Machine + Man: A field experiment on the role of discretion in augmenting AI-based lending models*

Anna M. Costello, Andrea K. Down and Mihir N. Mehta

University of Michigan, Ross School of Business

October 2019

Abstract

Does human discretion improve or diminish lending outcomes? We assess this question in the context of a randomized, controlled experiment using a large group of lenders that rely on machine-generated credit scoring models provided by a third party to make monthly credit decisions. Working with the credit scoring company, we design a new feature for their platform – the slider feature – which allows lenders to incorporate their discretion into the credit score. We randomly assign half of the lenders to the treatment group that gets the slider; the control group does not get the slider feature and thus makes credit decisions based primarily on the machine-generated model. Consistent with discretion aiding in loan decisions, we find that the treatment group's credit model adjustments are predictive of forward looking portfolio performance. However, we find that discretion is not useful in all cases. In fact, the control group does just as well as the treatment group in predicting credit risk for borrowers that have been traditionally classified as opaque. Our study highlights the growing prominence of AI-based lending models in crowding out some of the human's role.

JEL: G2; G21; G32; O33
Keywords: Relationship lending; Discretion; Machine-learning; Fintech

*We are grateful for financial support from the University of Michigan, Ross School of Business. We thank Phil Berger, Beth Blankespoor, Eric Floyd, Joao Granja, Bob Holthausen (Editor) Roby Lehavy, John List, Maria Loumioti (discussant), Greg Miller, Mike Minnis, Jim Omartian, Andrew Sutherland, Joe Weber, Jenna Wiens, Gwen Yu, an anonymous referee, and workshop participants at the London Business School Accounting Symposium, the University of Chicago (accounting group), the University of Chicago (working group on Field Experiments), the University of Illinois at Chicago, and the University of Michigan for valuable feedback. We are grateful to the management team at Credit2B-Billtrust, especially Shyarsh Desai and Irina Rabinovich for assistance in coordinating and setting up the field experiment. Costello serves on the Advisory Board for Credit2B-Billtrust. All errors are our own.

1 Introduction

Artificial intelligence (AI) is increasingly replacing humans in a wide range of tasks including diagnosing disease, recognizing human emotions, and even creating fine art. Global spending on cognitive and AI systems is projected to reach $77.6 billion by 2022, and many worry that AI will ultimately replace workers in many sectors of the economy.[1] One sector that has already seen rapid disruption from AI is the banking sector, where machine learning techniques are used for detecting fraud and money laundering, providing customer service in the form of chatbots, and automating credit profiling used for loan approvals. In fact, each year the banking sector spends about $4.7 billion in employee training, but is projected to invest over twice that amount – $10 billion – on artificial intelligence in 2020. With the availability of advanced machine-based credit models, many question whether humans still play an important role in lending decisions.

On the other hand, much of the academic literature suggests that loan officers play a critical role in capital allocation decisions, and their expertise may not be easily replicated by machine-based models. Specifically, autonomous loan officers exert effort to research potential projects and collect important decision-making inputs, and they may have preferences that differ from machine-based recommendations. This means that a range of discretionary factors, such as private information (Boot (2000); Diamond (1991)), competitive pressure (Ivashina and Sun (2011)), or gut instinct (Lipshitz and Shulimovitz (2007)) may play a role in loan officers' credit assessments. Yet the role of human discretion in assessing loan outcomes is not well understood. In particular, has the proliferation of AI-based lending threatened the role of relationship banking, or is there still a role for discretion in important lending decisions? If so, when is discretion useful or harmful in predicting credit risk? Our study seeks to answer these questions in the context of a randomized, controlled field experiment. Identifying the role of discretion in lending markets may help provide preliminary evidence on whether AI can completely replace the role of lending agents, or whether humans are still important in augmenting model-based approaches.

Gauging the usefulness of discretion in lending is difficult from an identification stand-

[1]Global spend is assessed by the International Data Corporation Worldwide Semiannual Cognitive Artificial Intelligence Systems Spending Guide.

point, because it requires the researcher to control for a multitude of confounding factors. For example, lenders that choose to adopt a relationship-based lending model also adopt appropriate incentive packages and organizational norms to facilitate the use of discretion. Indeed, observational studies show that the provision of relationship-based information is a function of the lender's organizational form and other supply and demand factors (i.e., Stein (2002); Petersen and Rajan (1994); Costello, Down, and Mehta (2019b)). These lender- and borrower- specific characteristics are also likely to have a direct impact on loan performance.

We employ a randomized field experiment on a cross-section of 428 lenders that rely on a machine-generated credit risk model as a major input to their credit decisions. Our intervention allows some lenders to augment the machine-based model with their discretionary preferences, whereas the control group does not get the intervention and thus relies primarily on the publicly observable, model-based credit score. Through randomization, our empirical tactic is to isolate the role of using discretion in improving or diminishing lending outcomes, while abstracting away from endogenous factors such as the existence of private information and organizational structure.

As our research setting we use a large sample of trade creditors that join a third party credit information platform called Credit2B-Billtrust.[2] Trade creditors (i.e., lenders) join the network in order to obtain credit-scoring information about their customers (i.e., borrowers).[3] Upon joining the network, the trade creditors are required to provide information on their receivables portfolio on a monthly basis. Credit2B aggregates millions of trade transactions, receivables performance data, financial statements, and other experimental and innovative data from various sources. A neural network classifier is then used to classify outcomes as "late" or "timely," which is then translated into a time-varying, customer-specific credit score and recommended credit line. Trade creditors in this network use the model-generated scores and credit lines provided by Credit2B as a major input to their lending decisions.[4]

We design and introduce a new feature to the scoring platform, called the slider feature.

[2] Credit2B was acquired by Billtrust in 2018, and changed its name to Credit2B-Billtrust, though for simplicity we refer to the company as Credit2B. Credit2B and the management team still operate with significant autonomy within the parent company.

[3] We use the terms suppliers, trade creditors, and lenders interchangeably, and we use the terms customers and borrowers interchangeably.

[4] In other words, by virtue of joining the network, these lenders have outsourced at least a portion of the inputs to their lending decision.

The slider allows the lender to use discretion to directly adjust the model-based risk scores and credit lines. It is referred to as a slider because the lender is presented with a continuum of possible credit lines and a toggle set to the machine-generated recommendation. The lenders in the treatment group are instructed to slide the toggle to "adjust your client's credit line upward or downward to incorporate your expertise and private information."

Our intervention randomly assigns half of the lenders in the sample the ability to use the slider, while the other half of the lenders in the sample do not get the new feature and thus continue to observe only the model-generated scores and credit lines for their borrowers. Our assumption is that, by randomizing treatment, on average the treatment and control groups are similar in their endowment of private information, organizational norms, and ex ante lending models. The intervention only varies the ability to *incorporate* discretion into the lending decision, allowing for a direct comparison of loan outcomes for decisions that incorporate more discretion versus those that rely primarily on machine-based recommendations. Our experiment thus allows us to provide evidence on whether augmenting machines with human judgement outperforms or underperforms the machine-based recommendation alone.

We use a difference-in-differences (DD) specification to capture the effect of our intervention. Our sample includes 428 unique lenders and close to 500,000 unique borrowers, thus covering a large cross-section of lending relationships.[5] The relatively large sample of loan observations allows us to also include time-invariant borrower and lender fixed effects as well as time varying borrower fixed effects to absorb any undesired variation in the supply of or demand for loans. Our DD estimator thus captures the differential effect of our intervention on changes in loan outcomes for the treatment group relative to the control group.

We begin by validating our assumption that the slider feature impacted lending decisions. One critical assumption in our experimental setup is that lenders in the network rely on the machine-generated scores and credit lines as a major input in deciding how much credit to extend. In other words, we assume that before intervention lenders faced constraints in their ability to deviate from the machine-generated recommendation, and therefore had more limited scope for incorporating their discretion in the loan decision. If lenders instead had

[5] Our sample includes 428 parent-company lenders, but over 1300 lender-establishments. Treatment was randomly assigned at the parent-company level.

unlimited scope to deviate from the model-based recommendation, then our intervention should have little impact on lending behavior. Therefore, we begin by measuring whether lenders 'deviate' from the model by lending more or less than the machine-based recommended credit line. Indeed, we find that on average the treatment group exhibits an 18% larger increase in their deviation from the machine-based recommendation after the intervention, relative to the control group. This finding helps to validate our assumption that the treatment relaxed lenders' constraints, and allowed for relatively more discretion in the lending decision.

We document two important empirical facts in our data. First, the slider allowed lenders to adjust their portfolio weightings but did not induce lenders to expand their overall supply of credit, since they likely faced budget constraints in the short-run. Thus, lenders in the treatment group used discretion to both expand and reduce access to credit, in order to balance their portfolio. However, we find that the average magnitude of negative adjustments is much smaller than the average magnitude of positive adjustments. This is a novel result in light of the prior literature that focuses primarily on the role of discretion in reducing Type II errors. This suggests that discretion may be particularly useful in identifying upside potential for a some clients, in order to win additional business.

Second, we show that the treated lenders make the largest adjustments to the model-generated credit lines for two groups of borrowers – very opaque borrowers and borrowers that lenders compete over. This insight helps to provide preliminary evidence on the nature of the discretion used in the lending decision. Specifically, the prior literature suggests that the scope for relationship lending is greatest for opaque borrowers, where information asymmetry is high (Petersen and Rajan (1994); Diamond (1991); Berger and Udell (1995)). Finding the largest deviations from the model in this subgroup of borrowers is consistent with the idea that discretion is used to incorporate private information in the lending decision. Additionally, prior studies hypothesize that deep-pocketed lenders may choose to relax credit terms to lower industry profits and accelerate the exit of rival firms (Telser (1966) and Bolton and Scharfstein (1990)). Consistent with this, we find that treated lenders use discretion as a competitive tool.

Next, we move to our main research question: whether augmenting the machine with additional discretion aids or hinders lending outcomes, relative to the benchmark of the machine-generated model alone. There are a number of reasons why discretion would be useful in lending. If discretion allows loan officers to incorporate private information, then it may help overcome information asymmetry problems that are prevalent in opaque borrowers and lead to more efficient capital allocation. Further, discretion may aid in competing over borrowers that present growth opportunities for the lender. On the other hand, there may be no scope for discretion because advances in machine learning techniques and 'big data' may have crowded out the scope for human judgement. Consistent with technology playing a role in reducing the need for in-person monitoring, Petersen and Rajan (2002) show that smaller firms are locating further from their lenders than in the past. Discretion may lead to sub-optimal allocation decisions if loan officers are subject to agency problems or cognitive biases (Banerjee, Cole, and Duflo (2009); Hertzberg, Liberti, and Paravisini (2010); Paravisini and Schoar (2013); Campbell, Loumioti, and Wittenberg-Moerman (2019)). Thus, whether discretion is useful or deleterious is an open empirical question.

If augmenting the machine with discretion leads to better decisions, then it should result in improvements in loan performance. There are two ways discretion might be useful in this setting: (1) it helps lenders identify poor quality borrowers that are not captured by the model (i.e., overvalued borrowers), thereby allowing lenders to cut lending to this group and (2) it helps lenders identify high quality borrowers that are not captured by the model (i.e., undervalued borrowers), thereby allowing lenders to expand lending to this group. In other words, discretion, if useful, should reduce the machine's Type I and Type II errors.

Since treatment lenders use the slider to reallocate capital within their portfolio, we study the impact of treatment on weighted-average lender-level portfolio outcomes. Using a DD specification, we investigate whether the use of discretion results in greater improvements in forward looking payment behavior and forward looking credit scores. Our results indicate that treatment lenders outperform control lenders along both of these dimensions. However, the improvements in overall credit risk are concentrated only in the subsample of the lenders' portfolios of very opaque private firms – those that do not have public financial statements or any social media accounts. In contrast, the treatment and control groups exhibit indistinguishable differences in performance for public borrowers and borrowers that are private but are on social media. We interpret our findings as follows: there is an important scope for augmenting machine-based models with human decision making, but it is limited to circumstances where borrowers are extremely opaque and therefore relationships matter most. Our finding highlights that machine-based models may go beyond incorporating traditional hard data like financial statements and have advanced to gathering and using alternative sources of information. It also suggests that the role of relationship lending may be declining in importance.[6]

We also find that treatment lenders perform worse – in terms of forward-looking credit risk – than the control group in the subsample of borrowers that they face high competition over. However, we also find that treatment lenders have a larger increase in future sales orders than the control group of lenders. This suggests that lenders may be willing to use discretion to take on more risk, in exchange for a boost to future sales. This is consistent with Heider and Inderst (2012) and Barrot (2015), who show that increased competition may cause lenders to be more aggressive with their lending strategy, which can lead to higher returns.

A few caveats are in order. First, our objective is not to run a horserace between machine and man. To do so, we would want to pit the best machine available against the most informed lender to see who wins. Instead, we are simply interested in whether augmenting the machine-based model with more human discretion performs better than the machine-generated model with less discretion. Since much of the literature focuses on *personal* interactions in lending decisions (e.g., Drexler and Schoar (2014); Khan, Li, Williams, and Wittenberg-Moerman (2018); Karolyi (2018); Engelberg, Gao, and Parsons (2012)), there's reason to believe that machines may not be able to mimic this type of specialized relationship. Thus, our empirical question is whether there is scope for discretion in the context of an AI era. A second caveat is that, as with any experiment, generalizability may be of concern. Because our setting includes a specific type of trade credit debt, our results may not generalize to the population of lending arrangements. Third, we caution that, since we can't observe the individual lender's incentives, we can't perfectly disentangle the vari-

[6] For example, most of the borrowers in Dealscan are large and public. Though we can't be sure our results extrapolate to this setting, our results call into question the role of relationships for large syndicated loans.

ous motivations for the use of discretion. Rather, our cross-sectional performance tests are intended to provide preliminary evidence that are consistent with discretion capturing both private information and competitive pressures. Fourth, the timeline of our experiment is relatively short; thus, we cannot speak to the long-run effect of discretion on loan portfolios.

Our primary contribution is that we impose an exogenous shock to the credit model that lenders employ, thus enabling us to draw a causal link between discretion and lending outcomes. The prior literature studying the role of discretion in lending outcomes is mixed; for example Degryse, Liberti, Mosk, and Ongena (2013) find that discretion helps to predict defaults over public information, though Puri, Rocholl, and Steffen (2011) find that loans approved based on the use of discretion do not perform any differently than other loans. Our randomized treatment, in combination with our large cross-section of borrowers, allows us to provide evidence on when discretion is useful and when it is not. We find that although the use of discretion is helpful in predicting credit quality for very opaque firms, it is not useful in predicting credit risk when competition for borrowers is high. Thus, our study offers novel insights on the uses and limitations of discretion in lending decisions.

Relatedly, we contribute to the growing literature studying the usefulness of machine-based models in the FinTech sector (i.e., Butaru, Chen, Clark, Das, Lo, and Siddique (2016); Ganeshapillai, Guttag, and Lo (2013); Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2018)). The advancement of machine-based learning algorithms may have changed the lending landscape, as these models are able to incorporate large volumes of data and use sophisticated techniques to predict loan outcomes. However, little is known about whether machine-based algorithms perform well in isolation, or whether augmenting these models with human judgment improves outcomes. We find that augmenting the machine with discretion improves outcomes for the subset of privately held firms that do not have social media, but not for those that do have social media accounts. This novel insight shows that current AI models go beyond incorporating traditional hard data like financial statements and have advanced to gathering and using alternative sources of information. Our result challenges the conventional assumption in the literature that all small, private firms benefit from relationship lending, and it highlights the growing prominence of AI-based models in crowding out some of the human's role. We also show that the treatment group does worse in terms of forward-looking credit quality than the control group in cases where there is competition for the borrower, but that the treatment lenders also gain future sales. This finding suggests that humans may be optimizing over multiple variables (i.e., maximizing future profits while minimizing credit risk), whereas models are typically trained on one outcome variable (i.e., minimizing credit risk). Thus, machine-based decisions alone may be too rigid and may not incorporate different incentives/preferences.

The remainder of our paper proceeds as follows. Section 2 proceeds with the conceptual framework, section 3 discusses the experimental setting, section 4 outlines the intervention, section 5 discusses the data, sections 6 and 7 present the validation and main results, respectively, section 8 discusses our robustness tests and caveats, and section 9 concludes.

2 Conceptual Framework

Lenders may choose to make discretionary adjustments to the machine-based model for a multitude of reasons. The use of discretion may include: (1) having a different information set than the model (i.e., private information); (2) competitive pressure to increase loan supply; and (3) gut instinct or other behavioral biases. Below, we discuss the theoretical predictions relating to these motives for using discretion.

2.1 Private Information

Lenders may make discretionary adjustments to the machine-based model because they have private information that is not incorporated in the model. The relationship lending literature highlights the lender's ability to alleviate information problems through screening and monitoring, thereby gathering important information to use in the lending decision. Consistent with prior literature (e.g., Diamond (1991); Boot (2000)) we think of the information gained through relationship lending as sharing three main characteristics: (1) the lender gathers information beyond readily available public information; (2) the lender exerts costly effort to obtain the information (through screening and monitoring), and (3) the information is often proprietary (available only to the intermediary and the borrower) in nature. We refer to this construct interchangeably as private or proprietary in nature.[7] In our setting,

[7]If the information is private and known only by the loan officer, then it is unlikely that the machine has access to the data, leaving scope for the lender's information. In contrast, if the information is soft but publicly available, machines are capable of processing it. See, e.g., Stuhlsatz, Meyer, Eyben, Zielke, Meier,

suppliers may obtain useful, borrower-specific information through cross-product sales, high frequency orders, personal interactions, or projected demand orders. Because the nature of the buyer-seller interaction, many of the relationships closely resemble the characteristics described in the relationship lending literature more broadly. In fact, the prior literature often takes the assumption that trade creditors are even more informed than banks, presumably because they transact more often and exchange proprietary information in the normal course of business (Emery (1984); Smith (1987); Biais and Gollier (1997)). Thus the lenders in our sample are likely to be particularly well-informed about borrower-specific information that is often proprietary in nature.

Private information may be useful in this setting because it allows lenders to better screen borrowers ex ante or because it serves to discipline their actions ex post. On the other hand, there may be limited scope for private information in the context of sophisticated AI models. Petersen and Rajan (2002) found that smaller firms, more likely to suffer from information asymmetry problems, were increasing the distance from their lenders. They suggest that the rising distance is due to the greater usage of "tools such as computers and communication equipment," which fundamentally altered the way loans are made. Advances in information technology may have crowded out the scope for private information, since public information is more timely and easy to transfer.

In addition to data modeling techniques crowding out the scope for private information, some argue that models are less prone to human bias in making loan decisions. Campbell et al. (2019) show that behavioral biases limit effective processing of soft information in lending, which leads to a reduction in loan quality. Further, in interpreting private information, loan officers may face agency problems, leading to worse lending outcomes (Banerjee et al. (2009); Hertzberg et al. (2010); Paravisini and Schoar (2013)). Thus, it is not clear ex ante whether using discretion to incorporate private information will lead to better or worse lending outcomes.

2.2 Competition

Prior work shows that competition may impact lending behavior. For example, Heider and Inderst (2012) analytically link the incentives to prospect for loans to a reduction

and Schuller (2011).

in lending standards. Empirically, Agarwal and Ben-David (2018) show that incentives for prospecting lead to a disregard for soft information and a greater reliance on a hard-information-based lending model; they link this behavior to a reduction in future loan performance. In contrast, Boot and Thakor (2000) and Degryse and Ongena (2007) show that higher competition between banks leads to more relationship-based lending and a greater reliance on soft information. These arguments suggest that the impact of competition on the use of discretionary information is ambiguous.

2.3 Behavioral biases

Prior literature shows that behavioral biases impact important decisions (e.g.,Cyert, March et al. (1963); Libby, Bloomfield, and Nelson (2002); Bloomfield (2002)) and individuals may rely on heuristics (e.g.,Tversky and Kahneman (1974); Hastie and Dawes (2010); Kahneman and Frederick (2002); Kahneman (2003)) or simply may be constrained in their ability to process large amounts of data (e.g., Hirshleifer and Teoh (2003); Lim and Teoh (2010)). Lenders often follow their gut instinct in assessing loan applicants, rather than hard indicators of creditworthiness (Lipshitz and Shulimovitz (2007)). If lenders in our sample use their discretion to incorporate behavioral biases, it should lead to adverse loan outcomes.

3 Setting

3.1 Trade Creditor versus Bank Lender

Our study relies on trade creditors to assess how lending outcomes respond to the addition of private information. Trade credit, or delayed payment for intermediate goods, plays a substantial and well-documented role in financing inter-firm trade. Indeed, the Bank of International Settlements (2014) estimates that two thirds of global trade is supported by trade credit, and trade credit outpaces bank credit as the main source of short-term financing for U.S. firms. Therefore, trade creditors represent important and substantial lenders in the world economy.

A large stream of literature questions why suppliers engage in lending when more specialized monitors like banks could fill that role. One prominent theory explaining trade credit suggests that it helps to alleviate frictions between customers and external financiers. The financing theory posits that trade credit substitutes for bank credit when trade partners face frictions in external financial markets. Prior studies show that trade credit can substitute for bank credit during periods of tight credit or financial crises (Nilsen, 2002; Choi and Kim, 2005; Love, Preve, and Sarria-Allende, 2007). Petersen and Rajan (1997) use the National Survey of Small Business Finance (NSSBF) to show that firms with better access to bank credit have higher levels of accounts receivable.

There are several reasons why suppliers might be willing to lend when banks are not. One hypothesis is that that suppliers have an informational advantage over banks, enabling them to screen and monitor loans more effectively. Smith (1987) and Biais and Gollier (1997) argue that suppliers have an informational advantage because of the frequency of interactions and the types of information exchanged. For example, the supplier may visit the buyer's premises more often than financial institutions or may obtain information about the buyer's creditworthiness through demand forecasts and other operational information. In other words, the financing theory of trade credit suggests that trade creditors, relative to bank lenders, are potentially *more informed* about borrower quality. This suggests that our setting may be particularly well suited to capture the use of discretion in the lending decision.

3.2 The Business Model: Credit Scoring and Recommended Credit Lines

Our research setting relies on large sample of lenders that are part of an information network called Credit2B, which produces credit information and risk scores for millions of borrowers worldwide. Lenders who join the network are required to produce a monthly report disclosing all transactions with borrowers; the report contains the receivables balance (i.e., loan amount) for each borrower along with aging reports which can be used to assess the borrower's payment performance. Credit2B assesses borrower performance by summarizing the receivables data provided by all of the lenders and augments it with other data including credit bureau information, financial filings, industry trends, and additional innovative sources of data to produce time-varying borrower risk scores and recommended credit lines, which are available to lenders in the network.

By virtue of joining the network, the lenders in the sample have outsourced at least a portion of their screening and monitoring activity to a third party (Credit2B). That is, the Credit2B model recommendation is a major input to their credit decision. This is unsurprising, since the average lender has over 7,000 borrowers. The value proposition offered by Credit2B is that aggregating peer-to-peer data with other relevant data into one summary measure provides better information about borrower credit quality than if the lender were to only rely on their own data.[8] Lenders use the scores and recommended credit lines as a major input into their credit decision. Most lenders use the Credit2B recommendation in the documentation process for the approval of a monthly trade loan.[9] Anecdotal evidence reveals that, before the rollout of our experiment, lenders faced some limitations in their ability to deviate from the Credit2B recommendation. This is because their employer has control systems in place to regulate and monitor the use of judgement.[10] An assumption in our experimental design is that lenders faced more constraints in their ability to deviate from the model in the pre-period, relative to after the slider was introduced. We discuss this assumption and how it may impact our research design in section 4.2.

3.3 The Machine-Generated Score and Credit Line

The score model used in this setting is proprietary to Credit2B.[11] The current version of the score (R-Score™ system Generation 2.0) was introduced in 2017,[12] and it includes millions of trade transactions, receivables performance data, financial statements, other public filings, relevant alerts (i.e., bankruptcy, UCC filings, downgrades, collections, etc.), and other data. In addition to traditional data inputs, Credit2B collects innovative and experimental data from various sources. A neural network classifier is then used to classify outcomes as "late" or "timely." Outcomes are then translated into a borrower-specific, time-varying score, ranging from 0 to 100, which is derived from the probability of a company paying on time. A corresponding recommended credit line is then calculated for each lender-borrower-month observation. Henceforth, we refer to these recommendations as the "machine-generated score" and the "machine-generated credit line," or more simply "the model." We assume that the

[8] This argument is similar to that in Sutherland (2018), who documents that information asymmetry problems in small commercial lending markets can be reduced by information sharing technologies that allow lender-to-lender reporting of borrower credit files.

[9] Lenders are typically part of the risk management team within their firm and are often required to follow formal internal loan approval processes that require an external credit score and credit line recommendation.

[10] Some lenders must strictly adhere to the guideline, while others are given scope for *limited* deviations without approval from more senior management.

[11] Costello is privy to the data inputs and modeling approach but has signed a nondisclosure agreement.

[12] The previous version of the score was calculated using a linear combination of a hand-picked number of elements, much like the Z-Score.

model-based scores and credit lines represent publicly-available information. For a given borrower-month, every lender observes the same model-generated output. Therefore, though the underlying inputs and the particulars of the model may not be commonly known, the output is equally observable by all of the lenders in the network and is therefore public in nature.[13]

The neural network approach thrives on large volumes of data, where more data points and larger training sets translate into more accurate models. Credit2B has over 15 years of historical data, and in 2018 alone, Credit2B recorded close to 23 million trade transactions, rendering the machine learning approach valuable. The model performs well in predicting borrower outcomes; when benchmarked against the models used by other large competitors, the neural network based model used by Credit2B achieves a 30% improvement in prediction accuracy (Trivedi, Rabinovich, and Desai (2017)). Nevertheless, the model does not predict future payment behavior perfectly, leaving sufficient scope for improvement. The training set may not be representative of future customer profiles, and for some borrowers the data inputs are sparse - the model is unlikely to capture private characteristics that may be important in predicting payment behavior. These limitations are not unique to our setting; the literature on machine learning cautions that models are a black box, may bias against race, gender or other minority groups, and often can't incorporate valuable private information (e.g., Rudin (2018); Chouldechova and Roth (2018)). Therefore, it is reasonable to predict that lending agents with expertise, preferences, and relationship-specific knowledge may be able to augment and improve the machine-generated approach.

4 The Intervention

Before intervention, each lender in the sample observed the same borrower-specific credit score. This machine-generated credit score is used to calculate a lender-borrower-month recommended credit line.[14] Users can login to their Credit2B online account where they can access each borrower's score page – which they typically do on a frequent basis before each

[13]Consistent with rational expectations models, we assume that the 'price' (i.e., the machine-generated output) is a sufficient statistic for all lenders' information.

[14]A maximum tolerable exposure value (MTEV) is calculated based on the size and scale of trade volume, where the hypothetical score is 100. The MTEV is then discounted based on the credit score to obtain a recommended credit line.

new order/loan is approved. Figure 1 illustrates the Credit2B user interface that the lender observes on a representative borrower/account. One can see that the machine-generated credit line is displayed as a single number without the ability to adjust within the system.

The objective of our experiment is to manipulate the inclusion of discretion in the lending decision, and to subsequently assess loan performance. The timeline of our experiment is outlined in Figure 2.[15] To begin, we gathered an assessment of the client's overall satisfaction with the model in order to ascertain whether there was scope for improvement, and specifically, to gather initial evidence on potential incremental information that is not reflected in the model. We sent a short survey to all active lenders in August 2018, and had a 14% response rate. The full survey is provided in the Appendix, A1.

On a 10 point scale (with 10 being the most satisfied), we found that lenders had an average satisfaction rate with the score model of 7.2 and an average satisfaction rate with the credit line model of 7.0. It is not surprising that these two responses were highly correlated, since the score is the most important input to the recommended credit line. While most respondents were satisfied with the scores, 17% of the respondents rated the scores below a 5 out of 10, responding that the scores were either too aggressive or too conservative. Of interest were the qualitative remarks; one respondent stated "There are numerous extenuating circumstances regarding a customer's credit limit. We'd like to account for our relationship with the customer and specific knowledge of the business." Another stated "I find the scores and CCLs excellent for well-established companies, but for certain companies they fall short. I've developed a rapport with certain clients which makes me more comfortable than the score and CCL reflects."

The evidence from the survey is far from conclusive. Instead, we view these anecdotes as consistent with the idea that the models perform well in most cases. However, clients seem to believe there is some scope for improvement based on their personal experiences and/or expertise.

In September 2018 we began the design of a new feature called the slider. Conceptually,

[15] Upon the recommendation of an Institutional Review Board representative, we submitted a full IRB application for our experiment in February 2018. On April 4, 2018, we received a letter that our research is considered Not Regulated and we did not need to proceed through the IRB process. That letter is provided in the Appendix, A4.

the slider is meant to augment the model-based scores and credit lines, with some flexibility to adjust them within the system. Visually, a line is presented with a toggle that is initially set to the model-based credit line. Users can slide the toggle away from the model recommendation, and are instructed to "adjust your client's credit line upward or downward to incorporate your expertise and private information." The lender has the ability to slide the toggle but is not forced to; therefore lenders without private information may opt to maintain the toggle at the model-recommended amount.

A preliminary version of the slider was introduced at the annual Product Advisory Council (PAC) meeting in New York in October 2018. The PAC meeting is attended by the senior management teams of Credit2B and the parent company, the advisory board members, and a target group of 25 Credit2B member firms (i.e., lenders that belong to the network).[16] The slider was discussed during a session on scores and credit lines, with the objective of learning (1) whether it may address the concerns raised in the September 2018 survey; (2) whether the mock-up was visually compelling; and (3) the proper wording to make it clear we'd like to capture private information.[17] The original mock-up can be seen in the Appendix, A2. Based on feedback, slight modifications were made to the slider to improve the user experience.

During the month of November 2018, the authors worked with the user interface team to design the final version of the slider. The slider that was ultimately placed in production has the following traits: (1) It is horizontal, where the far left end of the line is designated as 'Conservative' and the far right end of the line is designated as 'Aggressive,' (2) An adjustable toggle is set to the model-generated credit limit at the center of the line, and (3) Mouse-over explanations appear when a lender hovers over the words aggressive or conservative: "If your proprietary information about this customer suggests that you should take a more aggressive approach, move the slider to the right, and the credit limit will increase," and "If your proprietary information about this customer suggests that you should take a more conservative approach, move the slider to the left, and the credit limit will decrease." The final version of the slider that entered production is illustrated in Figure 3.

[16] Each year, the company invites a subset of clients to the PAC meeting, prioritizing those that contribute the most to revenues.

[17] Costello attended PAC and presented the score model and introduced the slider to get feedback.

In November of 2018 the Credit2B management team supplied us with a list of the current active lenders in the network. Using this list, we randomly assigned firms to treatment and control groups using the runiform function in Stata. If the random number given to the lender was less than or equal to 0.50, then it was assigned to the treatment group, else it was classified as a control lender. By randomizing treatment on subjects that are unaware they are taking part in an experiment, our design should naturally balance on unobservables (Floyd and List (2016)). To provide further evidence toward this assumption, Table 1 reports the covariate balance on characteristics of lenders as of treatment assignment (November 2018). So as not to bias toward large, public lenders, we obtain important descriptive characteristics from the Credit2B database, which includes information on both public and private firms. We investigate both lender firm-level characteristics, as well as the characteristics of their lending portfolio. One can see that there are no detectable differences across the treatment and control groups, suggesting that our randomization was successful.

On November 30, the slider was released to the treatment group only. Treatment lenders could now adjust each of their borrower's credit lines by sliding the toggle left or right; control lenders saw no change in their user interface.

4.1 Education on the Slider

In conjunction with the release of the slider, we employed a marketing campaign to inform the treatment group about the new feature. First, an email was sent to the treatment group on November 29, informing them about the coming release of the new slider feature. No such email was sent to the control group. Importantly, neither the treatment nor the control groups were aware that they were taking part in an experiment which helps alleviate concerns with selection bias from voluntary participation in experiments (Floyd and List (2016)).

The email announced the availability of the new slider feature, instructing clients: "If your proprietary information about a specific customer suggests that you should take a more aggressive approach, move the slider to the right, and the credit limit will increase. Similarly, moving the slider to the left will result in a credit limit decrease. The feature allows you more flexibility to incorporate your information into the calculated credit limits." The full email can be seen in the Appendix, A3. In addition to the email, a free webinar was conducted the first week of December to provide more clarity on the intended use of the slider. Lenders were specifically instructed to use the slider if they believed they had private information that could augment the model-based scores and credit lines. The webinar also provided an opportunity to field questions, and sample illustrations were offered to provide clarity. Individuals from the Credit2B sales and marketing teams were educated on the slider so that they could help individual clients, should questions arise. Finally, a banner was added to the treatment groups' Credit2B homepage that read: "NOW AVAILABLE! New slider feature which allows you to customize your client's credit line."

4.2 Empirical Approach: Main Specification

We use a difference-in-differences (DD) setting to capture the differential effects of using discretion in lending decisions. We study lending behavior over the period October 2018 through February 2019. The treatment period corresponds to the period beginning December 2018, which was the first month that treatment lenders were allowed to use the slider.

The DD approach compares the change in lending behavior for the treatment group relative to the control group and is estimated by OLS using the following specification:

$$y_{i,j,t} = \beta_1 After_{i,j,t} + \beta_2 Treatment_i * After_t + \alpha_{i,j} + \alpha_{t,j} + \varepsilon_{i,j,t} \quad (1)$$

where y is the characteristic of a loan from lender i to borrower j in month t. *After* is a dummy variable equal to one from the beginning of treatment through February 2019, and *Treatment* is a dummy indicating whether the lender has been randomly assigned the slider. We include an exhaustive set of fixed effects to capture time invariant characteristics of the lender and borrower as well as bilateral lender-borrower relationships. We also follow Khwaja and Mian (2008) and include a specification with borrowerXmonth fixed effects. This specification can be interpreted as the change in loan outcome for a particular borrower in a given month from a treated lender, relative to the change for that *same borrower* in that month from other lenders. Including these fixed effects requires that the borrower has multiple lenders, and it should absorb borrower-specific idiosyncratic shocks that might impact loan outcomes.

We make several important assumptions in our empirical approach. First, we assume that through randomization, the assignment to treatment is exogenous to the credit outcomes we study. It is possible, though unlikely, that treatment and control firms differ along dimensions that might be correlated with the outcome variables and would bias the DD estimation. Including a battery of fixed effects helps to alleviate this concern by absorbing a significant amount of the variation that we might be concerned about. Further, we investigate covariate balance between the treatment and control groups, and we assess trends in lending between the treatment and control groups (Table 1 and Figure 4). These tests are consistent with the assumption that the assignment to treatment is exogenous.

A second important assumption is that lenders rely on the machine-generated scores and credit lines as major inputs in their credit decision. In other words, we assume that lenders faced more limited scope for deviating to incorporate private information, personal preference, risk aversion, etc. If lenders have unlimited scope to deviate from the Credit2B recommendation, then the new slider feature should have little impact on lending which would render our experimental setup invalid. We assess our assumption in several ways. First, we note that by virtue of joining the network, lenders have revealed that they rely on the model to some extent, else there is no real benefit to paying for this service. Second, we spoke with several lenders in the network and found that most use the model-based recommendations in the documentation process for the approval of a monthly trade loan. In other words, the Credit2B credit report (with the model outputs) is a required document for the loan approval process. The slider allows for incorporating private information while still adhering to this approval process; by sliding the credit amount *within* the Credit2B system, the credit report itself is changed and thus still satisfies the documentation requirement. We do note, however, that there is significant variation in the extent to which the lender is allowed to deviate from the credit report. Some lenders must strictly adhere to the guideline, others are given limited scope for deviations, and others may deviate with senior management's approval. Our DD approach with lender fixed effects is important in addressing this concern. Specifically, we capture the *change* in behavior for the treatment group relative to the control group, while controlling for differences in approval policies across lenders. Thus, our empirical strategy should capture whether the intervention results in measurable changes in lending behavior.

Our third main assumption is that lenders in the treatment group and lenders in the control group are, on average, endowed with the same amount of information. Our intervention only varies the *ability to use* their discretion/information in the lending decision. This is a key assumption in the design of our experiment; through randomization, we aim to abstract away from factors that determine the provision of information to begin with. This is important, because prior work shows that the existence of private information is correlated with important borrower- and lender-specific qualities and the bank's organizational form (i.e., Costello et al. (2019b)), and it illustrates the importance of our experimental approach. Specifically, we impose an exogenous shift in the credit model employed by encouraging the use of more discretion, and thus we abstract away from other organizational choices that likely impact loan outcomes.

5 Data

A main innovation of the paper is the ability to observe how lending behavior and loan outcomes change with an exogenous increase in the use of discretion. To track this behavior, we rely on the proprietary data from Credit2B, which contains disaggregated data on transaction level loans and their performance. In this setting, loan decisions are made each time a sales transaction closes. Loans are generally short-term, are rolled over with high frequency, and are directly tied to the sale of the underlying good. Further, whereas bank loans contain multiple contractual features (i.e., covenants, collateral requirements, etc.), the loans in the trade credit sample typically only vary the *volume* of the credit line offered. Costello (2018, 2019) shows that these loans do not contain covenants or other explicit requirements and do not carry an interest rate.[18] This feature – that lenders only vary loan volume – makes our setting conducive to capturing the effects of discretion, since we don't have to simultaneously control for multiple terms in the loan package.

Credit2B has trade payment data collected from their member firms. In order to become a member firm, suppliers are required to produce a monthly report disclosing all transactions with customers; the report contains the current receivables and past due balances for their

[18] Some studies on trade credit calculate an implicit interest rate for those contracts that offer discounts for early payment. They do so by calculating the implicit interest for a forgone early discount. Costello (2018, 2019) shows that in practice, these discounts are rare and are concentrated in the retail sector. In fact, the duration of these loans cluster by industry and rarely vary.

universe of buyers. In other words, all members must make a monthly upload of every transaction made with every customer.[19] The data also includes a receivables aging report for each transaction. Past due balances are reported for each customer broken into 30 day buckets. The company relies on this data to produce time-varying customer risk scores, which are shared with all member firms.[20] In addition to the trade data, credit scores, and recommended credit lines, we have access to other data through Credit2B's affiliations with other data providers. This data includes information about loans placed for collection, sales orders, and other relevant data that these companies use to assess credit risk.

5.1 Sample and Descriptive Statistics

We use the transaction-level data to build a monthly panel of loan pairs over the period October 2018 through February 2019. Despite the importance of trade lenders in the world economy, we recognize that generalizability is a potential concern. In order to get a sense of potential issues of selection into the sample, Table 2 provides a comparison of our sample of public lenders relative to the Compustat universe. Though lenders in the sample are smaller and hold less cash on average than firms in the Compustat population, they are remarkably similar along all other dimensions. Perhaps, however, the Compustat population is not the appropriate benchmark since it includes non-financial firms. As a further comparison, we compare the lenders in our sample to the banks in the WRDS Bank Regulatory (Commercial Banks) file. Using the RCON Series file, we take each bank's total asset balance as reported in December 2017, which is the most recent quarter with available data.[21] The average size of these banks, as measured by total assets, is \$972M. Thus the *public* lenders in our sample appear to be larger on average than the average commercial bank. We note that the private lenders in our sample are likely to be smaller on average than the public lenders, which would bring our sample average down.

The public borrowers in the sample are of similar size to the Compustat population, though they have higher current liabilities and are older. Again, the sample of private

[19] Credit2B strictly enforces a rule that all member firms must upload receivables data at a minimum monthly frequency, and it must include all existing customers. If suppliers violate this rule, they lose access to all credit information.

[20] Identities of buyer-supplier pairs and their transaction balances are not disclosed to other member firms, which should mitigate suppliers' incentives to underreport due to proprietary concerns.

[21] We use unconsolidated numbers and omit bank holding companies.

borrowers is likely to impact the distribution of borrowers in the sample. Taken together, the results in Table 2 suggest that our sample of public lenders and borrowers are similar to the Compustat universe and specifically, to the sample of commercial banks.

An important assumption in our empirical setup is that the slider allows lenders to incorporate discretion in the loan decision and therefore to 'deviate' from the model-based recommended credit line. To gauge whether a lender deviates from the recommendation, we compare the model-based credit line to the total amount of loan outstanding.[22] We calculate the absolute value of the percentage difference between the actual loan outstanding compared to the model-based recommendation (we refer to this as the *deviation*, henceforth). On average, the lenders deviate from the recommendation by 62%, though the first quartile of loans exhibit virtually no deviation from the model-based credit line.

As discussed in section 3.2 we expect a baseline level of deviation, since some lenders have limited flexibility to deviate with or without approval from upper management.[23] However, if the intervention is meaningful, we expect an *increase* in the deviation for the treatment group after intervention, relative to the baseline control group. Figure 4 shows that in the two months leading up to the experiment, the treatment and control groups are remarkably similar. Immediately after the slider was available to the treatment group, one can see a divergence in the deviations for this group; the control group continues to maintain the baseline level of deviation. We view the evidence in Figure 4 as serving two purposes: (1) it confirms that the slider influenced loan activity which validates the power of our intervention, and (2) it illustrates that the treatment and control groups exhibit parallel trends in the pre-period, which helps to validate our research design.

[22] In this setting, customers borrow up to the credit line they are allowed since there is no interest charge on trade debt.

[23] The deviation in the pre-period highlights the importance of our difference-in-differences strategy, which should capture the differential changes in incrementally more discretion after the introduction of the slider. In robustness tests, we limit the sample to those lenders that exhibited only minor deviations in the preperiod (less than 10% average deviation), and our results are qualitatively similar.

6 Validation Results

6.1 The impact of the intervention on lending behavior

If the introduction of the slider had any meaningful impact on lending behavior, we should see that the treatment group increases the deviation from the model-based credit line more in the period after treatment, relative to the control group. The *actual* sliding behavior is unobservable, since lenders can slide and obtain the recommendation, but the Credit2B system does not save these adjustments. Thus we infer slider use based on the actual loan amount outstanding. Specifically, the realized amount lent should deviate from the recommended credit line if the lender uses the slider.

At this point we aim to validate the use of the slider, and thus any deviation – positive or negative – is of interest. We calculate the absolute value of the percentage deviation between the actual amount lent and the recommended credit line. While Figure 4 provides illustrative evidence, we also validate our intervention in regression analyses by estimating Equation 1, where the dependent variable is the absolute value of the percentage deviation from the model-based credit line (*|%Deviation|*). Results are reported in Table 3. We find that indeed, the treatment group exhibits an 18% larger increase in their deviation from the model after treatment, relative to the control group. Progressively adding more rigorous fixed effects does not alter our inferences. In the final column, we add borrower by month fixed effects. Though this specification dampens the treatment effect slightly, we still continue to find that the treatment groups' lending behavior changes more than the control group. Overall, we conclude that the slider was a significant enough intervention that it altered behavior.

The results in Table 3 show that lenders use discretion to adjust the loan amount. However, did the experiment change the overall supply of credit? We explore whether the overall lending portfolio changed after the introduction of the slider by estimating Equation 1, where the dependent variable is the log of the lender's total monthly credit portfolio. The regression is estimated at the lender's establishment-month level, and results are reported in Table 4, Panel A.[24] We fail to find that treated lenders changed their supply of credit more than the

[24] Treatment was assigned at the lender-parent level, but several of the parent companies are large and have multiple establishments. Because the establishments have autonomy in their trade and lending decisions, all

control lenders. Taken together, the results in Table 3 and Table 4, Panel A suggests that the slider caused treated lenders to reallocate their existing lending portfolio rather than to expand or cut the overall supply of credit. Given the short window of our study and the assumption that lending divisions face budget constraints, this result is unsurprising.

6.1.1 The magnitude of positive versus negative deviations

To this point, we have only tested whether the slider resulted in any deviation from the recommended credit line. However, private information may indicate a more negative outlook than the model suggests (which would produce negative deviations) or it may indicate a more positive outlook than the model suggests (which would produce positive deviations). The focus in the prior literature has been on using private information to reduce credit availability to lower-quality borrowers (i.e., avoid Type II errors), mostly because of empirical limitations in the ability to capture Type I errors.[25] In our setting, we can separately observe both upward and downward adjustments, thus we next ask whether lenders on average use positive or negative private information in the lending decision.

To provide some evidence on this question, we estimate Equation 1, where the dependent variable is the absolute value of the percentage deviation from the model-based credit line (*|%Deviation|*). We augment the model with an indicator for whether the deviation is positive, zero otherwise. The coefficient on the interaction term *Treatment*After* represents the DD estimator for for the magnitude of negative deviations, while the triple interaction term *Treatment*After*Positive* captures any incremental effect for the magnitude of positive deviations. Results are reported in Table 4, Panel B. The treatment group exhibits larger increases in both negative and positive deviations. However, the magnitude of the negative deviations is much smaller (and only marginally statistically significant) relative to the deviations on the positive side. Note that because the overall lending portfolio didn't change, this implies that treated lenders made a larger number of small negative adjustments, and a smaller number of large positive adjustments. We interpret this evidence as indication that in this setting, the content of positive private information is more significant than the content of negative private information. This is a novel result in light of the prior literature that

analyses are done at the establishment level.

[25] It's difficult to capture Type I errors because most observational data doesn't reveal high quality loans that have been rejected.

focuses primarily on the role of private information in identifying poor quality borrowers. As a caveat, we recognize that loan amounts are truncated at zero, which could mechanically influence our results in Table 4, Panel B. In robustness tests, we add a proxy for loan denials, and our results are similar.

To provide further evidence on the adjustments that treated lenders make after they are able to use private information, we plot their portfolio allocation in the period before the experiment relative to their portfolio allocation in the treatment period. Figure 5 plots the average percentage of the treatment lenders' portfolios that fall in each of five bins. The first bin includes borrowers with credit scores between 1 and 20; the second bin includes borrowers with credit scores between 21 and 40; the third bin includes borrowers with credit scores between 41 and 60; the fourth bin includes borrowers with credit scores between 61 and 80; and the final bin includes borrowers with credit scores between 81 and 100. Credit2B labels borrowers in the first bin as having a "remote" likelihood of repayment, and the borrowers in the fifth bin as an "almost certain" likelihood of repayment. The light gray bars represent the allocation before treatment, and the dark black bars represent the allocation upon treatment. One can see that before the intervention, treatment lenders allocated over 50% of their portfolio to the safest borrowers. Upon treatment, they reallocate their portfolio to borrowers with lower credit quality. This evidence is consistent with the idea that the treated lenders use their private information to identify potentially undervalued borrowers and to increase their line of credit.

6.2 What is causing the deviation?

Our experiment instructs lenders to use the slider to "incorporate your expertise and private information." However, we recognize that lenders may use their discretion to deviate for a number of reasons. Our empirical analyses, described below, are designed to provide some preliminary evidence on the motivation for using the slider.

6.2.1 Private Information

Lenders may make discretionary adjustments to the machine-based model because they have private information that is not incorporated in the model. To provide evidence on this front, we take advantage of the vast amount of data in the Credit2B database. Included in their files are specific comments made by lenders about their borrowers. The disclosure of these comments is voluntary; if a particular lender has a policy of recording comments about borrowers, then those comments are likely included in the receivables files that they upload to the Credit2B system.[26] As a validation test, we split the treatment lenders' loans into those with large deviations in December 2018 (above the median) and those with small deviations in December 2018 (below the median). We then assess the likelihood that a lender recorded at least one comment about a particular borrower over the course of the previous year. Table 5 presents the difference in the likelihood of making a comment for low deviation loans versus high deviation loans. Consistent with our expectation, for loans that exhibit larger deviations from the model, the lender was more likely to record a comment about that borrower than the loans that exhibit smaller deviations. Further, in the Appendix, A5, we provide examples of the comments we gather from this file. The descriptive anecdotes help to understand the content of information. The comments appear to be gathered primarily through relationships and are reflective of valuable information (i.e., "Rest assured, we will eventually pay you guys, as you were cool enough to give us 45 day terms from the get-go."). While far from conclusive, the evidence in Table 5 combined with the anecdotal examples in the Appendix are consistent with lenders deviating from the model-generated credit line due to their private information.

Second, we draw on the predictions from the relationship lending literature. In particular, we expect that the scope for relationship lending is greatest for opaque borrowers, where information asymmetry is high (Petersen and Rajan (1994); Diamond (1991); Berger and Udell (1995)). The literature also predicts that private information is gained through repeated interactions, thus lenders have likely gathered more private information in their more established relationships relative to their new ones. If the deviations from the model are due to private information, we should observe larger deviations in the cross-sections where private information is more likely to be present.[27]

[26]The comments in the database are assessed in detail in Costello et al. (2019b). Since they are voluntarily provided, there is not enough overlap with the lenders in the current study to conduct more detailed analyses.

[27]We investigate heterogenous treatment effects by splitting the sample based on borrower characteristics, rather than lender characteristics. Nearly all of the lenders use the tool – 99% of the lenders in the treatment group exhibited an average increase in deviation behavior of more than 20% in the first month of treatment. This indicates that the uptake of treatment was indeed successful and swift. We also investigate whether use of the tool was different based on lender size, consistent with predictions in Stein (2002). We fail to find any difference between large treated lenders and small treated lenders.

We use three cross-sectional variables to capture the existence of private information. First, we split lending relationships based on whether the borrower is public or private. Since private firms do not have public filings, they are relatively more opaque than public firms (Minnis (2011); Allee and Yohn (2009)). Consistent with private firms being relatively opaque, Minnis and Sutherland (2017) report that only 43% of private firms share tax returns with their bank, which is notable considering all private firms must prepare and file tax returns with the IRS. Second, we split the sample into lending relationships that are old (above the median relationship age) versus those that are new (below the median relationship age). Finally, we split the private firms into those that have social media accounts (Facebook, Twitter, or LinkedIn), and those that do not. Many less-visible firms use social media as a platform to disseminate important firm-specific information (Blankespoor, Miller, and White (2013); Blankespoor (2018)). Information that may be relevant in predicting credit risk includes information about new products, product recalls, and general information about the business (Zhou, Lei, Wang, Fan, and Wang (2014); Lee, Hutton, and Shu (2015)).

We determine public/private status by matching borrowing firms to Compustat based on name and location. We determine relationship age using a field reported in the Credit2B database which lists the date of the first sale between the borrower and lender. Social media presence is recorded by manually searching Facebook, LinkedIn, and Twitter for the 28,000 private borrowers in the sample. We estimate Equation 1 for each cross-sectional sample, where the dependent variable is the absolute value of the percentage deviation from the recommended credit line (*|%Deviation|*). Results are reported in Table 6, columns 1-6. Consistent with deviations capturing private information, we find that the treatment group has a larger increase in their deviation behavior relative to the control group for their portfolio of private borrowers. We do not capture a statistically significant difference in the change in deviation behavior between the treatment and control groups for public borrowers, though we note that the sample of public borrowers is small. In fact, we fail to find a statistically significant difference in the coefficients between columns (1) and (2).[28] When we decompose the portfolio of private borrowers into those that have social media and those that do not,

[28] We re-run Equation 1 and fully interact the model with each cross-sectional variable in order to assess statistical differences between the two sub-portfolios.

we find that the treatment group has a larger increase in the use of discretion than the control group only for the set of privately held firms that do not have social media. The coefficient of interest in column (5) is statistically different from the coefficient in column (6); showing that lenders do not augment the model for their private borrowers without social media accounts suggests that current AI-based lending models may go beyond incorporating traditional hard data like financial statements and have advanced to gathering and using alternative sources of information.

We also find that the treatment group exhibits a larger increase in the deviation than the control group for both old and young relationships, though the magnitude is larger for younger relationships. This result contrasts the relationship lending literature, which suggests that lenders gather information over time, thus should have more, not less, information about older borrower. Discussions with lenders reveal that they perform a lot of diligence at the *onset* of new relationships and in turn gather private information. In contrast, the model does not have a lot of repeated transactions for that borrower to train on. Thus, there is both higher access and higher scope for private information in these instances. Overall we interpret the evidence in Table 6, columns 1-6, as consistent with deviation activity capturing some components of *private information* in the lending decision.[29]

6.2.2 Competition

Lenders may deviate from the model to compete over borrowers/customers. To capture this construct, we calculate the median number of lenders that the borrowers have. If the borrower has above the median number of lenders, they are classified as a high competition borrower, else they are classified as a low competition borrower. We estimate Equation 1 for each cross-sectional sample, where the dependent variable is the absolute value of the percentage deviation from the recommended credit line (*|%Deviation|*). Results are reported in Table 6, columns 7-8. Consistent with lenders deviating from the model for competitive purposes, we find that the treatment group has a larger increase in their deviation behavior

[29] Another concern is that our measure is capturing lenders' adjustments for *public* information that is not properly captured by the model. If this were the case, we would expect all treatment lenders to adjust a given borrower's credit line and to do so by the same amount. Within each borrower-month we calculate the standard deviation of the adjustments that the treated lenders make. The dispersion is quite large: more than three times the mean adjustment, suggesting that the adjustments on average are not representative of publicly-available information.

relative to the control group for the sample of high competition borrowers. We do not capture a statistically significant difference in the change in deviation behavior between the treatment and control groups for the sample of low competition borrower, though we fail to find a statistically significant difference in the coefficients between columns (7) and (8).[30]

7 Main Results: The impact of private information on loan outcomes

Next, we move to investigating our main research question: whether augmenting a machine-based model with lender discretion leads to an improvement or deterioration in lending outcomes. If augmenting the machine with discretion leads to better decisions, then it should result in improvements in loan performance. There are two ways discretion might be useful in this setting: (1) it helps lenders identify poor quality borrowers that are not captured by the model (i.e., overvalued borrowers), thereby allowing lenders to cut lending to this group or (2) it helps lenders identify high quality borrowers that are not captured by the model (i.e., undervalued borrowers), thereby allowing lenders to expand lending to this group. In other words if used efficiently, discretion should reduce the machine's Type I and Type II errors. Alternatively, if lending agents are subject to agency problems or behavioral biases, the use of discretion may lead to worse outcomes, relative to the control group.

Since treatment lenders use the slider to reallocate their fixed portfolio, we measure performance at the lender-portfolio level.[31] Specifically, if discretion is valuable in the lending decision, we would expect that lenders in the treatment group see a larger improvement in their portfolio performance after the intervention, relative to the control group. Upon treatment, lenders should place more weight on borrowers that are undervalued by the model and less weight on borrowers that are overvalued by the model. Our DD approach compares the portfolio-level weighted average change in loan outcome for the treatment group relative to the control group and is estimated by OLS using the following specification:

$$y_{i,t} = \beta_1 Ahead_t + \beta_2 Treatment_i * Ahead_t + \alpha_i + \alpha_t + \varepsilon_{i,t} \quad (2)$$

[30] In untabulated tests, we find that there are more positive deviations than negative deviations for the sub-set of high competition borrowers, consistent with lenders expanding credit access to beat out rival lenders.

[31] As discussed in Section 6 the lender is measured at the establishment level. It is most appropriate to use the establishment-level data, since each establishment has autonomy in lending decisions and their own credit portfolio.

where y is the weighted average portfolio outcome for lender i in month t. *Ahead* is a dummy variable equal to one for the two months after treatment (January-February 2019), and *Treatment* is a dummy indicating whether the lender has been randomly assigned the slider. Because our analyses are at the portfolio month level, we include lender and month fixed effects. Note that the *Treatment* variable is defined on the basis of the randomization results, irrespective of whether the lender deviates from the machine or not, so our regressions estimate the intent-to-treat effects.[32]

We calculate the weighted average portfolio performance for each lender using the weights placed on each borrower as of the treatment month (December 2018).[33] The timeframe we use to capture effects is December 2018-February 2019, since our aim is to capture the change in forward-looking portfolio performance from the treatment date to the subsequent period. Thus, the DD specification should capture any incremental improvement in forward-looking portfolio performance for the treatment lenders relative to the control lenders. If using discretion allows lenders to reduce Type I and Type II errors, then the treatment lenders should outperform the benchmark model.

7.1 Outcome of Interest: Forward looking payment behavior

The first lending portfolio outcome we investigate is changes in the payment behavior of borrowers. If lenders are using private information to predict who will pay and who won't, we would expect them to place more weight on borrowers that improve their payment performance and less weight on borrowers that show a future decline in payment performance. Consistent with this type of information entering into the credit decision, Appendix A5, examples 2 and 7 show that the lender received private information that the borrower will pay, despite weak fundamentals.

To assess payment performance, we use two variables. The first is the percentage of the current period loan that the borrower defaults on. To calculate this, we compare the loan that classified as "current" in one month but is rolled into the "1-30 day late" bucket in the

[32]It is standard in the experimental literature to estimate intent-to-treat effects rather than per-protocol effects. In untabulated tests, we use a different specification: $y_{i,j,t} = \beta_1 Ahead_t + \beta_2 Treatment_i * Ahead_t * Deviation_{i,j} + \alpha_i + \alpha_j + \alpha_t + \varepsilon_{i,j,t}$. This is estimated at the loan-level, where $Deviation_{i,j}$ is fixed in the month of December 2018. Results are qualitatively similar. We thank the referee for this suggestion.

[33]Our weights remain fixed in December 2018 in order to capture how loan decisions as of the treatment date impact subsequent portfolio performance.

next month. We estimate Equation 2, where the dependent variable is the lender's weighted average portfolio default percentage. Therefore the coefficient on the variable of interest, *Treatment*Ahead*, represents the incremental improvement or decline in the future default percentage for treatment lenders relative to control lenders. Results are reported in Table 7, Panel A and indicate that treatment lenders have a 13% larger reduction in forward-looking payment defaults relative to control lenders. The first column captures the overall portfolio performance, and columns (2) through (7) capture the performance for the sub-portfolios of private, public, new, old, social=0, social=1, LowComp, HighComp borrowers. Interestingly, the result is primarily concentrated in the portfolio of private borrowers, and specifically those private borrowers that do not have social media. A statistical test of the difference in coefficients reveals that the coefficient of interest in column 6 is significantly different from that in column 7. These firms are particularly opaque and may be best suited for relationship loans. It is worth highlighting that the machine does just as well as the treatment group for the sample of private borrowers that are present on social media. The result challenges the conventional assumption in the literature that all small, private firms benefit from relationship lending, and it highlights the growing prominence of AI-based models in crowding out some of the human's role.[34] The result in column 9 shows that, in the sub-portfolio of borrowers that lenders face competition over, using discretion leads to *worse* payment performance relative to the control group. While preliminary, this result is consistent with humans may be optimizing over multiple variables (i.e., maximizing future profits while minimizing credit risk), rather than focusing solely on minimizing risk.

As a second measure of payment performance, we investigate the likelihood of a loan being transferred to a collection agency. This can be viewed as an extreme version of a default, since the receivable is so old that the lender has placed it for collections. Credit2B and their data partners obtain collections data from multiple sources. We estimate Equation 2, where the dependent variable is the lender's portfolio-level weighted average probability of a loan being placed for collection. Similar to the previous analysis, the coefficient on the variable of interest, *Treatment*Ahead*, represents the differential forward looking change in

[34]Though treatment lenders outperform control lenders for both New and Old relationships, they are not statistically different from each other.

the probability of a loan in the portfolio being placed for collection for treatment lenders relative to control lenders. Results are reported in Table 7, Panel B and indicate that treatment lenders have a larger reduction in the probability that their loans will be placed for collection relative to control lenders. Consistent with Panel A, the result is primarily concentrated in the portfolio of private borrowers without social media. Further, treatment lenders are willing to take on higher collection risk for their highly competitive borrowers.

The average treatment effects in Table 7 suggest that using additional discretion in the lending decision leads to better performance, relative to the machine-based model. What's more interesting is that this superior performance is not across the board; the treated lenders do not perform differentially to the model-based approach for public borrowers, or for private borrowers that advertise on social media. Further, in competitive landscapes discretion leads to worse outcomes – in terms of payment performance – relative to the machine-based model.

7.2 Outcome of Interest: Forward looking credit quality

The second lending portfolio outcome we investigate is changes in credit quality, as measured by the machine-generated credit score. Thus, the assumption made in looking at forward-looking changes in credit scores is that at time t (i.e., December 2018), the model does not incorporate the information relevant in the lender's use of discretion, but at time $t+1$, the information is impounded in the score.

The benefit of using credit scores to capture changes in performance is that they update on a monthly basis and thus capture short term changes in portfolio quality. The drawback is that the model may mechanically improve performance measures if previous changes in credit lines are an input to the future score model. Thus, lenders aren't using private information to predict future performance, rather their lending behavior is mechanically driving changes in credit scores. To mitigate such a concern, we use credit scores from Experian as our main outcome variable. The Experian scoring model is broader than the Credit2B scoring model and does not use trade lending behavior from Credit2B, thus their scoring model will not suffer from the mechanical concern raised above. The scores range from 0 to 100, and the sample average is a 71[35] Table 8, Panel A reports the results, where the dependent variable is

[35]Experian scores are not strictly trade credit based. Credit2B has access to these scores through data agreements.

the weighted average portfolio-level score. The interaction term, *Treatment*Ahead* captures the change in future portfolio performance for the treatment lenders relative to the same change for the control lenders. The first column captures the overall portfolio performance, and columns (2) through (9) capture the performance for the sub-portfolios of private, public, new, old, social=0, social=1, LowComp, HighComp borrowers. We find that, on average, treatment lenders predict forward-looking changes in credit quality. The treatment lenders' portfolios show a 2-3 point greater weighted-average increase in the score relative to control lenders. Consistent with prior results, the performance effect is concentrated in the portfolio of private borrowers, specifically those without social media accounts. Further, treatment lenders are willing to take on higher credit risk for their highly competitive borrowers, with a 1.2 point greater weighted average decline in score, relative to the control group.

Though we recognize that the Credit2B scores may suffer more from the mechanical relationship between credit volume and future credit scores, we nonetheless repeat the analysis with these scores as the dependent variable. Results are reported in Table 8, Panel B, and are largely consistent with Panel A in magnitude and significance. Overall, we recognize that the effects captured in Table 8 are somewhat modest in magnitude. However, our limited experimental period only captures two months of forward looking performance. Thus the effects we capture likely understate the full effects of private information in predicting forward looking credit quality.

Taken together, the evidence in Tables 7 and 8 indicate that, on average, additional discretion is useful in mitigating forward-looking credit risk. This evidence suggests that the adjustments have valuable information content, rather than simply capturing random adjustments or irrational biases.

7.3 Outcome of Interest: Sales orders

Thus far, we've show preliminary evidence that lenders use discretion for competitive purposes. However, if they use discretion rationally, they may choose to relax credit terms in exchange for higher sales. Consistent with this idea, Barrot (2015) shows that offering trade credit is often used as a competitive tool to win sales, and Cunat (2007) and Wilner (2000) show that suppliers use trade credit to lock-in important buyers. To test this theory, we estimate Equation 2, where the dependent variable is the log of the total order volume at the lender-month level. We obtain sales order volume from Credit2B's partnership with Experian. The primary source of this data is direct contact and projections from the borrowers. We note that because the data is aggregated to the lender level, we cannot investigate portfolio sub-groups as in the previous analyses.

The results are reported in Table 9. The effect is significant – lenders in the treatment group have a 68% larger increase in orders relative to the control lenders. Though our data do not allow us to drill down to borrower-level sales data, the results tell a potentially compelling story. Lenders use discretion to expand credit to attractive borrowers. By expanding credit, they may be better able to capture future sales, consistent with the role of credit as a competitive tool (Cunat (2007); Wilner (2000)).

8 Robustness Tests and Caveats

The goal of our empirical approach is to isolate the role of discretion in lending outcomes. Through randomization of treatment, we believe that our empirical results capture whether augmenting a machine with additional human discretion improves or impedes loan outcomes. It is possible, however, that the experiment also causes lenders to adjust other dials, which would confound our inferences.

First, one may be concerned that the experiment causes treated lenders to adjust their monitoring effort, which would have a direct impact on loan outcomes. We cannot observe monitoring effort, and therefore cannot fully rule out this possibility. In an effort to assess potential concerns, we reached out to suppliers in the treatment group to gather evidence on any potential changes in monitoring.[36] Anecdotally, we find that the treatment lenders did not adjust staffing or incentive plans over the short window of our experiment. Based on this, we make three reasonable assumptions about our setting: (1) Loan officers face constraints in the total loan portfolio in the short-run; (2) Correspondingly, total loan-officer labor hours are also sticky in the short run of our experiment; (3) Incentive plans remained unchanged during the short span of our experiment. Since our experiment only spanned two months and was largely unanticipated by the lenders, these assumptions seem valid. Therefore, *total* monitoring effort remained unchanged from the pre- to the post-period, though monitoring

[36] An email was sent out to a select group of lenders in September, 2019 that inquired about changes in staffing, incentive plans, and total credit-team hours. We obtained responses from 7 lenders.

effort was likely reallocated away from some borrowers towards others according to the portfolio adjustments made from the slider. To explain our results, monitoring effort would have to be *suboptimally* allocated in the pre-period, and then become *optimally* allocated in the post-period. It seems unlikely that the introduction of the slider resulted in more efficient allocation of fixed monitoring hours. If any of these assumptions are flawed or our anecdotal evidence does not extrapolate to the population, our inferences could be flawed.

Another concern may be that the presentation of the slider – specifically, the bounds offered to the treated lenders – cause the lender to infer that she should deviate more (e.g., she infers that a wide range means the model-based recommendation is very uncertain, while a tight range means the model-based recommendation is precise). The range is calculated based on Credit2B's proprietary formula, which we cannot disclose. We concur that if the range is correlated with borrower type, whereby the slider is tighter when the model is sharp and looser where the model is uncertain, this could hinder our inferences from the cross-sectional analyses in Table 6. However, the size of the slider bounds would be difficult to explain our performance results since lenders have to decide whether to increase or decrease credit in order to balance the overall portfolio. Nevertheless, in untabulated tests we control for the slider range and find qualitatively similar results.

A third caveat is that our primary analysis investigates loan adjustments on the intensive margin. While we cannot directly observe loan denials, we assume that a 'missing' transaction is equivalent to a loan rejection. If a borrower was in the sample in the pre-period but not in the post, we assume that the borrower is still transacting with the supplier, asked for credit, but got denied. This is a noisy measure of loan rejections, since it is possible that the borrower did not demand credit in that month (i.e., there was no sales transaction), but the measure should not be biased. Including these loan rejections in our tests, we find that our results are robust though slightly larger in magnitude (untabulated).

Another concern is that, because lenders have fixed portfolio of credit, the sequencing of loan requests may influence our results. For example, if lenders have incentives to expand credit access, those borrowers that arrive at the beginning of the month get more credit. By the end of the month, lenders are close to their budget constraint so those borrowers get less credit. If the timing of requests correlate with borrower type, this could confound our inferences. Though we don't observe the date of the loan request, each transaction is numbered by the lender/supplier. For each month, we split the transaction numbers into quartiles and control for the transaction number quartile in the regression. Results (untabulated) are similar in magnitude and significance.

Fifth, the focus of our study is whether discretion helps to assess credit quality. It is possible, though, that lenders may also use the slider to expand credit to borrowers that have an unexpected increase in their demand for products. The machine takes several steps to recommend the credit line: first, it predicts demand; then, it develops a risk score and discounts demand using the risk score to come to a recommended credit limit. We concede that the lender may use her discretion to adjust either the first step (demand prediction) or the second step (credit risk prediction), or both. Ultimately, these adjustments reflect information that the lender has that differs from the machine (discretion), but we agree that it isn't about credit quality per se. To address this empirically, we separate borrowers into two groups based on the past year's volatility in demand. Specifically, we calculate the borrower's coefficient of variation in the past year's monthly purchases, and separate borrowers into two groups – those with above median demand volatility and those with below median demand volatility. If the lenders deviate to account for unexpected shifts in the borrower's demand, we expect more deviations in the subset of high-purchase-volatility borrowers. Results (untabulated) show that there is no difference in the two groups. These results provide us with some assurance that the deviation behavior is not fully driven by unexpected shifts in demand.

Finally, like other randomized controlled trials, we face concerns about identifying the appropriate cross-sections to capture heterogeneous treatment effects. We did not submit a pre-registered plan due to the speed of implementation at the data provider. Choosing subgroups ex post opens a potential concern of data mining and p-hacking. This concern is exacerbated when there is a potentially large array of subgroups, and little guiding principles on which of those are likely to be relevant (Chernozhukov, Demirer, Duflo, and Fernandez-Val (2018)). Because our cross-sectional splits are guided by well-established theory in the lending literature, the concern in our setting is muted. Nonetheless, for completeness and transparency, we follow the recent RCT literature in economics and use a machine-learning

(ML) approach to estimate heterogeneous treatment effects. Specifically, we implement the Athey and Imbens (2016) causal tree methodology to identify the relevant cross-sectional splits and calculate the conditional average treatment effects. Using this methodology, we find that the largest treatment effects are realized in the portfolio of borrowers with the following characteristics: are privately held, have less than or equal to one social media account, and have least 4 suppliers.[37] The results of this machine-learning approach line up with our main cross-sectional tests, which show the largest effects for private borrowers, without social media, and with a high level of competition. Overall, we take comfort that our main cross-sectional tests capture appropriate heterogeneity in treatment.

9 Conclusion and Suggestions for Future Research

We study whether the use of human discretion is useful for improving lending decisions in the context of a randomized, controlled experiment. Using a large group of lenders that rely on machine-generated credit scoring models provided by a third party to make monthly credit decisions, we design an intervention that allows some lenders to augment the machine-generated model with additional discretion about a borrower. We then track how treatment lenders, who use both the model-based information and their discretion, perform relative to the benchmark of lenders who only use the model-based information.

In the cross-section we document that augmenting the machine with discretion is useful in some cases, but not all. In fact, the control group does just as well as the treatment group in predicting credit risk for borrowers that have been traditionally classified as opaque. Specifically, we find that additional human discretion used by the treatment group is only useful in the set of privately held firms that do not have social media. This shows that AI models may go beyond incorporating traditional hard data like financial statements and have advanced to gathering and using alternative sources of information. Our result challenges the conventional assumption in the literature that all small, private firms benefit from relationship lending, and it highlights the growing prominence of AI-based models in crowding out some of the human's role. Second, we find that the treatment group does worse in terms of forward-looking credit quality than the control group in cases where there is competition

[37] The CATE for this subgroup is estimated at 0.34, which is the difference between the treatment and control sample averages of deviations.

for the borrower. However, the treatment group also realizes a greater increase in sales volume. This finding suggests that humans may be optimizing over multiple variables (i.e., maximizing future profits while minimizing credit risk), whereas models are typically trained on one outcome variable (i.e., minimizing credit risk). An important insight may be that machine-based decisions are too rigid in that they pre-specify a limited number of outcome variables to optimize

Our empirical setting includes trade creditors, which represent an important source of financing for U.S. firms. Future research may explore the role of discretion in bank lending, where there is significantly more heterogeneity in lender-type, ranging from large commercial banks to small community banks. We also note that our study relies on a relatively sophisticated machine-learning model. However, these models are likely to become more prevalent and more sophisticated in the future. We leave it to future research to explore the evolution in scope for human discretion over time.

References

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

Allee, K. D. and T. L. Yohn (2009): "The demand for financial statements in an unregulated environment: An examination of the production and use of financial statements by privately held small businesses," *The Accounting Review*, 84, 1–25.

Athey, S. and G. Imbens (2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 113, 7353–7360.

Banerjee, A., S. Cole, and E. Duflo (2009): "Default and punishment: Incentives and lending behavior in Indian banks," *Manuscript. Harvard Business School.*

Barrot, J.-N. (2015): "Trade credit and industry dynamics: Evidence from trucking firms," *The Journal of Finance.*

Berger, A. N. and G. F. Udell (1995): "Relationship lending and lines of credit in small firm finance," *Journal of Business*, 351–381.

Biais, B. and C. Gollier (1997): "Trade credit and credit rationing," *Review of Financial Studies*, 10, 903–937.

Blankespoor, E. (2018): "Firm communication and investor response: A framework and discussion integrating social media," *Accounting, Organizations and Society*, 68, 80–87.

Blankespoor, E., G. S. Miller, and H. D. White (2013): "The role of dissemination in market liquidity: Evidence from firms' use of Twitter?" *The Accounting Review*, 89, 79–112.

Bloomfield, R. J. (2002): "The "incomplete revelation hypothesis" and financial reporting," *Accounting Horizons*, 16, 233–243.

Bolton, P. and D. S. Scharfstein (1990): "A theory of predation based on agency problems in financial contracting," *The American economic review*, 93–106.

Boot, A. W. (2000): "Relationship Banking: What Do We Know?" *Journal of Financial Intermediation*, 9, 7–25.

Boot, A. W. and A. V. Thakor (2000): "Can relationship banking survive competition?" *The journal of Finance*, 55, 679–713.

Butaru, F., Q. Chen, B. Clark, S. Das, A. W. Lo, and A. Siddique (2016): "Risk and risk management in the credit card industry," *Journal of Banking & Finance*, 72, 218–239.

Campbell, D., M. Loumioti, and R. Wittenberg-Moerman (2019): "Making sense of soft information: Interpretation bias and loan quality," *Journal of Accounting and Economics*, 101240.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018): "Generic machine learning inference on heterogenous treatment effects in randomized experiments," Tech. rep., National Bureau of Economic Research.

Choi, W. G. and Y. Kim (2005): "Trade credit and the effect of macro-financial shocks: Evidence from US panel data," *Journal of Financial and Quantitative Analysis*, 40, 897–925.

Chouldechova, A. and A. Roth (2018): "The frontiers of fairness in machine learning," *arXiv preprint arXiv:1810.08810.*

Costello, A. M. (2018): "Credit Market Disruptions and Liquidity Spillover Effects in the Supply Chain," *Working Paper, University of Michigan.*

——— (2019): "The Value of Collateral in Trade Finance," *Journal of Financial Economics*, 134, 70–90.

Costello, A. M., A. Down, and M. N. Mehta (2019b): "The Supply and Demand of Private Information in the Lending Decision," *Working Paper, University of Michigan.*

Cunat, V. (2007): "Trade credit: suppliers as debt collectors and insurance providers," *Review of Financial Studies*, 20, 491–527.

Cyert, R. M., J. G. March, et al. (1963): "A behavioral theory of the firm," *Englewood Cliffs, NJ*, 2, 169–187.

Degryse, H., J. M. Liberti, T. Mosk, and S. Ongena (2013): "Is loan officer discretion advised when viewing soft information?" *Working paper, Northwestern University.*

Degryse, H. and S. Ongena (2007): "The impact of competition on bank orientation," *Journal of Financial Intermediation*, 16, 399–424.

Diamond, D. W. (1991): "Monitoring and reputation: The choice between bank loans and directly placed debt," *Journal of Political Economy*, 99, 689–721.

Drexler, A. and A. Schoar (2014): "Do relationships matter? Evidence from loan officer turnover," *Management Science*, 60, 2722–2736.

Emery, G. W. (1984): "A pure financial explanation for trade credit," *Journal of financial and quantitative analysis*, 19, 271–285.

Engelberg, J., P. Gao, and C. A. Parsons (2012): "Friends with money," *Journal of Financial Economics*, 103, 169–188.

Floyd, E. and J. A. List (2016): "Using field experiments in accounting and finance," *Journal of Accounting Research*, 54, 437–475.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2018): "Predictably unequal? The effects of machine learning on credit markets," *The Effects of Machine Learning on Credit Markets.*

Ganeshapillai, G., J. Guttag, and A. Lo (2013): "Learning connections in financial time series," in *International Conference on Machine Learning*, 109–117.

Hastie, R. and R. M. Dawes (2010): *Rational choice in an uncertain world: The psychology of judgment and decision making*, Sage.

Heider, F. and R. Inderst (2012): "Loan prospecting," *The Review of Financial Studies*, 25, 2381–2415.

Hertzberg, A., J. M. Liberti, and D. Paravisini (2010): "Information and incentives inside the firm: Evidence from loan officer rotation," *The Journal of Finance*, 65, 795–828.

Hirshleifer, D. and S. H. Teoh (2003): "Limited attention, information disclosure, and financial reporting," *Journal of Accounting and Economics*, 36, 337–386.

Ivashina, V. and Z. Sun (2011): "Institutional demand pressure and the cost of corporate loans," *Journal of Financial Economics*, 99, 500–522.

Kahneman, D. (2003): "Maps of bounded rationality: A perspective on intuitive judgment and choice," *Nobel prize lecture*, 8, 351–401.

Kahneman, D. and S. Frederick (2002): "Representativeness revisited: Attribute substitution in intuitive judgment," *Heuristics and biases: The psychology of intuitive judgment*, 49, 81.

Karolyi, S. A. (2018): "Personal lending relationships," *The Journal of Finance*, 73, 5–49.

Khan, U., X. Li, C. D. Williams, and R. Wittenberg-Moerman (2018): "The Effect of Information Opacity and Accounting Irregularities on Personal Lending Relationships: Evidence from Lender and Manager Co-Migration," *The Accounting Review*.

Khwaja, A. I. and A. Mian (2008): "Tracing the impact of bank liquidity shocks: Evidence from an emerging market," *American Economic Review*, 98, 1413–42.

Lee, L. F., A. P. Hutton, and S. Shu (2015): "The role of social media in the capital market: Evidence from consumer product recalls," *Journal of Accounting Research*, 53, 367–404.

Libby, R., R. Bloomfield, and M. W. Nelson (2002): "Experimental research in financial accounting," *Accounting, Organizations and Society*, 27, 775–810.

Lim, S. S. and S. H. Teoh (2010): "Limited attention," *Behavioral finance: Investors, corporations, and markets*, 295–312.

Lipshitz, R. and N. Shulimovitz (2007): "Intuition and emotion in bank loan officers' credit decisions," *Journal of Cognitive Engineering and Decision Making*, 1, 212–233.

Love, I., L. A. Preve, and V. Sarria-Allende (2007): "Trade credit and bank credit: Evidence from recent financial crises," *Journal of Financial Economics*, 83, 453–469.

Minnis, M. (2011): "The value of financial statement verification in debt financing: Evidence from private US firms," *Journal of Accounting Research*, 49, 457–506.

Minnis, M. and A. Sutherland (2017): "Financial statements as monitoring mechanisms: Evidence from small commercial loans," *Journal of Accounting Research*, 55, 197–233.

Nilsen, J. H. (2002): "Trade Credit and the Bank Lending Channel," *Journal of Money, Credit, and Banking*, 34, 226–253.

Paravisini, D. and A. Schoar (2013): "The incentive effect of scores: Randomized evidence from credit committees," Tech. rep., National Bureau of Economic Research.

Petersen, M. A. and R. G. Rajan (1994): "The benefits of lending relationships: Evidence from small business data," *The Journal of Finance*, 49, 3–37.

——— (1997): "Trade credit: theories and evidence," *Review of Financial Studies*, 10, 661–691.

——— (2002): "Does distance still matter? The information revolution in small business lending," *The Journal of Finance*, 57, 2533–2570.

Puri, M., J. Rocholl, and S. Steffen (2011): "Rules versus discretion in bank lending decisions," *Available at SSRN 1786831*.

Rudin, C. (2018): "Please Stop Explaining Black Box Models for High Stakes Decisions," *arXiv preprint arXiv:1811.10154*.

Smith, J. K. (1987): "Trade credit and informational asymmetry," *Journal of Finance*, 863–872.

Stein, J. C. (2002): "Information production and capital allocation: Decentralized versus hierarchical firms," *The Journal of Finance*, 57, 1891–1921.

Stuhlsatz, A., C. Meyer, F. Eyben, T. Zielke, G. Meier, and B. Schuller (2011): "Deep neural networks for acoustic emotion recognition: raising the benchmarks," in *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 5688–5691.

Sutherland, A. (2018): "Does credit reporting lead to a decline in relationship lending? Evidence from information sharing technology," *Journal of Accounting and Economics*, 66, 123–141.

Telser, L. G. (1966): "Cutthroat competition and the long purse," *The Journal of Law and Economics*, 9, 259–277.

Trivedi, C., I. Rabinovich, and S. Desai (2017): "Applying Machine Learning to Leverage Big Data for Commercial Credit Scoring," *The Credit & Financial Management Review*, 23, 295–312.

Tversky, A. and D. Kahneman (1974): "Judgment under uncertainty: Heuristics and biases," *Science*, 185, 1124–1131.

Wilner, B. S. (2000): "The exploitation of relationships in financial distress: The case of trade credit," *Journal of Finance*, 153–178.

Zhou, M., L. Lei, J. Wang, W. Fan, and A. G. Wang (2014): "Social media adoption and corporate disclosure," *Journal of Information Systems*, 29, 23–50.

10 Figures

Figure 1. Machine-Generated Credit Limit without Slider
This figure illustrates the lender's view of a standard, machine-generated recommended credit limit for a given borrower.

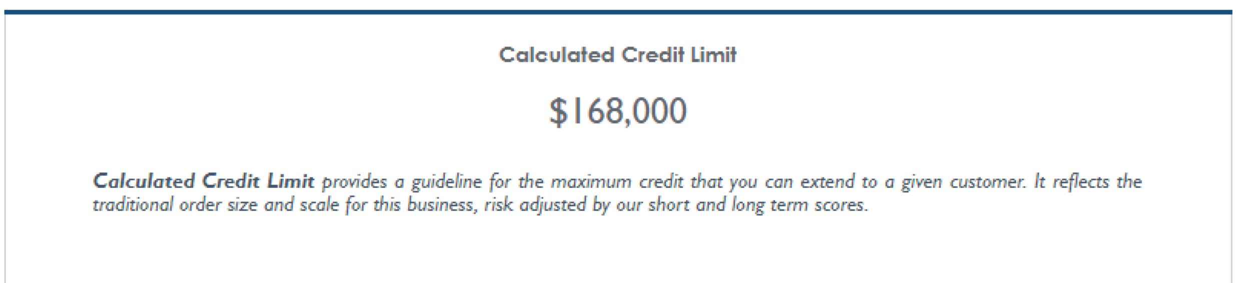


Figure 2. Timeline of Experiment

This figure displays the timeline of the experiment, from planning to execution.

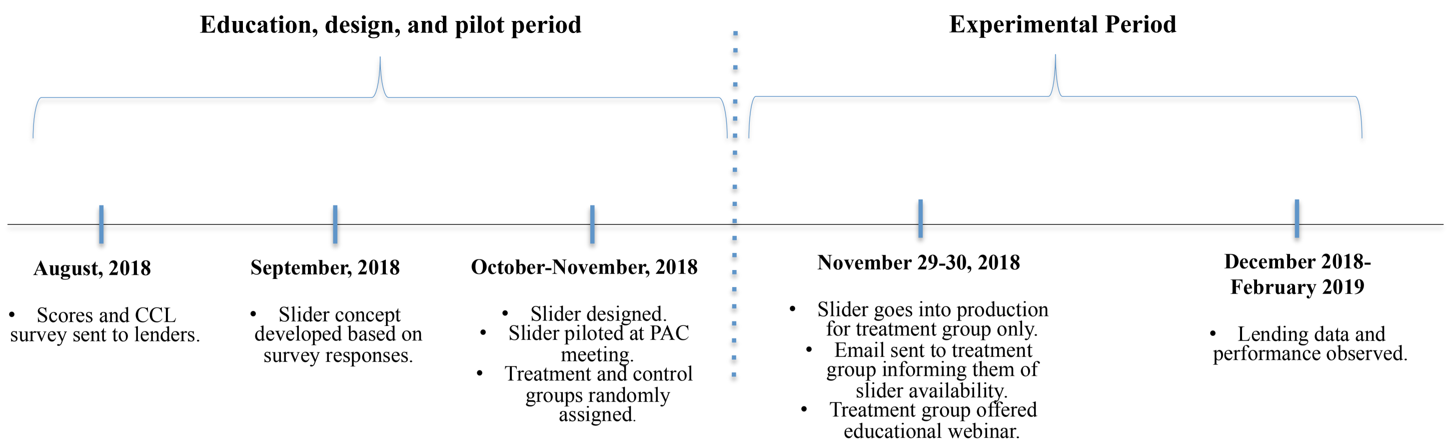


Figure 3. Slider Intervention

This figure illustrates slider intervention, which is only available to the treatment group. An adjustable toggle is set to the model-generated credit limit, which visually appears in the center of the line. Additionally, mouse-over explanations appear when a lender hovers over the words aggressive or conservative. Hovering over the word Aggressive reveals: "If your proprietary information about a specific customer suggests that you should take a more aggressive approach, move the slider to the right, and the credit limit will increase." Hovering over the word Conservative reveals "If your proprietary information about this customer suggests that you should take a more conservative approach, move the slider to the left, and the credit limit will decrease."

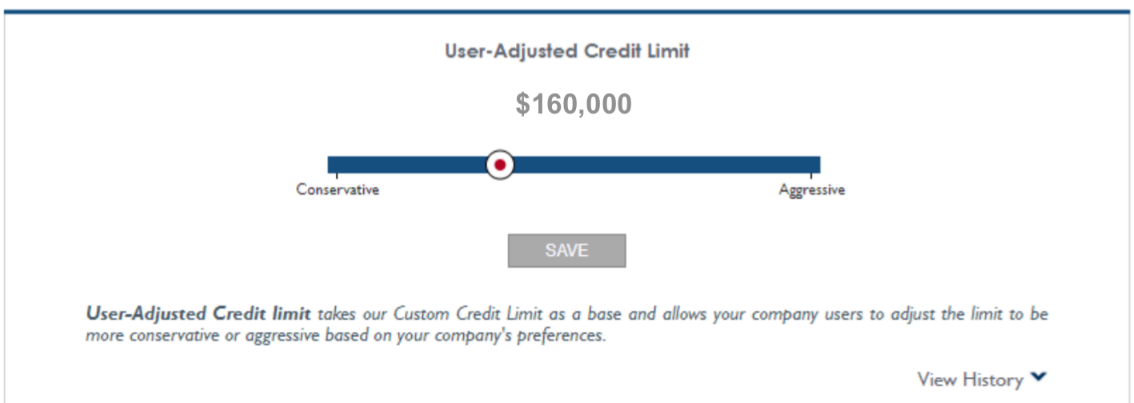


Figure 4. Deviation behavior over time for the treatment group and the control group

This figure plots the absolute value of the percentage difference between the actual loan outstanding and the model-generated credit line, for the treatment and control group over time. The lending behavior of the treatment group is indicated by the solid red line, and the lending behavior of the control group is indicated by the dotted blue line.

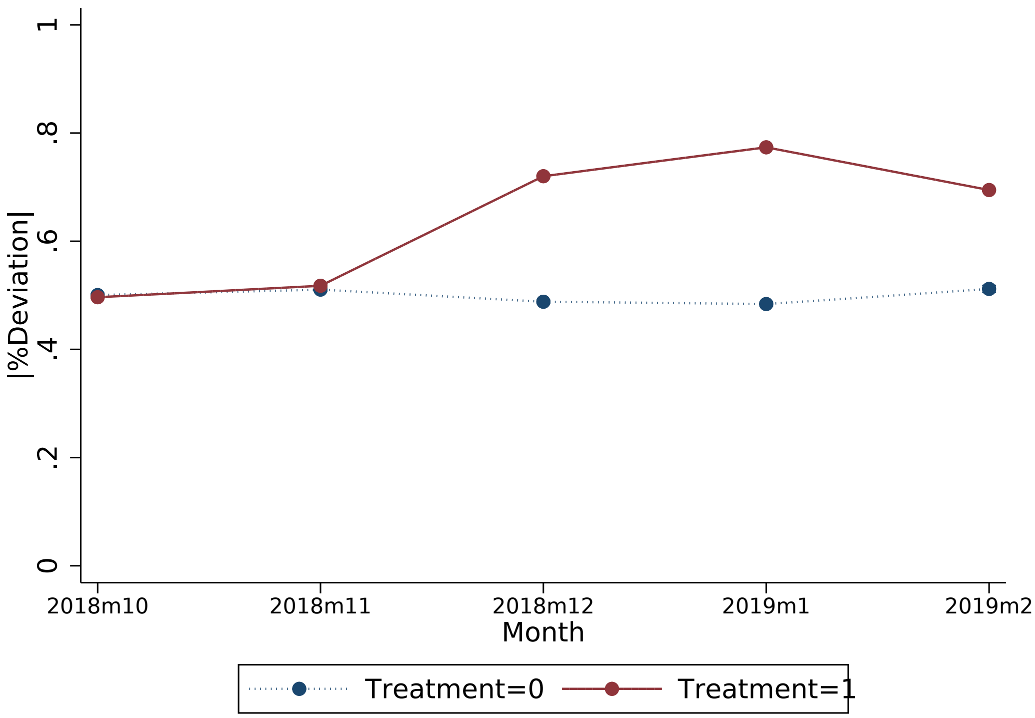


Figure 5. Portfolio Allocation
This figure displays the portfolio allocation for the treatment group in the period before treatment (November 2018), relative to the portfolio allocation upon treatment (December 2018). The x-axis includes bins of borrower credit quality. Category 1 includes borrowers with credit scores between 1 and 20; Category 2 includes borrowers with credit scores between 21 and 40; Category 3 includes borrowers with credit scores between 41 and 60; Category 4 includes borrowers with credit scores between 61 and 80; and Category 5 includes borrowers with credit scores between 81 and 100.

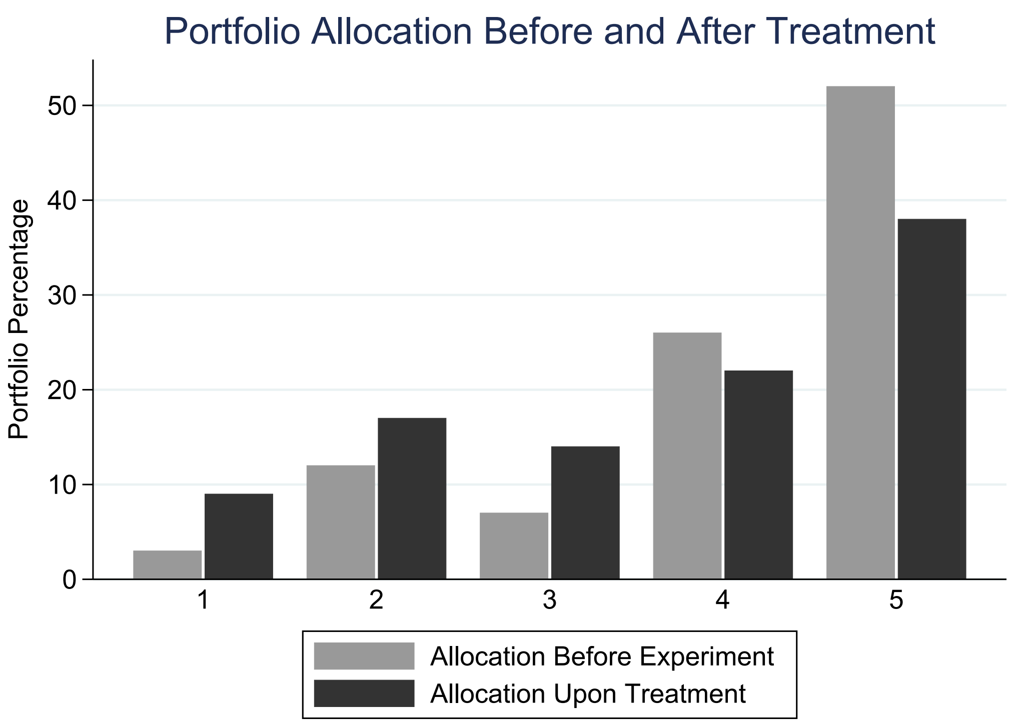


Table 1: Covariate Balance

This table reports the covariate balance between the treatment and control groups at the time of treatment assignment. *Number of Customers* is a count of the total number of customers for a given lender; *Annual Sales* is the dollar value of a lender's annual sales ;*Employee Count* is the number of employees; *Public* is an indicator variable if the lender has publicly traded equity; *% Big Customers* is the percentage of total customers in a lender's portfolio that have above the sample median total assets *% Slow Customers* is the percentage of total customers in a lender's portfolio that are past due on their trade loan; *% Domestic Customers* is the percentage of the total customers in a lender's portfolio that are located in the U.S.; *% Public Customers* is the percentage of the total customers in a lender's portfolio that have publicly traded equity. *Credit Line* is the dollar value of the machine-generated recommended credit line to a given borrower. *|%Deviation|* is the absolute value of the percentage difference between the actual dollar value of the loan and the *Credit Line*. *Experian (Credit2B) Score* is the machine-generated credit score for a given borrower; it ranges from 1 to 100; *Relationship Age* is the number of years since the lender and borrower's first transaction together. The *Difference* column calculates the difference in means for the treatment and the control group. *** , **, and * represent significant differences in the treatment and control groups at the 1%, 5%, and 10% level, respectively

		Control	*Treatment*	*Difference*	*T-Statistic*
Lender Characteristics					
	Number of Customers	1911.369	2248.084	-336.715	-0.380
	Annual Sales	353,000,000	222,000,000	131,000,000	0.956
	Employee Count	2,161	4,901	-2,740	-0.836
	Public	0.103	0.079	0.023	0.839
Portfolio Characteristics					
	% Big Customers	0.670	0.654	0.016	0.595
	% Slow Customers	0.395	0.402	-0.006	-0.182
	% Domestic Customers	0.914	0.919	-0.005	-0.265
	% Public Customers	0.053	0.049	0.004	0.265
	% Collection	0.004	0.006	-0.002	1.355
	Experian Score	62.433	62.659	-0.226	-0.475
	Credit2B Score	69.615	67.645	1.970	0.774
	Relationship Age	9.892	10.998	-1.106	1.691*
	Credit Line ($)	92,722	91,529	1,193	1.004
	|%Deviation|	0.513	0.509	0.004	1.089

Table 2: Descriptive Statistics and Compustat Benchmark

This table reports the sample descriptive statistics. Lenders and borrowers are compared to the Compustat population. To make this comparison, lenders and borrower are matched to Compustat by name and location. Thus, the statistics are presented only for the subsample of public lenders and borrowers, rather than the whole sample. *Total Assets, Long-Term Debt, Cash* and *Total Current Liabilities* are obtained from Compustat and are denoted in millions of dollars. *Age* is a calculation of the age of the firm based on the first year it appears in Compustat. The *Difference* column calculates the difference in means for the treatment and the control group. *** , **, and * represent significant differences in the treatment and control groups at the 1%, 5%, and 10% level, respectively.

	Mean	*Q1*	*Median*	*Q3*	*Compustat Mean*	*Difference*	*T-Statistic*
Lenders							
Total Assets ($M)	6,836.21	763.92	1,771.63	5,967.30	15,244.31	8,408.10	3.22***
Long-Term Debt ($M)	2,164.54	72.45	594.33	1,826.96	3,124.66	960.12	1.08
Cash ($M)	191.48	15.75	66.08	181.34	493.84	302.36	5.00***
Total Current Liabilities ($M)	1,113.45	158.19	560.66	1,299.74	1,292.95	179.50	0.58
Age	33.83	20.50	32.50	44.50	15.82	-18.01	-4.90
Borrowers							
Total Assets ($M)	13,043.40	316.18	1,542.59	6,654.97	15,244.31	2,200.91	1.19
Long-Term Debt ($M)	2,694.74	6.90	338.50	1,883.80	3,124.66	429.92	1.88*
Cash ($M)	593.04	18.22	77.50	295.62	493.84	-99.20	-1.57
Total Current Liabilities ($M)	1,816.31	43.08	239.27	992.51	1,292.95	-523.36	-2.74**
Age	27.62	13.00	23.00	37.00	15.82	-11.80	-24.06***

Table 3: Validation: Did Lenders Use the Slider?

This table reports the results of estimating Equation 1, where the dependent variable is *|%Deviation|*, which is calculated as the absolute value of the percentage difference between the actual dollar value of the loan and the *Credit Line*. *Treatment* is an indicator variable set equal to one if the lender was randomly chosen to receive the slider intervention, zero otherwise. *After* is an indicator variable set equal to one if the month is greater than or equal to December 2018, zero otherwise. Standard errors are clustered at the lender level. *** , **, and * represent significance at the 1%, 5%, and 10% level, respectively.

	|%Deviation|		
	(1)	(2)	(3)
Treatment*After	0.183***	0.185***	0.131***
	(0.038)	(0.037)	(0.031)
Observations	2,061,250	2,041,169	1,445,714
R-Square	0.758	0.805	0.889
Lender FE	Y	Y	Y
Borrower FE	Y	Y	Y
LenderxBorrower FE	N	Y	Y
BorrowerxMonth FE	N	N	Y

Table 4: Portfolio- and Loan-level Adjustments

Panel A of this table reports the results of estimating Equation 1, where the dependent variable is the natural log of the sum of the lender's monthly credit portfolio. The regression is estimated at the establishment-month level. Panel B of this table reports the results of estimating Equation 1, where the dependent variable is *|%Deviation|*, which is calculated as the absolute value of the percentage difference between the actual dollar value of the loan and the *Credit Line*. The regression is estimated at the loan-level. *Treatment* is an indicator variable set equal to one if the lender was randomly chosen to receive the slider intervention, zero otherwise. *After* is an indicator variable set equal to one if the month is greater than or equal to December 2018, zero otherwise. *Positive* is an indicator variable if the %Deviation is positive, zero otherwise. Standard errors are clustered at the lender level. *** , **, and * represent significance at the 1%, 5%, and 10% level, respectively.

(a) Panel A: Portfolio-level Changes

Treatment*After	-0.038
	(0.161)
Observations	6,680
R-Square	0.942
Lender FE	Y
Month FE	Y

(b) Panel B: Magnitude of Positive vs. Negative Adjustments

	(1)	(2)	(3)
		|%Deviation|	
Treatment*After	0.076*	0.072*	-0.010
	(0.045)	(0.043)	(0.034)
Treatment*After*Positive	0.411*	0.440**	0.589***
	(0.170)	(0.163)	(0.122)
Observations	2,061,250	2,041,169	1,445,714
R-Square	0.801	0.829	0.902
Lender FE	Y	Y	Y
Borrower FE	Y	Y	Y
LenderxBorrower FE	N	Y	Y
BorrowerxMonth FE	N	N	Y

Table 5: Correlation between *|%Deviation|* magnitude and the existence of personal comments
This table presents the correlation between the magnitude of *|%Deviation|* and the existence of a subjective comment in the loan file. Observations are grouped into *Small Deviations*, which are those below the sample median of *|%Deviation|*, and *Large Deviations*, which are those above the sample median of *|%Deviation|*. *Comment* is an indicator variable set equal to one if the lender records at least one subjective comment about the borrower over the course of 2018, zero otherwise. The T-test calculates the difference in the mean likelihood of recording a comment for these two groups of deviations.

	Large Deviations	*Small Deviations*	*Difference*	*T-Statistic*
Comment	0.179	0.062	0.117	9.409***

Table 6: Cross-Sectional use of Discretion Based on Customer Type

This table reports the results of estimating Equation 1, where the dependent variable *|%Deviation|*, as defined in the previous table. Column 1 reports the results from estimating Equation 1 on the subset of borrowers that are privately held, column 2 reports the results from estimating Equation 1 on the subset of borrowers that have publicly traded equity, column 3 reports the results from estimating Equation 1 on the subset of borrowers that have 'new' relationships with the lender (the relationship age is below the lender's median relationship), column 4 reports the results from estimating Equation 1 on the subset of borrowers that have 'old' relationships with the lender (the relationship age is above the lender's median relationship), column 5 reports the results from estimating Equation 1 on the subset of private borrowers that do not have social media (Facebook, LinkedIn, or Twitter), column 6 reports the results from estimating Equation 1 on the subset of private borrowers that are on social media, column 7 reports the results from estimating Equation 1 on the subset of borrowers lenders do not compete over (below the median number of suppliers), and column 8 reports the results from estimating Equation 1 on the subset of borrowers that lenders compete over (above the median number of suppliers). Standard errors are clustered at the lender level. *** , **, and * represent significance at the 1%, 5%, and 10% level, respectively.

	|%Deviation|							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Private	*Public*	*New*	*Old*	*Social=0*	*Social=1*	*LowComp.*	*HighComp.*
Treatment*After	0.129***	0.209	0.231***	0.085*	0.194***	0.073	0.120	0.167*
	(0.031)	(0.134)	(0.038)	(0.045)	(0.043)	(0.051)	(0.119)	(0.093)
Observations	1,443,929	1,661	76,149	83,482	154,771	1,289,158	1,148,104	285,413
R-Square	0.889	0.963	0.932	0.856	0.938	0.870	0.881	0.881
Lender FE	Y	Y	Y	Y	Y	Y	Y	Y
Borrower FE	Y	Y	Y	Y	Y	Y	Y	Y
LenderxBorrower FE	Y	Y	Y	Y	Y	Y	Y	Y
BorrowerxMonth FE	Y	Y	Y	Y	Y	Y	Y	Y

Table 7: Effect of Discretion on Portfolio Payment Performance

Panel A of this table reports the results of estimating Equation 2, where the dependent variable is the lender's weighted-average portfolio default percentage. Default percentage is calculated as the portion of a loan that is rolled into the past due bucket. Weights are held fixed as of the treatment date (December 2018). *Treatment* is an indicator variable set equal to one if the lender was randomly chosen to receive the slider intervention, zero otherwise. *Ahead* is an indicator variable set equal to one if the month is greater than or equal to January 2019, zero otherwise. Column 1 reports the results from estimating Equation 2 on portfolio of all of the lender's loans. Columns 2 through 9 report the results from estimating Equation 2 on the subset of the portfolio of *Private, Public, New, Old, Social=0, Social=1, LowComp* and *HighComp* clients, as defined in previous tables. Panel B of this table reports the results of estimating Equation 2, where the dependent variable is lender's weighted-average probability that a loan has been placed for collection. In all specifications standard errors are clustered at the lender level. *** , **, and * represent significance at the 1%, 5%, and 10% level, respectively.

(a) Panel A: Portfolio Aging

	% Aged								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Overall	*Private*	*Public*	*New*	*Old*	*Social=0*	*Social=1*	*LowComp.*	*HighComp.*
Treatment*Ahead	-0.130**	-0.162***	-0.013	-0.123**	-0.152***	-0.178**	-0.045	-0.051	0.249*
	(0.040)	(0.045)	(0.038)	(0.048)	(0.037)	(0.051)	(0.050)	(0.115)	(0.149)
Observations	4,008	2,548	1,460	840	1,736	512	2,036	1,336	2,532
R-Square	0.406	0.350	0.533	0.407	0.322	0.358	0.345	0.423	0.406
Lender FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

(b) Panel B: Likelihood of Collection

	Collection								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Overall	*Private*	*Public*	*New*	*Old*	*Social=0*	*Social=1*	*LowComp.*	*HighComp.*
Treatment*Ahead	-0.006**	-0.009**	0.006	0.000	-0.014**	-0.009***	-0.002	-0.008***	0.002**
	(0.002)	(0.004)	(0.005)	(0.000)	(0.005)	(0.002)	(0.003)	(0.002)	(0.001)
Observations	4,008	2,548	1,460	840	1,736	512	2,036	1,336	2,532
R-Square	0.239	0.243	0.446	0.073	0.227	0.198	0.116	0.248	0.212
Lender FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 8: Effect of Discretion on Lending Outcomes: Credit Scores

Panel A of this table reports the results of estimating Equation 2, where the dependent variable is the lender's weighted-average portfolio Experian credit score. Weights are held fixed as of the treatment date (December 2018). *Treatment* is an indicator variable set equal to one if the lender was randomly chosen to receive the slider intervention, zero otherwise. *Ahead* is an indicator variable set equal to one if the month is greater than or equal to January 2019, zero otherwise. Column 1 reports the results from estimating Equation 2 on portfolio of all of the lender's loans. Columns 2 through 9 report the results from estimating Equation 2 on the subset of the portfolio of *Private, Public, New, Old, Social=0, Social=1, LowComp* and *HighComp* clients, as defined in previous tables. Panel B of this table reports the results of estimating Equation 2, where the dependent variable is lender's weighted-average portfolio Credit2B credit score. In all specifications standard errors are clustered at the lender level. *** , **, and * represent significance at the 1%, 5%, and 10% level, respectively.

(a) Panel A: Experian scores

	Experian Score								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Overall	*Private*	*Public*	*New*	*Old*	*Social=0*	*Social=1*	*LowComp.*	*HighComp.*
Treatment*Ahead	2.763**	3.163**	0.417	1.431	0.765	3.892**	1.037	0.476	-0.825
	(0.962)	(1.328)	(0.470)	(1.013)	(0.497)	(1.091)	(0.924)	(1.063)	(0.787)
Observations	4,008	2,548	1,460	840	1,736	512	2,036	1,336	2,532
R-Square	0.425	0.378	0.348	0.356	0.391	0.441	0.346	0.430	0.407
Lender FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

(b) Panel B: Credit2B scores

	Credit2B Score								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Overall	*Private*	*Public*	*New*	*Old*	*Social=0*	*Social=1*	*LowComp.*	*HighComp.*
Treatment*Ahead	2.544***	1.537**	0.676*	1.616	1.280	1.697***	0.757	1.126	-1.246***
	(0.485)	(0.703)	(0.366)	(1.144)	(0.765)	(0.375)	(0.462)	(1.360)	(0.355)
Observations	4,008	2,548	1,460	840	1,736	512	2,036	1,336	2,532
R-Square	0.961	0.958	0.991	0.940	0.966	0.978	0.815	0.930	0.948
Lender FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 9: Effect of Private Information on Sales Orders

This table reports the results of Equation 2, where the dependent variable is the log of the total monthly volume of sales orders. Sales orders are available at the lender-month level (rather than the lender-borrower-month level), so the regression is estimated at this level. *Treatment* is an indicator variable set equal to one if the lender was randomly chosen to receive the slider intervention, zero otherwise. *Ahead* is an indicator variable set equal to one if the month is greater than or equal to January 2019, zero otherwise. Standard errors are clustered at the lender level. *** , **, and * represent significance at the 1%, 5%, and 10% level, respectively.

	Orders
Treatment*Ahead	0.679**
	(0.326)
Observations	1,775
R-Square	0.658
Lender FE	Y
Month FE	Y

11 Appendix

Figure A1. Pre-experiment Survey
This figure displays the original mock-up of the slider that was presented at PAC in October 2018.

Respondent ID:
IP Address:
First Name:
Last Name:
Organization:
Number of Customers:

1. On a scale of 1-10, (where 10 represents perfectly meets my expectations and 1 is fails to meet my expectations) please rate your OVERALL assessment of the scores.

2. If you rated our Scores below 5 in the question above, please give us your reason(s).
    a. Too aggressive
    b. Too conservative
    c. Mix of too aggressive or too conservative
    d. Other (please specify)

3. On a scale of 1-10, (where 10 represents perfectly meets my expectations and 1 is fails to meet my expectations) please rate your OVERALL assessment of the calculated credit limit.

4. If you rated our CCL below 5 in the question above, please give us your reason(s).
    a. Too aggressive
    b. Too conservative
    c. Mix of too aggressive or too conservative
    d. Other (please specify)

5. In addition to our scores and credit lines, what objective or subjective inputs would you like to include to assess your customers? Please provide detailed responses if appropriate.

6. Please feel free to provide any other comments on your experience with our scores and CCLs:

Figure A2. Slider Mock-up
This figure displays the original mock-up of the slider that was presented at PAC in October 2018.

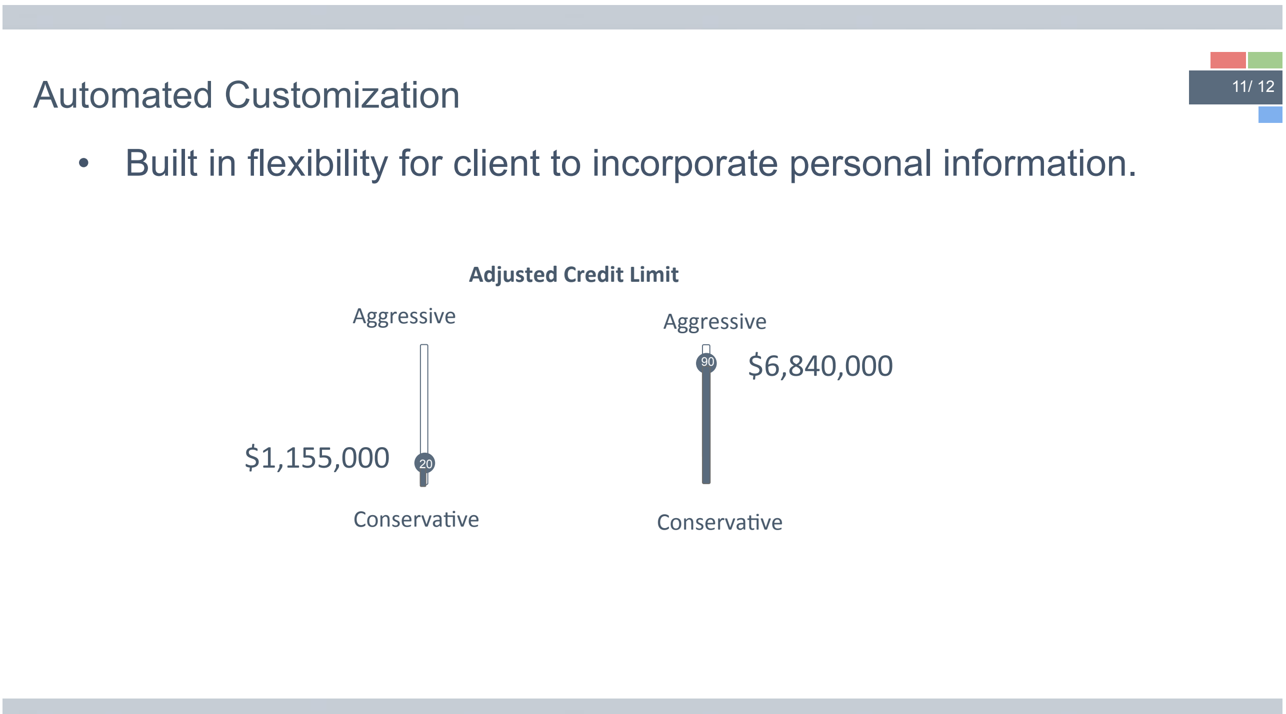


Figure A3. Email to Treatment Lenders

Below is the email that was sent to the treatment group on November 29, informing them of the new slider feature.

credit2b™
CREDIT REINVENTED
A division of billtrust

Dear Justin,

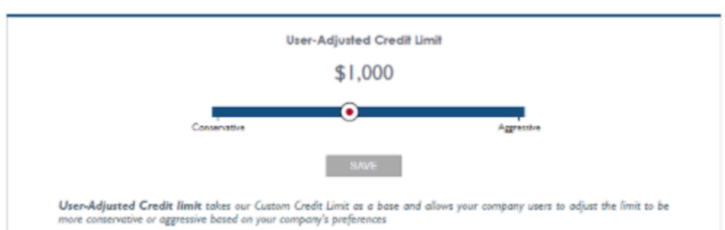
On the weekend of 11/30, Credit2B, a division of Billtrust, will complete a product release featuring enhancements to our Credit Limit Recommendations. It is our mission to continue to empower you with data driven insights so you can make quick and confident credit decisions.

What is new?

Credit Limit Slider within the Virtual Credit File:

With the upcoming release, you will be able to take advantage of the Credit Limit Slider, which gives you the ability to adjust each credit limit recommendation based on your experience with each customer. If your proprietary information about a specific customer suggests that you should take a more aggressive approach, move the slider to the right, and the credit limit will increase. Similarly, moving the slider to the left will result in a credit limit decrease. The feature allows you more flexibility to incorporate your information into the calculated credit limits. You can separately adjust each customer's credit line by going into the score page for each of your clients. Below is a visual example of the new slider feature.

For further guidance on how to use the slider, please sign up for a complementary Webinar HERE. Or feel free to contact your Credit2B sales representative directly.



Team Credit2B, a division of Billtrust

212-279-3300 | info@credit2b.com | http://www.credit2b.com

See what's happening on our social sites

Figure A4. IRB Correspondence

Below is the notification that our experiment is considered Not Regulated by the IRB.

3/16/2019 Activity Details

eRESEARCH | REGULATORY MANAGEMENT

Hello, **Andrea Down**

My Home | Human Subjects Studies | Biosafety / IBC | Repositories | Help

Human Subjects Studies > The Role of Soft Information in the Lending Decision

<< Return to Workspace **< Prev** **4 / 7** **Next >**

Activity Details (Changes Requested by Core Committee Staff)

Changes to the item were requested.

Author:	Li Morrow (OVPR-IRB Behavioral/Health Sci)
Logged For (Application):	The Role of Soft Information in the Lending Decision
Activity Date:	4/5/2018 8:16 AM

Activity Form | Property Changes | Documents | Notifications

The Role of Soft Information in the Lending Decision (HUM00143866)

When you submit this form, the Principal Investigator will be notified that changes are required to the application before it can be approved. If you do not see all of the issues below that need to be addressed, close this window and use the "Edit Open Issues" activity to add other issues that need to be addressed by the Study Team.

* ***Comments:***

Hi Andrea and Anna,

After careful consideration and discussing your study with my assistant director and then my director, it has been decided that this research is Not Regulated because it is organizational research. You do not need to have an IRB application as your study does not require IRB oversight. I am sorry for all of the work I asked of you to do a standard application but this decision required all of the detail you provided. Good luck with your research.

Best,

Li

Documents:

name	**version**
There are no items to display	

Close

<< Return to Workspace

https://errm.umich.edu/ERRM/sd/Rooms/RoomComponents/ProjectActivitiesView/ActivityDetailViewer?Activity=com.webridge.entity.Entity[OID[3B9C2… 1/1

A5. Examples of comments made about borrowers

1. I sent Dave an email asking a bunch of questions and why he hasn't responded or asked to meet with me when in NY recently. He called while I was on the phone and left the following: XXX will be handling the A/P of YYY for "a few more months." He thought the factor meetings went well (despite what we are hearing) and was "suprised that the factors would tell us due to the non-disclosure agreements." BULLSHIT.

   Capital stopped 2-3 months ago. They met with Dave 2 weeks ago and made him feel "uncomfortable!" They don't like the fact that ZZZ hasn't put more money in nor that the inventory is being reevaluated by the lenders that might cause a reduction in the borrowing base. They also saw payments slow down a few months ago.

   I spoke to a CIT client who says Joe gives her approvals on a very short leash and adds "I can't say that we won't pull this sometime soon."

2. Here is the email from John: Hey Rick, I apologize for the lack of communication. Judy and Brian often are swamped with calls and emails about bills, and can't get around to calling everyone back. I took it for granted that they'd at least given you some sort of feedback by now, but I guess not. Rest assured, we will eventually pay you guys, as you were cool enough to give us 45 day terms from the get-go. We're in a constant struggle to meet production/sales demands, versus paying BIG bank notes every month–which chews into payments to anyone. I'm often asked to stall cashing my paycheck until the next week; it's that dire of a situation. So, it's not a matter of "not wanting" to pay anyone, rather a "we don't have the liquid cash" to pay anyone. I've got to go up front and talk to Brian about a couple of things, so I'll for sure bring this matter to his attention. I'll let you know what our plan of action is.

3. Heard rumor that they lost their largest customer.

4. Peter advised that figures will be available soon. Said if we got them last year we will get them this year. Said sales were "very good" but refused all other info.

5. Heard rumor that the entire staff was fired and the business will be liquidating.

6. Mr. Gewirtz's daughters have actually been running the company for years and basically took over the operations.

7. Spoke to Lisa today. They had a big hit to cash flow with the XXX store closing at the headquarter address. Landlord will turn it into an indoor flea market and they have signed on for 12,000 sq ft of apparel space. Should re-open late march or early April. Her brother is looking at a possible new store in upstate NY. Are paying critical vendors first as expected. Will pay everyone. Not going away.