.006	−0.011	−0.013
	(0.006)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Exception x Male	−0.008	0.056***	0.040***	0.039***	0.045***	0.046***
	(0.006)	(0.009)	(0.010)	(0.009)	(0.008)	(0.008)
Aligned score			−0.014***	−0.014***	−0.014***	−0.015***
			(0.001)	(0.001)	(0.001)	(0.001)
Age				0.011	0.036**	0.021***
				(0.009)	(0.015)	(0.006)
Tenure					−0.107**	−0.068***
					(0.045)	(0.022)
Time effects	No	Yes	Yes	Yes	Yes	Yes
Zip code controls	No	No	No	No	No	Yes
Observations	362,660	362,660	362,660	362,660	362,660	362,633
Pseudo R^2	0.00	0.03	0.05	0.05	0.05	0.06

Table XVIII: **Loan approval rate by recommendation**

The table shows estimates of a logit specification of the impact of the recommendation of the application (classified into 5 categories), as well as its interaction with the gender of the loan officer, on the approval rate of the loan for all those loans originated before 2009. In specification (1) we control for the adjusted score, in (2) we add the age of the loan officer, in (3) we add the tenure of the loan officer and in (4) we add the geographical controls. Also, we control for the cohort of the loans using a time dummy in all the regressions. These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

	(1)	(2)	(3)	(4)
	2000-2008	2000-2008	2000-2008	2000-2008
Very positive (A1)	0.996***	0.996***	0.996***	0.996***
	(0.001)	(0.001)	(0.001)	(0.001)
Very positive × Male	0.001	0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Positive (A2)	0.988***	0.988***	0.988***	0.989***
	(0.001)	(0.001)	(0.001)	(0.001)
Positive × Male	0.002**	0.002**	0.002**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Neutral (A3)	0.962***	0.962***	0.962***	0.966***
	(0.003)	(0.003)	(0.003)	(0.003)
Neutral × Male	0.011***	0.011***	0.011***	0.007***
	(0.003)	(0.003)	(0.003)	(0.003)
Negative (D1)	0.895***	0.895***	0.896***	0.912***
	(0.006)	(0.006)	(0.006)	(0.005)
Negative × Male	0.033***	0.033***	0.033***	0.019***
	(0.006)	(0.006)	(0.006)	(0.006)
Very negative (D2)	0.675***	0.675***	0.675***	0.699***
	(0.015)	(0.015)	(0.015)	(0.016)
Very negative × Male	0.058***	0.058***	0.058***	0.041**
	(0.016)	(0.016)	(0.016)	(0.016)
Observations	350,518	350,518	350,518	349,739
Pseudo R^2	0.27	0.27	0.27	0.28

Table XIX: **Loan Override Rate**

The table shows estimates of a logit specification on the probability of overriding the negative recommendation of the expert system for the those loans originated before 2009. In specification (1) we control for the adjusted score, in (2) we add the age of the loan officer, in (3) we add the tenure of the loan officer and in (4) we add the geographical controls. Also, we control for the cohort of the loans using a time dummy in all the regressions. These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

	(1)	(2)	(3)	(4)
	2000-2008	2000-2008	2000-2008	2000-2008
Negative recommendation	0.849***	0.848***	0.850***	0.849***
	(0.006)	(0.006)	(0.006)	(0.006)
Negative recom. × Male	0.051***	0.052***	0.051***	0.051***
	(0.007)	(0.007)	(0.007)	(0.007)
Very negative recommendation	0.690***	0.690***	0.691***	0.692***
	(0.012)	(0.012)	(0.012)	(0.012)
Very negative recom. × Male	0.059***	0.059***	0.058***	0.056***
	(0.012)	(0.012)	(0.012)	(0.012)
Reject recommendation	0.801***	0.798***	0.800***	0.799***
	(0.006)	(0.006)	(0.006)	(0.006)
Reject recommendation × Male	0.055***	0.055***	0.054***	0.053***
	(0.006)	(0.006)	(0.006)	(0.006)
Observations	40,589	40,589	40,589	40,589
Pseudo R^2	0.31	0.31	0.31	0.32

Table XX: **Delinquency rate by recommendation**

The table shows estimates of a logit specification of the impact of the recommendation of the application, as well as its interaction with the gender of the loan officer, on the delinquency rate of the loan for all those loans originated before 2009. In specification (1) we control for the adjusted score, in (2) we add the age of the loan officer, in (3) we add the tenure of the loan officer and in (4) we add the geographical controls. Also, we control for the cohort of the loans using a time dummy in all the regressions. These time dummies are important since they control for the difference between the time of origination of the loans and the time of measurement of the delinquency dummy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

	(1) 2000-2008	(2) 2000-2008	(3) 2000-2008	(4) 2000-2008
Very positive (A1)	0.038***	0.038***	0.038***	0.038***
	(0.002)	(0.002)	(0.002)	(0.002)
Very positive × Male	−0.002	−0.002	−0.002	−0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Positive (A2)	0.104***	0.104***	0.104***	0.104***
	(0.003)	(0.003)	(0.003)	(0.003)
Positive × Male	0.011***	0.011***	0.011***	0.011***
	(0.003)	(0.003)	(0.003)	(0.003)
Neutral (A3)	0.139***	0.139***	0.139***	0.138***
	(0.006)	(0.006)	(0.006)	(0.006)
Neutral × Male	0.009	0.009	0.009	0.010
	(0.006)	(0.006)	(0.006)	(0.006)
Negative (D1)	0.157***	0.157***	0.157***	0.155***
	(0.009)	(0.009)	(0.009)	(0.009)
Negative × Male	0.029***	0.029***	0.029***	0.030***
	(0.009)	(0.009)	(0.009)	(0.009)
Very negative (D1)	0.170***	0.170***	0.169***	0.167***
	(0.015)	(0.015)	(0.015)	(0.015)
Very negative × Male	0.071***	0.071***	0.071***	0.072***
	(0.016)	(0.016)	(0.016)	(0.016)
Observations	338,994	338,994	338,994	338,976
Pseudo R^2	0.10	0.10	0.10	0.08

Table XXI: **Duration Model: Performance and Employment**

The table shows the estimation of the tenure as loan officer as a function of gender, age and the proportion of bad loans (accumulated delinquency rate by each loan officer). Columns (1) to (4) compare different duration model specifications. Notice that the interpretation of the coefficients in the Gamma specification is different from the previous three specifications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. t statistics in parentheses.

	(1) Proportional	(2) Weibull	(3) Gompertz	(4) Gamma
Male	-0.420^{***}	-0.430^{***}	-0.424^{***}	0.224^{***}
	(-3.34)	(-3.33)	(-3.45)	(3.33)
Age	-0.022^{***}	-0.021^{***}	-0.022^{***}	0.011^{***}
	(-3.02)	(-2.92)	(-2.97)	(2.60)
Bad loans	0.029^{***}	0.028^{***}	0.029^{***}	-0.014^{***}
	(5.95)	(5.62)	(5.92)	(-4.45)
Constant		-6.046^{***}	-3.476^{***}	3.152^{***}
		(-17.99)	(-11.17)	(15.05)
Observations	17,100	17,100	17,100	17,100
p		1.928		
γ			.033	
σ				.514

Table XXII: **Duration Model with Interaction: Performance and Employment**

The table shows the estimation of the tenure as loan officer as a function of gender, age, the proportion of bad loans (accumulated delinquency rate by each loan officer) and the interaction of the performance indicator with gender. Columns (1) to (4) compare different duration model specifications. Notice that the interpretation of the coefficients in the Gamma specification is different from the previous three specifications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. t statistics in parentheses.

	(1) Proportional	(2) Weibull	(3) Gompertz	(4) Gamma
Male	−0.175	−0.183	−0.157	0.0949
	(−1.09)	(−1.07)	(−1.02)	(1.08)
Age	−0.022***	−0.021***	−0.021***	0.011**
	(−2.96)	(−2.87)	(−2.89)	(2.52)
Bad loans	0.070***	0.070***	0.075***	−0.036***
	(4.61)	(4.21)	(5.27)	(−4.18)
Bad loans × Male	−0.042***	−0.042**	−0.047***	0.022**
	(−2.64)	(−2.44)	(−3.11)	(2.46)
Constant		−6.294***	−3.764***	3.299***
		(−17.69)	(−11.46)	(14.79)
Observations	17,100	17,100	17,100	17,100
p		1.923		
γ			.033	
σ				.513

Table XXIII: **Duration Model with interaction: Performance and employment**

The table shows the estimation of the tenure as loan officer as a function of gender, age, the proportion of bad loans (accumulated delinquency rate by each loan officer), the interaction of the performance indicator with gender and the average number of loans screened for each loan officer by month. Columns (1) to (4) compare different duration model specifications. Notice that the interpretation of the coefficients in the Gamma specification is different from the previous three specifications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. t statistics in parentheses.

	(1) Proportional	(2) Weibull	(3) Gompertz	(4) Gamma
Male	−0.195	−0.205	−0.185	0.107
	(−1.15)	(−1.14)	(−1.14)	(1.15)
Age	−0.020**	−0.018**	−0.020**	0.0093**
	(−2.54)	(−2.37)	(−2.55)	(2.08)
Bad loans	0.075***	0.075***	0.080***	−0.039***
	(4.91)	(4.44)	(5.62)	(−4.42)
Bad loans × Male	−0.044***	−0.044**	−0.048***	0.0231**
	(−2.68)	(−2.43)	(−3.12)	(2.46)
Average loans	0.005*	0.004	0.006**	−0.002
	(1.79)	(1.60)	(2.06)	(−1.64)
Constant		−6.426***	−3.922***	3.409***
		(−17.32)	(−11.40)	(14.11)
N	17,076	17,076	17,076	17,076
p		1.9		
γ			.033	
σ				.515