WORKING PAPER

DANMARKS NATIONALBANK

25 JUNE 2021 — NO. 179

Female business owners pay higher interest rates on corporate loans

Alessandro Tang-Andersen Martinello
alem@nationalbanken.dk
DANMARKS NATIONALBANK

Bjarke Mønsted
bmm@nationalbanken.dk
DANMARKS NATIONALBANK

Rastin Matin
rma@nationalbanken.dk
DANMARKS NATIONALBANK

Sigurd Anders Muus Steffensen
sams@nationalbanken.dk.dk
DANMARKS NATIONALBANK

Karen Holm Laursen
khl@nationalbanken.dk.dk
DANMARKS NATIONALBANK

WORKING PAPER — DANMARKS NATIONALBANK
25 JUNE 2021 — NO. 179

Female business owners pay higher interest rates on corporate loans

Abstract
We analyze micro-level data from the Danish credit register and find that female business owners pay higher interest rates on corporate loans than male owners. The gender gap is partly explained by differences in firm and loan characteristics. However, an economically and statistically significant gap persists even after flexible machine learning techniques are applied to the data. While the gender gap most likely arises during the negotiation process, we do not find that it depends on market power or the extent to which banks use data-driven approaches to determine interest rates.

Resumé
Vi analyserer mikrodata fra det danske kreditregister, og finder at kvindelige virksomhedsejere betaler højere renter af virksomhedslån end mandlige virksomhedsejere. Renteforskellen kan delvist forklares af forskelle i virksomheds- og lånekarakteristika, men en økonomisk og statistik signifikant restmængde består, selv når fleksible machine learning teknikker anvendes på data. Renteforskellen opstår sandsynligvis under forhandlingen af lånet, men vi finder ikke evidens for at bankernes markedskraft eller omfanget af deres brug af interne vurderingmodeller påvirker dens størrelse.

Key words
Financial sector; changes in interest rates; credit risk; statistical methods

JEL classification
G21, J16

Acknowledgements
We are grateful to Thais Lærkholm Jensen, Thomas Sangill, and Søren Korsgaard for comments and guidance, as well as internal reviewers, seminar participants, and members of the research unit at Danmarks Nationalbank for valuable feedback. We also thank representatives from the Danish banking sector for valuable discussions and suggestions. The authors alone are responsible for any remaining errors.

MICRO DATA SCIENCE

Data in new ways

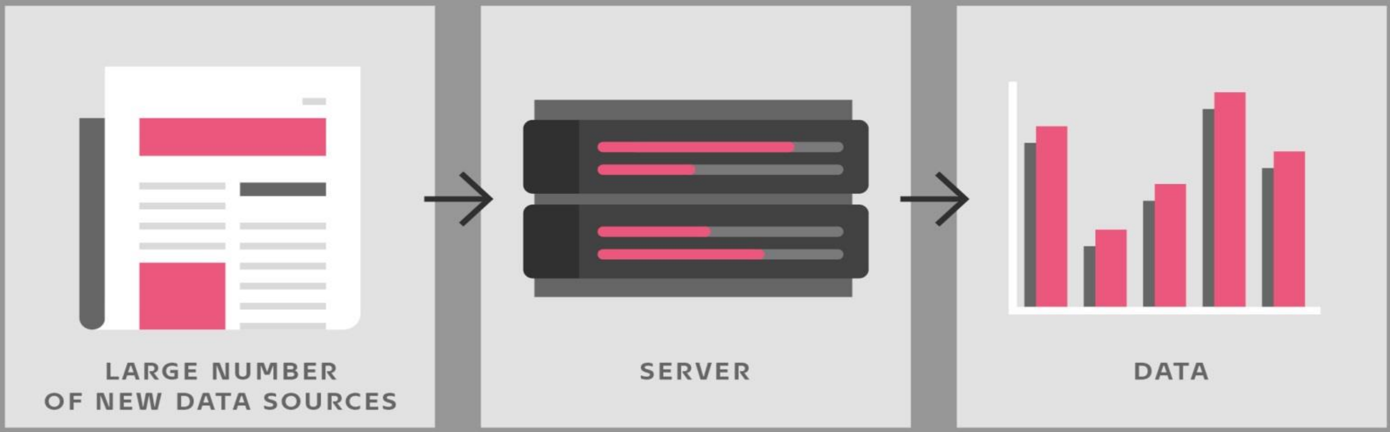
Data volumes have grown exponentially. By 2025, an estimated 450 exabytes of data will be created each day.

This is equivalent to hundreds of millions of personal computers being filled with data on a daily basis. The vast volumes of data are highly diverse, but new and sophisticated methods enable analysis of this data in new and more efficient ways.

New data types and new data collection methods may be used in various contexts in Danmarks Nationalbank's ongoing work.

In order to acquire more knowledge and a better basis for assessing the Danish economy, Danmarks Nationalbank focuses on new data types and methods in a series of publications of which this Working Paper is one.

New data creates new knowledge



PUBLICATIONS

NEWS

News offers quick and accessible insight into an Analysis, an Economic Memo, a Working Paper or a Report from Danmarks Nationalbank. News is published continuously.

ANALYSIS

Analyses from Danmarks Nationalbank focus on economic and financial matters. Some Analyses are published at regular intervals, e.g. *Outlook for the Danish economy* and *Financial stability*. Other Analyses are published continuously.

REPORT

Reports comprise recurring reports and reviews of the functioning of Danmarks Nationalbank and include, for instance, the *Annual report* and the annual publication *Danish government borrowing and debt.*

ECONOMIC MEMO

An Economic Memo is a cross between an Analysis and a Working Paper and often shows the ongoing study of the authors. The publication series is primarily aimed at professionals. Economic Memos are published continuously.

WORKING PAPER

Working Papers present research projects by economists in Danmarks Nationalbank and their associates. The series is primarily targeted at professionals and people with an interest in academia. Working Papers are published continuously.

DANMARKS NATIONALBANK
LANGELINIE ALLÉ 47
DK-2100 COPENHAGEN Ø
WWW.NATIONALBANKEN.DK

DANMARKS NATIONALBANK

1 Introduction

Credit institutions' ability to assess risk and efficiently price loans based on quantitative information has increased in recent years, driven by fast growth in the amount of available data as well as the power of statistical tools such as machine learning. This trend warrants a shift towards objective pricing of credit, and, in a fully competitive market, lower reliance on open negotiations and loan officers' subjective assessment and private information on borrowers (Liberti and Petersen, 2019).

Lower reliance on subjective assessment has the potential of mitigating the disadvantages that minorities face when accessing the credit market, disadvantages that have been documented in recent literature. For example, Cheng et al. (2015) document that black borrowers in the US pay a higher rate on their mortgage than comparable white borrowers. Alesina et al. (2013) show that between 2004 and 2007 Italian female-driven firms paid higher interest rates than their male-driven counterparts. Lower reliance on open negotiations, which might be influenced by implicit biases in the negotiation process, could reduce these differences.

This paper quantifies the adjusted gender gap in effective interest rates experienced by Danish small and medium-sized business owners in 2020, and assesses whether differential adoption of data-driven loan determination across credit institutions can explain differences in the observed gender gap. Exploiting rich micro-level data from the Danish credit register, we assess how interest rates on limited liability companies' loans vary by the gender of their owners, adjusting for a rich set of firm and loan characteristics. Uniquely, for large credit institutions this data also contains the probability of default computed by the credit institution issuing the loan, which should incorporate all information available to the debt issuer about the riskiness of a specific client.

We show that, on average, female business owners pay 98 basis point higher interest rates on their non-mortgage loans. Part of this difference is due to observable firm and loan characteristics. However, even when we exploit a uniquely rich micro-level dataset on loan and firm characteristics, and even when we apply flexible and powerful machine learning algorithms to extract as much information as possible for the data (Chernozhukov et al., 2018), a difference persists in interest rates between female- and male-owned corporate loans of between 28 ± 5 basis points. This difference is statistically significant, and about twice as large as that found by Alesina et al. (2013) in Italy. For the median female-owned firm in our sample, who has approximately €$46{,}000$ in outstanding non-mortgage loans, this gap corresponds to between €106 and €152 in additional interest charges per year paid by female business owners on their mortgage loans.

We provide evidence consistent with the gap occurring during the negotiation process by exploiting data on mortgages, for which we separately observe interest rates and administrative margins. While the administrative margins on Danish corporate mortgages are negotiable, interest rates are completely determined by the market rate of the underlying obligations. Consistently with the gender gap occurring during the negotiation process, we do not find any gender gap for interest rates, and a small yet significant gap for administrative margins (2.34 ± 1.62 basis points). For the median mortgage of €$275{,}000$, this gap corresponds to roughly €64 in additional administrative margins per year.

We do not find evidence that the gender gap depends on relative market power between banks and firms, or the extent to which credit institutions adopt data-driven approaches in their term setting processes.[1] Nonetheless, there can be multiple explanations for the emergence of this gender gap.

[1] We use two proxies for the degree to which credit providers employ data-driven approaches. First, we observe whether a credit institution is systemic or not, as systemic institutions have more resources to develop centralized systems (Berger et al., 2005;

In lab experiments, Borghans et al. (2009) and Croson and Gneezy (2009) show that women are generally more risk averse than men, and might therefore be willing to pay a premium for safer loan products, e.g. with fixed interest rates. Our data do not support this interpretation: once we flexibly account for differences in firm characteristics, female business owners are not systematically more likely to choose fixed over variable interest rate mortgages.

The economic literature also suggests that women are generally less successful bargainers than men, a factor often highlighted also in the vast literature on wage gender gap to which this paper naturally relates.[2] For example, Goldsmith-Pinkham and Shue (2020) show that men earn significantly higher returns on housing investments, partly because men tend to negotiate discounts relative to the list price more often than women. Niederle and Vesterlund (2007) show that women are more likely to avoid competitive situations than men. Accordingly, men might be more likely to consult multiple credit institutions and negotiate more aggressively before accepting a loan offer (Kaman and Hartel, 1994).[3]

Finally, part of the adjusted gender gap could be due to the fact that we do not observe the full relationship between banks and their clients in our data. Because we constrain our sample to limited liability companies, the bank's risk assessment should be based only on characteristics of the firm and not those of the owners. Nevertheless, credit institutions might reserve preferential treatment to clients with a large private economy. Offering convenient corporate loan terms conditional on owners transferring their private economy to the bank is a common strategy. If male business owners have significantly larger private economies so that banks are willing to offer them more aggressive discounts on corporate loans, this practice might contribute to the gender gap we observe in the data.[4]

The remainder of the paper is structured as follows. Section 2 presents our data, and describes how we choose and classify the firm and loan characteristics for which we control in our analysis. Section 3 shows that the gender gap persists in corporate loans and mortgage administrative margins after adjusting for relevant firm and loan characteristics. Section 4 shows that our main results do not depend on geography, bank and firm size, and proxies for how much a credit institution uses data-driven approaches for interest rate setting. Section 5 concludes.

2 Data

The data we use in this study are combined from multiple sources. The main data source is the Danish credit register, which contains detailed information on loans by commercial banks to

Qian et al., 2015). Second, we compute the predictability of interest rates of a specific credit institution from available observable characteristics. We assume that institutions for which we can explain a higher proportion of variance in interest rates from observable firm and loan characteristics are more likely to use data-driven, centralized approaches for loan term setting.

[2] See Blau and Kahn (2017) for a relatively recent review of this literature.

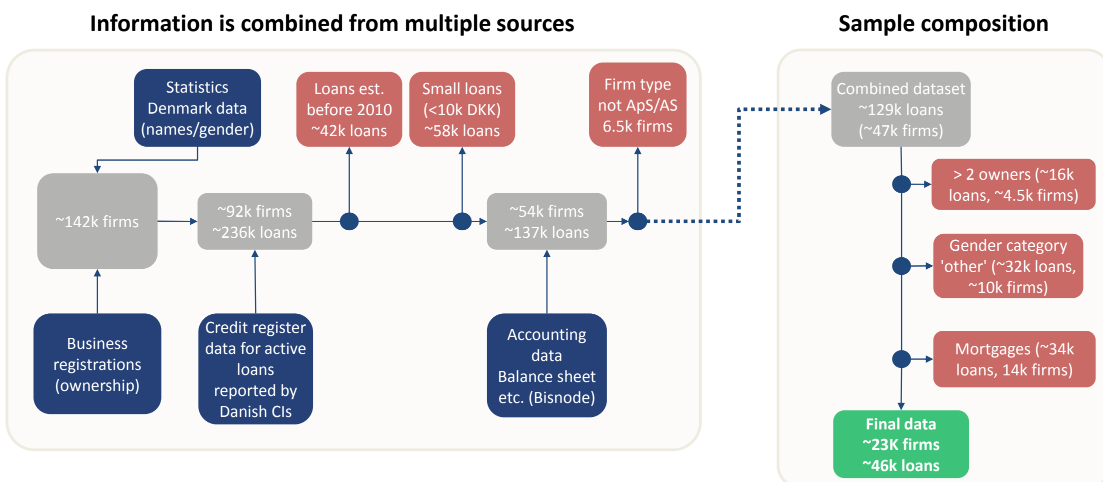
[3] Our data does not allow us to directly test this channel. One way to test this channel would be to assess whether gender gaps vary according to the gender of the bank financial advisor (see, e.g., Bellucci et al., 2010). We cannot, however, directly test this channel, as we do not observe the identity of financial advisors.

[4] Men are generally more likely to actively participate in the stock market, a profitable activity for banks (Almenberg and Dreber, 2015; Bianchi, 2018), and have larger wealth and pension savings in Denmark (Danmarks Nationalbank, 2019). Nonetheless, it is unlikely that this channel can explain the entirety of our results. First, the same bargaining process exists for private products as well, so that business owners can obtain perks on their private accounts by having their firms as clients of the same bank. Given the choice, most business owners would be more interested in pushing costs to a firm and benefits to personal finances in order to benefit from tax deductions. Second, firm economies are typically larger than personal ones, so to make a discount on firm-side products profitable, a bank would have to obtain very high marginal profits on their retail segment.

corporate borrowers.[5] We supplement data from the credit register with balance sheet data from the Danish Central Business Register ("CVR"), which contains accounting data on all businesses in Denmark. Although publicly available, we acquire these data through Bisnode, which provides the data in tabular format. Further, we obtain data on firm structure, such as number of employees and the names and ownership shares of the owners, through the Danish Business Authority. This information allows us to infer the gender of the owners by using public information from Statistics Denmark on gender-specific name frequencies. As name is the only owner-characteristics we observe, gender is the only variable we can infer.

The entire dataflow is summarized graphically in figure 1, and the method for inferring the genders of firm ownership is explained in appendix A.

Figure 1
Data Sources and Pipeline



NOTE: The figure illustrates the dataflow and sample selection. Gender breakdowns of business ownership shares are inferred using data from Statistics Denmark and the Danish Business Authority. The data are then merged with loan information from the credit register, as well as accounting data from Bisnode, retaining only data on large loans and loans by Danish public limited liability companies (*A/S*) and private liability companies (*ApS*). In the resulting combined dataset, we exclude data on mortgages, firms with more than 2 owners, and firms for which owner gender could not be unambiguously inferred. Se appendix appendix A for more details. Note that the figure shows approximate numbers of firms and loans in each category. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

From Statistics Denmark and the Danish Business Authority data we obtain a total of 142,000 firms with a valid gendered direct ownership. Out of these, only roughly half have an active loan. Our data consist of a snapshot of all active loans in June 2020. It is thus a study on cross-sectional data. Approximately 7,600 loans in our final sample were taken out in 2020, while the most of the loans in our sample were taken out before 2020.[6] We further focus on loans established after 2010 of at least DKK10,000 (approximately €1,350) to limited liability companies.

[5]The establishment of this register follows from the European Central Bank (ECB) requiring members of the euro area to establish central credit registers and to participate in a joint analytical credit database ("AnaCredit"), shared between the member states. The credit register, however, does not contain any information about the borrowers themselves, i.e., it contains neither accounting data nor firm data nor gender information.

[6]Since negotiation may explain the gender gap, 2020 could be considered a special year in the sense that the effect of negotiation could be exacerbated during the covid-19 pandemic. We explicitly provide a robustness test of this in table A.7, indicating that loans

Focusing on loans to limited liability companies ensures that the bank's assessments of the riskiness of the loan depends on firm characteristics and not on the unobserved wealth or private economy of the owners. While credit institutions might still require firm owners to provide collateral for loans from their personal properties, in these case the nature and size of the collateral is reported in the credit register, and we are therefore able to adjust for collateral characteristics. We also exclude firms with more than two direct owners, and firms for which a gender assessment based on their name is too imprecise, or firms owned by both male and female owners.[7]

Our final sample consists of 45,962 non-mortgage loans among 23,134 firms. The average effective interest rate for non-mortgage loans is 3.40 percent. While Alesina et al. (2013) focus on credit lines and overdraft facilities, our sample contains several loan types: 11.8 percent are credit lines, while financial leasing and overdraft facilities account for 39.1 percent and 22.6 percent of our sample, respectively. Other, more specific types beyond these account for 26.3 percent of the sample.[8] We also observe 34,237 mortgages, which we analyze separately. The average interest rate in the sample is 0.43 percent, excluding mortgage administrative margins.

Establishing direct gendered ownership in practice excludes all firms with complex company structures. As a consequence, the firms in our sample are representative of small and medium-sized enterprises: While the largest firms we observe have more than 200 employees, and more than one billion euros in assets, the median firm in our sample has three employees, and approximately €150,000 in assets. If we estimate gender as described in appendix A, women made up 19.4 percent of owners, holding 17.7 percent of total ownership shares, making the gap in firm ownership in our sample somewhat larger than in overall business ownership.[9] About 10 percent of firms in the sample are solely owned by women, while the fraction of loans to female-owned firms is lower, at about 8 percent. While female-owned businesses holding fewer loans than male-owned ones might imply a slightly higher denial ratio for female-owned businesses, we do not have the loan application data that would allow us to investigate this hypothesis. This pattern appears also in the data used by Alesina et al. (2013).

The bottom panel of figure 2 shows that female-owned firms tend to be smaller than male-owned firms in terms of both number of employees and asset value. Firm size is the single biggest difference between male- and female-owned firms. There are otherwise no particularly noticeable differences across key performance indicators. Yet, on average, female-owned firms pay 0.98 percentage points higher interest rates on their loans. This result is not due to pockets of outliers. Figure A.3 in the Appendix plots the distribution of our main outcome variables for male- and female-owned firms.[10] The figure demonstrates that for non-mortgage loans, the gender gap in interest rates is due to females owning fewer debts priced below approximately 2.5 percent, and relatively more debts prices at above five percent.

taken out during 2020 had a slightly higher gap, but the difference is not statistically significant. Loans taken up before 2020 alone still have a gender gap of 22 basis points.

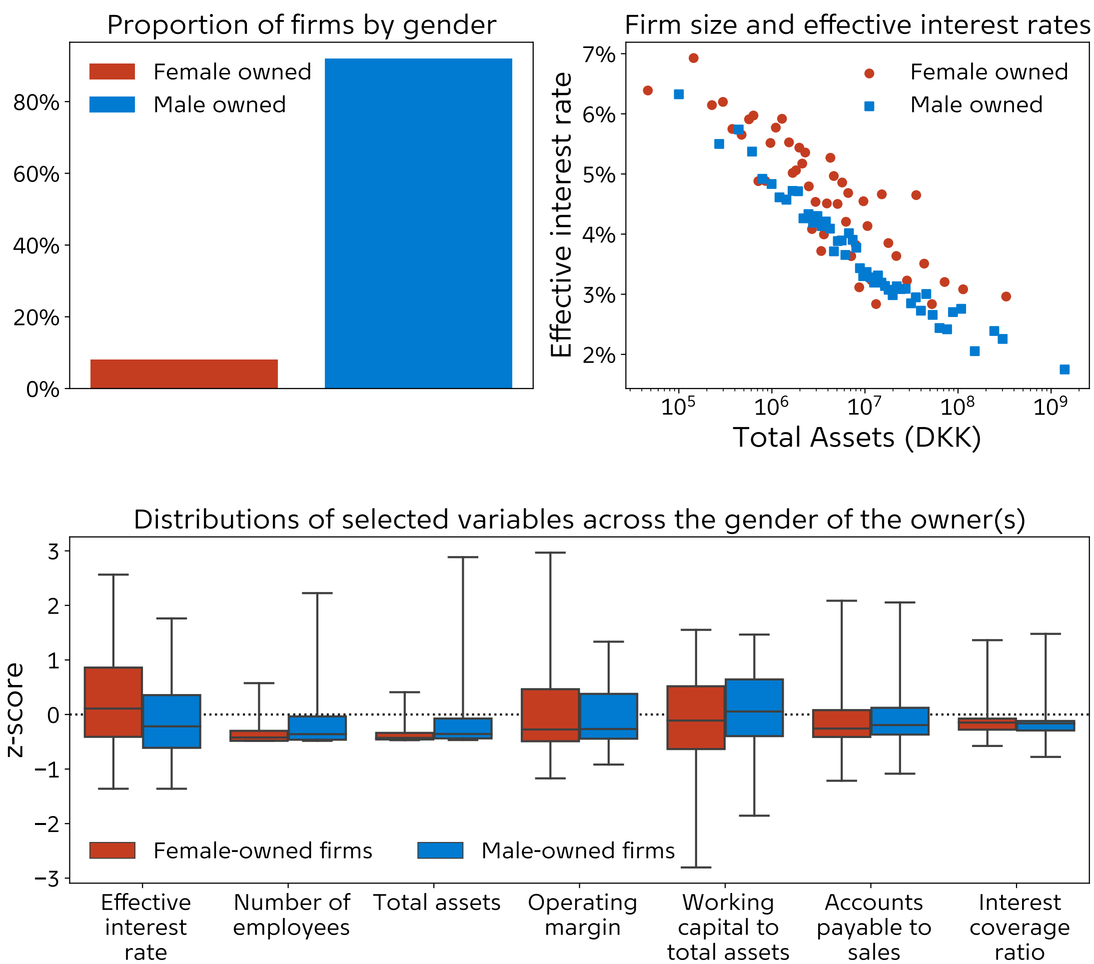
[7]As a threshold, we demand that 95 percent of shares must be estimated as held by male/female owners for the firm to be denoted as male-/female-owned. Details regarding this process are provided in appendix A.

[8]The loan types in this sample are summarized in table A.3 under "Loan type" in the appendix. We do not find any significant differences in the gender gap across loan types. These results appear in appendix table A.6.

[9]In 2019, 31 percent of all Danish self-employed persons were women. See table RAS300 from Statistics Denmark at https://www.statistikbanken.dk/.

[10]Figure A.4 replicates the firm for the raw sample of firms for which we can impute a gendered ownership, and thus includes debts held by non-limited liability companies, contracted before 2010, and smaller than DKK10,000. Particularly due to this last factor, the raw data is characterized by a much fatter right tail, as effective interest rates are not very meaningful measures for very small loans.

Figure 2
Overview of Key Descriptive Statistics



NOTE: Descriptives for selected variables. For the box plot, the variables have been standardized to their z-score. To avoid outliers, the whiskers in the box plot indicate the 5th and 95th percentiles. For the scatterplot, male- and female-owned firms have each been binned into 50 groups based on firm size, as measured by total assets. Each point represents the mean of the firm sizes and interest rates in each such bin. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

The gap is also not confined to small firms only. The top-right panel of figure 2 plots interest rates paid by male- and female-owned firms across firm total assets. Each dot in the graph represents 2 percent of the male- and female-owned firms. While smaller firms pay higher interest rates on average, a gender gap appears across all firm sizes. Figure A.2 in the appendix shows that the same is true for firms across geographical areas, ranging from the centers of large cities to rural areas, and that interest rates are higher in rural areas.

2.1 Measuring Credit Risk

A significant risk to banks is that debtors may be unable to meet their financial obligations, requiring banks to charge risk premia as compensation for the likelihood of default of individual loans. Measuring and modeling this expected default is therefore a key exercise in attempting to explain the observed gender gap, as one possible explanation for its existence is that women are generally more risky borrowers.

From a standard asset pricing perspective, the required return on debt is the risk-free rate for the term of the debt plus a default premium (Altman and Sabato, 2013). As such, the effective interest rate paid on a loan by firm i can be expressed as the risk-free rate for the term plus a firm-specific default premium, $r_i = r_{rf} + \mathbb{E}\left[DF_i\right]$. The default premium $\mathbb{E}[DF_i]$ measures the credit risk of the loan, i.e., the risk of not receiving timely interest and installment payments.

We employ a rich dataset of features to capture loan credit risk that closely reflects the actual data available to banks when setting interest rates. First, we consider a range of features that pertain broadly to firm characteristics such as age, sector and geography. We also uniquely observe the probability of default as estimated by the relevant credit institution. This feature likely incorporates a credit institution's private information about a firm, and thereby enables us to indirectly account for all information that credit institutions use to evaluate firms' credit risk.

We refer to Appendix B for a detailed overview of all features as well as descriptive statistics. Further, we supplement this information with a range of financial ratios computed using variables from firms' income statements and balance sheets. Table 1 shows the list of key ratios that we additionally construct as indicators of firm default risk from accounting variables, consistent with Penman (2010) and Altman and Sabato (2013). All these financial ratios have been applied in previous studies of firm defaults, and their expected effect on credit risk is both intuitive and discussed thoroughly in the literature, see, e.g., Ohlson (1980), Shumway (2001), Lando and Nielsen (2010), and Jensen et al. (2016).

Table 1 groups these features into the following three categories: (i) short-term liquidity, (ii) long-term solvency, and (iii) operating ratios, as is common in the literature. Short-term liquidity ratios signal the firm's ability to have enough cash to repay its short-term obligations. For example, current assets are those which are expected to generate cash within one year while current liabilities are obligations that mature within one year. As such, current ratio measures the ability of the firm to meet its short-term obligations using short-term assets. Long-term solvency ratios signal the firm's ability to meet its obligations in the more distant future. These ratios do not only signal the firm's solvency, but also the firm's debt capacity, i.e., whether the firm has capacity to take on more debt in its capital structure. Finally, operating ratios are included as a measure of debt risk, as liquidity and solvency are largely driven by the outcome of firms' operations.[11]

Table 1
Credit Risk Variables

Name	Definition	Description
Panel A: Short-Term Liquidity		
Current ratio	Current assets / current liabilities	The ability of current assets to pay off current liabilities
Cash ratio	Cash / current liabilities	Ability of cash to pay off near-term liabilities
Quick ratio (acid test)	(Cash + accounts receivable) / current liabilities	Ability of near-term assets to pay off current liabilities
Working capital to total assets	(Current assets - current liabilities) / total assets	Net liquid assets to total assets of the firm
Short-term debt ratio	Short-term debt / total equity	Measures short-term debt relative to equity
Panel B: Long-Term Solvency		
Debt-to-equity	Total debt / total equity	Measures the proportion of debt and equity used to finance the firm's assets
Long-term debt ratio	Long-term debt / (long-term debt + total equity)	Ability to meet long-term debt obligations
Interest coverage ratio	EBIT / net interest expense	How many times operating earnings cover interest requirements
Accounts payable to sales	Accounts payable / sales	How much of suppliers' money does the firm use to support its sales
Cash asset ratio	Cash / total assets	Measures the proportion of firm assets held as liquid funds
Panel C: Profitability		
Return on assets	(Net income + interest expense) / total assets	The effectiveness of the firm in generating return on investments in assets
Operating margin	EBIT / sales	Measures the profitability of the firm's sales
Net profit margin	Net income / sales	The amount of profit generated from revenue
Return on equity	Net income / total equity	Measures how much profit the firm generates relative to its capital
Cash turnover ratio	Cash / sales	How many times cash is turned over in an accounting period
Asset turnover ratio	Sales / total assets	The efficiency with which the firm uses assets to produce sales

Overview of credit risk indicators. Panels A and B show the ratios capturing the short-term liquidity and long-term solvency of the firm, while Panel C gives the ratios capturing overall firm profitability. The ratios capture crucial information from firm financial statements and balance sheets that signal the future likelihood of defaults.

[11] As an example, we include return on assets (ROA), which measures the effectiveness of firms in generating return on investments in assets. A higher ROA entails greater profitability of the firm, i.e., higher revenue entails a better ability to meet financial obligations and thus a lower default risk.

2.2 Classifying Controls According to a Causal Inference Approach

Through our data, we can observe a rich set of both firm and loan characteristics. Yet, a kitchen sink approach that simultaneously attempts to adjust for all these characteristics at once would harm our results' transparency and consistency. First, such an approach would not allow us to pinpoint the background features responsible for the majority of the gender gap and shed light on its sources. Second, blindly including additional variables in a regression would not guarantee a bias reduction. On the contrary, a situation where the added variables acted as colliders/bad controls would bias our estimate (Hünermund and Bareinboim, 2019; Pearl, 1995).

Our goal is to estimate the adjusted gender bias from a positive, descriptive perspective. In this respect, the terminology and language from the field of causal inference helps clarify the role of observable characteristics in the process of interest rate determination (Pearl, 1995; Hünermund and Bareinboim, 2019).

Confounders represent underlying factors that affect both outcome and treatment variables of interest. Adjusting for these factors mitigates (and eliminates, if all confounding factors can be accounted for) omitted variable bias. One example of this variable is firm denomination. In our sample, 87 percent of female-owned firms are private limited liability company (*ApS*), compared to 68 percent of male-owned firms. At the same time, interest rates applied to private limited liability companies are on average much higher than those applied to public limited liability companies (*A/S*) in our sample, likely due to the fact that private limited liability companies tend to be smaller. As firm legal denomination is exogenous to the interest rate setting process, legal denomination is a potential confounder for which we want to adjust. The gender gap within private and public limited liability companies is 76 and 47 basis points, respectively, each lower than the unadjusted gender gap of 98 basis points.

Mediators represent channels through which the effect materializes. One example is the type of interest rate of the loan, or other loan characteristics determined during the negotiation process. Gender might have an effect on interest rates because women hold a disproportionate amount of fixed interest rates. The direction of causality goes through a mediator. While we can adjust for these characteristics, the interpretation of results changes. Explaining 100 percent of a gender gap through loan characteristics does not mean the gap does not exist for comparable business owners and is due to external background factors. It merely means that we have identified the channels through which it occurs.

Finally, colliders (sometimes referred to as "bad controls") represent information that, if adjusted for, can introduce additional bias to an estimate. These are typically variables which are a direct consequence of the outcome of interest, such as the balance of an overdraft facility. As an example, imagine two otherwise identical male and female clients. For whatever reason, the female client pays a higher interest rate on her outstanding debt, the same as the interest paid by a third male client with a higher background risk. In this example, there is a positive gender gap in interest rates. Yet, as a consequence of the higher interest rates, the low-risk woman and high-risk man rely less on the overdraft, and at any given time have a lower debt on the balance than the low-risk man. Were we to stratify male and female clients based on their debt balance on their overdraft facility on the assumption that we should compare clients with the same debt level, we would erroneously conclude that there is no gender gap in this example.

While the nature of a specific variable depends crucially on the underlying causal structural model, in order to aid the interpretation of the results, we divide variables in likely confounders, likely mediators, and likely colliders. Within each group we order them according to their explanatory power for non-mortgage interest rates. This approach, novel in the finance literature, ensures the transparency of our results while carefully considering the nature of controls.

Confounders are chiefly firm characteristics, including all variables listed in table 1, as these are determined independently of the loan term negotiation process. Mediators are chiefly loan characteristics, as these are an integral part of negotiations and are often determined at the same time as interest rates.[12] Nonetheless, as our underlying causal model might differ from that of the reader, we use these groupings for guidance only. To maximize the transparency of our approach, we iteratively add all observable controls in the resulting order, and examine how the estimated adjusted gender gap changes as we adjust for more information. In this way, a reader can observe the impact of each control separately, and construct her own interpretation of the results based on her own underlying assumptions and causal model.

3 The Adjusted Gender Gap in Corporate Loan Interest Rates

In this section, we use the micro-level data in an attempt to explain the observed difference in the unconditional average interest rate paid by men and women for corporate loans. We first show that even a standard OLS approach can adjust for approximately half of the unadjusted gender gap. Second, we show that while approaches incorporating machine learning techniques that flexibly capture nonlinearities and interaction terms in the data can account for a higher fraction of the gender gap using the same data, an adjusted gender gap of about 28 basis points remains on non-mortgage loans. Finally, we expand our analysis and show that a gender gap also exists in the administrative margins paid on firms' mortgage loans.

3.1 Basic OLS Results

In this section, we test if banks' expected default, as captured by the features introduced in section 2.1, can explain the gender gap. That is, we assume that the effective interest rate paid by each individual firm is linearly related to various firm-specific characteristics as

$$Y = T\beta + \gamma W + \varepsilon, \qquad \text{with} \quad \mathbb{E}\left[\varepsilon \mid W\right] = 0, \tag{1}$$

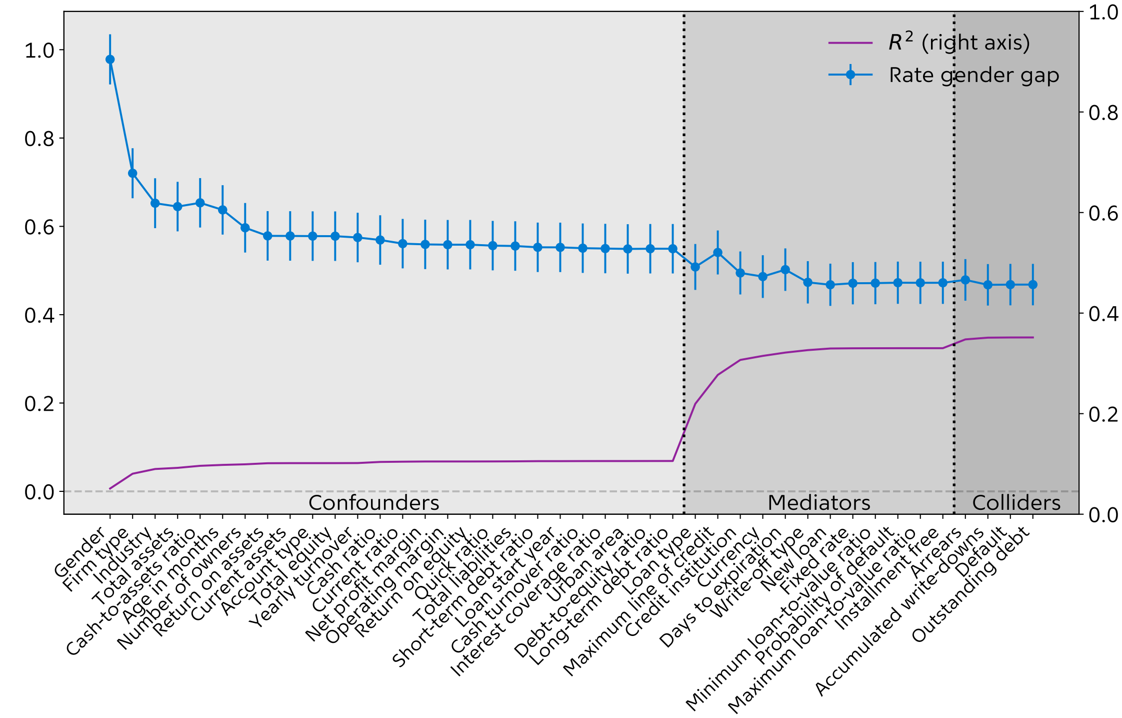
where Y is a matrix containing the effective interest rates on corporate loans, β is a gender dummy variable, and W is a matrix containing the relevant covariates introduced in table 1. We include these variables[13] one at a time in the estimation of the model given by equation (1) and examine how the gender gap evolves as the model accounts for more and more background variables. Within each category (confounders, mediators, and colliders), variables are added according to their marginal contribution in explanatory power for interest rates, measured by R^2. This approach allows identifying the marginal contribution of each variable in adjusting the gender gap in interest rates.

The results from the estimation are presented in figure 3. The inclusion of new variables goes from left to right - initially, we include only a dummy for gender, yielding a difference of about 1

[12] We consider probability of default as a potential mediator, as this variable is computed by the credit institution itself. We have no information on how this quantity is computed for each credit institution, and we therefore cannot exclude that it is influenced by the negotiation process.

[13] Categorical variables are included as dummies.

Figure 3
Adjusted Gender Gap Computed though OLS Regressions



NOTE: The figure shows the size of the gender gap when incrementally including covariates in the linear OLS model in equation (1), as well as the explained variance of the effective interest rate when including each given covariate. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

percentage point in the average interest rate paid by men and women, respectively. This is expressed by the blue, dotted line with the size of the gender gap expressed on the left-hand y-axis. The R^2 statistic shown on the right-hand axis represents the share of total interest rate variance that is explained by the model when including a given covariate as well as all those appearing to the left.

We initially control for firm type and see that this causes a sizeable drop of around 30 basis points in the gender gap. We include in our sample private and public limited liability companies (*ApS* and *A/S*). A noticeable amount of the gender gap is caused by women being significantly more present in the private limited liability companies, which are typically smaller and riskier than public limited liability companies.

Including all the features into the model explains a considerable amount of the gender gap, and drastically improves the predictive performance of the model. Yet, even after controlling for expected defaults, a noticeable gender gap in interest rates of around 50 basis points persists. This amount is not only economically sizeable but also statistically significant.

3.1.1 Kitagawa-Oaxaca-Blinder Decomposition

Our approach of estimating the gender gap through a dummy indicator, adjusting the relationship linearly by a series of control variable, is a common approach in the gender wage gap literature. This approach is taken, for example, by Cook et al. (2018), who compute the adjusted wage gap across male and female Uber drivers. Nevertheless, many economists prefer to identify the adjusted wage gap through Kitagawa-Oaxaca-Blinder decompositions (Kitagawa, 1955; Blau and Kahn, 2017). Interest rates paid by firms can be written as

$$y_j = X_j \beta_j + \varepsilon_j$$

for $j \in \{m, f\}$ indicating whether a firm has male or female owners. The difference in expected interest rates can thus be written as

$$E[y_f] - E[y_m] = E[X_f \beta_f - X_m \beta_m + X_f \beta_m - X_f \beta_m] \tag{2}$$

and rearranging

$$E[y_f] - E[y_m] = \underbrace{E[X_f \beta_m - X_m \beta_m]}_{\text{“Explainable”}} + \underbrace{E[X_f \beta_f - X_f \beta_m]}_{\text{“Unexplainable”}} \tag{3}$$

The first term of equation (3) can be interpreted as the share of the gap due to the difference between female- and male-owned characteristics, if these were priced into interest rates for both categories as if they were owned by firms. The second term of the equation represents the difference by which the same characteristics of female-owned firms are priced into interest rates by the male and female model, and can be interpreted as the adjusted gender gap. We show in table A.4 in the appendix that this approach and that visualized in equation (1) produce virtually identical results.

Nonetheless, a potential weakness of the OLS approach is that control variables enter the model linearly. As a consequence, this approach cannot capture non-linear relationships between controls, gender, and effective interest rates, just as interactions between such variables cannot be captured in this simple linear model.[14] In the following section, we therefore turn to more advanced methods to control for potential non-linear combinations of features.

3.2 Accounting for Non-Linear Effects and Interactions

In this section, we take into account interaction terms and nonlinearities in the data that can potentially explain the presence of the gender gap. While a fully saturated OLS model would be able to flexibly adjust for controls, the sheer number of controls and the fact that many of these are continuous would make such a model computationally unfeasible and prone to overfitting. To avoid this problem, we take the following three approaches. That is, we: (i) hard-code interactions between specific features in the previous OLS model, (ii) apply gradient boosted trees regressions, and (iii) use double machine Learning.

As a first step, we hard-code interactions across the control variables that we expect *a priori* to be relevant for explaining the gender gap. More specifically, we hard-code interactions with the issuing bank and loan type, both of which are both mediating covariates. We interact each of these two control variables with all other confounding and mediating variables. This is denoted by the control variable name appended with * in the following sections. We also interact each of these two control variables with the colliding variables in addition to confounding and mediating variables, and denote them by the control variable name appended with **.

Second, we use gradient boosted trees (XGBoost) regressions (Chen and Guestrin, 2016) to predict effective interest rates without relying on a priori assumptions about the features' relations.

[14] For example, if specific banks have different pricing plans for different types of loans, and female business owners are represented differently in these categories, a standard linear model would not be able to fully capture how loan type and issuing bank affect the gender gap.

XGBoost relies on an ensemble of decision trees, a standard method for classification and regression in machine learning.[15] As decision trees are prone to overfitting, and sensitive to small changes in the input data, the machine learning literature has seen the development of ensemble methods such as random forests (Breiman, 2001), in which a number of trees are fitted independently, using random subsets of the available data, and computing its prediction as a weighted average over the individual decision trees. XGBoost works by fitting decision trees not independently, but iteratively, such that each new tree is fitted to reduce the errors of the previous one.

To assess the effect of gender on interest rates, we first fit the model to part of the data, and then predict the interest rates in the remaining data in two counterfactual scenarios where all firms are male- and female-owned, respectively. Subtracting these predictions gives a measure of the model's estimate of the effect of gender in the out-of-sample data. We run a 5-fold cross validation scheme, in which the model is repeatedly fit to 80 percent of the data, then predict on the remaining 20 percent, covering all of the data. The average of the gender gap in these predictions is the model gender gap, depicted in figure 5. This approach has two disadvantages, however. First, no confidence intervals can be directly obtained from the method, although a measure of the volatility of the inferred gap can be obtained using bootstrap methods. Second, inferring the effect of a single variable using regularized approaches such as XGBoost can result in a downward bias due to regularization (Chernozhukov et al., 2018).

Therefore, we turn to double machine learning (DML) for our third approach (Chernozhukov et al., 2018). Standard machine learning approaches are targeted towards the problem of accurate out-of-sample predictions and forecasting. In contrast, DML is useful when we are interested in a particular model parameter such as a coefficient or a treatment effect while wanting at the same time to adjust for confounding covariates as flexibly as possible. The intuition behind DML is to exploit powerful and flexible machine learning models to orthogonalize the outcome and the main variable of interest, and then estimate the parameters of interest on the orthogonalized residuals in a second stage. Formally, the underlying structural model can be written as:

$$Y = T\beta + g(W) + \varepsilon \qquad \text{with} \quad \mathbb{E}\left[\varepsilon \mid W\right] = 0, \tag{4a}$$

$$T = f(W) + \nu \qquad \text{with} \quad \mathbb{E}\left[\nu \mid W\right] = 0 \quad \text{and} \quad \mathbb{E}\left[\nu \cdot \varepsilon \mid W\right] = 0, \tag{4b}$$

where $f(W)$ and $g(W)$ are arbitrarily complex functions of a set of covariates W, and Y and T, the outcome and treatment variable, respectively (in our case, effective interest rates and gender). We estimate the two first-stage functions $f(\cdot)$ and $g(\cdot)$ via XGBoost. To avoid overfitting, we split our data in two random subsets, and we use models trained on one subset to orthogonalize the other.

DML is in a sense an extension of a standard OLS model. DML would produce the same results as the standard OLS approach if we used unregularized OLS models to estimate $f(W)$ and $g(W)$. However, by splitting the estimation in two stages and using machine learning models to approximate $f(W)$ and $g(W)$, we can flexibly adjust for confounders. Specifically, tree-based machine learning models such as XGBoost perform extremely well when interactions across W can be used to predict T and Y effectively. The DML approach helps us adjust for the set of covariates in our data as flexibly and thoroughly as possible.

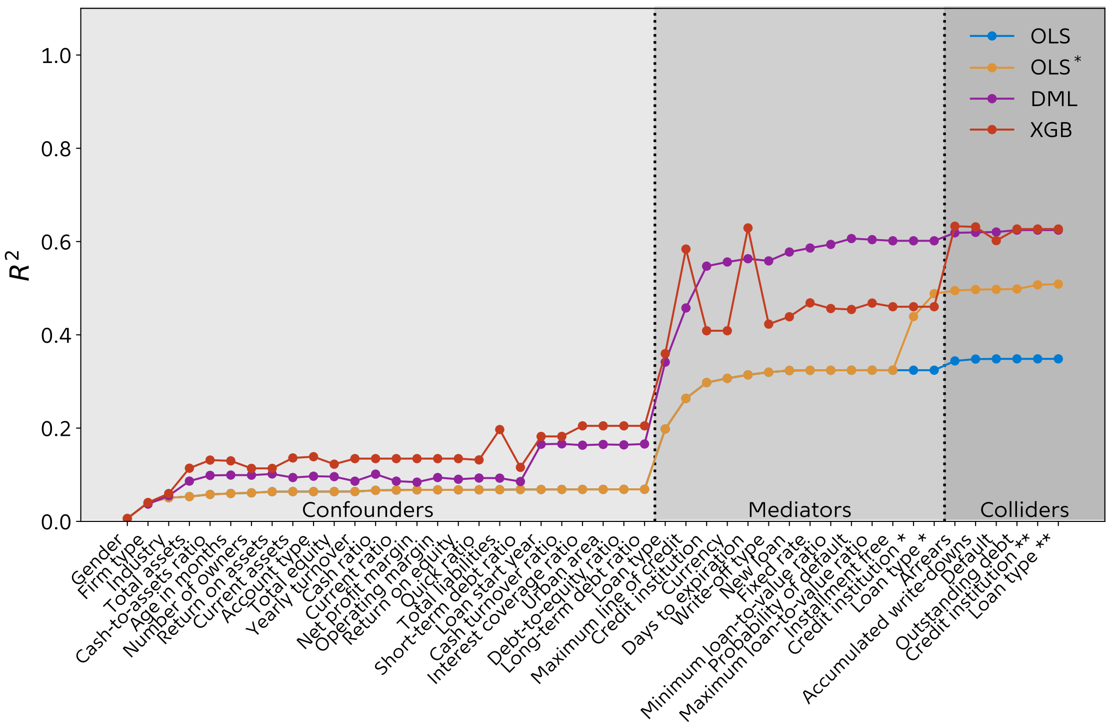
Figure 4 shows that by using the flexible non-linear models we are able to extract much more information from the same set of covariates relative to the simple OLS approach. The figure shows

[15] Regression with decision trees works by iteratively splitting the data into subgroups based on whether selected features exceed certain threshold values. The threshold values and the features based on which the data are split are determined so as to maximize the prediction accuracy when predicting the mean of each subgroup (Breiman et al., 1984).

the development of the R^2 of the estimated models as we add covariates, and shows that for any given set of variables non-linear models are able to explain a much higher fraction of variance in effective interest rates, even when including a large set of interactions in an OLS model.[16]

Overall, non-linear models are able to explain approximately 60 percent of the variance in effective interest rates, compared with approximately 35 percent for a simple OLS model.

Figure 4
Explained Interest Rate Variance, Linear and Non-linear Approaches



NOTE: The figure compares the explained variance (R^2) in interest rates for four different methods: Simple linear regression (OLS), linear regression with interactions (OLS*), XGBoost (XGB), and double machine learning (DML). The y-axis shows the R^2 for each model, as the number of covariates controlled for increases. For the XGB and DML models, the R^2 is computed out-of-sample. For OLS and OLS*, the R^2 is computed on all data. The adjusted R^2 corresponding to the most extensive model is 50.2 percent.

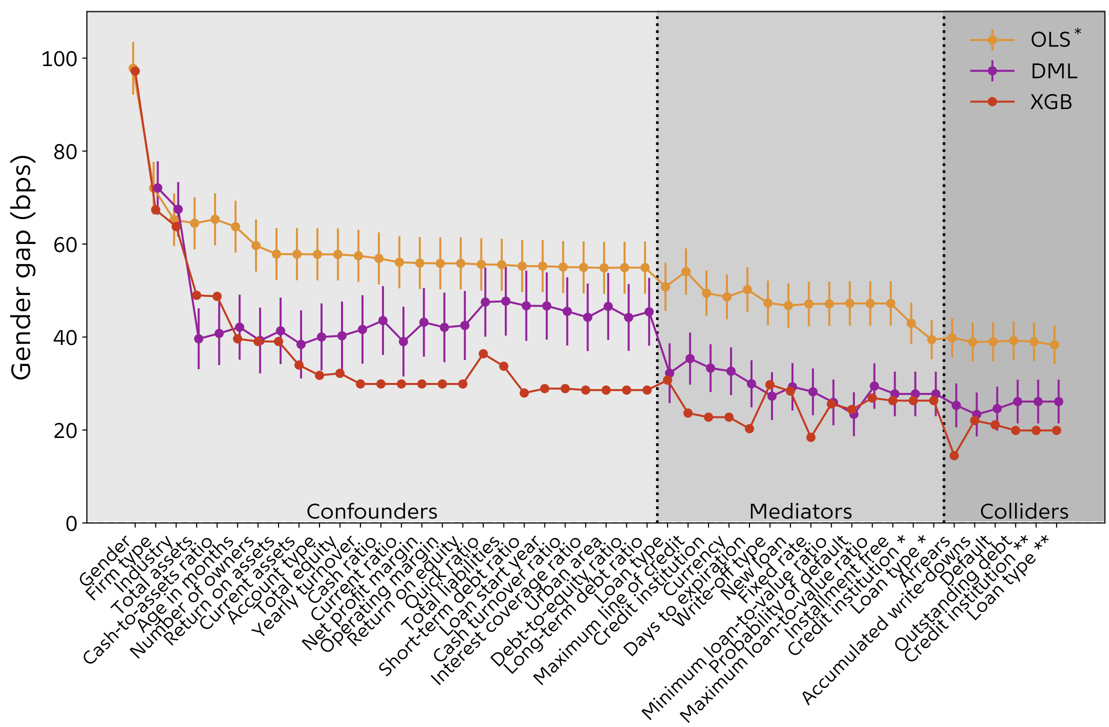
Although explaining a large fraction of variance out-of-sample does not have any direct impact on the ability of the model to adjust the gender gap by itself, these results indicate that the process of setting interest rates by credit institution is complex, and take into account multiple combinations of observable characteristics. If male and female business owners are represented differently across these combinations of observable variables, a linear model would not be able to adjust the gender gap accordingly.

Figure 5 shows that this is indeed the case. For any given combination of controls, the gender gap adjusted non-linearly is smaller than the one obtained by adjusting for covariates linearly, even when

[16]For the OLS models, the reported values for R^2 are simply calculated on all data due to the low number of free parameters in the model relative to the amount of data points. The DML and XGBoost models have a large number of free parameters, so the R^2 values are computed out-of-sample during cross-validation to avoid overfitting. The hard-coded interactions entering the OLS model are denoted by * and **, respectively. This notation signals that for a specific variable, we include all interactions between the starred variable and features appearing to the left of it in the graph.

including a flexible set of interactions. As expected, the gap estimated directly through an XGBoost model is lowest, as it is likely to suffer from regularization bias. However, a DML approach, which both accounts for non-linear adjustments and does not suffer from regularization bias, also estimates a significantly smaller adjusted gender gap than a standard OLS approach does.[17] For all three models, we observe that the gap stagnates once all variables have been included.

Figure 5
Adjusted Gender Gap, Non-linear Approaches



NOTE: Comparison of the magnitude of the interest rate gender gap estimated with three different methods: Linear regression with interactions (OLS*), XGBoost (XGB), and double machine learning (DML). The y-axis shows for each model the estimated rate difference in basis points between firms with male and female owners, as the number of covariates controlled for increases. For OLS*, the gender gap is simply the weight of the gender variable in the OLS model. For the XGB and DML models, the gender gap is computed out-of-sample during cross-validation, as explained in section 3.2. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

Crucially, non-linear models are able to extract enough information to account for the majority of the necessary adjustment of the raw gender gap from a handful of covariates. Firm type, industry, and firm size alone (all characteristics which are arguably exogenous inputs to the loan term setting process) are enough to reduce the adjusted gender gap from 98 to 40 basis points using a DML approach. Inclusion of additional potential confounders does not reduce the adjusted gap further, and inclusion of mediators (which represent the channels through which this gap takes place) only reduces it by additional 12 basis points, bringing it to 28 basis point in total due to difference in loan types by gender. Inclusion of potential colliders that may introduce biases in the estimates does not substantially change our estimates and brings the adjusted gender gap to 26 basis points.

[17]Point estimates and estimated confidence interval appear in table A.5 in appendix C.

This result mirrors the descriptive evidence presented in figure 2, which shows that female and male firms are remarkably similar in terms of key performance indicators, and differ primarily in terms of their size and sector. No other variables can account for adjustments of this magnitude to the gender gap, even though they might be important determinants for interest rates. For example, maximum line of credit alone can explain 12 percent of the variation in interest rates even after accounting for all confounders and loan types. However, as there are no large differences between female and male-owned firms in terms of maximum line of credit, adjusting for this variable has virtually no impact on the adjusted gender gap.

Nonetheless, while the non-linear models are able to capture a much larger fraction of the cross-sectional variation in interest rates relative to the simple OLS model, a small but economically and statistically significant gender gap of around 28 basis points persists. This difference in interest rates cannot be accounted for even by exploiting to the fullest our rich dataset on firm and loan characteristics, which includes most of the information banks use to assess the riskiness of their clients.

3.3 Robustness Using Mortgage Loans

Our main results focus on non-mortgage loans as these contracts are less strictly regulated than mortgage loans and because their interest rates are set partly through bilateral negotiations. However, corporate mortgage loans provide a unique opportunity to test the validity of our empirical results.

For mortgage loans provided to Danish firms, we are able to observe interest rate and administrative margins separately, both expressed as a percentage of the size of the outstanding mortgage. In the Danish system, the interest rate is entirely determined by the market, as it is solely a function of the price of the underlying bonds. On the other hand, mortgage institutions can set different administrative margins according to the type of product, amount of collateral provided, and the underlying risk of the clients. However, for the corporate segment, the administrative margins can to some extent be negotiated in Denmark. To be consistent with our results for non-mortgage loans, we would therefore expect to observe a gender gap in administrative margins and no gender gap for interest rates.

Figure 6 confirms our hypotheses. The figure illustrates our results for mortgage administrative margins and interest rates in the top and bottom panel, respectively, using the DML approach.[18] The top panel shows that even by flexibly adjusting for the same set of covariates used in our previous analysis, there exists a gender gap in administrative margins that is significantly larger than zero.[19] Similarly to the results for loan interest rates, shown in figure 5, the DML method estimates the lowest gender gap relative to OLS and XGBoost, because it can flexibly control for non-linear combinations of covariates, and a few key observable characteristics - firm size in particular - are enough for a non-linear approach to adjust the gap for background characteristics. Using this more sophisticated method does not, however, eliminate the gender gap.

The bottom panel shows that the small initial gender gap in mortgage interest rates disappears once we adjust for observable characteristics using either empirical approach. Specifically, the

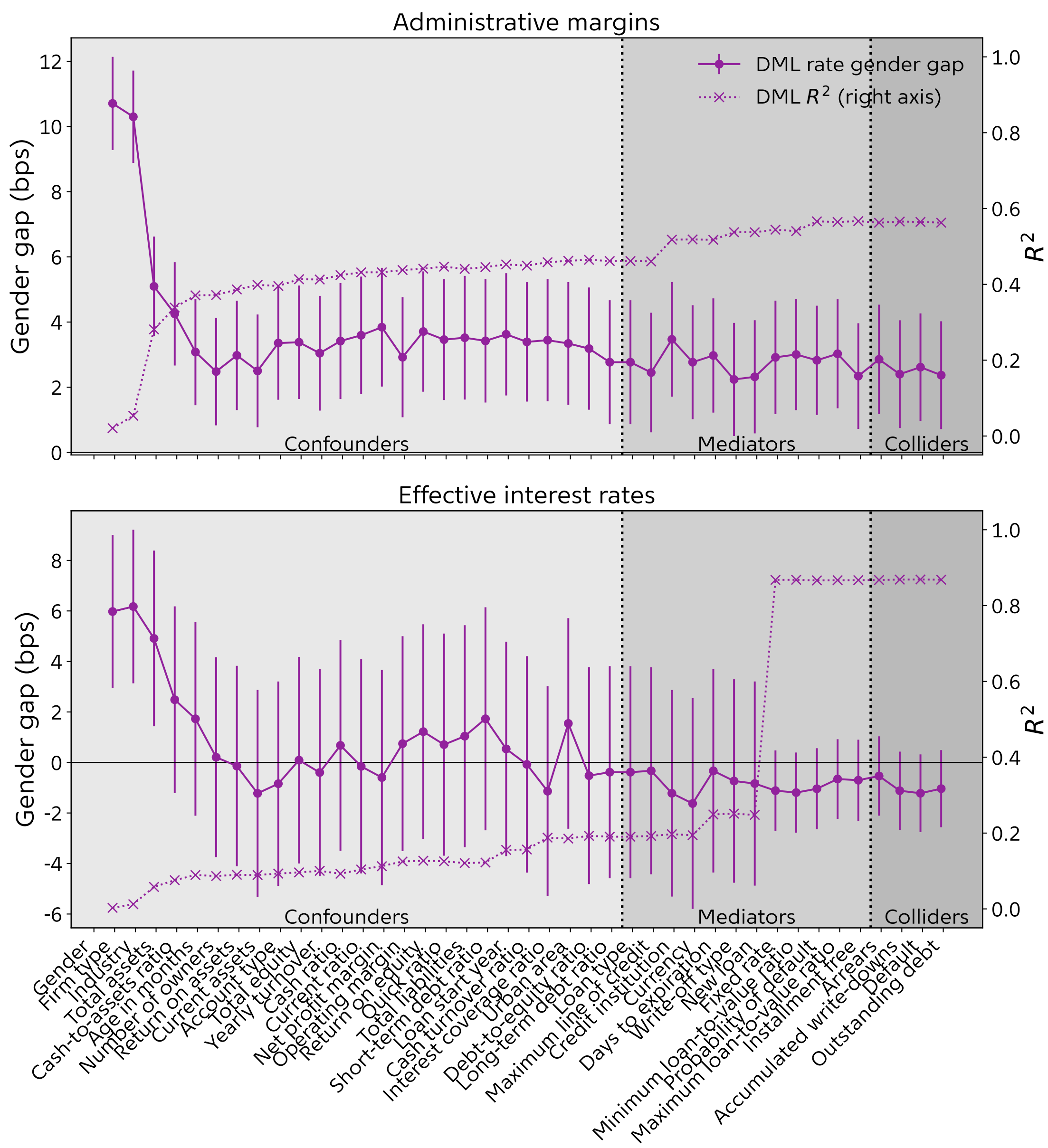
[18] We provide results for DML, as the previous section showed that this approach is the most effective. See appendix F for a comparison of results using the various different methods.

[19] The gender gap in mortgage administrative margins is around 2 basis points and, as such, considerably lower than that discovered in the previous analysis. This does not mean, however, that the gap in mortgage administrative margins is economically insignificant relative to that in corporate loans. Instead, it reflects the fact that mortgage rates tend to be lower and that the products are more highly regulated. As such, any observable difference between the rates paid by men and women in these products should be smaller than that discovered in the previous analysis.

gender gap disappears once we adjust for the fact that female business owners are more likely to hold fixed interest rate mortgages than male owners. As women have been shown to be more risk averse in lab experiments (Croson and Gneezy, 2009), it might be tempting to interpret this result as an indication that the gender gap in mortgage interest rates occurs because female borrowers have a higher degree of risk aversion.

However, while the type of interest applied to a mortgage is certainly the variable responsible for most of the variation in interest rates (the model R^2 increases by more than 60 percentage points for both approaches as soon as this variable is included, and the standard errors associated with the estimates shrink accordingly), our DML results shows that this interpretation would likely be incorrect. Flexibly adjusting for firm characteristics is enough to explain the totality of the gender difference in interest rates, even without accounting for any differences in the characteristics of loans taken across male- and female-owned firms. In other words, within comparable groups of firms, female business owners do not have a particular predisposition for choosing fixed-rate mortgages with respect to males, showing no particular aversion towards interest rate risk in our sample.

Figure 6
Adjusted Gender Gap in Administrative Margins and Interest Rates for Mortgage Loans



NOTE: The effective interest rates on mortgages also exhibit a slight gender gap. However, this disappears entirely when controlling for covariates, in particular whether the mortgage is a fixed rate loan. While the interest rates of mortgages are determined entirely by the market value of the underlying obligation, the administrative margins on mortgages are not fixed in the same manner. The administrative margins on mortgages exhibit a small adjusted gender gap of 2.34 basis points. While lower than the gap in corporate loan interest rates, the median mortgage is approximately 6 times the median corporate loan. Our sample consists of 34,237 mortgage loans among 13,894 firms, 1,213 of which are female-owned.

4 Heterogeneity Analysis

In the previous sections, we have shown that female business owners pay more in effective interest rates for both non-mortgage and mortgage products in Denmark, even when we flexibly adjust for a wide range of information about firm and loan characteristics. However, the gender gap itself might differ between different groups of firms. Identifying observable differences in the gender gap allows for a better understanding of the channels and mechanisms leading to this result.

In this section, we investigate two potential channels for the presence of the gender gap in non-mortgage loans. First, relative negotiation power between credit institutions and clients may play a role in the size of the gender gap. If our results are due to credit institutions being able to disproportionately extract more profits from female clients at the negotiation stage, the effect should be stronger when rent extraction opportunities are highest. In other words, we expect gaps in interest rates to be exacerbated in situations where credit institutions hold relatively more market power, and flattened otherwise. Consistent with the literature, we expect credit institutions to wield relatively more market power when dealing with small firms that are located in rural areas (Agarwal and Hauswald, 2010).

Second, we expect differences across genders to be smaller when banks adopt a greater degree of automated credit assessment procedures based on hard, i.e. quantifiable, background information, so that negotiation with bank advisors are less central for the determination of interest rates. One might expect large credit providers to exhibit greater model precision (R^2), and smaller gender gaps, as they tend to rely more on hard information compared to small providers (Liberti and Petersen, 2019). We assess this in table 2 and the subsequent discussion.

We also construct a proxy for the usage of hard information for single credit institutions according to the fraction of the variation in interest rates that we are able to explain for each single credit provider with all the data at our disposal through a machine learning model (XGBoost). The more variance we can explain, the more we expect that interest rates are set primarily using deterministic procedures.

According to these channels, we would expect the gender gap to be higher for small firms that are located in urban areas, and smaller for large systemic credit institutions and for credit institutions for which we compute a high predictability proxy with respect to the baseline. Table 2 shows the difference in gender gap within each of these comparisons, estimated both by a simple OLS and through a flexible DML approach where we control for potential confounders and mediators. While the results obtained through OLS confirm our priors, with the exception of credit institution size, the estimated differences become much smaller and statistically insignificant once we adjust flexibly for covariates.

As shown in table 2, the magnitude of the gender gap does not significantly depend on whether the credit provider is systemic, or exhibits an above-median predictability of interest rates. A more granular analysis confirms this result. For each of the 15 largest credit providers, we estimate the gender gap and predictability (R^2) using the XGBoost method. Comparing the gender gaps and predictability scores using Spearman's rank-order correlation test (Kokoska and Zwillinger, 2000, sec. 14.7), showed no significant correlation ($r = -0.01, p = 0.97$), nor was the gender gap correlated with credit provider size, as proxied by the number of loans ($r = 0.06, p = 0.84$). Surprisingly, a similar analysis shows that predictability is not significantly related to credit provider size ($r = -0.03, p = 0.95$).

Table 2
Differences in Gender Gaps Across Observable Groups Disappear Once We Flexibly Control for Covariates

	OLS			DML		
	Difference to baseline	Standard error	t-stat	Difference to baseline	Standard error	z-stat
	Over debitor (firm) size					
Firm assets above median	-0.2517	0.103	-2.442	-0.005	0.119	-0.045
	Over geography (baseline: rural areas)					
Central Copenhagen and Aarhus	-0.1749	0.154	-1.136	0.150	0.182	0.827
Copenhagen and Aarhus suburbs	-0.5120	0.152	-3.376	-0.233	0.174	-1.342
Other towns above 85K population	-0.3072	0.118	-2.608	-0.198	0.141	-1.406
	Over credit provider size					
Small (not systemic) credit provider	-0.3309	0.104	-3.169	-0.041	0.136	-0.300
	Over how much interest rates are explainable					
Bank predictability above median	0.2602	0.097	2.683	-0.060	0.120	-0.500

NOTE: The first and fourth columns of the table show the point estimate of the difference in gender gaps between the group described in the stub and the baseline, which varies for each section. In the first section, the baseline is the interest rate gender gap estimated in rural areas; in the second, that estimated for systemic financial institutions; in the third, that estimated for firms whose assets are below the sample median; in the fourth, that estimated for credit institutions for which we can explain a lower-than-the-median amount of variation in the spread of interest rates through an xgboost model. The first three columns describe the difference in gaps estimated through a simple linear OLS, estimated separately for each group. The last three columns describe the difference in gaps estimated through a DML model, where the second-stage model includes group-level dummies. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

The fact that large and highly predictable credit providers do not exhibit different gender gaps with respect to the rest of the sample is consistent with at least three explanations. First, while one would intuitively expect large institutions to rely less on soft information, data-driven processes might in turn introduce bias through systemic statistical discrimination (Fuster et al., 2017). These two effects might counteract one another, so that no overall interaction occurs between gender gap magnitude and credit provider size and predictability. Second, different reliance on centralized processes for loan term setting might not have any impact on gender gaps, which occurs during the negotiating process regardless the guidance from centralized data-driven systems. Third, there might be no meaningful difference between credit institutions in their reliance on centralized data-driven systems for guidance in loan term settings, with our proxies being too imperfect to capture meaningful signals in the data.

5 Conclusions

In 2020, Danish female business owners paid 98 basis points higher interest rates than male owners on the loans undertaken by their firms. Using a rich dataset of firm and loan characteristics which includes the probability of default computed by the credit institution issuing the loan, we find that even flexible machine learning approaches can only account for a part of this gap. We estimate an adjusted gender gap of 26 basis points on non-mortgage loans, more than twice as large as that documented by Alesina et al. (2013) for overdraft facilities in Italy between 2004 and 2007. This gap results in between €100 and €150 of additional interest costs a year for the median female-owned firm in our sample.

We do not find evidence that the size of the gender gap in non-mortgage loans is related to relative market power or to the use of data-driven approaches for interest rate setting by Danish credit institutions. Specifically, the gender gap does not differ between urban and rural areas, by firm or credit institution size, or by the predictability of interest rates offered by specific credit institutions.

However, we show that this gap is likely to occur during the negotiation process. For mortgage loans, we are able to observe separately interest rates components: One which is determined by the market and depends on the price of the underlying collateralized bonds, and another, the administrative margins, which are typically negotiable for the corporate segment. While the same set of observable characteristics can account for the totality of the gender gap in mortgage interest rates, a sizeable gender gap in administrative margins persists. This gap implies an additional expense of about €64 in mortgage administrative margins for the median female-owned firm in our sample.

Nonetheless, a number of possible explanations might lie behind the gender gap we document. For example, men might be more highly profitable retail clients than women due to e.g. higher and more active participation in the stock market. In this case, banks might offer heavier discounts to male firm owners, conditional on them moving their entire private economy to the credit institution, thus profiting on them as retail clients as well. Alternatively, there might be gender differences in the propensity to negotiate terms and eventually shop around multiple credit institutions for the best offers.

It is beyond the scope of this paper, and beyond the reach of our available data, to test for these channels. In the future, as credit registers grow with time and longer time series of bank-client relationships become available, it will be possible to test whether these gaps are due to men exploiting competition between banks to a higher extent. Nonetheless, this paper provides data-driven evidence on the continued existence of a gender gap in loan lending in the corporate segments, and shows that female business owners might be leaving a disproportionate share of money on the table during negotiations with credit institutions.

References

Agarwal, S. and R. Hauswald (2010). Distance and private information in lending. *The Review of Financial Studies 23*(7), 2757–2788.

Alesina, A. F., F. Lotti, and P. E. Mistrulli (2013). Do women pay more for credit? Evidence from italy. *Journal of the European Economic Association 11*(suppl_1), 45–66.

Almenberg, J. and A. Dreber (2015). Gender, stock market participation and financial literacy. *Economics Letters 137*, 140–142.

Altman, E. I. and G. Sabato (2013). Modeling credit risk for SMEs: Evidence from the US market. In *Managing and Measuring Risk: Emerging Global Standards and Regulations After the Financial Crisis*, pp. 251–279. World Scientific.

Bellucci, A., A. Borisov, and A. Zazzaro (2010). Does gender matter in bank–firm relationships? Evidence from small business lending. *Journal of Banking & Finance 34*(12), 2968–2984.

Berger, A. N., N. H. Miller, M. A. Petersen, R. G. Rajan, and J. C. Stein (2005). Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial economics 76*(2), 237–269.

Bianchi, M. (2018). Financial literacy and portfolio dynamics. *The Journal of Finance 73*(2), 831–859.

Blau, F. D. and L. M. Kahn (2017, September). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Borghans, L., J. J. Heckman, B. H. H. Golsteyn, and H. Meijers (2009, 05). Gender Differences in Risk Aversion and Ambiguity Aversion. *Journal of the European Economic Association 7*(2-3), 649–658.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and regression trees*. CRC press.

Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 785–794.

Cheng, P., Z. Lin, and Y. Liu (2015). Racial discrepancy in mortgage interest rates. *The Journal of Real Estate Finance and Economics 51*(1), 101–120.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Cook, C., R. Diamond, J. Hall, J. A. List, and P. Oyer (2018, June). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Working Paper 24732, National Bureau of Economic Research.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic literature 47*(2), 448–74.

Danmarks Nationalbank (2019, March). Women have lower pension wealth than men. Individual pension statistics, Danmarks Nationalbank.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2017). Predictably unequal? The effects of machine learning on credit markets. CEPR Discussion Papers 12448, C.E.P.R. Discussion Papers.

Goldsmith-Pinkham, P. and K. Shue (2020, March). The gender gap in housing returns. Working Paper 26914, National Bureau of Economic Research.

Hünermund, P. and E. Bareinboim (2019). Causal inference and data-fusion in econometrics. Technical report, arxiv.org.

Jensen, T., D. Lando, and M. Medhat (2016). Cyclicality and firm size in private firm defaults. *International Journal of Central Banking 13*(4), 97–145.

Kaman, V. S. and C. E. Hartel (1994). Gender differences in anticipated pay negotiation strategies and outcomes. *Journal of Business and Psychology 9*(2), 183–197.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association 50*(272), 1168–1194.

Kokoska, S. and D. Zwillinger (2000). *CRC standard probability and statistics tables and formulae*. CRC Press.

Lando, D. and M. S. Nielsen (2010). Correlation in corporate defaults: Contagion or conditional independence? *Journal of Financial Intermediation 19*(3), 355–372.

Liberti, J. M. and M. A. Petersen (2019). Information: Hard and soft. *Review of Corporate Finance Studies 8*(1), 1–41.

Niederle, M. and L. Vesterlund (2007, 08). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research 18*(1), 109–131.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika 82*(4), 669–688.

Penman, S. H. (2010). *Financial statement analysis and security valuation*, Volume 4. McGraw-Hill New York.

Qian, J., P. E. Strahan, and Z. Yang (2015). The impact of incentives and communication costs on information production and use: Evidence from bank lending. *The Journal of Finance 70*(4), 1457–1493.

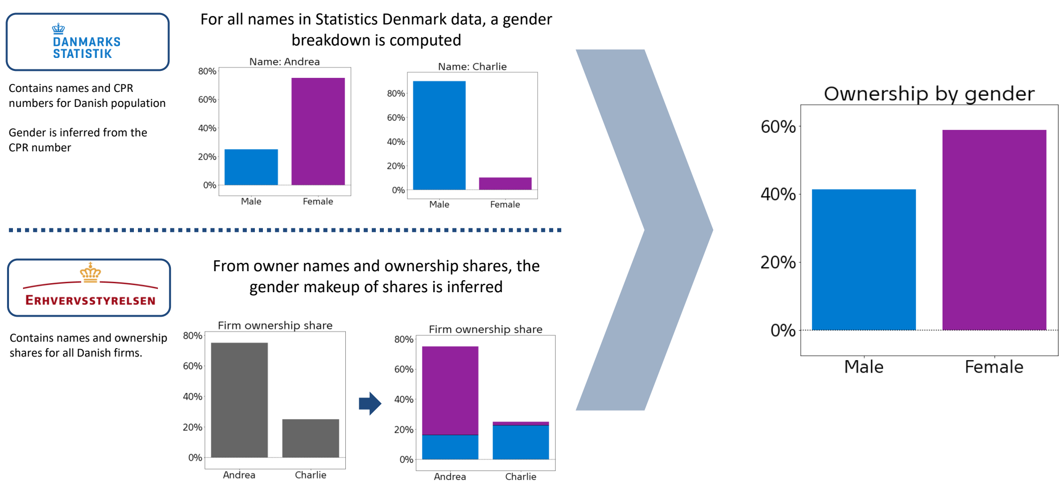
Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business 74*(1), 101–124.

A Appendix: Inferring Gender From Owner Names

The procedure for inferring gender from owner names consists of two steps. The first step is to compute the probability of belonging to each gender for all names in Statistics Denmark, the central authority on Statistics Denmark. This is accomplished by collecting[20] the names and civil registration numbers called ("CPR number") for the entire Danish population. The gender can be inferred from the latter. With this probability distribution at hand, the next step is to extract information on the names and ownership shares of each borrower from the Danish Business Authority ("Erhvervsstyrelsen"). By combining these two components as illustrated in figure A.1, we are able to estimate for each borrower the percentage of total ownership held by male and female owners. Based on this, we partition borrowers into four gender categories:

- "Female owned" for firms with at least 95 percent female ownership
- "Male owned" for firms with at least 95 percent male ownership.
- "Mixed" for firms where both male and female owners hold at least 45 percent
- "Other" for the remaining cases, as well as cases where owner gender could not be inferred

Figure A.1
Inferring Gender From Owner Names



NOTE: Using data from Statistics Denmark, a max-likelihood probability distribution over genders is inferred from names occurring at least five times in the data. Using owner names and ownership shares obtained from the Danish Business Authority, an inferred gender makeup of firm ownership is inferred.

B Appendix: Feature Overview and Descriptive Statistics

This appendix contains an overview of the features included in the models in sections 3 and 3.2. For each variable, some simple descriptive statistics are shown for all firms in our dataset, as well as for male- and female-owned separately. Table A.1 describes the numerical features except accounting variables, which are provided instead in table A.2. Categorical features are shown in table A.3.

[20]Due to privacy issues, we are only allowed to extract this information for names that appear more than five times in the entire population.

Table A.1
Descriptive statistics for non-accounting numerical variables.

Variable	Control type	Gender	NaN	Std	Mean	p10	p50	p90
Effective interest rate	n/a	All	0.0%	3.32	3.73	1.04	3.03	6.7
Effective interest rate	n/a	Male	0.0%	3.24	3.65	1.03	3.03	6.59
Effective interest rate	n/a	Female	0.0%	4.02	4.63	1.2	3.89	8.24
Outstanding debt	colliders	All	0.0%	2e+07	1.7e+06	3.99e+04	2.4e+05	2.64e+06
Outstanding debt	colliders	Male	0.0%	2.08e+07	1.77e+06	4.04e+04	2.42e+05	2.7e+06
Outstanding debt	colliders	Female	0.0%	2.83e+06	9.49e+05	3.35e+04	2.22e+05	1.99e+06
Maximum line of credit	mediators	All	4.5%	3.32e+07	2.24e+06	9.53e+04	4e+05	3.36e+06
Maximum line of credit	mediators	Male	4.6%	3.46e+07	2.33e+06	9.64e+04	4e+05	3.5e+06
Maximum line of credit	mediators	Female	4.1%	3.5e+06	1.25e+06	8.43e+04	3.75e+05	2.58e+06
Accumulated write-downs	colliders	All	1.9%	7.55e+05	7.38e+04	0	471	4.55e+04
Accumulated write-downs	colliders	Male	1.9%	7.68e+05	7.43e+04	0	466	4.45e+04
Accumulated write-downs	colliders	Female	1.7%	5.82e+05	6.88e+04	0	536	6.38e+04
Current assets	confounders	All	0.0%	1.65e+08	2.45e+07	9.1e+04	2.72e+06	3.35e+07
Current assets	confounders	Male	0.0%	1.71e+08	2.61e+07	9.7e+04	3e+06	3.38e+07
Current assets	confounders	Female	0.0%	5.82e+07	6.47e+06	5.2e+04	9.62e+05	1.13e+07
Total liabilities	confounders	All	0.0%	2.94e+08	5.74e+07	8.15e+05	8.52e+06	9.04e+07
Total liabilities	confounders	Male	0.0%	3.05e+08	6.09e+07	8.94e+05	9.15e+06	9.82e+07
Total liabilities	confounders	Female	0.0%	7.18e+07	1.75e+07	4.34e+05	3.51e+06	3.89e+07
Total equity	confounders	All	0.0%	1.38e+08	2.01e+07	-3.6e+04	2.18e+06	2.53e+07
Total equity	confounders	Male	0.0%	1.43e+08	2.13e+07	-9e+03	2.4e+06	2.69e+07
Total equity	confounders	Female	0.0%	5.91e+07	7.08e+06	-2.76e+05	6.42e+05	1.36e+07
Total assets	confounders	All	1.0%	2.68e+08	5.51e+07	8.1e+05	8.53e+06	9.04e+07
Total assets	confounders	Male	1.0%	2.78e+08	5.84e+07	8.9e+05	9.12e+06	9.83e+07
Total assets	confounders	Female	1.2%	7.18e+07	1.74e+07	4.11e+05	3.48e+06	3.89e+07
Yearly turnover	confounders	All	1.0%	2.77e+08	3.16e+07	0	0	0
Yearly turnover	confounders	Male	1.0%	2.88e+08	3.43e+07	0	0	0

Yearly turnover	confounders	Female	1.2%	2.04e+07	1.43e+06	0	0	0
Age in months	confounders	All	0.0%	134	1.21e+03	1.01e+03	1.23e+03	1.35e+03
Age in months	confounders	Male	0.0%	135	1.21e+03	1.01e+03	1.23e+03	1.34e+03
Age in months	confounders	Female	0.0%	126	1.24e+03	1.06e+03	1.28e+03	1.35e+03
Arrears	colliders	All	0.0%	3.69e+05	1.35e+04	0	0	0
Arrears	colliders	Male	0.0%	3.83e+05	1.4e+04	0	0	0
Arrears	colliders	Female	0.0%	1.2e+05	8.37e+03	0	0	0
Probability of default	mediators	All	51.1%	0.196	0.063	0.0005	0.0053	0.107
Probability of default	mediators	Male	50.3%	0.189	0.0589	0.0005	0.005	0.0801
Probability of default	mediators	Female	60.2%	0.273	0.122	0.0009	0.011	0.362
Maximum loan-to-value ratio	mediators	All	10.8%	5e+05	1.38e+04	0	1	3.18
Maximum loan-to-value ratio	mediators	Male	10.5%	5.12e+05	1.33e+04	0	1	3.18
Maximum loan-to-value ratio	mediators	Female	14.5%	3.17e+05	2.04e+04	0	1	3.13
Minimum loan-to-value ratio	mediators	All	10.8%	4.47e+05	1.19e+04	0	0	2.16
Minimum loan-to-value ratio	mediators	Male	10.5%	4.59e+05	1.16e+04	0	0	2.18
Minimum loan-to-value ratio	mediators	Female	14.5%	2.72e+05	1.58e+04	0	0	1.97
Days to expiration	mediators	All	34.5%	1.98e+03	4.36e+04	4.22e+04	4.31e+04	4.52e+04
Days to expiration	mediators	Male	34.1%	1.96e+03	4.36e+04	4.22e+04	4.31e+04	4.51e+04
Days to expiration	mediators	Female	38.4%	2.22e+03	4.39e+04	4.22e+04	4.33e+04	4.65e+04
Loan start year	confounders	All	0.0%	52.7	2.02e+03	2.02e+03	2.02e+03	2.02e+03
Loan start year	confounders	Male	0.0%	54.9	2.02e+03	2.02e+03	2.02e+03	2.02e+03
Loan start year	confounders	Female	0.0%	2.23	2.02e+03	2.01e+03	2.02e+03	2.02e+03
Number of owners	confounders	All	0.0%	0.41	1.21	1	1	2
Number of owners	confounders	Male	0.0%	0.418	1.23	1	1	2
Number of owners	confounders	Female	0.0%	0.258	1.07	1	1	1

Table A.2
Descriptive statistics for accounting numerical variables.

Variable	Control type	Gender	NaN	Std	Mean	p10	p50	p90
Quick ratio	confounders	All	0.2%	35.4	2.81	0.154	1.55	4.41
Quick ratio	confounders	Male	0.2%	36.8	2.84	0.156	1.56	4.39
Quick ratio	confounders	Female	0.2%	10.4	2.42	0.112	1.34	4.71
Cash-to-assets ratio	confounders	All	0.0%	0.152	0.072	0	0.00436	0.241
Cash-to-assets ratio	confounders	Male	0.0%	0.151	0.0712	0	0.00414	0.239
Cash-to-assets ratio	confounders	Female	0.0%	0.161	0.0812	0	0.00665	0.253
Current ratio	confounders	All	0.2%	66.1	3.09	0.0701	1.01	2.39
Current ratio	confounders	Male	0.2%	68.7	3.14	0.072	1.02	2.39
Current ratio	confounders	Female	0.2%	20.6	2.53	0.0594	0.853	2.36
Cash ratio	confounders	All	0.2%	11.6	0.507	0	0.0152	0.618
Cash ratio	confounders	Male	0.2%	12	0.514	0	0.0153	0.61
Cash ratio	confounders	Female	0.2%	3.93	0.421	0	0.015	0.705
Debt-to-equity ratio	confounders	All	0.1%	974	0.557	-1.33	1.99	11.4
Debt-to-equity ratio	confounders	Male	0.1%	1.01e+03	-0.165	-1.07	1.99	11.4
Debt-to-equity ratio	confounders	Female	0.0%	162	8.8	-3.63	1.76	11.8
Long-term debt ratio	confounders	All	0.1%	315	0.435	0	0.191	3.9
Long-term debt ratio	confounders	Male	0.1%	327	0.0621	0	0.204	3.89
Long-term debt ratio	confounders	Female	0.0%	114	4.69	0	0.0366	3.99
Operating margin	confounders	All	1.6%	13.4	0.43	-0.0831	0.203	1
Operating margin	confounders	Male	1.6%	13.4	0.44	-0.0726	0.203	1
Operating margin	confounders	Female	1.7%	12.9	0.317	-0.226	0.197	1
Net profit margin	confounders	All	1.6%	190	-5	-0.478	0.0981	1.03
Net profit margin	confounders	Male	1.6%	198	-5.23	-0.473	0.0997	1.01
Net profit margin	confounders	Female	1.7%	57.6	-2.31	-0.546	0.0769	1.19
Cash turnover ratio	confounders	All	1.6%	44.9	-1.28	0	0.0108	0.424
Cash turnover ratio	confounders	Male	1.6%	45.4	-1.31	0	0.0108	0.424
Cash turnover ratio	confounders	Female	1.7%	38	-0.906	0	0.011	0.424
Return on equity	confounders	All	0.1%	9.32	0.118	-0.182	0.156	0.761
Return on equity	confounders	Male	0.1%	9.6	0.119	-0.166	0.157	0.741

Return on equity	confounders	Female	0.0%	5.09	0.108	-0.398	0.153	0.982
Short-term debt ratio	confounders	All	0.1%	665	0.223	-1	1.21	6.03
Short-term debt ratio	confounders	Male	0.1%	693	-0.108	-0.333	1.23	5.94
Short-term debt ratio	confounders	Female	0.0%	56.7	4	-2.81	0.993	7.39
Return on assets	confounders	All	0.0%	2.04	0.0224	-0.0695	0.0441	0.219
Return on assets	confounders	Male	0.0%	1.13	0.0398	-0.0591	0.0451	0.218
Return on assets	confounders	Female	0.0%	6.1	-0.177	-0.209	0.0368	0.245
Interest coverage ratio	confounders	All	0.3%	29.6	1.14	-0.884	-0.099	0.926
Interest coverage ratio	confounders	Male	0.3%	30.4	1.14	-0.893	-0.101	0.908
Interest coverage ratio	confounders	Female	0.4%	18.6	1.24	-0.744	-0.0665	1.04

Table A.3
Descriptive statistics for categorical variables.

Installment free

Value	Frequency Male	Frequency Female	Frequency All
0	98.2%	97.7%	98.2%
Other	1.8%	2.3%	1.8%

Currency

Value	Frequency Male	Frequency Female	Frequency All
DKK	98.3%	99.0%	98.3%
Other	1.7%	1.0%	1.7%

Fixed rate

Value	Frequency Male	Frequency Female	Frequency All
0	87.7%	90.9%	87.9%
1	12.3%	9.1%	12.1%

Gender

Value	Frequency Male	Frequency Female	Frequency All
Female	0.0%	100.0%	8.1%
Male	100.0%	0.0%	91.9%

Loan type

Value	Frequency Male	Frequency Female	Frequency All
Financial leasing	40.0%	28.0%	39.1%
Other loans	21.8%	28.2%	22.3%
Overdraft facilities	22.2%	26.8%	22.6%
Other	4.2%	4.5%	4.3%

Industry

Value	Frequency Male	Frequency Female	Frequency All
Construction sector	20.0%	11.1%	19.2%
Non-financial holding firms	7.2%	5.7%	7.1%

Real estate	13.5%	13.0%	13.4%
Rental and leasing of motor vehicles	5.8%	2.9%	5.5%
Transportation by land	7.0%	4.3%	6.8%
Wholesale	7.1%	6.9%	7.1%
Other	39.6%	56.1%	40.9%

Default

Value	Frequency Male	 Female	 All
0	93.6%	89.8%	93.3%
1	6.4%	10.2%	6.7%

Write-off type

Value	Frequency Male	 Female	 All
Stage 1 (IFRS/NGAAP)	74.3%	68.7%	73.8%
Stage 2 (IFRS/NGAAP)	17.4%	17.9%	17.5%
Stage 3 (IFRS/NGAAP)	6.3%	11.7%	6.8%
N/A	1.9%	1.7%	1.9%

New loan

Value	Frequency Male	 Female	 All
0	87.3%	86.9%	87.3%
1	12.7%	13.1%	12.7%

Account type

Value	Frequency Male	 Female	 All
Group	2.1%	0.3%	2.0%
Normal	97.9%	99.7%	98.0%

Urban area

Value	Frequency Male	 Female	 All
Lightly populated areas	55.7%	54.9%	55.6%
Central Copenhagen and Aarhus	25.3%	22.0%	25.0%
Copenhagen and Aarhus suburbs	11.7%	11.4%	11.7%
Other towns above 85K population	7.3%	11.7%	7.6%

Firm type

Value	Frequency Male	Female	All
Private limited liability companies (*ApS*)	68.0%	87.3%	69.5%
Public limited liability companies (*A/S*)	32.0%	12.7%	30.5%

C Appendix: Data tables

Point estimates, standard errors, and explained variances for all three methods depicted in figures 4 and 5 are shown in table A.5.

Table A.4
Adjusted Gender Gaps Estimated Through Linear Methods

	OLS		Oaxaca-Blinder decomposition
Controls	Gender gap	Standard error	"Unexplained" component
		Confounders	
Gender	0.978128	0.056780	0.978128
Firm type	0.720278	0.056625	0.722733
Industry	0.652417	0.056496	0.658070
Total assets	0.644764	0.056079	0.651124
Cash-to-assets ratio	0.653322	0.055945	0.657027
Age in months	0.637286	0.055908	0.639715
Number of owners	0.596815	0.056088	0.597642
Return on assets	0.578475	0.056041	0.584810
Current assets	0.578290	0.056035	0.584694
Account type	0.577945	0.056049	0.583883
Total equity	0.577780	0.056049	0.583893
Yearly turnover	0.574771	0.056054	0.581026
Cash ratio	0.569134	0.055981	0.580775
Current ratio	0.560884	0.055986	0.580573
Net profit margin	0.559166	0.055975	0.580133
Operating margin	0.558485	0.055977	0.580041
Return on equity	0.558506	0.055979	0.579478
Quick ratio	0.556364	0.055985	0.579471
Total liabilities	0.555497	0.055980	0.579333
Short-term debt ratio	0.552589	0.055972	0.579273
Loan start year	0.552467	0.055973	0.579137
Cash turnover ratio	0.550597	0.055973	0.579379
Interest coverage ratio	0.549789	0.055974	0.579571
Urban area	0.548914	0.055987	0.578254
Debt-to-equity ratio	0.549364	0.055985	0.578419
Long-term debt ratio	0.549350	0.055984	0.578435
		Mediators	
Loan type	0.508099	0.051966	0.544621
Maximum line of credit	0.540916	0.049804	0.579716
Credit institution	0.494504	0.048669	0.525726
Currency	0.486291	0.048353	0.514922
Days to expiration	0.502005	0.048097	0.529519
Write-off type	0.473285	0.047940	0.493427
New loan	0.467763	0.047813	0.489175
Fixed rate	0.471201	0.047802	0.491905
Minimum loan-to-value ratio	0.471484	0.047797	0.492178
Probability of default	0.472432	0.047796	0.493506
Maximum loan-to-value ratio	0.472194	0.047797	0.493174
Installment free	0.472256	0.047798	0.493199
		Colliders	
Arrears	0.478785	0.047090	0.501434
Accumulated write-downs	0.467635	0.046956	0.490341
Default	0.468108	0.046941	0.490561
Outstanding debt	0.468152	0.046939	0.490577

NOTE: The table shows the effect of gender on interest rates computed as the number of variables we adjust for increases. Each row of the table corresponds to a model in which the corresponding variable and all those appearing above it are adjusted for. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

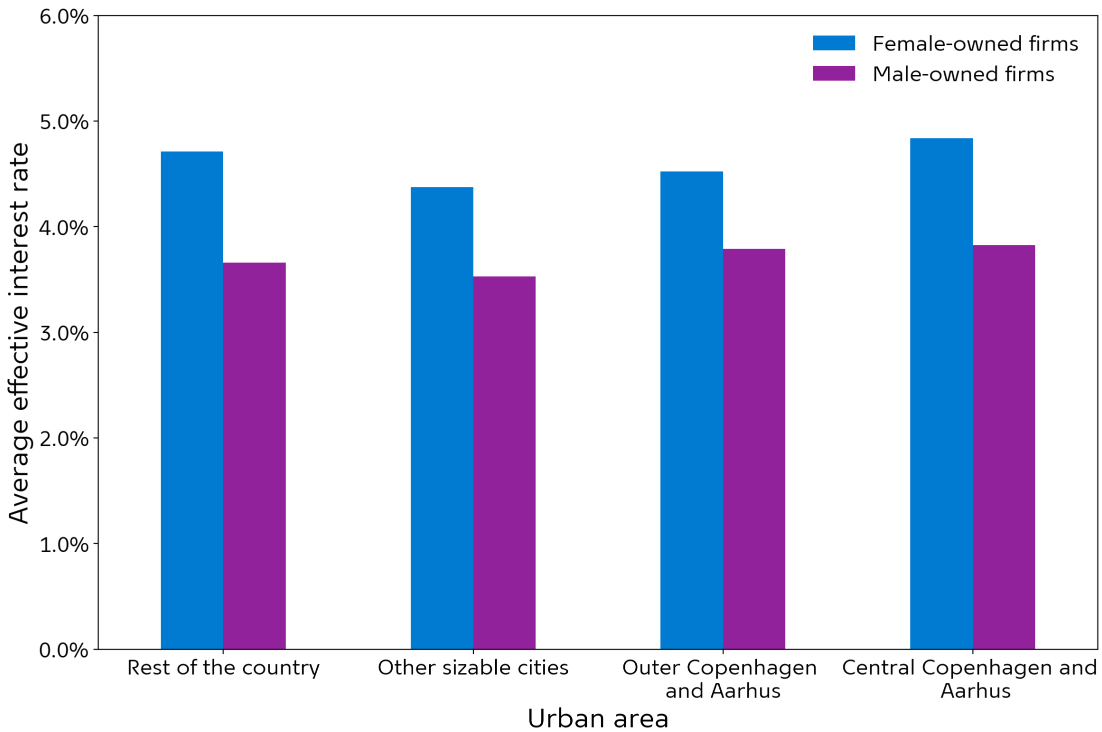
Table A.5
Adjusted Gender Gaps Estimated Through Non-Linear Methods

	OLS*			XGB			DML		
Controls	Gender gap	Standard error	R^2	Gender gap	Standard error	R^2	Gender gap	Standard error	R^2
				Confounders					
Gender	0.98	0.06	0.01	0.97	-	0.01	nan	nan	nan
Firm type	0.72	0.06	0.04	0.67	-	0.04	0.72	0.06	0.04
Industry	0.65	0.06	0.05	0.64	-	0.06	0.67	0.06	0.06
Total assets	0.64	0.06	0.05	0.49	-	0.11	0.40	0.07	0.09
Cash-to-assets ratio	0.65	0.06	0.06	0.49	-	0.13	0.41	0.07	0.10
Age in months	0.64	0.06	0.06	0.40	-	0.13	0.42	0.07	0.10
Number of owners	0.60	0.06	0.06	0.39	-	0.11	0.39	0.07	0.10
Return on assets	0.58	0.06	0.06	0.39	-	0.11	0.41	0.07	0.10
Current assets	0.58	0.06	0.06	0.34	-	0.14	0.38	0.07	0.09
Account type	0.58	0.06	0.06	0.32	-	0.14	0.40	0.07	0.10
Total equity	0.58	0.06	0.06	0.32	-	0.12	0.40	0.07	0.10
Yearly turnover	0.57	0.06	0.06	0.30	-	0.13	0.42	0.07	0.09
Cash ratio	0.57	0.06	0.07	0.30	-	0.13	0.44	0.07	0.10
Current ratio	0.56	0.06	0.07	0.30	-	0.13	0.39	0.08	0.09
Net profit margin	0.56	0.06	0.07	0.30	-	0.13	0.43	0.07	0.08
Operating margin	0.56	0.06	0.07	0.30	-	0.13	0.42	0.07	0.09
Return on equity	0.56	0.06	0.07	0.30	-	0.13	0.42	0.07	0.09
Quick ratio	0.56	0.06	0.07	0.36	-	0.13	0.48	0.07	0.09
Total liabilities	0.56	0.06	0.07	0.34	-	0.20	0.48	0.07	0.09
Short-term debt ratio	0.55	0.06	0.07	0.28	-	0.12	0.47	0.08	0.09
Loan start year	0.55	0.06	0.07	0.29	-	0.18	0.47	0.07	0.17
Cash turnover ratio	0.55	0.06	0.07	0.29	-	0.18	0.46	0.07	0.17
Interest coverage ratio	0.55	0.06	0.07	0.29	-	0.20	0.44	0.07	0.16
Urban area	0.55	0.06	0.07	0.29	-	0.20	0.47	0.07	0.17
Debt-to-equity ratio	0.55	0.06	0.07	0.29	-	0.20	0.44	0.07	0.16
Long-term debt ratio	0.55	0.06	0.07	0.29	-	0.20	0.45	0.07	0.17
				Mediators					
Loan type	0.51	0.05	0.20	0.31	-	0.36	0.32	0.06	0.34
Maximum line of credit	0.54	0.05	0.26	0.24	-	0.58	0.35	0.06	0.46
Credit institution	0.49	0.05	0.30	0.23	-	0.41	0.33	0.05	0.55
Currency	0.49	0.05	0.31	0.23	-	0.41	0.33	0.05	0.56
Days to expiration	0.50	0.05	0.31	0.20	-	0.63	0.30	0.05	0.56
Write-off type	0.47	0.05	0.32	0.30	-	0.42	0.27	0.05	0.56
New loan	0.47	0.05	0.32	0.28	-	0.44	0.29	0.05	0.58
Fixed rate	0.47	0.05	0.32	0.18	-	0.47	0.28	0.05	0.59
Minimum loan-to-value ratio	0.47	0.05	0.32	0.26	-	0.46	0.26	0.05	0.59
Probability of default	0.47	0.05	0.32	0.24	-	0.45	0.23	0.05	0.61
Maximum loan-to-value ratio	0.47	0.05	0.32	0.27	-	0.47	0.29	0.05	0.60
Installment free	0.47	0.05	0.32	0.26	-	0.46	0.28	0.05	0.60
Credit institution *	0.43	0.04	0.44	0.26	-	0.46	0.28	0.05	0.60
Loan type *	0.39	0.04	0.49	0.26	-	0.46	0.28	0.05	0.60
				Colliders					
Arrears	0.40	0.04	0.49	0.14	-	0.63	0.25	0.05	0.62
Accumulated write-downs	0.39	0.04	0.50	0.22	-	0.63	0.23	0.05	0.62
Default	0.39	0.04	0.50	0.21	-	0.60	0.25	0.05	0.62
Outstanding debt	0.39	0.04	0.50	0.20	-	0.63	0.26	0.05	0.62
Credit institution **	0.39	0.04	0.51	0.20	-	0.63	0.26	0.05	0.62
Loan type **	0.38	0.04	0.51	0.20	-	0.63	0.26	0.05	0.62

NOTE: The table shows the effect of gender on interest rates computed as the number of variables we adjust for increases. Each row of the table corresponds to a model in which the corresponding variable and all those appearing above it are adjusted for. For the XGB and DML models, the R^2 is computed out-of-sample. For OLS and OLS*, the R^2 is computed on all data. The adjusted R^2 corresponding to the most extensive model is 50.2 percent. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

D Appendix: Gender Gap Heterogeneities

Figure A.2
Heterogeneity by Geographic Area



NOTE: Comparison of the interest rates for male- and female-owned firms in different regions of Denmark. The y-axis shows the average interest rates for male- and female-owned firms in 4 different types of urban regions, ranging from the centres of larger cities to rural areas. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

Table A.6
Differences in Gender Gap Across Loan Types

	Point estimate	Standard error	z-stat
Baseline: Other loans (n=12209)	0.358	0.093	3.831
	Difference with respect to baseline		
Financial leasing (n=17949)	-0.149	0.106	-1.400
Direct line of credit (n=10390)	-0.016	0.161	-0.097
Revolving credit and credit cards (n=5414)	-0.171	0.172	-0.992

NOTE: The table shows the difference in gaps estimated through a DML model, where the second-stage model includes group-level dummies. The model includes all potential confounders and mediators, as highlighted in section 2. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

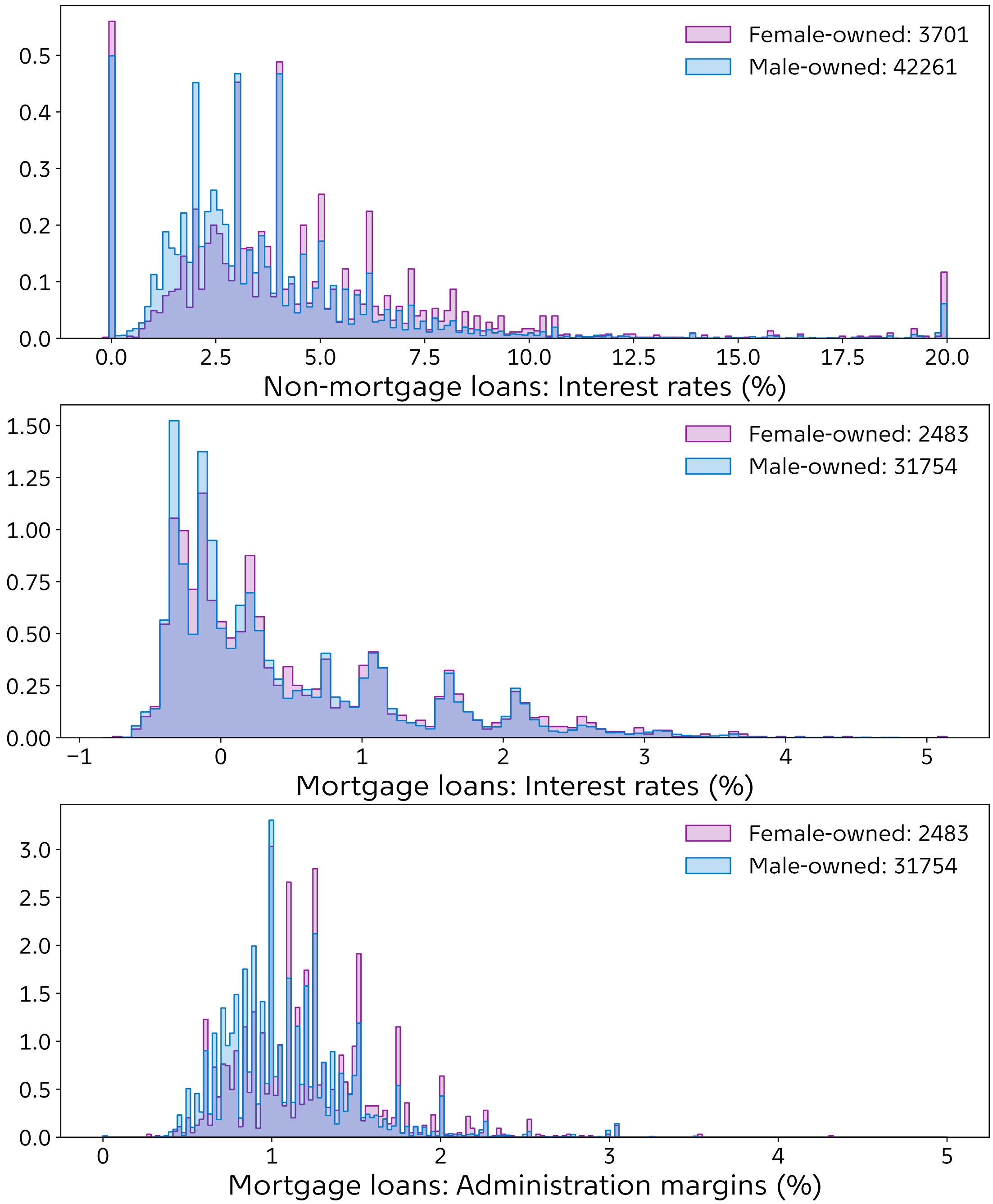
Table A.7
Differences in Gender Gap Across Year of Origination

	Point estimate	Standard error	z-stat
Baseline: Loans established before 2020 (n=38328)	0.221	0.051	4.354
	Difference with respect to baseline		
Loans established during 2020 (n=7634)	0.149	0.231	0.644

NOTE: The table shows the difference in gaps estimated through a DML model, where the second-stage model includes group-level dummies. The model includes all potential confounders and mediators, as highlighted in section 2. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.

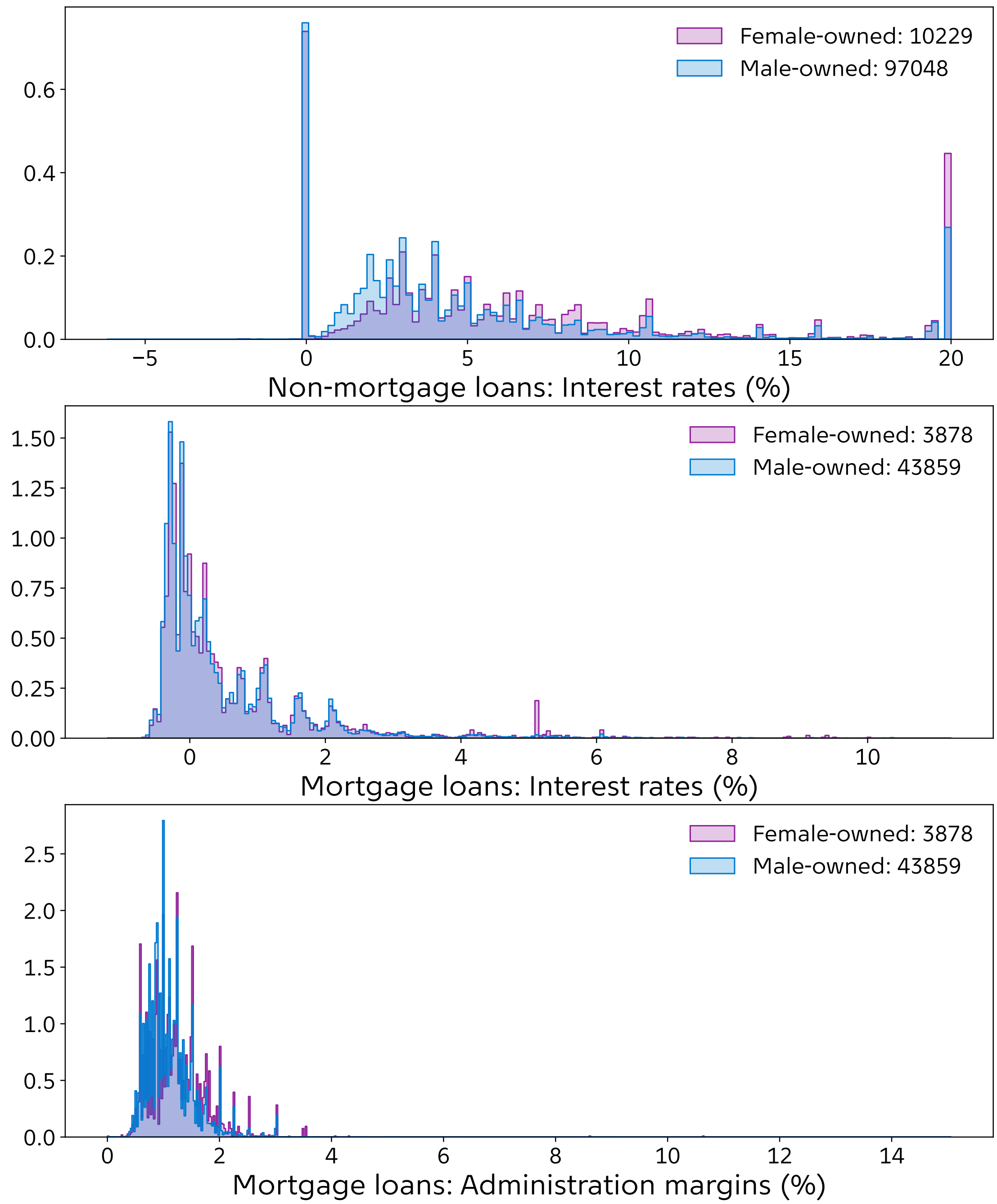
E Appendix: Distribution of Administrative Margins and Effective Interest Rates

Figure A.3
Distribution of Administrative Margins and Effective Interest Rates by Gender



NOTE: Distribution of administrative margins for mortgages and the effective interest rate for mortgages and non-mortgage loans. Effective interest rates for non-mortgage loans is capped at 20 for ease of visualization. The y-axis represents density.

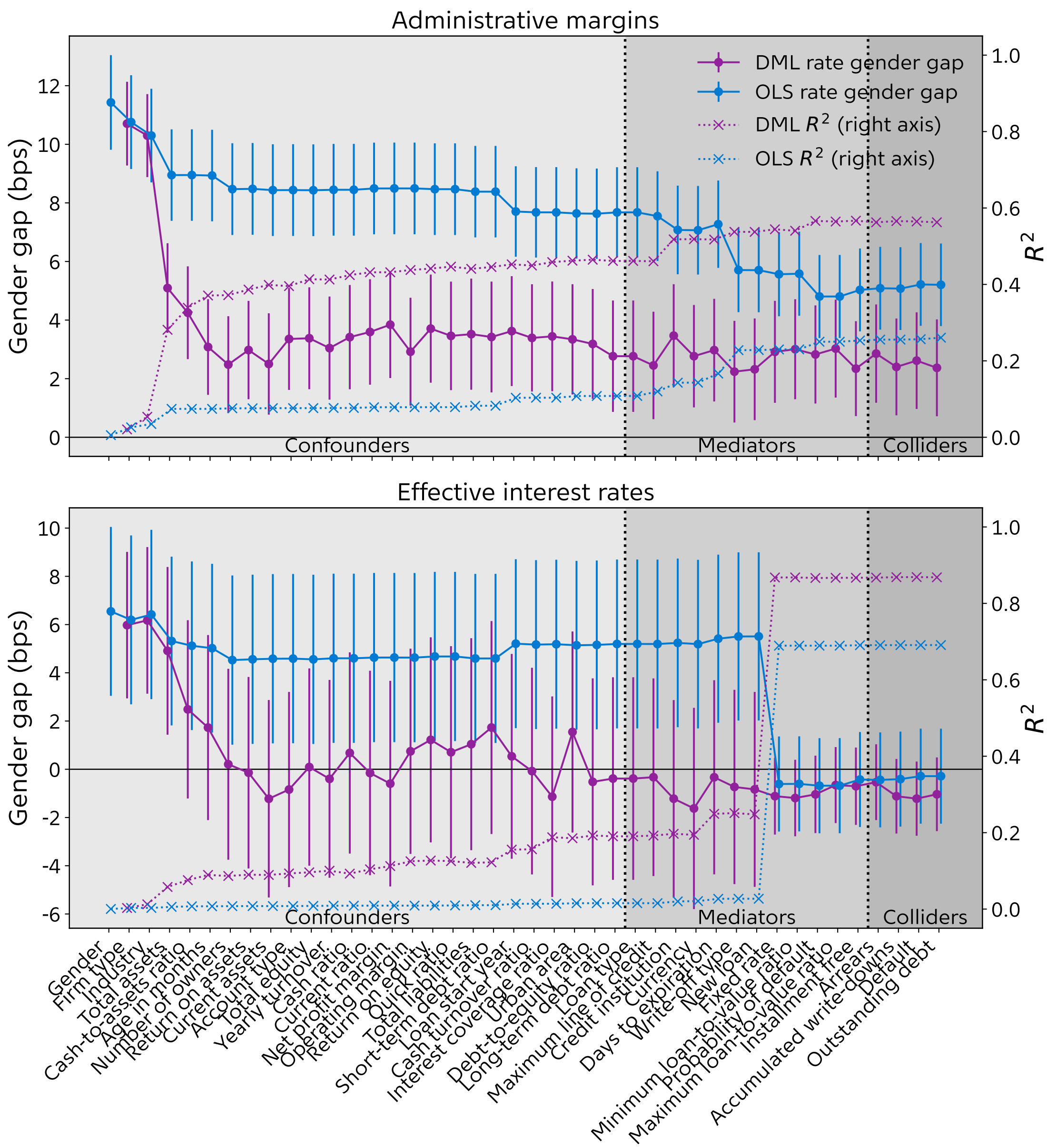
Figure A.4
Distribution of Administrative Margins and Effective Interest Rates, Including Loans to Non-Limited Liability Companies (Below DKK10,000) and Originating Before 2010



NOTE: Distribution of administrative margins for mortgages and the effective interest rate for mortgages and non-mortgage loans, for the sample of firms for which we can impute gendered ownership. This sample is not what we use in the analysis, and it includes loans to non-limited liability companies, below DKK10,000, and originating before 2010. Effective interest rates for non-mortgage loans are capped at 20 for ease of visualization. The y-axis represents density.

F Appendix: Mortgage Analysis Comparing OLS and DML Results

Figure A.5
Gender Gap in Mortgage Interest Rates and Administrative Margins; Method Comparison



NOTE: Analysis of the gender gap in mortgage interest rates and administrative margins, respectively, using both the DML and OLS methods. Our sample consists of 45,962 non-mortgage loans among 23,134 firms, 3,703 of which are female-owned.