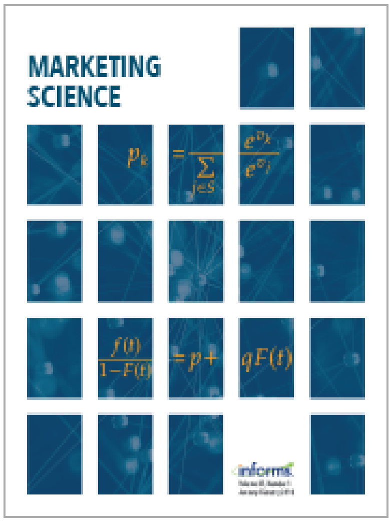


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Value of Verified Employment Data for Consumer Lending: Evidence from Equifax

Tat Chan, Naser Hamdi, Xiang Hui, Zhenling Jiang

MARKETING SCIENCE
Articles in Advance, pp. 1–20
ISSN 0732-2399 (print), ISSN 1526-548X (online)
http://pubsonline.informs.org/journal/mksc

The Value of Verified Employment Data for Consumer Lending: Evidence from Equifax

Tat Chan,[a] Naser Hamdi,[b] Xiang Hui,[a] Zhenling Jiang[c,*]

[a] Washington University, St. Louis, Missouri 63130; [b] Equifax Inc, St. Louis, Missouri 63146; [c] University of Pennsylvania, Philadelphia, Pennsylvania 19104
*Corresponding author
Contact: chan@wustl.edu (TC); hui@wustl.edu (XH); zhenling@wharton.upenn.edu, https://orcid.org/0000-0002-4581-0125 (ZJ)

Received: September 23, 2020
Revised: July 19, 2021; October 2, 2021
Accepted: October 5, 2021
Published Online in Articles in Advance: December 28, 2021

https://doi.org/10.1287/mksc.2021.1335

Abstract. What is the value of *verified* employment data in consumer lending? We study this question using a data set covering all employment verification inquiries to Equifax. Using a difference-in-differences approach, we analyze the changes in applicants' loan outcomes after their employers join Equifax's digital verification system, which provides lenders with an efficient way of accessing the (employer-) verified employment data in auto loan applications. Holding the employment status constant, we find that the availability of digitally verified data significantly expands credit access: the loan origination rate increases by 35.5% on average, and is more significant among deep subprime (146%) and subprime consumers (44%). The interest rates charged on these loans rise only slightly. The expanded credit access also benefits lenders, with an estimated 19.6% increase in profit. This is because the benefit of the market expansion effect dominates the cost of a higher delinquency risk among the expanded loan portfolio. Our results suggest that, besides seeking new data sources, managers and policy makers should also consider ways to extract more value from existing data.

History: Avi Goldfarb served as the senior editor and served as associate editor for this article.
Supplemental Material: The online appendix and data are available at https://doi.org/10.1287/mksc.2021.1335.

Keywords: value of verified data • digital verification • inclusive access to credit

1. Introduction

An important foundation of modern-day marketing activities is the increasing availability of consumer data. These data are valuable to firms and consumers in many settings, such as offline retailing (e.g., Rossi et al. 1996), online platforms (e.g., Lambrecht and Tucker 2013, Jiang et al. 2021), and business-to-business markets (e.g., Dubé and Misra 2017). Consumer data are also valuable in the auto loan market, which is our empirical setting, as lenders heavily rely on credit scores and other types of data to make lending decisions. Adopting credit scores in consumer lending has been shown to reduce information asymmetry and improve loan outcomes (e.g., De Janvry et al. 2010, Einav et al. 2013). Despite the benefits, relying only on credit scores has limitations, because credit scoring is based on historical data and may not accurately reflect an applicant's current risk profile (Li and Ching 2019). Furthermore, as many as 45 million American households have no credit score, and half of the remaining 190 million consumers are subprime.[1] These consumers may be excluded from the credit market if lenders rely exclusively on credit scores when issuing loans.

Both policy makers and lenders are interested in additional data to complement credit scores. To enable greater access to credit for consumers with low credit scores, the Credit Access and Inclusion Act was proposed to leverage applicants' on-time payment data from utility companies in loan decisions.[2] Banks and FinTech lenders leverage alternative data, such as applicants' digital footprint, for risk profiling. One potential difficulty of this approach is that wide usage of new data sources may not be a feasible option for policy makers and lenders.

In this paper, we study the value of *digitally verifiable* employment and income data in the market for auto loans to consumers and lenders. When consumers apply for an auto loan, lenders commonly request employment and income information. These self-reported data, however, are subject to mis-reporting and hard to verify. The digital verification service provides an efficient way of accessing (employer-) verified employment information. We study the impact of having digitally verifiable employment information when applying for an auto loan. In particular, we focus on subprime consumers (credit scores between 500 and 600) and deep subprime consumers (credit scores less than 500), who typically have difficulty obtaining a loan. The focus on digitally verifiable information differentiates our paper from prior work, which often utilizes data that reflects new dimensions of applicants' behavior. Our key insight is that if we make the verification and transmission of information less costly, large value can be extracted from "old-fashioned" data. Additionally, understanding the impact of such digitally verified employment data also has important social and policy implications. This is because enabling access to credit in the auto loan market is crucial for economically disadvantaged customers, as it directly impacts their mobility and the ability to reach their workplace, get groceries, and benefit from childcare.[3] Moreover, the auto loan market has high economic significance: Americans held $1.36 trillion in auto loan debt in 2020, equivalent to nearly 10% of all household debt.[4]

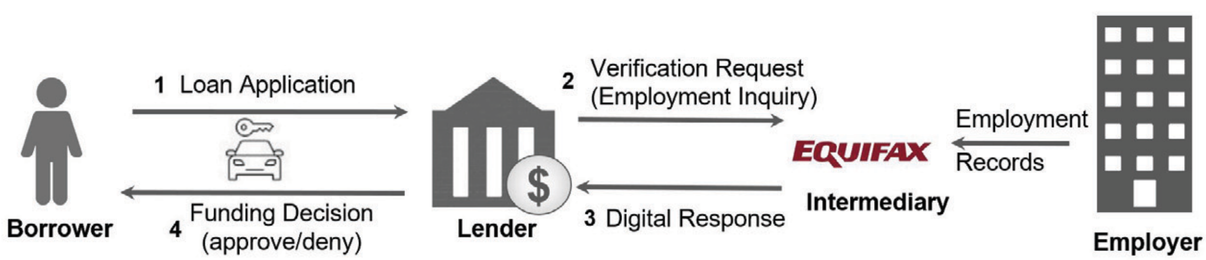
In our study, lenders' access to verified employment information is achieved by a digital employment verification system created by Equifax. Besides being one of the three major credit bureaus in the United States, Equifax also manages the largest employment and income database in the country and offers financial institutions a digital verification service. The structure of the employment verification system is as follows (also shown in Figure 1). Employers outsource the handling of employment-related inquiries to Equifax, and thereby reduce their HR-related costs. To do so, employers join the digital verification system by providing employment data to Equifax.[5] Lenders can submit an employment inquiry to Equifax digitally and instantaneously obtain the employment information of loan applicants if their records belong to the system. Based on the verification results, lenders decide whether to approve the loan origination and the loan terms. Compared with manual verification, this digital system lowers the information cost for both employers and lenders.

We use a large-scale data set that covers all employment inquiries to Equifax for auto loan applications in 2016 and 2017. The data set covers whether the employment records of the inquired consumers are in the database or not and, if so, their employment and income information. In addition, we observe whether the loan applicant obtained an auto loan after being inquired and, if so, the loan characteristics and loan repayments. The data set also includes some consumer characteristics including consumers' credit scores and geographic location.

To estimate the causal effects of access to verified employment data, our identification strategy exploits instances where employers joined Equifax's verification system (henceforth the system) *during* our sample period. Before employers join the system, lenders cannot retrieve any information on the employees from the system, and therefore have to either proceed without the information or use manual verification. The difference in loan outcomes for similar individuals who applied for a loan before and after their employers joined the system reflects the effect of digital employment verification. We exploit the different timing of employers' joining the system via a difference-in-differences (DiD) approach, which compares the outcomes centered at the time when an employer joins the system. The loan outcomes with the same timing for employers who have not yet joined the system and those who have already joined it act as the baseline control group. Our data and analysis are conditional on lenders making inquiries and employers joining the system, so our results should be viewed as the average treatment effect on the treated.

Our results show a very significant increase in the auto loan origination rate among consumers after their employers joined the system. Overall, the probability of obtaining an auto loan increases by 35.5% (from a loan origination rate of 35.5% to 48.1%). The most significant increase comes from consumers with low credit scores: the loan origination probability increases by 146.4% (from a loan origination rate of 12.4% to 30.4%) for deep subprime consumers (i.e., with credit scores lower than 500), and 43.5% (from a

Figure 1. (Color online) Illustration of Digital Employment Verification in Loan Application



loan origination rate of 31.3% to 44.9%) for subprime consumers (i.e., with credit scores between 500 and 600). These findings suggest that many consumers who are employed but have a poor credit history are excluded from the loan market but can significantly benefit from having their employment information made digitally verifiable when they apply for credit. Given that 69% of the loan inquiries in the data come from deep subprime and subprime consumers, allowing access to verified employment data has a major economic and societal consequence for the auto loan industry. Besides loan origination, we further test whether the interest rate significantly increases among the expanded loan portfolio after lenders get digital access to verified employment data. We find that the increase in loan origination does not come with a much higher interest rate for consumers across all credit segments, which ranges from 2%–4%. This suggests that the employment information allows lenders to adjust their risk evaluation of low-credit-score consumers with current employment and therefore are able to offer more loans to these consumers without increasing the interest rates too much. These results suggest that consumer data can increase welfare for some consumer groups while not hurting others, consistent with some of the previous literature (e.g., Sun et al. 2020, Jiang et al. 2021).

For lenders, our findings are consistent with those from the extensive empirical marketing literature that identified the benefits of more consumer data for firms (see a list of these papers in Section 1.1). We find that, although the delinquency rate is significantly higher among the expanded loan portfolio, the lenders' overall profit is estimated to increase by 19.6%.[6] This is because the benefit from the increase in loan origination outweighs the cost from a higher loan delinquency rate. In particular, the profit from deep subprime and subprime consumers grows by 77.2% and 26.8%, respectively, consistent with the large increase in loan origination in these consumer segments. These results suggest that the market expansion enabled by the availability of digitally verified employment information benefits not only consumers but also lenders in the auto loan industry.

Our identification strategy relies on comparing the outcomes of loan applications from consumers after their employers join the system to those before joining (treatment group) while using the loan applications from other employers (control group) to account for time trends in loan origination. The identification strategy relies on two main assumptions. The first assumption is that the treatment and control groups have common time trends. Although this assumption is not directly testable, we show evidence consistent with it. First, the pretreatment time trends of loan origination are parallel for the treatment and control groups. We also perform a leads-and-lags analysis and find similar results. The second assumption is that the types of loan applicants as well as lenders' inquiry behavior do not change after the treatment. If different types of employees choose to apply for loans after their employment records become digitally verifiable, or lenders change their propensity of making an inquiry to Equifax, our results may be biased due to the systematic difference between inquiries before and after the treatment. In support of the second assumption, we find that the types of loan applicants, in terms of observables (average credit score and income), do not change after the treatment. Moreover, the percentage of employees with loan inquiries also stays the same after the treatment. This evidence suggests there is little change in the types of loan applicants or lenders' inquiry propensity. Finally, we use a matching approach to explicitly control for possible differences in the type of inquiries (due to changes in the types of loan applicants) by matching inquiries from loan applicants who have similar credit scores and income, and work for the same employer. The matching analysis shows very similar results to those of the main analysis, implying that the effects are not driven by the differences in inquiries before and after employers join the system.

Our results have important managerial and policy implications. From a managerial viewpoint, our results show that digital access to verified employment information enhances lenders' information and therefore increases their profit. From a policy viewpoint, our results suggest that while seeking novel data sources for risk profiling can be fruitful, we can benefit from ways of extracting more value from existing data. In our setting, the value extraction is achieved through a digital infrastructure for verifying and transmitting data that have been available but used to be costly to access and verify. More generally, the verification can be achieved in decentralized ways (e.g., through blockchain), and there could be other ways of more efficient use of data, such as from improved analytics capacity (e.g., Brynjolfsson and McElheran 2019, Berman and Israeli 2021, Du et al. 2021). Our results suggest that more efficient use of data can disproportionately help disadvantaged consumers, such as consumers with subprime credit scores, access the credit market.

1.1. Literature Review

Our paper is related to three strands of literature. First, our paper relates to a large empirical literature in marketing that studies the value of data. Better information enables firms to tailor their decision variables to different consumers or consumer segments. One key area of customized decisions is targeted pricing, which has been shown to increase overall market efficiency in various settings, including offline retailing (Bult and Wansbeek 1995, Rossi et al. 1996, Besanko et al. 2003, Simester et al. 2006, Pancras and Sudhir 2007), online platforms for goods and services (Ansari and Mela 2003, Zhang and Krishnamurthi 2004, Khan et al. 2009, Zhang and Wedel 2009, Lambrecht and Tucker 2013, Ghose et al. 2014, Sahni et al. 2018, Rafieian and Yoganarasimhan 2020, Jiang et al. 2021), business-to-business relationships (Zhang et al. 2014, Dubé and Misra 2017), and market for digital content (Shiller and Waldfogel 2011). Consistent with the qualitative findings in this literature, we find that giving lenders access to verified employment data increases the total number of loans and lenders' profit, and it has a disproportionately larger benefit for consumers with low credit scores. We expand the prior literature in evaluating the value of data in the auto loan industry, which is economically significant. More importantly, digitally verified employment data promote inclusiveness in this market, which has significant implications for disadvantaged customers in terms of their mobility and the ability to reach their workplace, get groceries, and use childcare.

Second, our paper contributes to the literature that studies the trade-off between the benefit and privacy cost of using personal data. Because consumers' privacy concerns have increased over time (Goldfarb and Tucker 2012) and many marketing decisions are powered by consumer data, understanding ways of ameliorating the aforementioned trade-off is important for firms' long-term profitability and consumer welfare. In the advertising space, a plethora of evidence shows that ads' effectiveness critically depends on consumers' belief about whether their privacy is infringed and whether they have control over their own data (Goldfarb and Tucker 2011a; Tucker 2012, 2014). A few papers estimate consumers' value for privacy using experiments (Athey et al. 2017, Lin 2019, Tang 2019). Consumers' value for privacy can also be short-sighted and highly context-specific (Prince and Wallsten 2020), and therefore a one-size-fits-all policy will unlikely protect consumers adequately and can create unintended consequences (Acquisti et al. 2016). Specifically, data regulations such as the GDPR can have unintended consequences (Goldfarb and Tucker 2011b, Campbell et al. 2015, Johnson et al. 2019) and can reduce matching efficiency (Sun et al. 2020). Although we do not measure privacy cost in our paper, consumers may find a more efficient use of existing personal data less intrusive. If this is the case, our results suggest that extracting more value from existing data (as opposed to asking consumers to share new data) may bring out the benefit of data sharing without incurring too much privacy cost.

Lastly, our paper contributes to the household finance literature that analyzes alternative data sources that complement credit scores in consumer lending. In intermediated lending institutions, relationship-specific information can be highly predictive of the borrowers' risk (Norden and Weber 2010). The recent rise of peer-to-peer (P2P) lending platforms highlights the benefits of using alternative data, such as lending information from peers (Zhang and Liu 2012) and from professional investors (Catalini and Hui 2018), social network information (Lin et al. 2013), text description in loan applications (Netzer et al. 2019), and consumers' digital footprint on the internet (Berg et al. 2018). All these alternative data sources can help predict the borrowers' credit-worthiness, but as some authors warn, the use of soft information could also induce borrowers' strategic behavior to bias lenders' decisions. In comparison, our paper suggests that making more efficient use of existing data can increase loan outcomes while circumventing the above-mentioned strategic behavior because employment data verified by employers is hard to manipulate.

2. Background and Data

Although lenders typically ask for employment information when consumers apply for secured loans, these data may not have achieved its fullest potential because of high verification costs for both lenders and employers. Although lenders need to verify the self-reported information with applicants' employers, employers may not always have the dedicated HR resources to facilitate the verification.[7] Equifax encourages employers to sign up for the database to enjoy several key benefits. First, by joining the database, employers can reduce the amount of HR resources needed to fulfill employment verification requests.[8] Second, outsourcing employment verification ensures compliance and data security.[9] The HR team does not need to face the risk and potential liability of releasing confidential employee information to parties who should not get it. Furthermore, there is no charge for employers to join the system. Instead, Equifax monetizes the database by charging verifiers (e.g., auto lenders) who obtain employment information from the database. Specifically, verifiers pay a fee to Equifax only if an inquiry is fulfilled.

Our empirical analysis leverages anonymized data on individual employment and credit databases from Equifax Inc. Just as lenders can make an inquiry to Equifax to find out an applicant's credit score, they can also obtain instantaneous access to the applicant's employment records through the employment verification service. The employment database covers *a subset of employers* who chose to report the information to Equifax. With the selected coverage, one of two scenarios can happen when a lender submits an employment inquiry: Equifax's database returns the loan applicant's verified employment information and the inquiry is considered "fulfilled," or no such information is returned and the inquiry is considered "unfulfilled." The fulfillment status depends on whether the loan applicant is covered by Equifax's employment database. If the inquiry is unfulfilled, the loan applicant can be either unemployed or employed by an employer who is not in the system. For fulfilled inquiries, lenders see the employment information for the consumer, including the name and location of the employer and the consumer's income, job title, and duration of employment tenure.

An important feature of our data is that we can back out consumers' employment information during the period *before* their employers joined the system. This is because Equifax asks employers upon their joining the system to provide the start day of each employee. This feature allows us to map inquiries to employers even for inquiries submitted before an employer joined the system. We elaborate on this point in Section 3 when discussing the identification strategy.

Our starting point is the full sample of 12 million auto loan inquiries submitted to Equifax during a two-year period from 2016 to 2017. We know whether each of these inquiries is fulfilled or not. We observe the employment information for all inquired consumers if their employers are in the employment database. We match the consumers in the employment database with their credit profile, based on which we identify whether a consumer had an auto loan originated after the inquiry. In addition, the credit profile provides other consumer characteristics at the time of inquiry, including credit score and geographic location. For all the originated loans (from consumers with and without fulfilled inquiries), we observe the loan characteristics, such as loan amount, length, and interest rate. Furthermore, we also track the repayment behaviors for each loan (i.e., whether the monthly payment is paid to the lender).

We first get a sense of how the inquired loan applicants in our full sample compare with all customers who applied and originated an auto loan in 2016– 2017. Figure 2 plots the distribution of credit scores across the two populations. The sample with employment inquiries is heavily skewed toward lower-credit-score consumers, because for higher-credit-score consumers, lenders typically do not need to verify employment status to originate loans. Conversely, for lower-credit-score consumers, lenders are much more likely to verify employment status, and not all inquiries would lead to an originated auto loan.

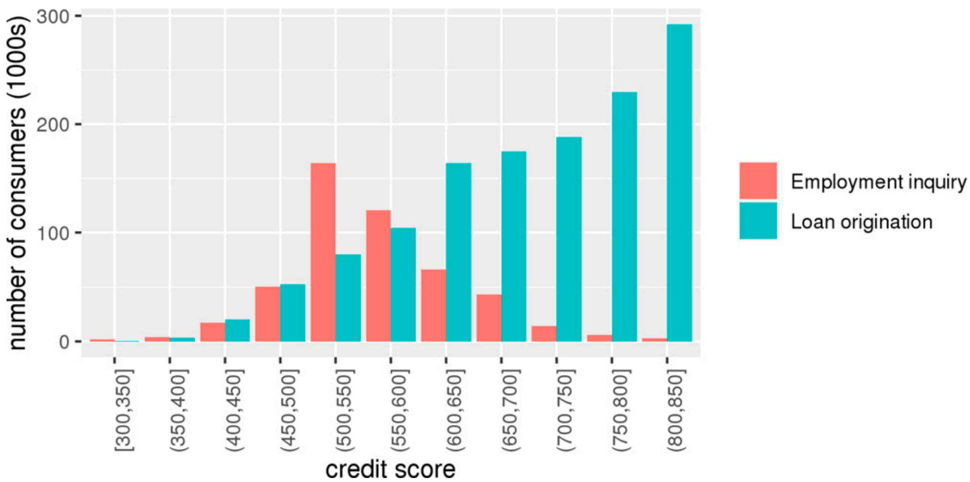
Table 1, Panel A reports summary statistics for the full sample. We start by presenting inquiry-level statistics in Panel A1: The average credit score of the loan applicants is 568, and their average annual income is 38,387. The loan origination rate for fulfilled inquiries is 52.3%, much higher than the 33.7% rate for unfulfilled inquiries.

To visualize the correlation between the changes in the loan origination rate and an inquiry's fulfillment status, in Figure 3 we plot the loan origination rate by each credit score, grouped by whether inquiries are fulfilled (represented by the dots) using the full sample. The solid line fits a smooth curve across the origination rate for each credit score. We see three patterns. First, conditional on fulfillment status, the loan origination rate is higher for borrowers with a higher credit score. Specifically, the loan origination rate is 10%–15% for the loan applicants with a credit score of 500 or less and unfulfilled employment inquiries, indicating that these consumers are excluded from the loan market. However, many of these individuals could be employed at the time of the inquiry, but their employment status could not be verified because their employers had not joined the system. Second, the loan origination rate is consistently lower for unfulfilled inquiries. Finally, the slope of the loan origination curve is steeper among unfulfilled inquiries. As a result, the difference in the loan origination rate between fulfilled and unfulfilled inquiries for high-credit-score consumers is much smaller than that for low-credit-score consumers, indicating that the employment information is more important for the latter consumers in obtaining a loan.

The positive correlation between the loan origination rate and an inquiry's fulfillment status is a key motivation for our study. However, this correlation may not be causal because it essentially compares the outcomes of inquiries of loan applicants who are covered in the system and those who are not. The two types of applicants can be different in terms of their employment status or the types of firms that they work for, which can contribute to the difference in loan outcomes.

To mitigate this concern, we construct our estimation sample, which uses 495 employers who joined the system at some point during the data observation period. We refer to this sample as the main sample in the rest of the paper. The benefit of using these employers is that we observe inquiries of their employees both before and after the employer join date, which creates temporal variation in whether the records are available in the system while holding fixed the actual employment status. Importantly, we are able to estimate *within-employer* changes across time.[10] About 176,000 inquiries are associated with employees from these 495 employers. For these inquiries, all consumers are employed at the time of inquiry: we know they joined the employer prior to the inquiry based on their start date. We do not include the inquiries where we do not observe the actual employment information

Figure 2. (Color online) Distribution of Credit Scores: Inquired Consumers vs. Overall Consumers



Note. "Employment inquiry" corresponds to the sample of consumers for whom lenders inquired about employment status; "Loan origination" corresponds to all customers who applied and originated an auto loan in 2016–2017.

(e.g., inquiry happened prior to their start date). Panel B1 of Table 1, which reports the inquiry-level statistics of our main sample, shows that the magnitudes are mostly similar to those of the full sample. In particular, the loan origination rate increases from 35.4% to 51.5% when inquiries are fulfilled.

Having discussed inquiry-level characteristics, we now report summary statistics for the inquiries that led to a loan. Panel A2 shows that conditional on loan origination, the average credit score of the borrower is 589 and the borrower's average income is $40,096, both of which are unsurprisingly larger than their counterparts in Panel A1. The average loan amount is $21,231, which is similar to the national average loan size for used cars. The loan length and interest rate given to the consumers are 5.65 years and 13.8%, respectively. For loan repayment behaviors, we follow the industry practice and use the 90-day delinquency rate, which captures whether a loan was 90 days past due (90 DPD), and this variable is 21.9% on average. To represent the potential loss for banks in dollar terms, we use the past due amount as another measure for delinquency. The average past due amount one year after loan origination is $2,162. In Panel B2, we report summary statistics for the loans in our main sample. The average income is about 10% higher than that in the full sample, and the consumers have a better repayment behavior than those in the full sample. Apart from these differences, the average credit score of the consumers in our main sample and the loan terms they receive are similar to those in the full sample.

Lastly, we report summary statistics at the employer level in Panels A3 and B3. The loan applicants with a fulfilled inquiry are associated with a total of 2,638 employers. For each employer, we calculate the number of employees who had an auto inquiry during the two-year period and their average income, average percentage of hourly employees, and average job tenure. Regarding the employer-level observations, the averages are taken first within each employer and then across employers. On average, the number of employees who had auto loan inquiries is 374 across different employers, and the employers in the bottom and top quartiles had 7 and 190 employees inquired, respectively. The distribution is highly skewed because there are several very large employers with more than 20,000 employees inquired during this two-year period. In addition, the average income is $51,095, the average percentage of hourly employees (employees paid by the hour) is 33.6%, and the average job tenure (number of months worked for the employer) is 61.7 months. We observe a large variation in these measures across employers. This observation suggests that there could be a large employer heterogeneity, and a convincing empirical strategy should take this into account. Next, comparing across the two panels, we see that the average number of employees inquired is much larger on average and more skewed for the early adopters in panel A3 than for the new adopters in panel B3. This pattern could be driven by a higher need to outsource HR-related inquiries for employers with a larger number of employees, who are more likely to join the system earlier.

3. Empirical Strategy

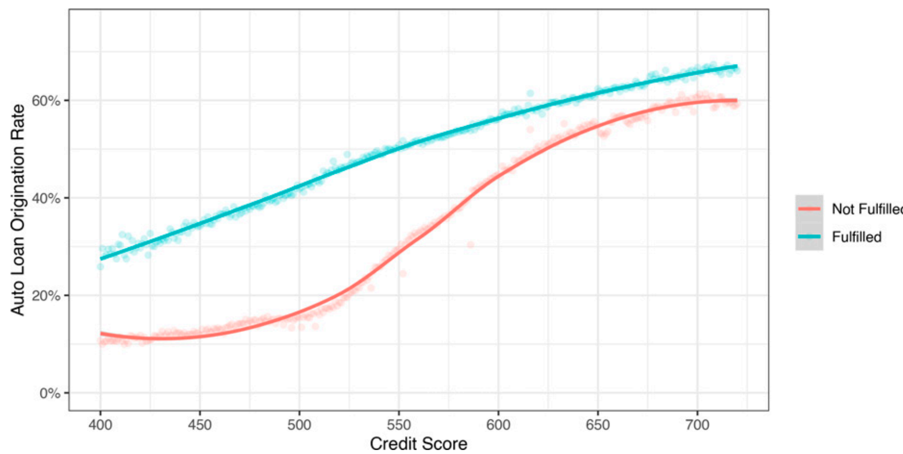
Our main empirical strategy uses the 495 employers who joined the system at some point during our sample period. To estimate the effect of having digital access to employment and income data on loan

Table 1. Summary Statistics

	Mean	25th Percentile	Median	75th Percentile
Panel A. Full sample				
A1. Inquiry-level statistics				
Credit score	568	522	568	616
Income, $	38,387	22,880	29,681	41,672
Auto loan origination				
(fulfilled inquiries)	52.3%	0	1	1
(unfulfilled inquiries)	33.7%	0	0	1
A2. Loan-level statistics				
Credit score	589	548	591	634
Income, $	40,096	23,660	31,200	44,179
Loan amount, $	21,231	14,206	19,371	26,328
Loan length, year	5.65	5.50	6.00	6.08
Interest rate	13.8%	8.5%	13.2%	18.7%
90-day delinquency	21.9%	0	0	0
Past due amount, $	2,162	0	0	78
A3. Employer-level statistics				
No. employees inquired	374	7	39	190
Income, $	51,095	30,742	51,285	53,826
Hourly employee, %	33.6	0	24.3	63.2
Tenure, months	61.7	37.2	57.4	81.1
Panel B. Main sample (employers who joined during the sample period)				
	Mean	25th Percentile	Median	75th Percentile
B1. Inquiry-level statistics				
Credit score	569	523	570	616
Income, $	41,191	23,982	33,194	49,421
Auto loan origination				
(fulfilled inquiries)	51.5%	0	1	1
(unfulfilled inquiries)	35.4%	0	0	1
B2. Loan-level statistics				
Credit score	588	548	589	631
Income, $	44,534	26,270	36,675	53,649
Loan amount, $	21,383	14,393	19,571	26,606
Loan length, year	5.66	5.50	6.00	6.08
Interest rate	13.8%	8.7%	13.2%	18.5%
90-day delinquency	19.8%	0	0	0
Past due amount	1,838	0	0	0
B3. Employer-level statistics				
No. employees inquired	172	23	78	192
Income, $	49,586	30,621	41,089	53,508
Hourly employee, %	31.4	0	19.2	58.7
Tenure, months	62.5	35.7	60.2	85.5

outcomes, we exploit instances in which employers joined the verification system during our sample period. Figure 4 plots the number of employers that joined the system each month in 2016 and 2017.[11] This staggered enrollment enables our difference-in-difference (DiD) analysis. As shown in Goodman-Bacon (2018), the DiD estimate with variation in the treatment time gives us a weighted average of (1) comparisons between inquiries of consumers who work for early adopters (i.e., employers) over the periods when the later adopters have not yet joined the system, in which inquiries of consumers from later adopters are used as the control group for those from early adopters, and (2) comparisons between inquiries of consumers from early adopters and those from later adopters over the periods when the early adopters have joined the system, in which inquiries of consumers from early adopters are used as the control group for inquiries of consumers from later adopters.[12] Our analysis compares the outcomes of inquiries before and after the employers joined the system with the outcomes of inquiries of those employers who had joined the system earlier and those who have not yet joined it, in the control group. Lenders cannot retrieve any information from the verification system before employers join the system, and therefore need to proceed either without the employment information or with manual verification. The estimated results represent the average treatment effect on the treated, because our analysis focuses only on employers who joined the verification system during the sample period. The results may differ in magnitude for those who never join the system (e.g., very small businesses).

Figure 3. (Color online) Difference in Loan Origination Rate Across Fulfillment Status



Notes. The average percentage difference in the loan origination rate across fulfillment status is computed at each credit score. The solid line fits a third-order polynomial across the points.

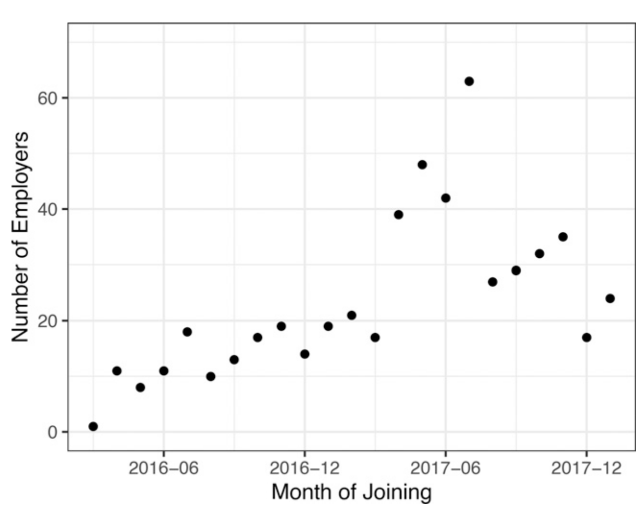
The regression equation is given as follows:

$$Y_{ijt} = \beta Digital_{jt} + \gamma X_i + \eta_j + \xi_t + \epsilon_{ijt}. \tag{1}$$

The level of observation in the regression is an inquiry. Specifically, Y_{ijt} is an outcome variable, such as the loan origination status associated with an employment inquiry i, whose underlying loan applicant works for employer j at time t; $Digital_{jt}$ equals 1 if the employer j had joined the digital verification system by time t; and X_i is a list of characteristics of inquiry i's underlying loan application, such as the applicant's credit score, to control for the observed heterogeneity among loan applications. In addition, η_j represents the employer fixed effect, which captures the potential unobserved heterogeneity across inquiries that are attributable to an applicant's workplace; ξ_t is the year-month fixed effect, which captures time trends in loan outcomes; and ϵ_{ijt} is an idiosyncratic error. The coefficient β represents the effect of having a fulfilled inquiry across all loan applications. In all the analyses, we cluster standard errors at employer level to allow for arbitrary within-employer correlation in the error term.

The DiD methodology allows us to control for two types of unobservables: (1) time-invariant employer-specific loan outcomes (e.g., government employees may have higher loan approval rates than other employers) and (2) time-specific loan outcomes that are the same across employers (e.g., seasonality in auto loan origination rates). This identification strategy relies on two assumptions. The first is the parallel trends assumption: employers who joined the system and those who did not should have parallel trends in loan origination rates without any treatment. In particular, we need the exact timing of treatment to be random with respect to potential loan originations. Theoretically, we do not think this is a concern given that the

Figure 4. Number of Employers that Joined the System Each Month in the Main Sample



main reason that firms join the Equifax employment database is to save HR-related costs. Therefore, the decision and timing of joining the database is unlikely endogenous to employees' loan application decisions. We provide evidence consistent of parallel trends assumption in Section 5.1. The second is that selection into having an inquiry is not correlated with treatment. That is, the unobserved types of borrowers (e.g. higher or lower quality) is not affected by the treatment. Failure of this assumption would constitute collider bias. Although this assumption is fundamentally untestable, we will show that the types of loan applicants and the propensity of lenders to make an inquiry are similar before and after treatment in Section 5.

4. Results

Using the empirical strategy discussed in Section 3, in this section we report the results showing how having access to verified employment data impacts the auto loan outcomes for both borrowers and lenders. We start by presenting a significant increase in the auto loan origination rate with fulfilled inquiries, which happen after employers join the system. This increase is disproportionately larger for subprime borrowers. Furthermore, although the average interest rate increases across consumer segments, these increases are economically small. These results suggest that the digital employment verification enables access to the credit market for low-credit-score consumers. Next, we study the impact of having access to verified employment data on lenders' profit. We find that despite the average delinquency rate being higher among the expanded loan portfolio, lenders' gross profit (without accounting for service fees) increases significantly. This is because the market expansion effect dominates the decrease in profit per loan due to the increase in the delinquency rate.

4.1. Impact on Borrowers: Loan Origination and Interest Rate

We run a linear probability model as shown in Equation 1, where Y_{ijt} is an indicator variable that equals 1 if there is an auto loan originated after the inquiry (in the same month or the month after inquiry). The unit of observation is an inquiry. We use a linear probability model instead of a logit model because we estimate a large number of fixed effects in the model. The main parameter of interest is β, which measures the impact of having a fulfilled inquiry on the likelihood of loan origination.

The results are shown in Table 2, column 1. We find that the probability of having an auto loan originated significantly increases by 12.6% with a fulfilled inquiry, that is, after employers join the database. To put the estimate in perspective, the baseline loan origination rate without a fulfilled inquiry is 35.5%. Therefore, having a fulfilled inquiry increases the auto loan origination rate by 35.5% from the baseline. We have also performed robustness checks by controlling for credit score fixed effects, instead of a linear term of the credit score, and the results are very similar.

Table 2. Loan Origination Rate

Dependent variable: loan origination (0/1)	(1)	(2)
Digital	0.12551***	
	(0.00720)	
Digital: Deep subprime		0.18151***
		(0.00759)
Digital: Subprime		0.13573***
		(0.00672)
Digital: Near prime		0.08579***
		(0.00813)
Digital: Semiprime		0.04219***
		(0.01002)
Credit score	0.14575***	0.17658***
(in 100)	(0.00325)	(0.00432)
State FE	✓	✓
Year-month FE	✓	✓
Employer FE	✓	✓
Observations	175,952	175,952
R^2	0.13497	0.13669

Notes. Regressions based on the main sample. One observation is an inquiry. The credit score cutoffs for deep subprime, subprime, near prime, and semiprime consumers are 500, 600, 660, and 720, respectively. Standard errors clustered at the employer level.
***Indicates significance at $p = 0.01$; **$p = 0.05$; *$p = 0.1$.

Next, we investigate how the increase in the auto loan origination rate varies across consumers with different credit scores. We run a similar regression with the variable $Digital_{jt}$ interacting with four credit score segments. The credit score cutoffs for deep subprime, subprime, near prime, and semiprime consumers are 500, 600, 660, and 720, respectively. The results are shown in column 2. Consistent with the data pattern shown in Figure 3, the increase in the auto loan origination rate is significantly higher for consumers with lower credit scores than for those with higher credit scores. We compare the increase with the benchmark origination rate without fulfilled inquiries for each group to calculate the percentage increase. Having a fulfilled inquiry increases the auto loan origination rate by 146.4% for deep subprime consumers (an increase of 18% from the baseline of 12.4%), which suggests that the use of employment information dramatically helps deep subprime borrowers, who typically face the most difficulty in obtaining a loan. There is also a substantial 43.5% increase for subprime consumers (an increase of 13.6% from the baseline of 31.3%). The increase is much smaller for near prime consumers (14.9%) and for semiprime consumers (6.4%). The much larger magnitude of increase in the auto loan origination rate for subprime consumers is likely due to the employment information being more valuable for these consumers in obtaining a loan. This is consistent with the observation in Figure 2 that lenders are much more likely to verify the employment information of subprime consumers than of those with high credit scores.

Besides loan origination, we also check whether the interest rate is different among the expanded loan portfolio. It is possible that the inclusion of risky consumers who would not get a loan without the employment verification can lead to an increase in interest rates. We compare the interest rates for loans with fulfilled and unfulfilled inquiries, conditional on loan origination. We estimate Equation 1, where the dependent variable represents the interest rate for the loan, and X_i is a list of control variables that may influence the loan interest rate, including credit score, loan amount, loan length, and state fixed effects. Similarly, η_j and ξ_t represent the employer fixed effects and year–month fixed effects, respectively.

The results are shown in Table 3, column 1. Among all consumers, the average interest rate is 0.42% higher among loans with fulfilled inquiries. This represents a 3.1% increase from the baseline average interest rate for loans without fulfilled inquiries (13.4%). Compared with the increase in the loan origination rate, the magnitude of the interest rate increase is much smaller. Column 2 then compares the increase in interest rates among borrowers in different credit score segments. The increase in interest rate from deep subprime borrowers is the smallest among all four segments. For the other segments, the relative increases are 2.8% (a 0.43% increase from the baseline of 15.42%) for subprime, 4.1% (a 0.46% increase from the baseline of 11.23%) for near prime, and 3.0% (a 0.26% increase from the baseline of 8.80%) for semiprime consumers.

Table 3. Interest Rate

Dependent variable: interest rate	(1)	(2)
Digital	0.00417***	
	(0.00075)	
Digital: Deep subprime		0.00389**
		(0.00173)
Digital: Subprime		0.00428***
		(0.00091)
Digital: Near prime		0.00458***
		(0.00093)
Digital: Semiprime		0.00265**
		(0.00119)
Loan amount, \$1K	−0.00199***	−0.00198***
	(0.00005)	(0.00005)
Loan length, year	−0.00517***	−0.00546***
	(0.00050)	(0.00050)
Credit score	−0.04037***	−0.03861***
(in 100)	(0.00043)	(0.00087)
State FE	✓	✓
Year-month FE	✓	✓
Employer FE	✓	✓
Observations	75,458	75,458
R^2	0.38758	0.39143

Notes. Regressions based on the main sample conditional on loan origination. Each observation is an inquiry. The credit score cutoffs for deep subprime, subprime, near prime, and semiprime consumers are 500, 600, 660, and 720, respectively. Standard errors clustered at the employer level.

***Indicates significance at $p = 0.01$; ** $p = 0.05$; * $p = 0.1$.

4.2. Impact on Lenders: Past Due Amount and Profit

A large literature in marketing has shown the value of consumer data for firms. We investigate how digital access to verified employment data affects lenders' profitability in the auto loan industry. To do so, we first compare the repayment behavior for loans with fulfilled and unfulfilled inquiries, conditional on loan origination. Recall that the loans in the main sample were originated either in 2016 or 2017. We measure the repayment behaviors by April 2020. Specifically, we measure the past due amount, which captures the loss for lenders, of the loans in 2016–2017 until April 2020.

We run Equation 1 using the past due amount as the dependent variable on the main sample conditional on loan origination. The results are shown in Table 4. Column 1 shows that the loans with fulfilled inquiries have a \$273 higher past due amount by April 2020. In Column 2, we further control for loan characteristics including the loan amount, loan length, and interest rate, which help control for the observed heterogeneity across loans. The estimate decreases to \$227. Next, we study how the increase in the past due amount differs by credit score segments. Columns 3 and 4 show that the past due amount increases for all credit score segments. The increases are largest for deep subprime and subprime consumers, which is intuitive because lenders are extending more loans to these riskier consumers. For near-prime and semiprime consumers, the increases in the past due amount are also significantly positive. As a robustness check, we study the delinquency behavior by using the delinquency rate as the target outcome instead of the past due amount and find similar results. The estimates are reported in Online Appendix A.

The lenders' profit is impacted by digital verification in three ways. First, there is a higher probability of loan origination from the market expansion effect. Second, the higher interest rate leads to a higher monthly payment for any given loan amount. Third, the lenders' profit is negatively impacted by an increase in the past due amount. To quantify the impact of digital verification on the lenders' profit, we construct a gross profit variable (without taking account

Table 4. Past Due Amount

Dependent variable: past due amount	(1)	(2)	(3)	(4)
Digital	273.25***	227.04***		
	(61.92)	(59.24)		
Digital			276.00**	397.60***
Deep subprime			(130.58)	(127.45)
Digital			338.41***	244.97***
Subprime			(71.40)	(69.11)
Digital			193.25***	142.51**
Near prime			(74.65)	(72.48)
Digital			217.29**	270.34***
Semiprime			(93.47)	(90.18))
Credit score	−871.69***	−369.90***	−823.07***	−323.57***
(100)	(33.40)	(36.35)	(54.05)	(58.10)
Loan amount		64.44***		64.61***
($1K)		(3.78)		(3.79)
Loan length		111.23***		112.69***
(Year)		(23.47)		(23.69)
Interest rate		15,427.99***		15,455.06***
		(525.71)		(521.96)
State FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
Employer FE	✓	✓	✓	✓
Observations	75,458	75,458	75,458	75,458
R^2	0.04586	0.07306	0.04594	0.07315

Notes. Regressions based on the main sample conditional on loan origination. One observation is an inquiry. The credit score cutoffs for deep subprime, subprime, near prime, and semiprime consumers are 500, 600, 660, and 720, respectively. Standard errors clustered at the employer level.

***Indicates significance at $p = 0.01$; **$p = 0.05$; *$p = 0.1$.

of the service fee to Equifax) for each loan inquiry. If an inquiry does not lead to a loan, this variable is zero. Conditional on loan origination, the gross profit is calculated as $(P_l \cdot n_l - LM_l * (n_l/N_l) - PD_l)$, where P_l is the monthly payment of loan l; n_l is the observed number of months since loan origination; N_l is the loan length in months; LM_l is the loan amount, that is, the amount that lenders extend to borrowers; and PD_l is the past due amount, which is taken in April 2020. Because 96% of the loans have not reached full terms by the end of our sample, we prorate the loan amount as $LM_l * (n_l/N_l)$.[13] If a borrower has made all the scheduled payments each month, the profit is calculated as the interest payment. On the other extreme, if the borrower does not make any payment, lenders suffer the loss of the loan amount extended.

With the calculated profit for each inquiry, we use Equation 1 to estimate the change in lenders' gross profit on the main sample. Table 5, column 1 shows that the loan profit increases by $226 on average, which represents a 19.6% increase given a baseline profit of $1,156.28. Column 2 shows the heterogeneous impact on profit for different types of consumers. We find that the increases are largest for deep subprime ($300, or a 77.2% increase) and subprime ($293, or a 26.8% increase) consumers,

Table 5. Lenders' Overall Profit Increase

Dependent variable: profit	(1)	(2)
Digital	226.44***	
	(35.71)	
Digital: Deep subprime		300.40***
		(57.98)
Digital: Subprime		292.81***
		(42.12)
Digital: Near prime		50.11
		(53.36)
Digital: Semiprime		16.96
		(77.32)
Credit score	307.84***	381.39***
(100)	(18.78)	(33.80)
State FE	✓	✓
Year-month FE	✓	✓
Employer FE	✓	✓
Observations	175,952	175,952
R^2	0.03689	0.03798

Notes. Regressions based on the main sample. The dependent variable is the gross profit of an inquiry (more details in the paper), which equals zero if the inquiry does not lead to a loan. One observation is an inquiry. The credit score cutoffs for deep subprime, subprime, near prime, and semiprime consumers are 500, 600, 660, and 720, respectively. Standard errors clustered at the employer level.

***Indicates significance at $p = 0.01$; **$p = 0.05$; *$p = 0.1$.

whereas the increase becomes statistically insignificant for near-prime and semiprime borrowers. Because the changes in interest rates are small, the profit increases are primarily driven by the market expansion effect across different credit score segments.

To sum up, in this section we find that digital access to verified employment data allows lenders to extend more loans to consumers, disproportionately for those with lower credit scores. This is achieved without a very large increase in interest rates among the expanded loan portfolio. These results suggest that the digital verification can significantly benefit consumers with low credit scores by enabling access to the credit market. In addition, the lenders' gross profit significantly increases, despite the average past due amount being higher among the expanded loan portfolio. This profit increase is driven by the significant market expansion effect, especially for lower-credit-score consumers. These results are consistent with the hypothesis that due to the cost of verifying employment data, lenders do not always take advantage of these data as provided by consumers. Therefore, low-credit-score consumers with a good employment history are excluded from access to credit. After employers join the employment database, the records become digitally verifiable for lenders, allowing them to profitably extend more loans.[14]

5. Validation Tests

Our identification strategy relies on two main assumptions. The first assumption is that the treatment and control groups have parallel trends in loan origination without the treatment. We provide indirect tests of this assumption graphically and through a leads-and-lags analysis in Section 5.1. The second assumption is that the types of loan applicants and the propensity of lenders to make an inquiry to Equifax do not change. We test this assumption in Section 5.2. Lastly, in Section 5.3, we discuss the potential generalizability of our findings.

5.1. Test of Parallel Trends

To provide evidence on the parallel trends assumption, in Figure 5 we plot the average auto loan origination rate every month for each of the four consumer groups: consumers from employers who had joined the system before the start of the data observation period ("always fulfilled"), consumers who never had a fulfilled inquiry ("always unfulfilled"), consumers from newly joined employers after they joined the system ("after joining"), and consumers from newly joined employers before they joined the system ("before joining"). The first clear pattern from the figure is that the four groups have very similar time-varying fluctuations, perhaps due to seasonality or changes in macroeconomic factors. This observation is consistent with the parallel trends assumption. Additionally, the two groups with fulfilled inquiries have very similar loan origination rates, which are much higher than those of the groups with unfulfilled inquiries. This observation is consistent with the regression result that the auto loan origination rate increases significantly with a fulfilled inquiry.

We follow-up with a more rigorous leads-and-lags analysis to formally test whether the parallel trends assumption holds. Specifically, we run a regression using the following specification:

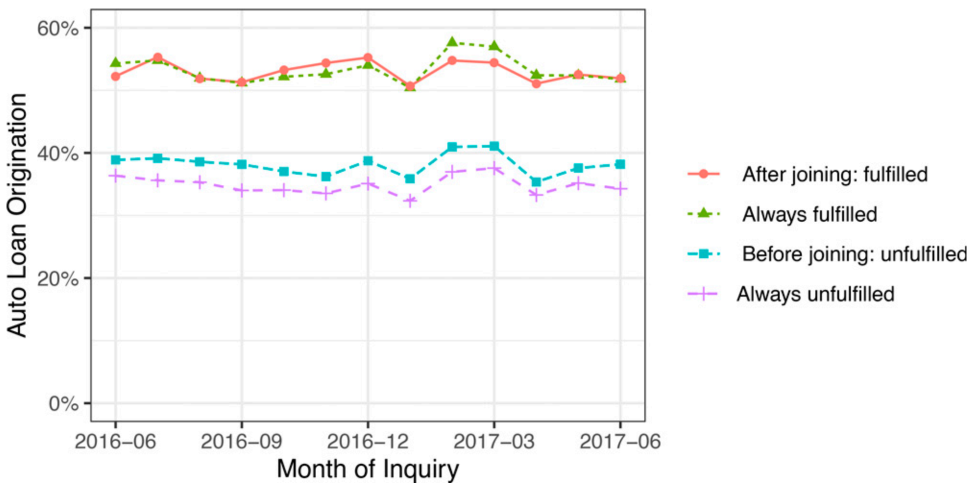
$$Y_{ijt} = \sum_{l=-m}^{n} \beta_l T_{jt}(t = k_j + l) + X_i + \eta_j + \xi_t + \epsilon_{ijt}, \qquad (2)$$

where k_j is the month when employer j joins the system, $T_{jt}(t = k_j + l)$ are time dummies for whether the current period is $k_j + l$, and l represents the l-th lead ($l > 0$) or lag ($l \leq 0$) relative to k_j. If the borrower is in the control condition, $T_{jt}(t = k_j + l) = 0$, indicating that the individual's employer joined the system outside the $(k_j - m, k_j + n)$ window. The coefficient of interest is β_l, which measures the difference in average outcome between the treatment and control groups for each period. Under the parallel trends assumption, $\beta_l = 0$ for $l < 0$. That is, there should be no pretreatment differences between the treatment and control groups after controlling for employer fixed effects and time fixed effects. Besides testing for the necessary condition for the parallel trends assumption, the leads-and-lags specification also allows us to test for the potential dynamic effect of the treatment, which is represented by β_l for $l \geq 0$.

We run the regression for inquiries in the six months before and the six months after the month when employers join the system ($m = n = 6$). We normalize the effect for $l = -6$ to zero. The results are shown in Table 6. The parameter estimates and the corresponding 95% confidence intervals are also plotted in Figure 6. We see that during the six months prior to employers' joining the system, there is no statistically significant difference in the probability of loan origination between loan inquiries in the treatment and control groups. This result is consistent with the validity of the parallel trends assumption.

There is a clear jump in the origination probability after employers join the verification system ($l = 1$ to 6). During the month they joined the system ($l = 0$), the origination rate experiences a partial increase as the employer joins the system at some point during that month. The impact is fully realized one month after joining the system, as there is no significant difference between that month and later months. The lack of a dynamic effect suggests that there is little change in the types and behaviors of lenders and borrowers after the employers join the digital verification system. As

Figure 5. (Color online) Parallel Trends: Loan Origination Rate



long as the employment information becomes digitally verifiable, whether the employer joins the system for just one month or six months makes no difference.

Table 6. Leads-Lags Analysis

Dependent variable: Loan origination (0/1)	
Month: β_{-5}	0.00931
	(0.00808)
Month: β_{-4}	0.01443
	(0.00902)
Month: β_{-3}	0.01915*
	(0.01017)
Month: β_{-2}	0.00972
	(0.01082)
Month: β_{-1}	−0.00537
	(0.01128)
Month: β_0	0.04945***
	(0.01267)
Month: β_1	0.09919***
	(0.01308)
Month: β_2	0.11221***
	(0.01380)
Month: β_3	0.10873***
	(0.01378)
Month: β_4	0.11359***
	(0.01476)
Month: β_5	0.12392***
	(0.01485)
Month: β_6	0.11381***
	(0.01500)
Credit score	0.16453***
(in 100)	(0.00432)
State FE	✓
Year-month FE	✓
Employer FE	✓
Observations	89,607
R^2	0.08648

Notes. Regressions based on inquiries from the set of employers for which we have observations from both six months before and six months after their joining the system. One observation is an inquiry. The βs are the leads-and-lags coefficients in Equation 2. Standard errors clustered at the employer level.

***Indicates significance at $p = 0.01$; **$p = 0.05$; *$p = 0.1$.

Lastly, we also conduct a robustness check by allowing for employer-specific time trends, which would not be captured by the year-month fixed effects in the main analysis. Of course, we cannot have year-month fixed effects for each employer because doing so would coincide with the time of their joining the system. Instead, we add a flexible time trend that is unique for each employer:

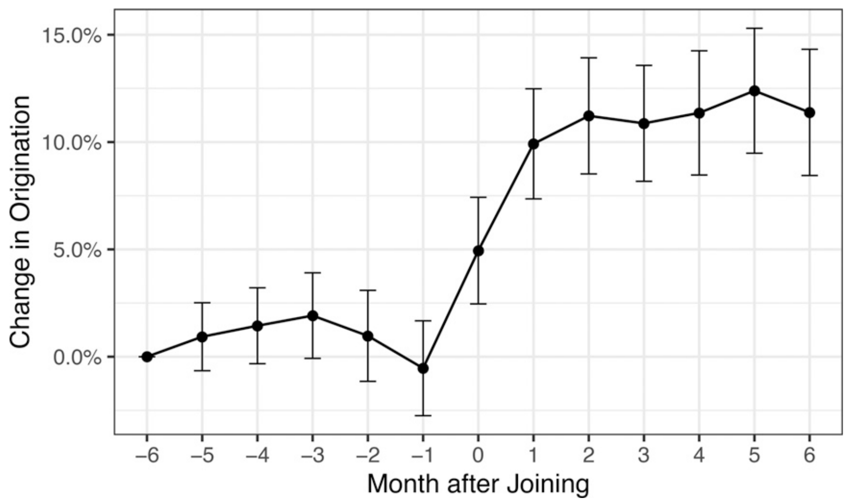
$$Y_{ijt} = \beta Digital_{jt} + X_i + \eta_j + f_j(t) + \epsilon_{ijt}. \tag{3}$$

We use two specifications to capture the time trend. In the first specification, $f_j(t) = \gamma_j t$ and γ_j captures the employer-specific loading on the time trend. In the second specification, we use a second-order polynomial where $f_j(t) = \gamma_j^1 t + \gamma_j^2 t^2$. The results are reported in Table 7. Across both specifications (columns 1 and 2), the results are very close to the main specification where we see a significant increase in the auto loan origination with fulfilled inquiries. This analysis adds reassuring evidence for the validity of our parallel trends assumption.

5.2. Changes in Loan Applicants and Lenders' Inquiries

We have established that the increase in the loan origination rate takes place right after the introduction of the digital verification system. We ascribe this change to a reduction in the time and effort required to obtain verified employment information as a result of lenders gaining digital access to such data. One might be concerned that selection into the sample might be caused by the treatment, which would constitute collider bias, rendering estimates difficult to interpret. This would happen, for example, if certain types of employees were more likely to apply for loans after their employers joined the system. Other examples include that after joining the system, employers may hire different types of employees, or employers may

Figure 6. Leads–Lags Analysis



decide to join the system because they have hired more employees who may benefit from such a system. Another concern is that the increase can come from changes in lenders' propensity of making inquiries to Equifax about loan applicants, which could arise if lenders observe when employers join the verification system and condition their inquiry decision based on where loan applicants work. Note that both changes in applicants' types and lenders' inquiries will manifest in our inquiry-level data, as the decision to inquire is a function of both. We conduct two analyses to test against these concerns.

First, we plot the employee characteristics and the number of employees in the weeks before and after their employers joined the system, together with the 95% confidence interval around the weekly averages. The left diagram of Figure 7 shows that among inquired employees, there is no statistically significant change in their average credit score, and the middle diagram shows that there is no statistically significant change in the loan applicants' average income after their employers joined the system (during month 0). These results suggest that our estimates are not driven by possible changes in the type of loan applicants. To test against possible changes in lenders' propensity of making inquiries, we would ideally test if the percentage of inquiries among all loan applicants changes over time. However, we do not observe the loan applications unless there is an employment inquiry. Instead, we study whether there is a change in the percentage of employees getting employment inquiries. Because in the first two diagrams we have provided evidence that the type of loan applicants stays the same, this suggests that their overall demand for loans should stay the same. In this case, our measure is a good proxy for the ideal measure. In the right diagram of Figure 7, we see that the percentage of employees getting inquiries stays statistically the same over time, suggesting that lenders' propensity of making inquiries does not change.[15] Formal test results on whether the weekly averages are statistically different from each other are reported in Online Appendix B.

Table 7. Loan Origination Rate with Employer-Specific Time Trend

Dependent variable: Auto loan origination		
	(1)	(2)
Fulfilled	0.13581***	0.14521***
	(0.00829)	(0.00864)
Credit score	0.14405***	0.14393***
(in 100)	(0.00327)	(0.00330)
T Employer	✓	✓
T^2 Employer		✓
State FE	✓	✓
Employer FE	✓	✓
Observations	175,952	175,952
R^2	0.13838	0.14170

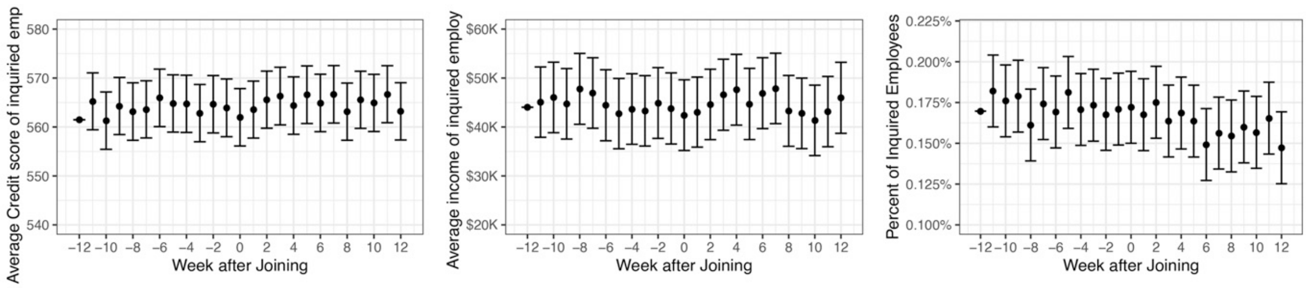
Notes. Regressions based on the main sample. One observation is an inquiry. "T Employer" is an employer-specific time trend. "T^2 Employer" is an employer-specific second-order polynomial on time. Standard errors clustered at the employer level.

***Indicates significance at $p = 0.01$.

5.2.1. Matched Inquiries Before and After Employers Join the System.

To further show that the estimated effect does not come from changes in the types of loan applicants, we conduct a matching analysis. We match each inquiry after the change with an inquiry before the change. Each pair of matched applicants have (1) the same employer, (2) credit scores not different by more than 10 points, and (3) an annual income not different by more than \$5,000 (version 1) or \$2,500 (version 2). The matching is done without replacement.

Figure 7. Share of Loan Applicants, Average Credit Score, and Average Income Over Time



Notes. The x-axis plots the number of months after employers join the verification system, where 0 denotes the month of employers' joining the system. The y-axis reports the average credit score of inquired employees, their average income, and percentage of employees that have an employment inquiry, respectively.

We first check whether the matching procedure is effective in balancing the core characteristics of loan inquiries. Figure 8 shows a Q-Q plot on the credit score and income of the underlying loan applicants in the matched treatment (after joining) and control (before joining) groups. All points are very close to the 45-degree line.

Using the matched sample, we estimate the effect of having digitally verifiable employment data on loan origination and report the results in Table 8. Columns (1) and (2) correspond to version 1 of the matching criterion, and the last two columns correspond to version 2. In both cases, the estimated effects for all consumers and across credit score groups are very close to those of the main analysis shown in Table 2. The results suggest that the main effects are not driven by possible changes in loan applicants before and after their employers join the system.

Although we have shown that inquired employees do not differ in the key aspects, we cannot empirically rule out the possibility that they differ in dimensions unobservable to us researchers. Here we appeal to the fact that participation in a program like this is typically not highlighted in HR materials and it is unlikely that employees would base the decision of loan application on whether their employers participate in the employment database.

5.3. Generalizability of Findings

In this section, we investigate the potential generalizability of our results. In particular, we study how the employers in our sample compare with an average employer in the U.S. economy. Moreover, we also study the characteristics of the employees in our sample compared with the general population in the United States. The high-level findings suggest that our results on

Figure 8. Q-Q Plot for Credit Score and Income

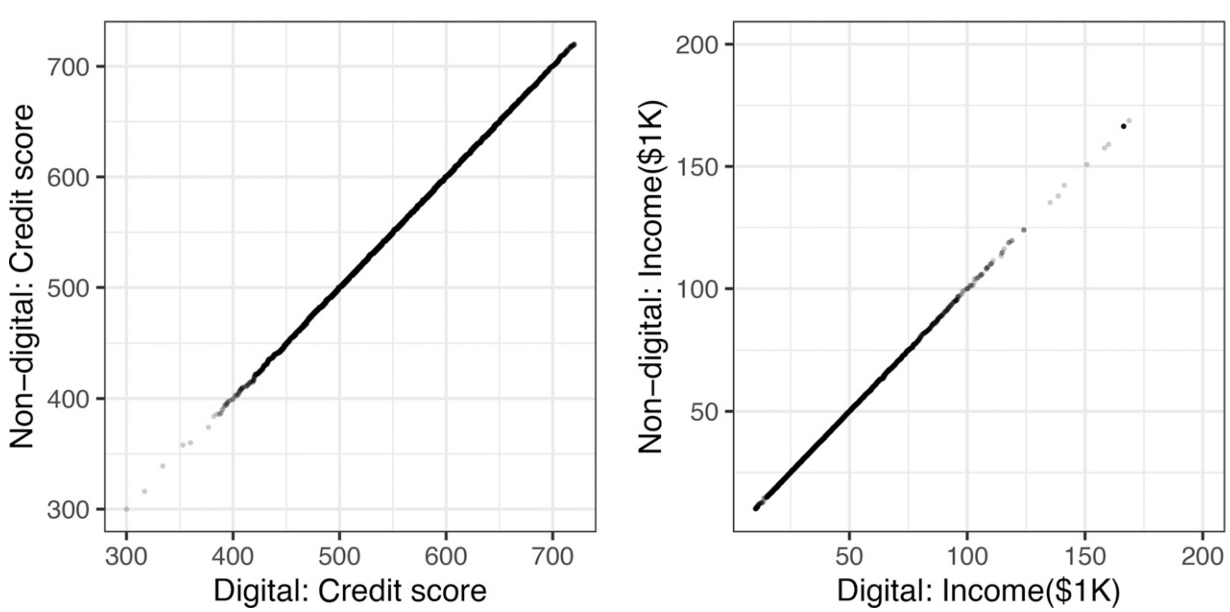


Table 8. Loan Origination Rate with Matched Inquiries

Dependent variable: Loan origination (0/1)	(1)	(2)	(3)	(4)
Digital	0.10439***		0.09462***	
	(0.01099)		(0.01243)	
Digital: Deep subprime		0.15356***		0.14091***
		(0.01332)		(0.01397)
Digital: Subprime		0.11742***		0.10991***
		(0.01220)		(0.01362)
Digital: Near prime		0.04509***		0.02868*
		(0.01367)		(0.01689)
Digital: Semiprime		0.03338***		0.01998
		(0.02051)		(0.02168)
Matched pairs FE	✓	✓	✓	✓
State FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
Observations	39,386	39,386	32,654	32,654
R^2	0.36600	0.36747	0.41643	0.41809

Notes. Regressions based on the main sample after matching. We match an inquiry before an employer joins the system to one after it joins the system based on (1) the same employer, (2) credit scores not different by more than 10 points, and (3) an annual income not different by more than \$5,000 (version 1) or \$2,500 (version 2). The matching is done without replacement. Each observation is an inquiry. Standard errors are clustered at the employer level.

***Indicates significance at $p = 0.01$; ** $p = 0.05$; * $p = 0.1$.

the impact of verified employment data have external validity.

We first examine the pool of employers covered by the Equifax database. Compared with the universe of private firms sampled by the Bureau of Labor Statistics (BLS), the employers in the Equifax system in our sample period tend to be larger employers. In Table 9, column 1, we see that a large percentage of the employers in our sample have at least 1,000 employees. The distribution is quite different from the BLS private sector firms, which are dominated by small firms with fewer than 50 employees. Part of the difference can be attributed to public sector employers, which can be larger in size and are included in our sample but not in the BLS sample.

Given the discrepancy in employer size, we analyze whether our estimated effects differ much for employers of different sizes. Specifically, we separately estimate the main effects (Table 2) by the inquiries from smaller employers (fewer than 1,000 employers) and larger employers (at least 1,000 employers). Results are shown in Table 10, where columns 1 and 2 contain observations from smaller employers, and columns 3 and 4 contain observations from larger employers. We see that the effect size is almost identical (13.0% and 12.5%), and the distributional effect on consumers in different credit score groups is qualitatively the same across samples. The latter finding is important and suggests that our main finding (i.e., access to verified employment information promotes inclusive access to credit) does not depend on employer size.

Table 9. Employer Size: Our Sample vs. BLS Sample

	Share of employers	
Number of employees	Our sample	BLS private sector firms
	(1)	(2)
1–49	2.22%	94.88%
50–499	6.90%	4.66%
500–999	7.47%	0.23%
1,000 or more	83.38%	0.21%

Next, we compare key consumer characteristics, namely income levels and credit scores, of the employees in the Equifax employment database to those of the general population in the United States. The income data for the general population comes from the American Community Survey by the Census Bureau. In Figure 9, we plot the distribution of the different income groups in the general population and the employment database. The employment database is slightly better represented in the higher-income groups, but the two distributions are quite close. Note that for the employment database, we take all employees in the database, many of whom do not have an employment inquiry from auto lenders.

Lastly, Figure 10 plots the distribution of credit scores across the two populations. The credit score data for the general population comes from the credit database of Equifax, which contains all consumers in the United States who have a credit history. We consider all employees in the employment database, who may or may not have an employment inquiry. We find that the two distributions of credit scores are quite close. Together, Figures 9 and 10 suggest that the income levels and credit scores of the employees

Table 10. Loan Origination Rate by Employer Size

Dependent variable: Loan origination (0/1)	(1)	(2)	(3)	(4)
	Small employers		Large employers	
Digital	0.13005***		0.12508***	
	(0.02952)		(0.00734)	
Digital: Deep subprime		0.21613***		0.17976***
		(0.0374)		(0.00776)
Digital: Subprime		0.11337***		0.13596***
		(0.02923)		(0.00685)
Digital: Near prime		0.09447**		0.08538***
		(0.03795)		(0.00825)
Digital: Semiprime		0.09910*		0.04117***
		(0.05889)		(0.01014)
Credit score	0.13598***	0.16342***	0.14592***	0.17676***
(in 100)	(0.01276)	(0.02044)	(0.00331)	(0.00439)
State FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
Employer FE	✓	✓	✓	✓
Observations	3,894	3,894	172,058	172,058
R^2	0.21716	0.2194	0.13329	0.13501

Notes. Regressions based on the main sample split by whether the employer has at least 1,000 employees. One observation is an inquiry. The credit score cutoffs for deep subprime, subprime, near prime, and semiprime consumers are 500, 600, 660, and 720, respectively. Standard errors clustered at the employer level.

***Indicates significance at $p = 0.01$; **$p = 0.05$; *$p = 0.1$.

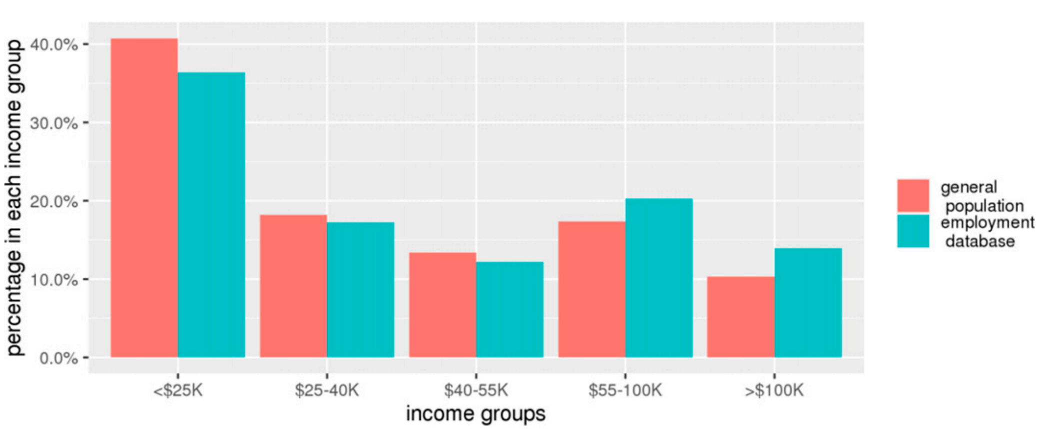
in the employment database are broadly representative of the general population.

6. Discussion

In this paper, we study how the digital access to verified employment and income data affects lenders and borrowers in the auto loan market. Using the sample of employers who joined the employment verification system during our sample period, we conduct a diff-in-diffs analysis to show that the auto loan origination rate increases by 35.5% after the employers of loan applicants joined the system. The increase among deep subprime borrowers is as high as 146.4%. The expanded access to credit does not come with a much higher interest rate. Lenders also experience an increase in their gross profit after gaining digital access to verified employment data. The increase is driven by the ability of lenders to extend loans to applicants with a low credit score but a digitally verifiable employment record. Although the delinquency risk is higher among the expanded loan portfolio, the benefit from the expansion outweighs the loss from the loans.

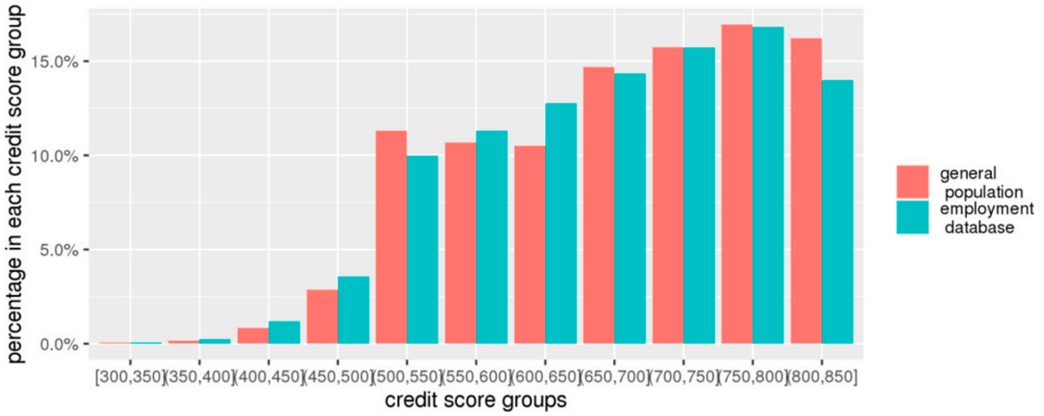
Our paper provides evidence that access to verified employment and income information promotes inclusive access to credit for consumers with low credit scores while increasing lenders' profit. Our findings shed light on recent public and private efforts on utilizing new data sources to complement the credit score in assessing loan eligibility. In the recent S.1828 - Credit

Figure 9. (Color online) Income Level: Our Sample vs. General Population



Note. Statistics for the general population based on the American Community Survey by the Census Bureau.

Figure 10. (Color online) Credit Score: Our Sample vs. General Population



Note. Statistics for the general population based on the credit database of Equifax, which contains all consumers who have a credit history.

Access and Inclusion Act, Senator Scott argues in favor of including other consumer credit-related information in loan underwriting decisions, such as whether applicants pay their lease, utility bills, or phone bills on time. In the private sector, financiers are also considering leveraging new data, such as whether applicants shop at discount stores or subscribe to magazines, to complement the credit score in issuing loans. What our paper shows is that, besides the use of creative and new data sources, we can benefit from ways of extracting more value from existing data. The value extraction in our setting is achieved through a digital infrastructure for verifying and transmitting data that have been available but is costly to access, but more generally, this could be achieved in other ways such as through better analytics of existing data. The main benefit of having digitally verifiable data are that the information is available almost instantaneously, therefore increasing the value of this information. Another potential benefit is that consumers may perceive lower privacy cost from firms making more efficient use of existing data, compared with getting more data from them.

Our study has several limitations. The measured impact of having a fulfilled employment inquiry is taken at a particular point in time where the employment database has partial coverage of the U.S. labor force. We are not able to study the counterfactual scenarios where the employment database did not exist or had full coverage of all employed individuals. Under those scenarios, the lenders' equilibrium behavior in terms of loan origination may change. For example, it is plausible that if the coverage of the employment database gets so good, lenders may interpret the absence of information as bad information (e.g., unemployed), in which case the individuals inquired may be worse off compared with the world where no one's employment information is available. Furthermore, our measured impact is also limited to the group of lenders that currently utilize the Equifax employment database. It is possible that the magnitude of the effects could be different for other lenders. We invite future research to investigate the long-term equilibrium behavior of lenders as well as lenders who participate early versus later.

Endnotes

[1] These numbers are published by the Consumer Financial Protection Bureau, https://prosperitynow.org/blog/4-key-questions-about-inclusive-credit-scoring-answered. Accessed September 25, 2019.

[2] The bill was first introduced in 2017 and re-introduced in 2019. See https://www.scott.senate.gov/media-center/press-releases/scott-manchin-introduce-legislation-to-expand-credit-access.

[3] For example, Gautier and Zenou (2010) show that car ownership allows whites to reach more jobs per unit of time, which gives them a better bargaining position in the labor market than minorities. Goldberg (2001) shows that government supported car ownership programs can help low-income families get and keep jobs.

[4] Additionally, auto loan is the third largest debt category, after mortgages and student loans. Source: https://www.investopedia.com/personal-finance/american-debt-auto-loan-debt/. Accessed March 26, 2021.

[5] See https://www.theworknumber.com/employer/. Accessed September 25, 2019.

[6] This calculation does not account for the service fee lenders pay to Equifax.

[7] Employment verification can be either required by law (e.g., government inquiries) or helpful to employees (e.g., consumer lending, rental application). Therefore, it is common for employers to fulfill employment verification requests as a service to their employees.

[8] The descriptions on the Work Number database website confirm this statement: https://workforce.equifax.com/solutions/employment-verifications.

[9] The Equifax database complies with the Fair Credit Reporting Act (FCRA): verifiers must provide permissible purpose to verify a person's employment status (e.g., the verification is indeed coming from an auto lender for a loan request).

[10] For the main analysis, we do not rely on observations of the employers who are either always or never in the Equifax database. We have repeated our analyses on the loan origination rate using the full sample and found a slightly larger main effect and the same distributional pattern across borrowers with different credit scores.

[11] Despite our conversation with a senior executive at Equifax, we could not find any specific reason that contributed to the relatively large number of participating employers around June 2017. However, this does not affect our identification as long as the parallel trends assumption holds true.

[12] Results are similar if we include in the regressions those employers who had joined the system before the sample period and those who never joined the system.

[13] One limitation of using this proration is that a borrower who is late on payments by the end of our sample period can repay them in a later period. In this case, our profit estimate is conservative.

[14] Equifax charges fees to lenders for each fulfilled inquiry. Although we may not disclose the exact change in Equifax's profit, it will increase with a larger number of records in the database, which occurs when more employers join the system.

[15] This statement is true unless lenders change their propensity of inquiries and borrowers change their propensity of applying for loans at the same time, and the two changes cancel each other out.

References

Acquisti A, Taylor C, Wagman L (2016) The economics of privacy. *J. Econom. Lit.* 54(2):442–492.

Ansari A, Mela CF (2003) E-customization. *J. Marketing Res.* 40(2): 131–145.

Athey S, Catalini C, Tucker C (2017) *The digital privacy paradox: Small money, small costs, small talk.* Technical report. National Bureau of Economic Research, Cambridge, MA.

Berg T, Burg V, Gombović A, Puri M (2018) *On the rise of FinTechs–credit scoring using digital footprints.* Technical report. National Bureau of Economic Research, Cambridge, MA.

Berman R, Israeli A (2021) The added value of data-analytics: Evidence from online retailers. Technical report, working paper, University of Pennsylvania, Huntsman Hall.

Besanko D, Dubé J-P, Gupta S (2003) Competitive price discrimination strategies in a vertical channel using aggregate retail data. *Management Sci.* 49(9):1121–1138.

Brynjolfsson E, McElheran K (2019) Data in action: Data-driven decision making and predictive analytics in us manufacturing. Rotman School of Management Working Paper (3422397), Stanford University, Stanford, CA.

Bult JR, Wansbeek T (1995) Optimal selection for direct mail. *Marketing Sci.* 14(4):378–394.

Campbell J, Goldfarb A, Tucker C (2015) Privacy regulation and market structure. *J. Econom. Management Strategy* 24(1):47–73.

Catalini C, Hui X (2018) Online syndicates and startup investment. Technical report. National Bureau of Economic Research, Cambridge, MA.

De Janvry A, McIntosh C, Sadoulet E (2010) The supply-and demand-side impacts of credit market information. *J. Dev. Econom.* 93(2):173–188.

Du RY, Netzer O, Schweidel DA, Mitra D (2021) Capturing marketing information to fuel growth. *J. Marketing* 85(1):163–183.

Dubé J-P, Misra S (2017) *Scalable price targeting*. Technical report. National Bureau of Economic Research, Cambridge, MA.

Einav L, Jenkins M, Levin J (2013) The impact of credit scoring on consumer lending. *RAND J. Econom.* 44(2):249–274.

Gautier PA, Zenou Y (2010) Car ownership and the labor market of ethnic minorities. *J. Urban Econom.* 67(3):392–403.

Ghose A, Ipeirotis PG, Beibei Li (2014) Examining the impact of ranking on consumer behavior and search engine revenue. *Management Sci.* 60(7):1632–1654.

Goldberg H (2001) *State and County Supported Car Ownership Programs Can Help Low-Income Families Secure and Keep Jobs* (Center on Budget and Policy Priorities, Washington, DC).

Goldfarb A, Tucker C (2011a) Online display advertising: Targeting and obtrusiveness. *Marketing Sci.* 30(3):389–404.

Goldfarb A, Tucker C (2012) Shifts in privacy concerns. *Amer. Econom. Rev.* 102(3):349–353.

Goldfarb A, Tucker CE (2011b) Privacy regulation and online advertising. *Management Sci.* 57(1):57–71.

Goodman-Bacon A (2018) Difference-in-differences with variation in treatment timing. Technical report. National Bureau of Economic Research, Cambridge, MA.

Jiang Z, Chan T, Che H, Wang Y (2021) Consumer search and purchase: An empirical investigation of retargeting based on consumer online behaviors. *Marketing Sci.* 40(2):219–240.

Johnson G, Shriver S, Goldberg S (2019) Privacy & market concentration: Intended & unintended consequences of the GDPR. Preprint, submitted November 15, http://dx.doi.org/10.2139/ssrn.3477686.

Khan R, Lewis M, Singh V (2009) Dynamic customer management and the value of one-to-one marketing. *Marketing Sci.* 28(6): 1063–1079.

Lambrecht A, Tucker C (2013) When does retargeting work? Information specificity in online advertising. *J. Marketing Res.* 50(5): 561–576.

Li X, Ching AT (2019) How does a firm adapt in a changing world? The case of prosper marketplace. Preprint, submitted June 26, http://dx.doi.org/10.2139/ssrn.3403404.

Lin T (2019) Valuing intrinsic and instrumental preferences for privacy. Preprint, submitted June 25, http://dx.doi.org/10.2139/ssrn.3406412.

Lin M, Prabhala NR, Viswanathan S (2013) Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Sci.* 59(1):17–35.

Netzer O, Lemaire A, Herzenstein M (2019) When words sweat: Identifying signals for loan default in the text of loan applications. *J. Marketing Res.* 56(6):960–980.

Norden L, Weber M (2010) Credit line usage, checking account activity, and default risk of bank borrowers. *Rev. Financial Stud.* 23(10):3665–3699.

Pancras J, Sudhir K (2007) Optimal marketing strategies for a customer data intermediary. *J. Marketing Res.* 44(4):560–578.

Prince J, Wallsten S (2020) How much is privacy worth around the world and across platforms? Preprint, submitted March 3, http://dx.doi.org/10.2139/ssrn.3528386.

Rafieian O, Yoganarasimhan H (2020) Targeting and privacy in mobile advertising. *Marketing Sci.* 40(2):193–218.

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4): 321–340.

Sahni NS, Wheeler SC, Chintagunta P (2018) Personalization in email marketing: The role of noninformative advertising content. *Marketing Sci.* 37(2):236–258.

Shiller B, Waldfogel J (2011) Music for a song: an empirical look at uniform pricing and its alternatives. *J. Indust. Econom.* 59(4):630–660.

Simester DI, Sun P, Tsitsiklis JN (2006) Dynamic catalog mailing policies. *Management Sci.* 52(5):683–696.

Sun T, Yuan Z, Li C, Zhang K, Xu J (2020) The value of personal data in Internet commerce: A high-stake field experiment on data regulation policy. Preprint, submitted June 3, https://ssrn.com/abstract=3566758.

Tang H (2019) The value of privacy: Evidence from online borrowers. Technical report, Working Paper, London School of Economics, London.

Tucker CE (2012) The economics of advertising and privacy. *Internat. J. Indust. Organ.* 30(3):326–329.

Tucker CE (2014) Social networks, personalized advertising, and privacy controls. *J. Marketing Res.* 51(5):546–562.

Zhang J, Krishnamurthi L (2004) Customizing promotions in online stores. *Marketing Sci.* 23(4):561–578.

Zhang J, Liu P (2012) Rational herding in microloan markets. *Management Sci.* 58(5):892–912.

Zhang J, Wedel M (2009) The effectiveness of customized promotions in online and offline stores. *J. Marketing Res.* 46(2):190–206.

Zhang J, Netzer O, Ansari A (2014) Dynamic targeted pricing in b2b relationships. *Marketing Sci.* 33(3):317–337.