The Direct and Indirect Effects of Information Deletion: Evidence from Medical Collections*

Victor Duarte
University of Illinois

Julia Fonseca
University of Illinois

Divij Kohli
University of Illinois

Julian Reif
University of Illinois and NBER

October 2024

PRELIMINARY AND INCOMPLETE

Abstract

We study the effect of deleting medical collections from credit reports on consumer credit access and financial health. In April 2023, the three major credit bureaus announced they were no longer reporting information about medical collections below $500. Using a regression discontinuity design to compare individuals on either side of this threshold, we find no direct benefits of this information deletion, ruling out even small improvements in access to credit, repayment behavior, or subprime borrowing. To study the potential indirect effects of information deletion, we build a prediction model of default behavior and split consumers by whether their predicted probability of default increases or decreases when medical collections are removed from the model. We then estimate a difference-in-differences model that compares outcome changes in these two groups before and after the April 2023 decision. Consistent with our regression discontinuity results, we find no evidence of negative spillover effects from the information deletion. Finally, we show that medical collections are not meaningfully predictive of default, which explains why past interventions to remove this information from reports did not have the intended benefit and suggests that further deletion—such as the CFPB's proposal to remove remaining medical collections from credit reports—would be similarly ineffective.

*We thank Ray Kluender and Neale Mahoney for helpful comments and discussions.

1 Introduction

Around one in seven Americans had a medical account in collections between 2019 and 2022. Most medical collections arise when third-party debt collectors report unpaid medical debts to credit bureaus.[1] Policymakers have increasingly scrutinized these reporting practices, expressing concerns that making medical debt visible to lenders could impair access to credit following unexpected medical shocks. In response to these concerns, the three major U.S. credit bureaus announced in April 2023 that they were no longer reporting medical collections below $500. Building on this announcement, the CFPB proposed a rule in June 2024 to remove all remaining medical collections from credit reports, arguing that this measure would enhance credit access and improve terms for consumers burdened with medical debt. To address concerns that deleting this information could harm other consumers, the CFPB cited evidence that medical collections are poor predictors of future default, and thus have limited relevance for credit origination decisions (Consumer Financial Protection Bureau, 2024).

However, these arguments reveal a potential contradiction. If medical collections are not meaningfully predictive of default, removing them should have no impact on risk-based pricing or loan terms offered by lenders. This suggests that while deleting medical collections is unlikely to cause negative spillover effects across credit markets, it also should not lead to greater credit access or improved terms for consumers whose medical debts are removed from reports. In other words, the absence of negative *indirect* effects implies the absence of positive *direct* effects. Clarifying this relationship is crucial for making informed policy decisions in this area.

Against this backdrop, this paper studies the direct and indirect effects of deleting medical collections from credit reports, focusing both on the direct benefits to consumers for whom this information is removed and the potential negative spillovers on other consumers. We begin by training two credit scoring models using 2019 data and an XGBoost algorithm: one that includes data on medical collections below $500 and one that does not. Comparing their outputs, we find that small medical collections are indeed not meaningfully predictive

[1] Hospitals can also report medical collections, but this is less common (Brevoort and Kambara, 2014).

of future default, indicating that they have little relevance for credit origination decisions. Thus, the April 2023 intervention that deleted this information is unlikely to have any direct or indirect effects on credit access and terms.

We confirm this insight using two separate research designs. First, we use Regression Discontinuity (RD) around the $500 threshold of information deletion. Comparing individuals just above and below this threshold, we find no evidence of direct benefits in terms of credit access, repayment behavior, or subprime borrowing. Our estimates are precise, allowing us to rule out small effects.

Next, we use our two credit scoring models—one with and one without information on medical collections below $500—to identify consumers potentially harmed by the removal of this information. Specifically, we focus on consumers whose predicted probability of default increases by 2 percentage points (the 95th percentile of the distribution) or more when medical collections below $500 are removed from the model. We use a difference-in-differences research design to compare them to consumers whose predicted probability falls by at least 2 percentage points. We show that the two groups are observationally similar across a wide range of characteristics and evolved in close parallel prior the the 2023 information deletion.[2] Consistent with our RD results, we find no evidence of negative spillover effects from the information deletion, with precise estimates that again rule out small effects.

We conclude that the 2023 decision to delete small medical collections had no negative indirect effects, but also provided no benefits to those whose information was removed. Additionally, we show that medical collections above $500 are also not predictive of default behavior, suggesting that the CFPB's 2024 proposed rule is unlikely to significantly affect credit access. Our findings suggest that policies promoting information deletion do not have the intended benefit for consumers burdened by medical debt.

Our paper relates to several strands of literature. First, we contribute to the literature on information deletion in credit and other markets. Liberman et al. (2019) study the deletion of default information in Chile and show that it led to large increases in borrowing

[2]Both groups consist primarily of low-income consumers with thin credit files. As we show in the main text, when reliable information is scarce, even randomly generated variables can influence default predictions. Thus, a noisy predictor like medical collections effectively partitions this population into two randomly assigned groups.

for consumers for whom the predicted probability of default declines with the information deletion, but also large declines for those for who become a higher default risk. Agan and Starr (2017) studies the removal of criminal history information from job applications and finds that Black applicants receive fewer callbacks. We complement this work by focusing on an ongoing policy discussion surrounding the deletion of medical collections in the United States. Rather than highlighting the unintended consequences of this information deletion, we focus on the conditions under which this deletion has any consequences—intended or otherwise. In our context, we show that the information being deleted is not relevant for the decisions of the key market players (lenders) and thus its removal has no consequences in this market.

We also relate to work on medical debt forgiveness, as debt forgiveness is viewed as a policy alternative to deleting information about the debts. Kluender et al. (2024) conduct two large-scale randomized experiments forgiving medical debt and find no impact of debt relief on credit access, utilization, and financial distress. Taken together, the findings of our two papers suggest that neither deleting information about medical debt nor purchasing debt in the secondary market and extinguishing it are effective at alleviating the distress of people who are burdened by medical debt. These findings are in contrast with work on other types of debt forgiveness, which generally find positive results (Dobbie and Song, 2015; Di Maggio et al., 2020; Cespedes et al., 2024).

This paper also contributes to the literature using machine learning techniques for credit scoring and, more broadly, to better understand credit markets. Our finding that removing medical collections from credit scores leads to non-random sorting of consumers by predicted default probabilities despite their lack of predictive power is closely related to Blattner and Nelson (2022), who show that default probability estimates for low-income consumers with thin credit files are noisier than those of other consumers. Similar to our paper, Liberman et al. (2019) uses machine-learning algorithms to study the effects of information deletion. Also related are works using machine learning for credit scoring (e.g. Agarwal et al., 2019; Fuster et al., 2022). We contribute to these works by using machine-learning credit scoring models to identify consumers at risk of negative spillover effects following the deletion of medical collections.

Finally, we also add to a growing literature that studies debt collections (e.g. Fedaseyeu, 2020; Cheng et al., 2020; Kluender et al., 2021; Fonseca, 2023). We add to this literature by studying the deletion of information about the most prevalent type of accounts in collections: those arising from medical debts. Similar to Fonseca (2023), we study both mainstream and subprime credit outcomes by linking traditional credit reports from one of the major credit bureaus to reports from Clarity, a bureau specializing in alternative financial services.

The remainder of this paper is structured as follows. Section 2 describes the data used in our analysis. Section 3 investigates whether medical collections are predictive of default. Section 4 presents our RD results on the direct effect of deleting medical collections. Section 5 reports estimates of the indirect effect from a differences-in-differences analysis. Section 6 concludes.

2 Data

We use the Gies Consumer and Small Business Credit Panel (GCCP), a panel with anonymized credit record data on consumers and small businesses obtained from Experian, one of the three major U.S. credit reporting agencies. The GCCP consists of a one percent random sample of individuals with a credit report, linked to alternative credit records from Experian's alternative credit bureau, Clarity Services, and business credit records for individuals who own a business.[3] The panel spans the 2004–2024 period, with each year consisting of a snapshot of credit records of all consumers in the sample as of the end of the first quarter. Consumers are sampled based on two randomly assigned digits of an individual identifier, allowing for flows into the panel as new consumers are randomly assigned those digits and out of the panel due to death or inactivity, mimicking flows in and out of the population of records.

The GCCP provides detailed debt information at the credit account ("tradeline") level, such as amounts owed for mortgages, student loans, and credit cards. We also observe

[3]Alternative credit records include information not reported to the major credit bureaus, such as payday loans and title loans. See Fonseca (2023) and Correia et al. (2023) for a discussion of the link between mainstream and alternative credit records in the GCCP, Fonseca and Wang (2023) on the link between consumer and business credit records, and Fonseca and Liu (2024), Howard and Shao (2022), and Fonseca et al. (2024) for other papers using the GCCP.

individuals' credit scores according to the VantageScore model and payment histories, as well as bankruptcies, judgments, and other public records. Finally, the data include information on the individual's 5-digit zip code of residence, age, gender and—for a subset of observations—the individual's education, marital status, homeownership status, and occupation. We identify a medical collection as a collection in which either the creditor is categorized as Medical/Health Care or the furnisher is categorized as a medical or health-related business.[4]

We restrict our sample to years between 2019 and 2024. We also restrict our sample to consumers aged 18 and older and exclude those with missing age, credit score, or income information and those whose age increases by 10 years or more during a 12-month period. We are left with 15,313,702 observations, which we summarize in Table 1. The first three columns refer to the full sample while the next three columns show summary statistics for individuals with medical collections. Approximately half of our sample is female and consumers in the full sample have, on average, a credit score of 702, an annual income of $51,960, and a total balance (across all credit products) of $76,460. Approximately 20% of consumers have a subprime Clarity record and the average number of collections in the full sample is 0.25. In contrast, consumers with medical collections have, on average, lower credit scores, lower income, and lower balances, and are substantially more likely to have a subprime credit record. Conditional on having a medical collection, consumers have 2.44 medical collections on average and 1.45 medical collections below $500 between 2019 and 2024.

2.1 Regression Discontinuity Sample

For our RD analysis, we focus on the period between 2022 and 2024. We restrict the sample to consumers with a medical collection in 2022 and a non-missing credit score between 2022 and 2024. The resulting sample has 266,213 consumers per year, for a total of 798,693 observations. We show summary statistics for the RD sample in 2022 in Table 2. On

[4]These category labels are Dentists, Chiropractors, Doctors, Medical group, Hospitals and clinics, Osteopaths, Pharmacies and drugstore, Optometrists and optical outlets, and Medical and related health-nonspecific.

average consumers with medical collections have a lower credit score of 626, compared to the consumers in the full GCCP sample. Consumers in this sample have an average of 3.53 accounts in collections in 2022, of which 2.67 accounts are medical collections and 1.85 are medical collections below $500 on average.

3 Do Medical Collections Predict Default?

The effect of deleting medical collections hinges on their predictive power for future default. If medical collections predict default, they should be relevant for credit origination decisions, and deleting this information may restrict the access to credit of some consumers. On the other hand, if they are not meaningfully predictive of future default, deleting this information should not have consequences for risk pricing, including for consumers for whom the information is deleted. Accordingly, we start our analysis by investigating the predictive power of medical collections. To mimic the consequences of the April 2023 decision to delete medical collections below $500, we train two credit scoring models: one that includes data on medical collections below $500 and one that does not. As we describe below, we find that medical collections below $500 are not meaningfully predictive of default. This implies that the April 2023 intervention should have no direct or indirect effects on credit access and terms, which we show is the case in Sections 4 and 5, respectively. Finally, we show that medical collections above $500 are also not predictive of default, suggesting that the CFPB's proposed rule to delete remaning medical collections should also not lead to significant direct or indirect effects on credit access and terms.

3.1 Credit scoring with and without medical collections

Traditional credit scoring models, such as the FICO and the VantageScore, are typically logit models estimated using person-level data (Federal Reserve Board, 2007). The typical model's objective is to predict "default," defined as any account that becomes 90 days or more past due within the next 18–24 months, based on observed variables relating to payment history, amounts owed, length of credit history, new credit, and credit mix (Federal Reserve Board, 2007). Consistent with these traditional models, we select 48 features with

information on current accounts and balances past due, medical and non-medical collections, bankruptcies and other public records, balances and accounts of different credit types, average account age, age of oldest account, and new inquiries and accounts.[5] Unlike traditional models, our baseline algorithm is XGBoost, a state-of-the-art machine learning algorithm for classification problems.[6] We train two person-level credit scoring models: one including the number of medical collections below $500 and one excluding this information. To mimic the effect of the 2023 intervention, we keep information on medical collections above $500 and non-medical collections. We train both models using data from 2019 to 2021, a period preceding the removal of information on medical collections below $500. Specifically, we use features in 2019 data on over 2.8 million consumers to predict defaults occurring between 2020 and 2021, reserving 90% of observations to train the model and 10% to assess model performance out-of-sample.

We report model performance across a range of metrics in Table 3. The first column shows results for models with medical collections below $500, and the second column without. In our context, the Accuracy score—measured as the share of correct predictions—is not particularly meaningful because default is a relatively uncommon outcome. For example, a model that predicts that consumers never default would have an Accuracy score of nearly 88%, equal to the share of consumers in the 2019 sample that do not default in 2020–2021. Similarly, the area under the Receiver Operating Characteristic curve (AUC), which measures the probability that the model will assign a higher default probability to a true defaulter than to a non-defaulter, is also not as informative as the Precision or Recall scores in this setting (Davis and Goadrich, 2006).

Precision and Recall scores are defined as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{1}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2}$$

[5]For more information on the features included in traditional credit scoring models, see https://www.myfico.com/credit-education/whats-in-your-credit-score.

[6]We assess the robustness of our results to a logit model in Section 3.3.

Precision measures the proportion of positive identifications (in our case, predicted defaults) that were correctly classified, while recall measures the proportion of actual positives that were correctly classified as such. Because both Precision and Recall matter, our preferred metric is the F1 score, which is a harmonic mean of Precision and Recall:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

Table 3 shows that our algorithm does a good job of predicting default. In the baseline model with medical collections below $500, we obtain an F1 Score of 0.5517. This result outperforms the 0.115 score reported by Agarwal et al. (2019), who use XGBoost and data including traditional credit scores, mobile and social footprint, and other consumer characteristics. Although our AUC of 0.7100 is lower than the 0.8602 obtained by Fuster et al. (2022) with a Random Forest model, our Precision is substantially higher (0.7339 vs. 0.0620).[7]

Comparing the two columns in Table 3 shows that removing medical collections has minimal impact on model performance, with the F1 Score only slightly declining from 0.5517 to 0.5513. In Section 3.2 below, we benchmark this difference against the effect of excluding other types of information and confirm that this difference is negligible. This provides strong evidence that medical collections below $500 are not meaningfully predictive of default. Further analysis, illustrated in Figure 1, shows that predicted default probabilities change little when removing information on medical collections below $500, with only approximately 10% of consumers seeing a change greater than 2 p.p. in absolute value.[8]

We categorize consumers into three groups based on the changes in their predicted default probabilities across the two models:

> **Negatively treated**: Consumers above the 95th percentile in the distribution of probability difference, whose predicted probability of default increases by 2 p.p. or more when medical collections below $500 are removed from the model.

[7] Blattner and Nelson (2022) obtain a similar AUC as Fuster et al. (2022) with an XGBoost model, ranging from 0.808 to 0.881 across different consumer groups, but do not report Recall or Precision scores.

[8] For comparison, 2 p.p. is the difference in 2022Q2–2024Q1 90 days or more delinquencies between consumers with a VantageScore of 300–500 and those with scores of 501–520 (https://www.vantagescore.com/lenders/risk-ratio/).

Positively treated: Consumers below the 5th percentile in the distribution of probability difference, whose predicted probability of default decreases by 2 p.p. or more when medical collections below $500 are removed from the model.

Unaffected: Consumers between the 25th and 75th percentiles, who experience a change in predicted default probabilities (in absolute value) of 0.0015 p.p. or less.

If medical collections were meaningfully predictive of default, we would expect significant differences between those consumers whose predicted probabilities increase versus those whose probabilities decrease when medical collections are removed. Specifically, we would expect consumers with medical debt to see lower predicted default probabilities once that information is eliminated from the credit scoring model. In that case, consumers in the positively treated group should have a significantly higher number of medical collections than those in the negatively treated group.

Table 4 shows summary statistics for all three groups. Negatively and positively treated consumers are similar in demographics and credit characteristics, but have lower credit scores, lower income, and lower balances than unaffected consumers. Figure 2 visualizes these results using balancing regressions. All variables are standardized and each dot is a regression coefficient of the variable labeled on the y-axis on either the positive (blue) or negatively treated flag (red). While both the positively and negatively treated groups are meaningfully different from the unaffected group, they are similar to each other. If medical collections were strongly predictive of default, we would expect a pronounced sorting effect along that variable. Figure 2 shows some evidence of sorting, with the positively treated slightly more likely to have medical collections, but these differences are very small.

While the positively and negatively treated consumers are quite similar to each other, they differ significantly from the unaffected group, indicating that comparing predicted probabilities with and without medical collections under $500 results in non-random sorting of consumers. This sorting arises because default probability estimates for low-income consumers with thin credit files are noisy (Blattner and Nelson, 2022). Consequently, even a feature with little to no predictive power such as small medical collections may receive positive loadings and influence predictions for these consumers. Excluding this predictor

will then distinguish between the less noisy (Unaffected) and more noisy (Positively Treated and Negatively Treated) groups, but fail to further differentiate between positively and negatively treated consumers, effectively assigning them across the two groups randomly. To confirm this explanation, we later introduce a randomly generated variable into the model and demonstrate that excluding it reproduces the same pattern observed when excluding medical collections under $500, underscoring their limited predictive value (see Section 3.2).[9]

3.2 Comparison to deleting other variables

3.2.1 Other variables in consumer credit reports

We benchmark the effect on model performance of removing medical collections below $500 by comparing it to the effect of removing other information. Table A.1 presents the results. The first two columns reproduce the results presented in Table 3. The subsequent columns show results when, starting from the full set of 48 features in our baseline model, we alternatively drop: (i) medical collections above $500, (ii) non-medical collection trades and balance, (iii) age of the oldest account and average account age, (iv) age of newest trade and number of inquiries in the past 12 months, (v) 90 days or more past due trades and balances and share of accounts that are 90 days or more past due.

Removing medical collections below $500 causes a substantially smaller decline in model performance than all other variables other than medical collections above $500, whose removal slightly improves model performance. According to our preferred metric, the F1 Score, the performance decrease from removing non-medical collections is over 7 times larger than from removing small medical collections, while removing all information about accounts that are 90 days or more past due produces a decline nearly 40 times larger.

Note that removing information on larger medical collections actually improves model performance. As we demonstrate in Section 3.2.2, the same effect occurs when we introduce and remove a randomly generated from our feature set, suggesting that larger medical collections are even less predictive of default than smaller collections. In Sections 4 and 5, we

[9] In theory, models estimated using state-of-the-art algorithms such as XGBoost should learn to disregard variables with no predictive value during training. In practice, these variables can still receive non-zero loadings due to limitations of finite sample sizes.

show that the 2023 intervention, which deleted medical collections below $500 from credit reports, had no meaningful direct or indirect effects on access to credit and consumer financial health. Given that medical collections above $500 appear to be even less predictive of default, we expect that the CFPB's proposed rule to erase all remaining medical collections will also have little or no meaningful effect on consumers.

3.2.2 Random noise

Next, we compare the effect of deleting medical collections below $500 with deleting pure noise. To do so, we train a version of our model that includes a feature drawn randomly from a uniform distribution. We then compare the performance of that model to our baseline model, which does not include this random variable. We show performance metrics in Table A.2 in the Appendix. Removing small medical collections reduces the F1 Score from 0.5517 to 0.5513, whereas removing a random variable increases it from 0.5510 to 0.5517. This difference suggests that small medical collections are not purely noise and do have non-zero predictive power. However, the changes in predicted default probabilities that we obtain by removing the noise variable are strikingly similar to those from removing small medical collections. In Figure A.1, we overlay the histogram of probability differences from Figure 1 with the analogous histogram obtained from removing the noise variable. The patterns are very similar in both cases.

Moreover, Figure A.2 shows that removing noise sorts consumers across our three groups (unaffected, positively treated, and negatively treated) in a manner that mirrors the effects of removing medical collections under $500. Panel A reproduces the balance plot from Figure 2, where each dot represents the coefficient from a regression of the (standardized) variable labeled on the y-axis on either the positive (blue) or negatively treated flag (red). In panel B, we present an analogous plot by comparing two models, one with and without the noise variable, and re-sorting consumers into three groups as described in Section 3.1. We again find that "treated" consumers have, on average, lower credit scores, lower income, and lower balances, reinforcing the idea that estimates of default probabilities are substantially noisier for these consumers (Blattner and Nelson, 2022).

3.3 Robustness

3.3.1 Logit Model

Next, we show that our results above are not specific to our choice of algorithm. We train a logit model on the same data, using the same set of 48 features and again reserving 90% of observations to train the model and 10% to assess model performance out-of-sample. Table A.3 shows our results. The logit model with medical collections has an F1 score of 0.4577, substantially lower than the 0.5517 achieved with XGBoost (Table 3). Importantly, the decline in performance from deleting small medical collections is identical across the two algorithms: $0.4577 - 0.4573 = 0.0004$ (Table A.3) with a logit and $0.5517 - 0.5513 = 0.0004$ with XGBoost (Table 3).

Moreover, deleting small medical collections from the logit model sorts consumers across our three groups (unaffected, positively treated, and negatively treated) along very similar dimensions as in XGBoost. Figure A.3 presents a balance plot analogous to the one shown in Figure 2. Once again, we find that positive and negatively treated consumers have lower credit scores, lower income, and thinner credit files, showing that the noisier estimates of default probabilities that we obtain for these consumers are not sensitive to our choice of algorithm.[10]

4 Direct Effect: Regression Discontinuity

4.1 Empirical strategy

We employ an RD design to identify the effect of medical debt deletion on consumer credit outcomes. Accounts with less than $500 of medical debt were deleted. For analytical convenience, we recenter the medical debt running variable by measuring its value as distance (in dollars) from the $500 cutoff. Although treatment occurs at the account level, our main analysis is conducted at the individual level, which necessitates aggregating the running variable across an individual's accounts. We aggregate by taking the maximum debt value among all

[10] This result aligns with Blattner and Nelson (2022), who show that the finding of noisier credit scores for low-income, thin-file, minority consumers is robust to different algorithms, including XGBoost and logit.

accounts. Alternative methods of aggregation and their interpretations are discussed below.

Our main identifying assumption is that assignment to either side of the medical debt cutoff is as good as random. We believe this assumption is reasonable: it is difficult for individuals to precisely manipulate their medical debt, which depends on fixed prices that often are unknown in advance to the individual. Our data come from Experian's administrative records, which minimizes concerns about sample selection bias.

We estimate the following model at the individual level:

$$Y_{it} = \alpha DEBT_i^{2022} + \beta ABOVE_i^{2022} + \gamma(ABOVE_i^{2022} \times DEBT_i^{2022}) + \epsilon_{it} \tag{4}$$

The dependent variable, Y_{it}, is an outcome for individual i in year t. The running variable, $DEBT_i^{2022}$, is measured as distance of the individual's maximum-medical-debt account from the \$500 cutoff, and $ABOVE_i^{2022}$ is an indicator equal to one if $DEBT_i^{2022} \geq 0$. Our model employs a local linear approximation to the unknown regression functions underlying the average causal effect at the threshold. We allow the slope of our linear approximation to vary on either side of the cutoff.

We use a triangular kernel in all regressions. Our preferred specification uses a mean-squared error optimal bandwidth that remains constant on either side of the cutoff but can vary across outcomes. We report robust bias-corrected confidence intervals that accommodate potential misspecification of the estimating equation (Calonico et al., 2014).

Our focal parameter of interest is β, which we interpret as the intent-to-treat effect of not having any accounts deleted. Equivalently, we interpret $-\beta$ as the effect of having at least one account deleted. In additional, unreported analyses, we estimated a model using the minimum rather than the maximum debt value across all accounts, as well as a model estimated at the account level instead of the individual level. While these models estimate different treatment effects, the results remain qualitatively similar.

4.2 Results

We begin by estimating the effect of the 2023 deletion on the number of accounts in collections. Panel A of Figure 3 indicates that this intervention reduced the total number of

collections in 2024 by 0.24. Focusing specifically on medical collections, Panel B shows that the intervention reduced medical collections by 0.3 (107%). These findings confirm that the credit reporting agencies largely removed small collections accounts from their reports.

Figure 4 shows the effect of the 2023 deletion on measures of credit access and utilization. We do not observe any significant changes in the outcomes around the debt threshold. Table 5 presents formal estimates. Our 95% confidence intervals can rule out that deletion improved credit scores by 0.43 (0.07%), balances by $1,940 (3.7%), and revolving utilization by 0.55 (1.82%). Finally, Figure 5 shows the effect on different measures of past-due trades and subprime credit.

Under the assumptions of our RD design, outcomes unrelated to medical debt should remain unchanged at the threshold. We present a series of covariate smoothness tests in Figure A.4. In line with our expectations, we find no significant differences in average age, average income, or the proportion of individuals married at the threshold, for both 2022 and 2024.

5 Indirect Effect: Difference-in-differences

5.1 Empirical strategy

We use a differences-in-differences research design to study the indirect effects of deleting medical collections below $500 from credit reports. To do so, we restrict our sample to consumers whose predicted probability of default increases when medical collections below $500 are removed from our credit scoring model (negatively treated) and those whose probability of default decreases (positively treated), as described in Section 3. Because these groups are sorted based on their 2019 characteristics, we restrict our sample for this analysis to 2020–2024.

If lenders make origination decisions using credit scoring models that load on medical collections below $500 and that information ceases to be available, consumers in the negatively treated group would see lower credit availability as their predicted probability of default rises. As we discuss in Section 3 and show in Figure 2, positively and negatively

treated groups are similar across a wide range of observables, making our key identifying assumption—that outcomes for the two groups would have evolved similarly in the absence of the information deletion—more plausible.

We estimate the following regression model at the individual level:

$$Y_{ict} = \alpha + \beta TREATED_i \times POST_t + \lambda_i + \delta_{ct} + \epsilon_{it}, \tag{5}$$

where Y_{ict} is an outcome for consumer i, residing in county c, in year t; $TREATED_i$ is a binary variable equal to one if the consumer belongs to the negatively treated group and zero if they belong to the positively treated group; $POST_t$ equals zero before 2023 and one after; and λ_i and δ_{ct} are vectors of consumer and county-year fixed effects, respectively. Because we assign treatment based on differences in the predicted probability of default with and without medical collection, we create 1,000 equal-sized bins of the difference in predicted probability of default in our full sample and cluster our standard errors at the bin level. The negative and positively treated groups correspond to approximately 10% of the full sample, and we are thus left with approximately 100 clusters.

Our coefficient of interest is β which, under the assumption of parallel trends, is an estimate of the average effect of deleting medical collections on outcomes of consumers whose predicted probability of default increases when medical collections are removed, relative to consumers whose predicted probability of default decreases. Note that, to the extent that removing medical collections affects the credit outcomes that we measure, the control group—consumers whose predicted probability of default decreases when medical collections are deleted—would benefit. That means that β is an overestimate of what we would obtain by comparing consumers in our negatively treated group to a control group that is unaffected by the information deletion. Thus, the true indirect of deleting medical collections is likely even smaller than the estimates we report below, which rule out with 95% confidence even a small impact.

To show evidence in favor of the parallel trends assumption, we also estimate event-study versions of Equation (5):

$$Y_{ict} = \alpha + \sum_{\kappa=2020}^{2024} \beta_\kappa TREATED_i \times \mathbb{I}_{\kappa=t} + \lambda_i + \delta_{ct} + \epsilon_{it}, \tag{6}$$

where $\mathbb{I}_{\kappa=t}$ is a binary variable equal to one in year t and zero otherwise. We omit the binary variable for the year prior to the intervention (2022), which serves as the reference year.

5.2 Results

Figure A.6 plots coefficient estimates and 95% confidence intervals for Equation (6) for outcomes relating to credit availability, including credit scores (panel A), total balance across all credit products (panel B), revolving credit limit (panel C), and accounts opened in the past 6 months (panel D). Across all outcomes, we find that consumers in the negative and positively treated groups evolved in close parallel before 2023, supporting the parallel trends assumption. After 2023, we see no differential effects between the two groups in these measures.[11]

Figure 7 shows similar results for measures of payment history and subprime borrowing, including the number of trades 90 days or more past due (panel A), the number of subprime trades (panel B), total subprime credit (panel C), and a binary variable equal to one if the consumer has a subprime record (panel D). We again find no evidence of pre-trends or of an effect of the intervention.[12] Taken together, these results suggest that the positive treated group is a good counterfactual for the negatively treated group and that deleting medical collections from credit reports does not have negative spillover effects on other consumers.

We report pooled estimates from Equation (5) in Table 6, for the same set of outcomes as Figures A.6 and 7. Not only are estimates insignificant at 5%, they are also small. Magnitudes range from 0.04% of the sample mean for credit scores (column 1) to -3.34% for the presence of a subprime record (column 8). Note that estimates for subprime trades (column 6) and having a subprime record (column 8) are significant at 10% but estimates are small relative to the sample mean and negative, suggesting, if anything, a positive spillover effect of reducing subprime borrowing which is hard to reconcile with a causal interpretation.

[11] Note that we should not expect a mechanical effect on credit scores from removing small medical collections since medical collections have not been used in the VantageScore model since 2022.

[12] We show additional outcomes in Appendix Figure A.5.

We conclude that, despite the fact that we may be overstating the true effect by comparing the negatively treated group to a group that is positively treated rather than unaffected, we can rule out even small negative spillovers from the deletion of small medical collections with 95% confidence.

6 Conclusion

This paper studies the direct and indirect effects of deleting medical collections from credit reports using a combination of machine-learning algorithms for predicting default, regression discontinuity, and differences-in-differences analysis. Contrary to intentions, we demonstrate that deleting medical collections from credit reports likely has no meaningful impact on credit access or terms. Our analysis focuses on medical collections under $500, which were removed from reports by all three major credit bureaus in 2023. First, we show that this information is not meaningfully predictive of default by comparing the output of two machine-learning credit scoring models: one with and one without this information.

We then confirm our hypothesis: if small medical records are not relevant for risk pricing, their deletion should not affect credit access and terms. Using a regression discontinuity design, we find no evidence that consumers benefit from the removal of derogatory information in terms of credit access, repayment behavior, or subprime borrowing, ruling out even small effects. Next, we analyze potential spillover effects using a differences-in-differences framework. We compare consumers whose predicted default probability rises when small medical collections are removed from the credit scoring model with observationally similar consumers whose predicted default probability declines. We find no evidence of negative spillover effects from the information deletion, again ruling out small effects.

Our findings contribute to the policy debate on alleviating the burden of medical debt. While standard economic models advocate for ex-ante solutions like health insurance, this approach has proven difficult to achieve in practice: about 30 million Americans remain without formal insurance, and even the insured often face substantial out-of-pocket expenses (Einav and Finkelstein, 2023). Recent policy efforts have centered around alternative, ex-post solutions such as debt forgiveness and deleting medical debt from credit reports. However, these

alternatives may not deliver their intended benefit to consumers. For example, Kluender et al. (2024) find that purchasing medical debt in the secondary market and forgiving it has no impact on credit access and financial distress in two large-scale randomized experiments. Our findings add to this evidence, suggesting that deleting records of these medical debts is similarly ineffective.

References

Agan, A. and S. Starr (2017, 08). Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*. *The Quarterly Journal of Economics 133*(1), 191–235.

Agarwal, S., S. Alok, P. Ghosh, and S. Gupta (2019). Financial inclusion and alternate credit scoring: Role of big data and machine learning in fintech. *Working paper*.

Blattner, L. and S. Nelson (2022). How costly is noise? data and disparities in consumer credit. *Working paper*.

Brevoort, K. and M. Kambara (2014). Data point: Medical debt and credit scores. *Technical Report*. Consumer Financial Protection Bureau.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Cespedes, J., C. Parra, and C. Sialm (2024). The effect of principal reduction on household distress: Evidence from mortgage cramdown. *Review of Financial Studies Forthcoming*.

Cheng, I.-H., F. Severino, and R. R. Townsend (2020, 07). How Do Consumers Fare When Dealing with Debt Collectors? Evidence from Out-of-Court Settlements. *The Review of Financial Studies 34*(4), 1617–1660.

Consumer Financial Protection Bureau (2024, June). Prohibition on creditors and consumer reporting agencies concerning medical information (Regulation V). Federal Register. Docket No. CFPB-2024-0023, RIN 3170-AA54.

Correia, F., P. Han, and J. Wang (2023). The online payday loan premium. *Working paper*.

Davis, J. and M. Goadrich (2006). The relationship between precision-recall and roc curves. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, New York, NY, USA, pp. 233–240. Association for Computing Machinery.

Di Maggio, M., A. Kalda, and V. Yao (2020). Second chance: Life without student debt. *Working paper*.

Dobbie, W. and J. Song (2015, March). Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *American Economic Review 105*(3), 1272–1311.

Einav, L. and A. Finkelstein (2023). *We've Got You Covered: Rebooting American Health Care*. New York: Portfolio.

Fedaseyeu, V. (2020). Debt collection agencies and the supply of consumer credit. *Journal of Financial Economics 138*(1), 193–221.

Federal Reserve Board (2007). Report to the congress on credit scoring and its effects on the availability and affordability of credit.

Fonseca, J. (2023). Less Mainstream Credit, More Payday Borrowing? Evidence from Debt Collection Restrictions. *The Journal of Finance 78*(1), 63–103.

Fonseca, J. and L. Liu (2024). Mortgage lock-in, mobility, and labor reallocation. *Journal of Finance Forthcoming*.

Fonseca, J., L. Liu, and P. Mabille (2024). Unlocking mortgage lock-in: Evidence from a spatial housing ladder model. *Working paper*.

Fonseca, J. and J. Wang (2023). How much do small businesses rely on personal credit? *Working paper*.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance 77*(1), 5–47.

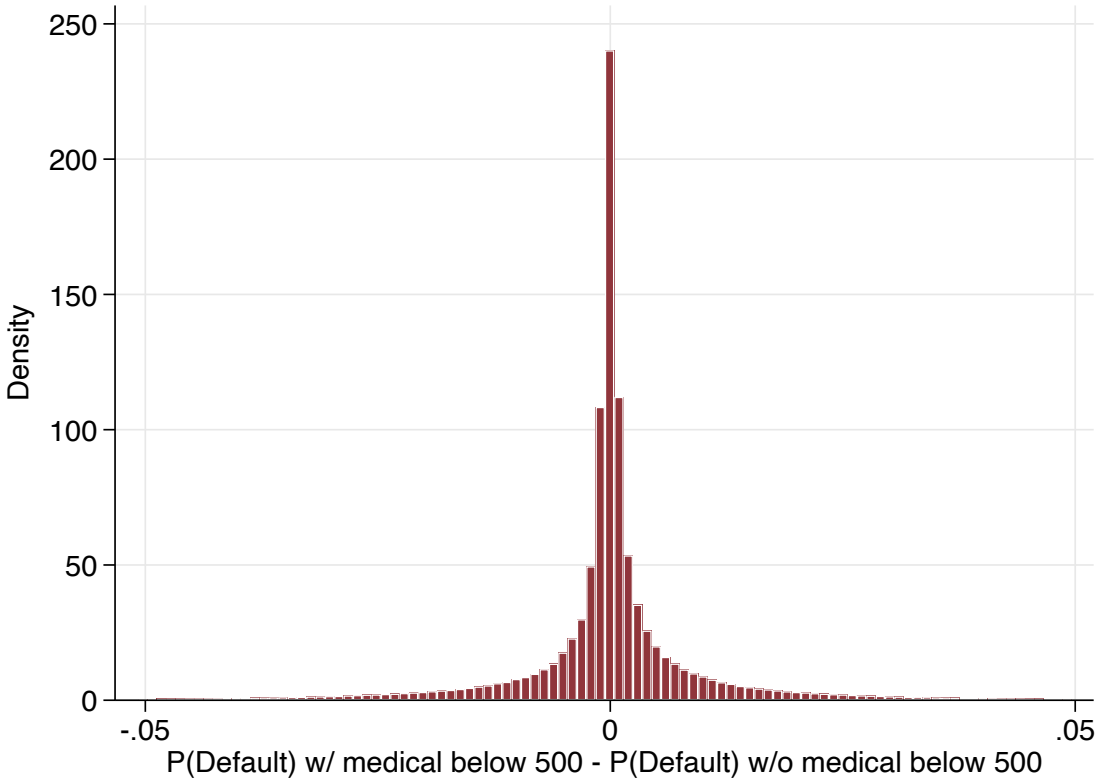
Howard, G. and H. Shao (2022). Internal migration and the microfoundations of gravity. *Working paper*.

Kluender, R., N. Mahoney, F. Wong, and W. Yin (2021). Medical debt in the US, 2009-2020. *JAMA 326*(3), 250–256.

Kluender, R., N. Mahoney, F. Wong, and W. Yin (2024). The effects of medical debt relief: Evidence from two randomized experiments. *Working paper*.

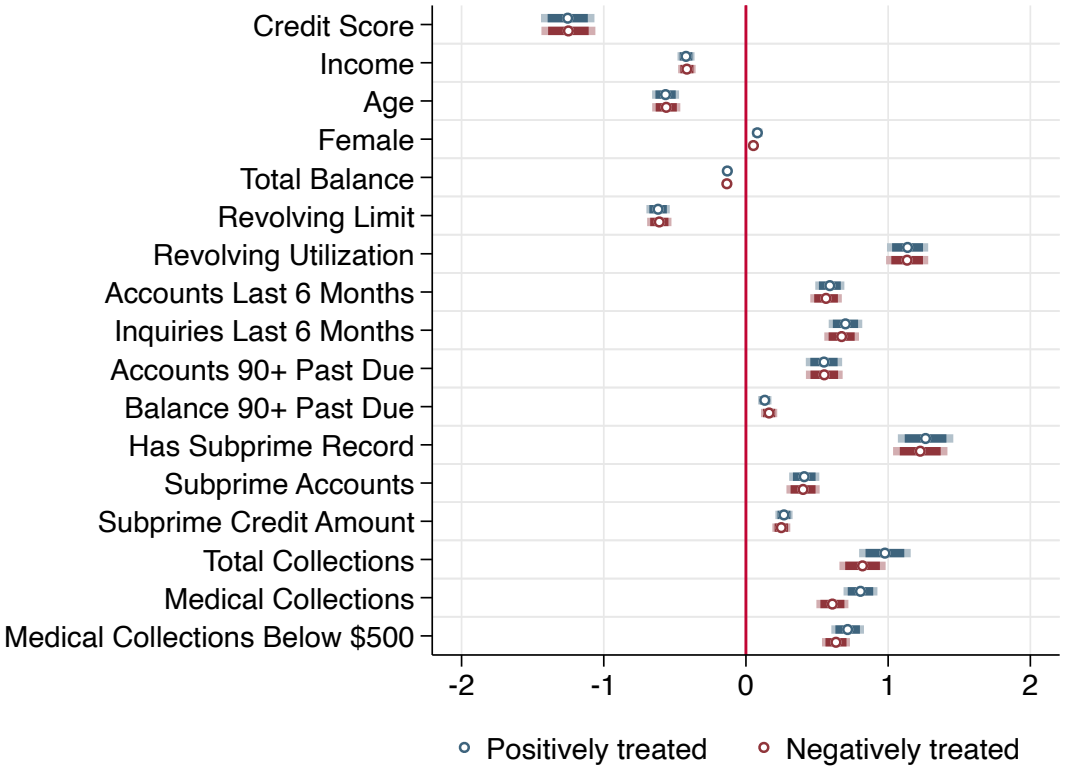
Liberman, A., C. Neilson, L. Opazo, and S. Zimmerman (2019). The equilibrium effects of information deletion: Evidence from consumer credit markets. *Working paper*.

Figure 1: Changes in Predicted Default Probabilities from Removing Medical Collections



This figure shows a histogram of the difference in predicted probability of default under our baseline credit scoring model with medical collections below $500 and in the model without this information. In both versions of the model, default is predicted using 2019 features.

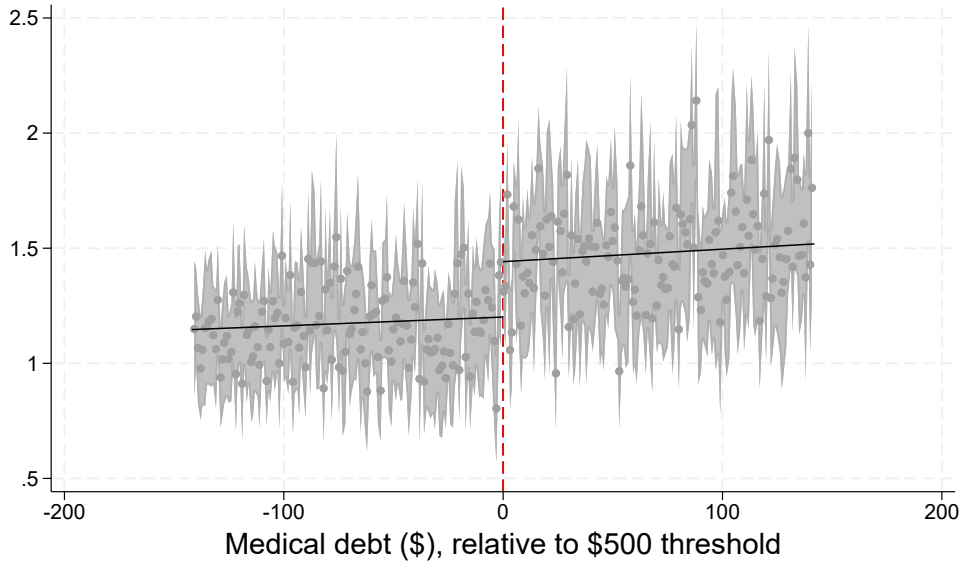
Figure 2: Covariate Balance by Changes in Default Probabilities



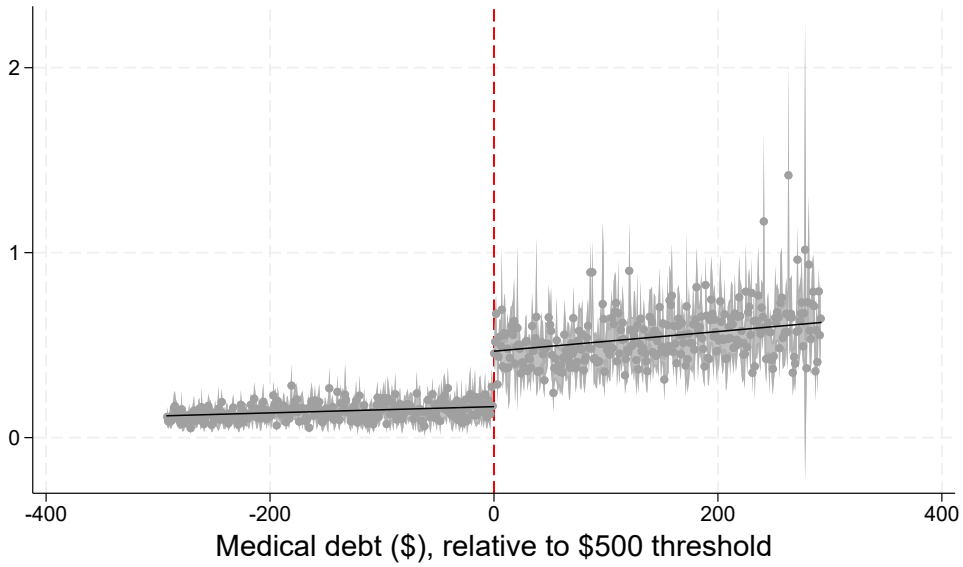
This figure shows coefficients of balancing regressions for a range of covariates, denoted in the y-axis. Negatively Treated refers to consumers above the 95th percentile in the distribution of probability difference, whose predicted probability of default increases by 2 p.p. or more when medical collections below $500 are removed from our baseline credit scoring model. Positively Treated refers to consumers below the 5th percentile in the distribution of probability difference, whose predicted probability of default decreases by 2 p.p. or more when medical collections below $500 are removed from our baseline credit scoring model. The omitted category consists of consumers between the 25th and 75th percentiles, who experience a change in predicted default probabilities (in absolute value) of 0.0015 p.p. or less. We construct 100 equal-sized bins of the difference in predicted probability of default and cluster standard errors at the bin level.

Figure 3: Number of Accounts in Collections, 2024

(a) Total Collections (#)

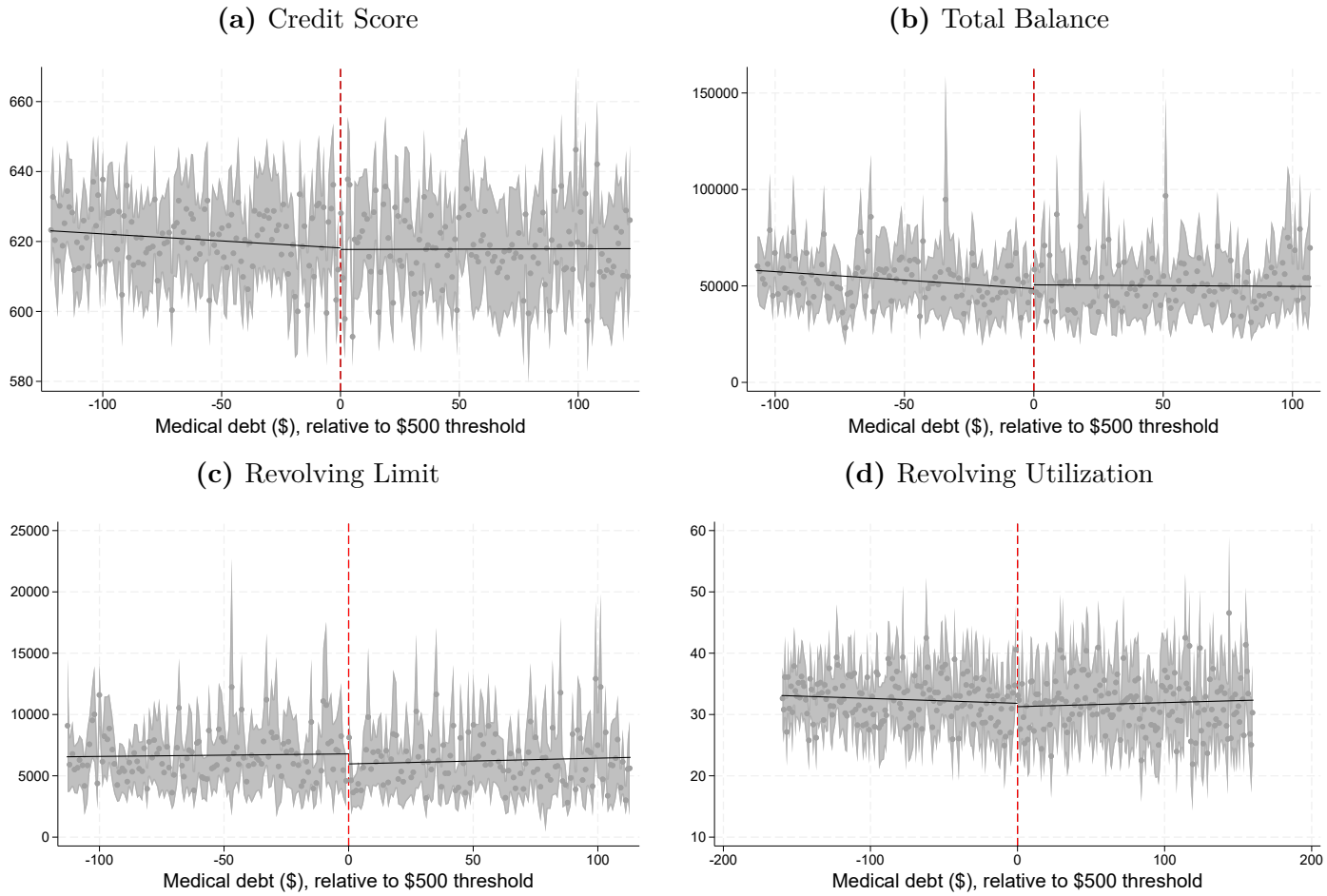


(b) Medical Collections (#)



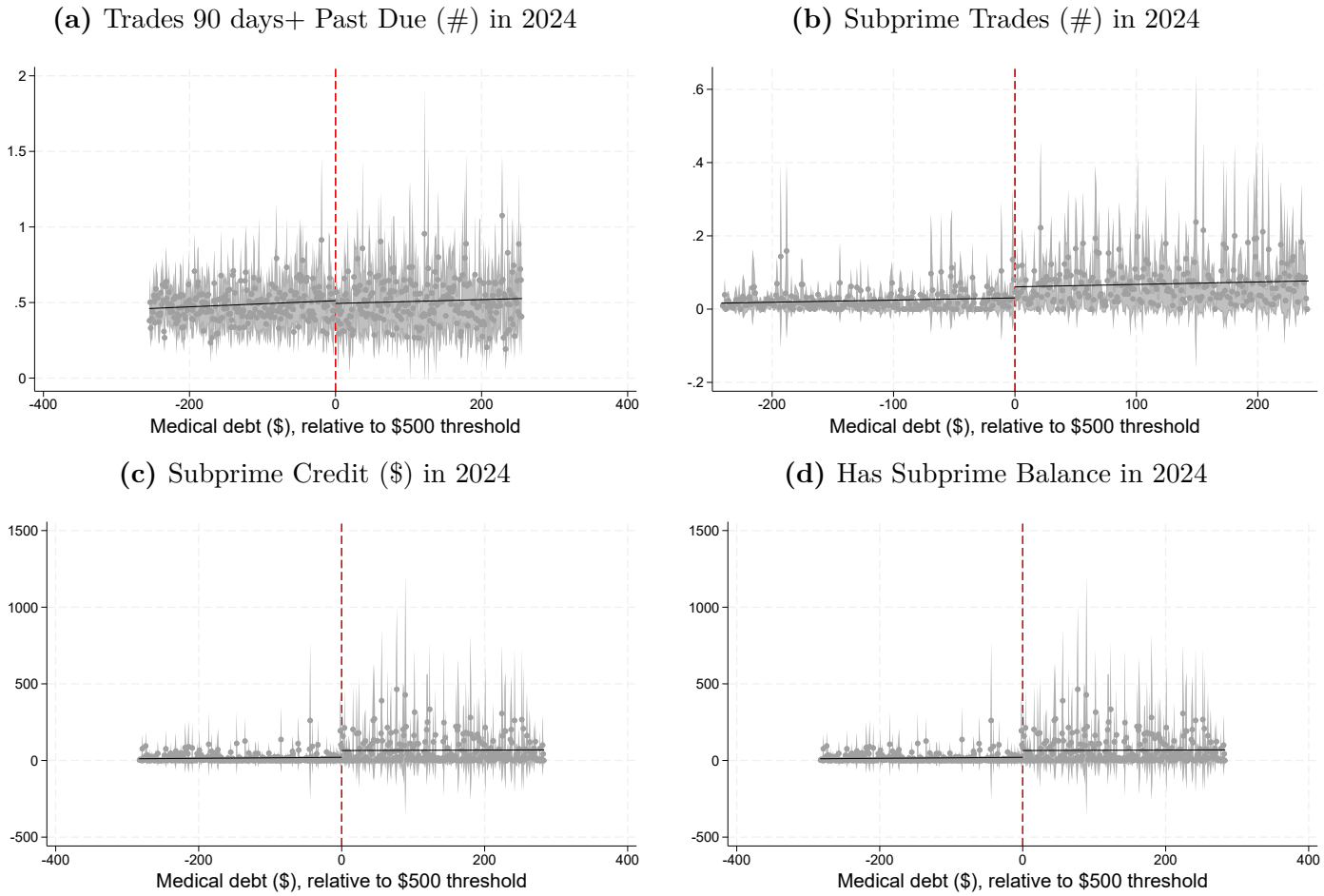
Notes: This figure shows the average number of consumer accounts in collections by medical debt, relative to the $500 threshold. The running variable for medical debt corresponds to the highest debt amount across the consumer's medical collections accounts. Total Collections is the total number of trades in collection. Medical Collections is the total number of trades in collection under Medical or Healthcare category. The fitted lines are estimated using Equation (4).

Figure 4: Credit Access and Utilization, 2024



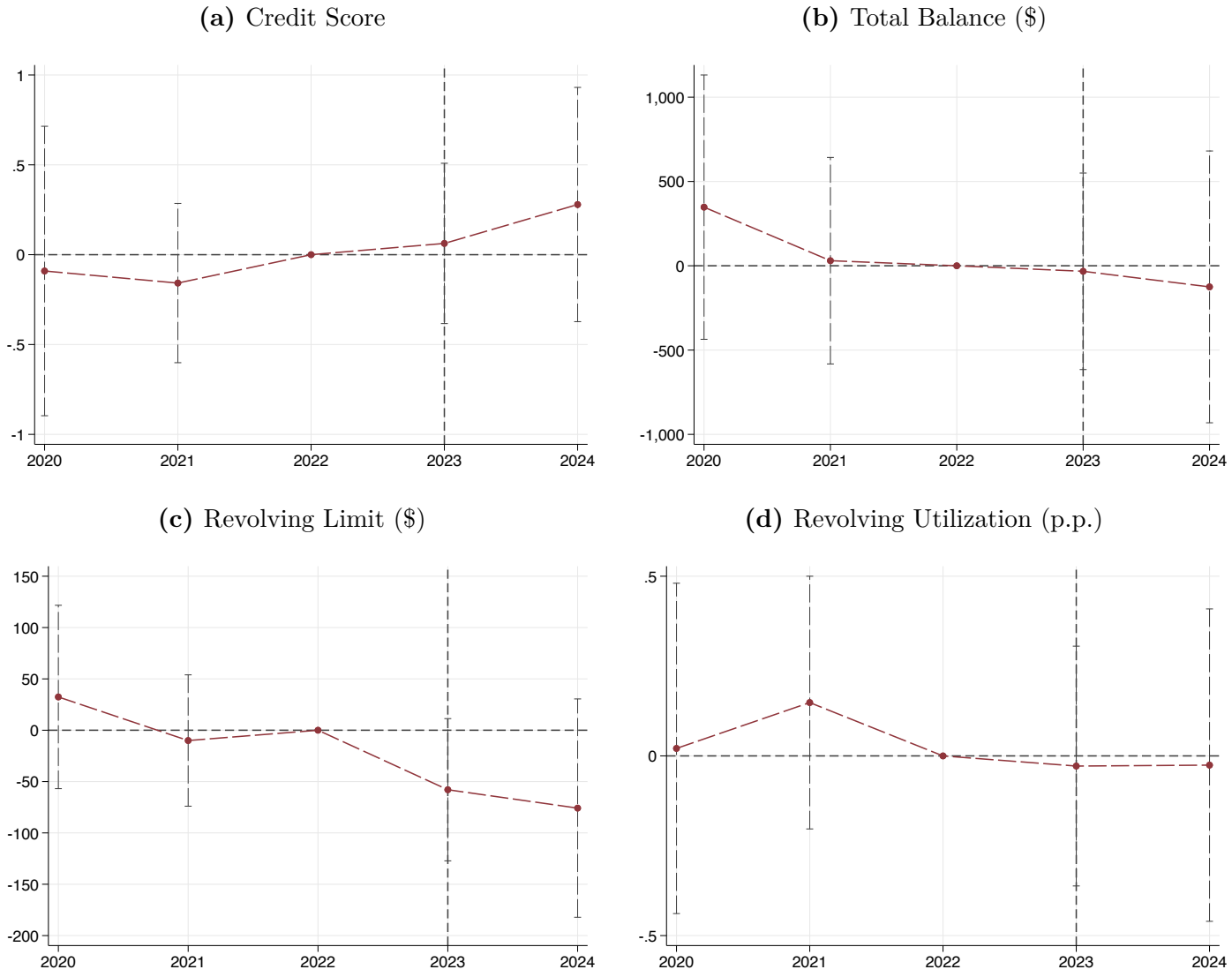
Notes: This figure shows the average consumer credit score, total balance, revolving limit, and revolving utilization by medical debt, relative to the $500 threshold. The running variable for medical debt corresponds to the highest debt amount across the consumer's medical collections accounts. Credit Score are credit scores according to the VantageScore model. Total Balance is the total balance across all credit accounts. Revolving Limit is the total limit across all revolving credit accounts. Revolving Utilization is the balance-to-limit ratio across all revolving credit accounts. The fitted lines are estimated using Equation (4).

Figure 5: Past-due Trades and Subprime Credit, 2024



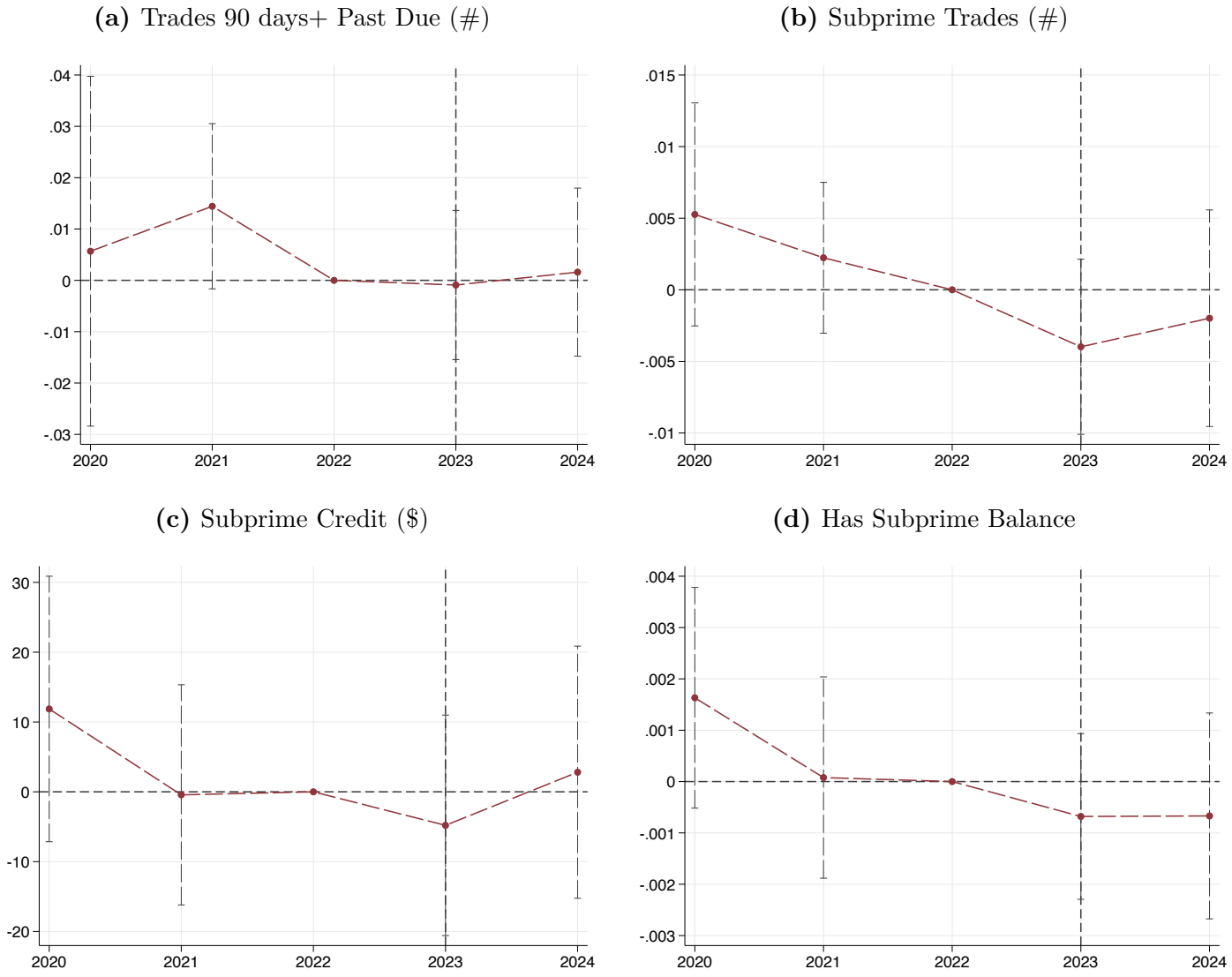
Notes: This figure shows the average consumer number of past-due trades, subprime trades, and subprime credit by medical debt, relative to the $500 threshold. The running variable for medical debt corresponds to the highest debt amount across the consumer's medical collections accounts. Trades 90 days+ Past Due is the total number of trades 90 days or more past-due. Subprime Trades is the total number of subprime and alternative credit accounts in the consumer's Clarity report. Subprime Credit is the total credit amount across all subprime trades.Has Subprime Balance is a binary variable that equals one if the consumer has a positive subprime credit amount. The fitted lines are estimated using Equation (4).

Figure 6: Event Study: Credit Access and Utilization



This figure shows coefficient estimates and 95% confidence intervals of Equation (6) for the outcomes denoted in panel captions. Credit Score are credit scores according to the VantageScore model. Total Balance is the total balance across all credit accounts. Revolving Limit is the total limit across all revolving credit accounts. Revolving Utilization is the balance-to-limit ratio across all revolving credit accounts. We we create 1,000 equal-sized bins of the difference in predicted probability of default in our full sample and cluster our standard errors at the bin level. Our differences-in-differences sample corresponds to approximately 10% of the full sample, and we are thus left with approximately 100 clusters.

Figure 7: Event Study: Past-due Trades and Subprime Credit



This figure shows coefficient estimates and 95% confidence intervals of Equation (6) for the outcomes denoted in panel captions. Trades 90 days+ Past Due is the total number of trades 90 days or more past-due. Subprime Trades is the total number of subprime and alternative credit accounts in the consumer's Clarity report. Subprime Credit is the total credit amount across all subprime trades. Has Subprime Balance is a binary variable that equals one if the consumer has a positive subprime credit amount. We we create 1,000 equal-sized bins of the difference in predicted probability of default in our full sample and cluster our standard errors at the bin level. Our differences-in-differences sample corresponds to approximately 10% of the full sample, and we are thus left with approximately 100 clusters.

Table 1: Summary Statistics

	Full Sample			With Medical Collection		
	Mean	Med.	St. Dev.	Mean	Med.	St. Dev.
Credit Score	702.26	715.00	100.85	611.21	601.00	86.69
Income ($1,000)	51.96	41.00	32.81	38.69	34.00	19.51
Age	50.56	49.00	19.41	44.90	43.00	15.16
Female (p.p.)	50.02	100.00	50.00	54.06	100.00	49.83
Total Balance ($1,000)	76.46	10.14	140.15	41.91	6.75	86.90
Revolving Limit ($1,000)	21.47	6.42	33.72	4.92	0.00	14.46
Revolving Utilization (p.p.)	28.09	13.00	32.84	49.86	47.00	39.50
Accounts Opened Last 6 Months	0.41	0.00	0.83	0.47	0.00	0.95
Inquiries Last 6 Months	0.41	0.00	0.74	0.62	0.00	0.90
Accounts 90+ Past Due	0.19	0.00	0.90	0.46	0.00	1.33
Balance 90+ Past Due ($1,000)	0.81	0.00	12.16	1.95	0.00	15.90
Total Collections	0.60	0.00	1.87	3.81	3.00	4.11
Total Collection Balance ($1,000)	0.56	0.00	3.65	3.14	1.39	9.90
Has Subprime Record (p.p.)	19.99	0.00	39.99	47.93	0.00	49.96
Subprime Accounts	0.05	0.00	0.43	0.13	0.00	0.68
Subprime Credit Amount ($1,000)	0.05	0.00	0.79	0.13	0.00	1.19
Medical Collections	0.25	0.00	1.02	2.44	2.00	2.17
Medical Collections Below $500	0.15	0.00	0.66	1.45	1.00	1.53
Observations	15,313,700			1,585,485		

Notes: This table shows descriptive statistics between 2019 and 2024 for our full sample in the first three columns and of the sample conditional on at least one medical collection on the next three columns. All variables come from the Gies Consumer and Small Business Credit Panel.

Table 2: Summary Statistics for Consumers with Medical Collections, 2022

	Mean	Med.	St. Dev.	Observations
Credit Score	625.22	618.00	87.95	276,976
Income ($1000)	40.69	35.00	20.54	276,976
Age	45.14	43.00	15.18	276,976
Female (p.p.)	0.55	1.00	0.50	266,022
Total Balance ($1,000)	49.62	10.63	106.28	276,976
Revolving Limit ($1000)	6.15	0.22	18.13	276,976
Revolving Utilization (p.p.)	29.52	6.00	38.59	276,976
Accounts Opened Last 6 Months	0.54	0.00	1.00	276,976
Inquiries Last 6 Months	0.67	0.00	0.93	276,976
Accounts 90+ Past Due	0.32	0.00	0.92	276,976
Balance ($1,000) 90+ Past Due	0.20	0.00	3.00	276,976
Accounts in Collections	3.54	2.00	3.78	276,976
Collection Balance ($1,000)	2.69	1.10	6.49	276,976
Has Alternative Credit (p.p.)	0.08	0.00	0.27	276,976
Alternative Credit Accounts	0.14	0.00	0.67	276,976
Alternative Credit Amount ($1,000)	0.18	0.00	1.44	276,976
Medical Collections	2.67	1.00	3.57	276,976
Medical Collections Below $500	1.84	1.00	2.77	276,976

Notes: This table provides summary statistics from the Gies Consumer and Small Business Credit Panel (GCCP). The statistics are calculated using data from 2022, the year preceding the removal of information on medical collections below $500. The unit of observation is the consumer. The sample is limited to consumers with a non-missing credit score during the years 2022–2024 and who have at least one medical collection account.

Table 3: Credit Scoring Metrics

Metric	With Medical < $500	Without Medical < $500
Accuracy	0.9133	0.9132
Recall	0.4420	0.4418
Precision	0.7339	0.7330
F1 Score	0.5517	0.5513
AUC	0.7100	0.7098

Notes: This table reports the performance of our two baseline credit scoring models: with and without information of medical collections below $500. Accuracy score is the share of correct predictions. Precision is the proportion of predicted defaults that were correctly classified. Recall is the proportion of actual defaults that were correctly classified. F1 score is a harmonic mean of Precision and Recall. AUC if the area under the Receiver Operating Characteristic curve, which measures the probability that the model will assign a higher default probability to a true defaulter than to a non-defaulter.

Table 4: Summary Statistics by Changes in Default Probabilities

	Unaffected			Positively Treated			Negatively Treated		
	Mean	Med.	St. Dev.	Mean	Med.	St. Dev.	Mean	Med.	St. Dev.
Credit Score	747.53	784.00	84.25	613.53	612.00	80.12	613.38	612.00	80.61
Income ($1,000)	57.94	47.00	34.81	44.07	38.00	22.69	44.31	38.00	23.08
Age	59.06	60.00	18.73	46.50	45.00	14.36	46.68	45.00	14.61
Female (p.p.)	50.33	100.00	50.00	54.42	100.00	49.80	53.24	100.00	49.89
Total Balance ($1,000)	82.28	4.51	150.07	65.10	16.88	115.61	64.98	17.87	115.36
Revolving Limit ($1,000)	30.20	17.20	38.88	6.06	0.48	16.21	6.31	0.43	16.95
Revolving Utilization (p.p.)	14.73	6.00	21.82	52.41	52.00	38.36	52.58	52.00	38.47
Accounts Opened Last 6 Months	0.26	0.00	0.62	0.72	0.00	1.25	0.71	0.00	1.23
Inquiries Last 6 Months	0.25	0.00	0.56	0.79	0.00	1.00	0.77	0.00	1.00
Accounts 90+ Past Due	0.03	0.00	0.36	0.58	0.00	1.59	0.59	0.00	1.59
Balance 90+ Past Due ($1,000)	0.16	0.00	5.16	2.65	0.00	23.49	3.00	0.00	28.50
Total Collections	0.15	0.00	0.92	2.29	1.00	3.86	1.92	1.00	3.15
Total Collection Balance ($1,000)	0.12	0.00	1.49	2.14	0.28	6.02	1.66	0.15	4.64
Has Subprime Record (p.p.)	5.27	0.00	22.35	54.61	100.00	49.79	53.71	100.00	49.86
Subprime Accounts	0.01	0.00	0.16	0.18	0.00	0.82	0.18	0.00	0.83
Subprime Credit Amount ($1,000)	0.01	0.00	0.31	0.22	0.00	1.60	0.21	0.00	1.57
Medical Collections	0.07	0.00	0.51	1.01	0.00	1.90	0.79	0.00	1.63
Medical Collections Below $500	0.05	0.00	0.39	0.65	0.00	1.38	0.59	0.00	1.29
Observations	3,400,235			334,209			316,078		

Notes: This table shows descriptive statistics between 2019 and 2024 for Unaffected consumers in the first three columns, Positively Treated consumers in the next three columns, and Negatively Treated consumers in the last three columns. Negatively Treated refers to consumers above the 95th percentile in the distribution of probability difference, whose predicted probability of default increases by 2 p.p. or more when medical collections below $500 are removed from our baseline credit scoring model. Positively Treated refers to consumers below the 5th percentile in the distribution of probability difference, whose predicted probability of default decreases by 2 p.p. or more when medical collections below $500 are removed from our baseline credit scoring model. Unaffected consists of consumers between the 25th and 75th percentiles, who experience a change in predicted default probabilities (in absolute value) of 0.0015 p.p. or less. All variables come from the Gies Consumer and Small Business Credit Panel.

Table 5: Effect of removing small medical collections from credit report

Outcome variable	(1) Mean	(2) RD Estimate
A. Number of Credit Records		
Total Credit Records	1.31	0.24***
		[0.16, 0.34]
Medical Credit Records (#)	0.28	0.30***
		[0.27, 0.32]
B. Credit Access and Utilization		
Credit Score	619	-0.43
		[-3.56, 5.89]
Total Balance ($)	52,160	1,940
		[-3,195, 8,669]
Revolving Utilization (p.p.)	30.08	-0.55
		[-2.56, 1.32]
C. Past-due Trades and Subprime Credit		
Trades 90 days+ Past Due (#)	0.49	-0.02
		[-0.07, 0.03]
Subprime Trades (#)	0.04	0.03***
		[0.01, 0.04]
Subprime Credit ($)	33.17	44.51***
		[16.54, 65.78]
Has Subprime Balance	0.02	0.02***
		[0.02, 0.03]

Notes: This table shows the coefficient (β) estimates and 95% confidence intervals of Equation (4). The running variable for medical debt corresponds to the highest debt amount across the consumer's medical collections accounts. Column (1) reports means of the dependent variable in 2024. Column (2) reports MSE-optimal estimates with robust, bias-corrected 95% confidence intervals in brackets. A *** indicates significance at the 1% level using conventional inference.

Table 6: Indirect Effect: Difference-in-differences

Dependent Variable:	Credit Score	Balance	Limit	Utilization	Del. Trades	Sub. Trades	Sub. Credit	Has Subprime
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated × Post	0.253	-204.162	-74.272	-0.081	-0.006	-0.005*	-4.818	-0.001*
	(0.276)	(383.784)	(53.150)	(0.148)	(0.007)	(0.003)	(6.158)	(0.001)
Sample Mean (p.p.)	616.42	68,280.15	6,920.53	52.98	0.58	0.19	214.29	0.04
% of Mean	0.04	-0.30	-1.07	-0.15	-1.09	-2.88	-2.25	-3.24
Consumer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,132,156	1,132,156	1,132,156	780,593	1,132,156	1,132,156	1,132,156	1,132,156

Notes: This table shows estimates Equation (5). Credit Score are credit scores according to the VantageScore model. Balance is the total balance across all credit accounts, Limit is the total limit across all revolving credit accounts. Utilization is the balance-to-limit ratio across all revolving credit accounts. Del. Trades is the total number of trades 90 days or more past-due. Sub. Trades is the total number of subprime and alternative credit accounts in the consumer's Clarity report. Sub. Credit is the total credit amount across all subprime trades. Has Subprime is a binary variable that equals one if the consumer has a positive subprime credit amount. We we create 1,000 equal-sized bins of the difference in predicted probability of default in our full sample and cluster our standard errors, reported in parentheses, at the bin level. Our differences-in-differences sample corresponds to approximately 10% of the full sample, and we are thus left with approximately 100 clusters.

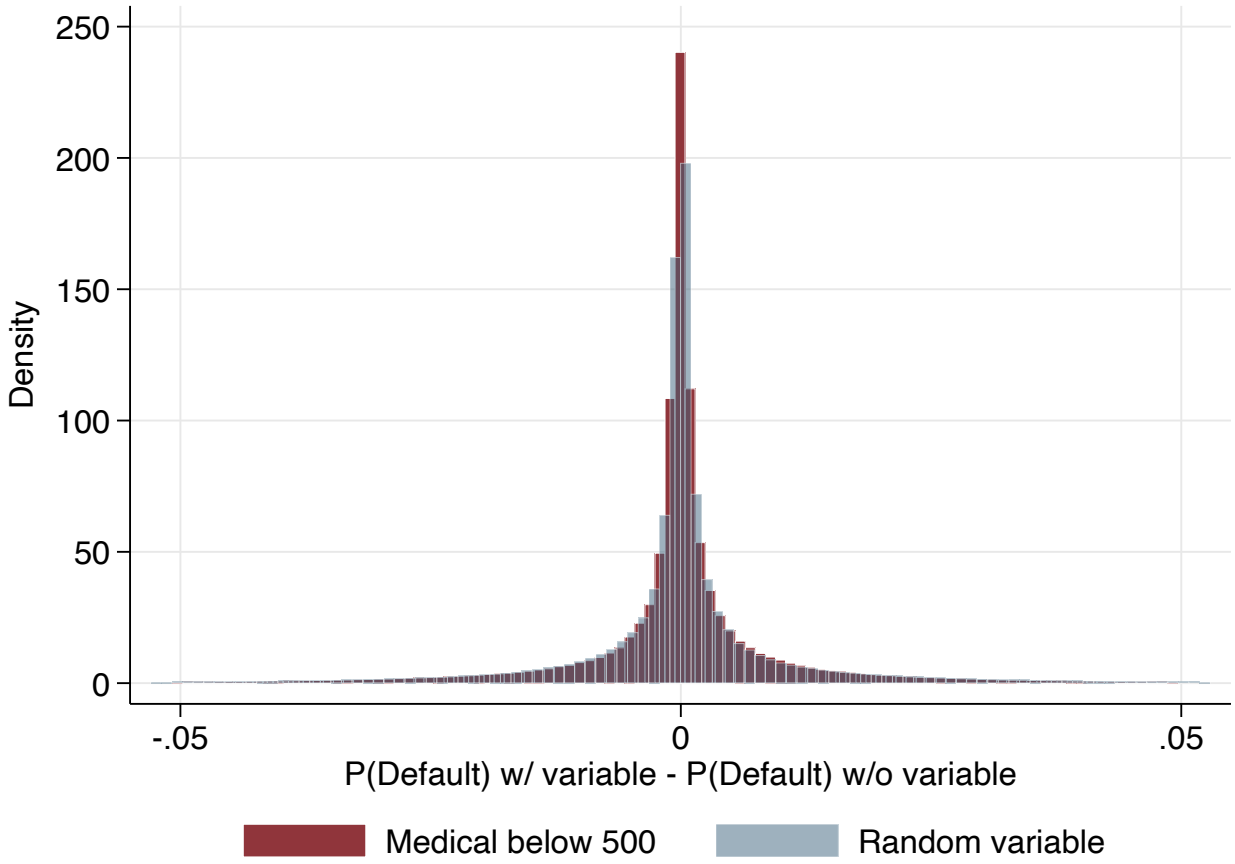
Online Appendix

"The Direct and Indirect Effects of Information Deletion: Evidence from Medical Collections"

Victor Duarte, Julia Fonseca, Divij Kohli, Julian Reif

A Additional Figures and Tables

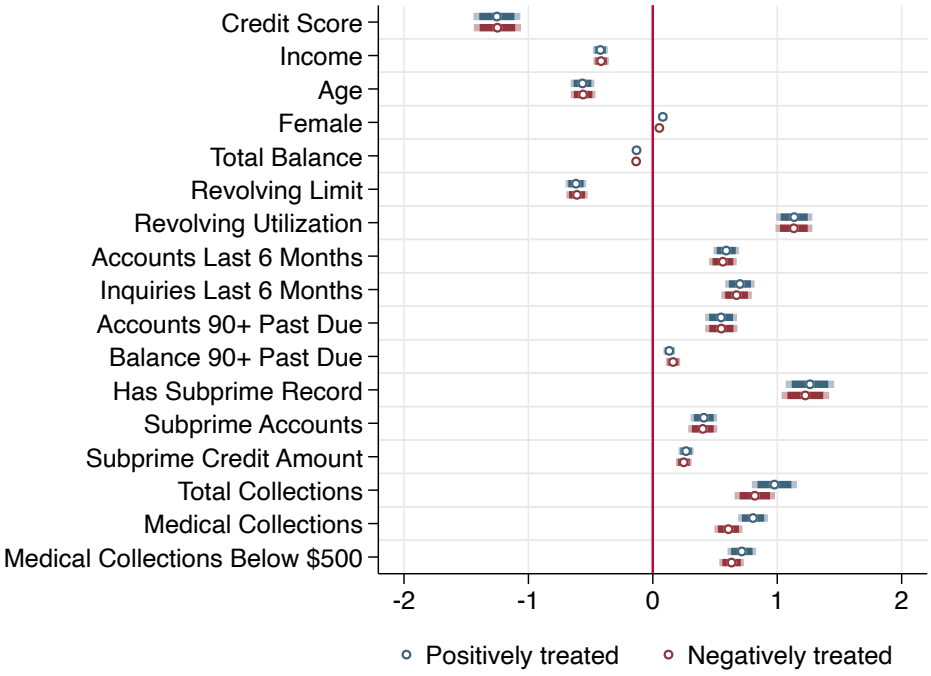
Figure A.1: Changes in Predicted Default Probabilities: Medical vs. Noise



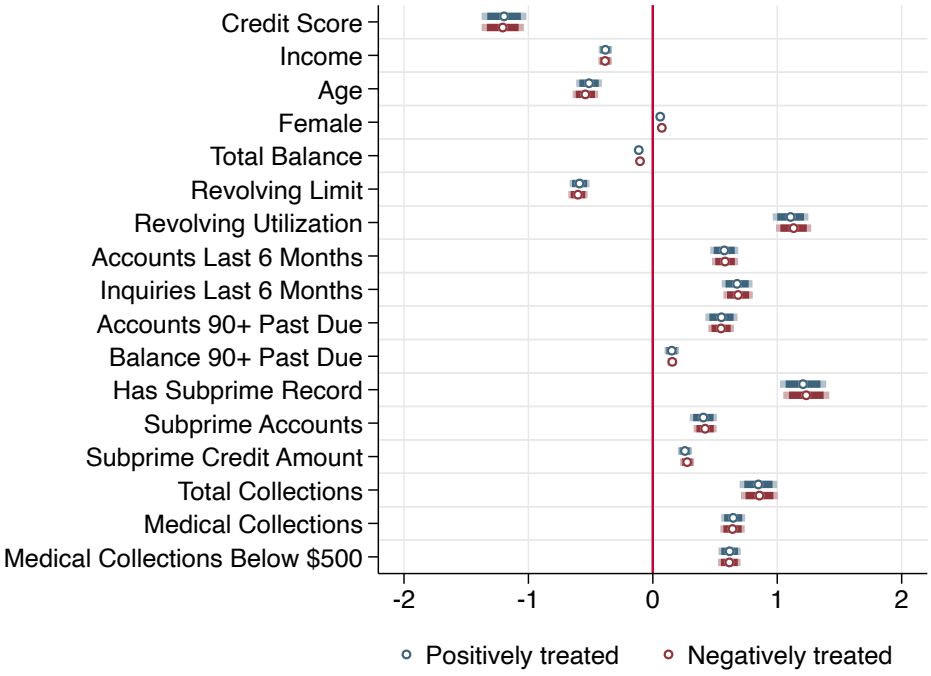
This figure shows a histogram of the difference in predicted probability of default under our baseline credit scoring model with medical collections below $500 and in the model without this information in red. In blue, it shows an analogous histogram for credit scoring models with and without a feature drawn randomly from a uniform distribution. In all versions of the model, default is predicted using 2019 features.

Figure A.2: Covariate Balance by Changes in Default Probabilities: Medical vs. Noise

(a) Medical Collections Below $500

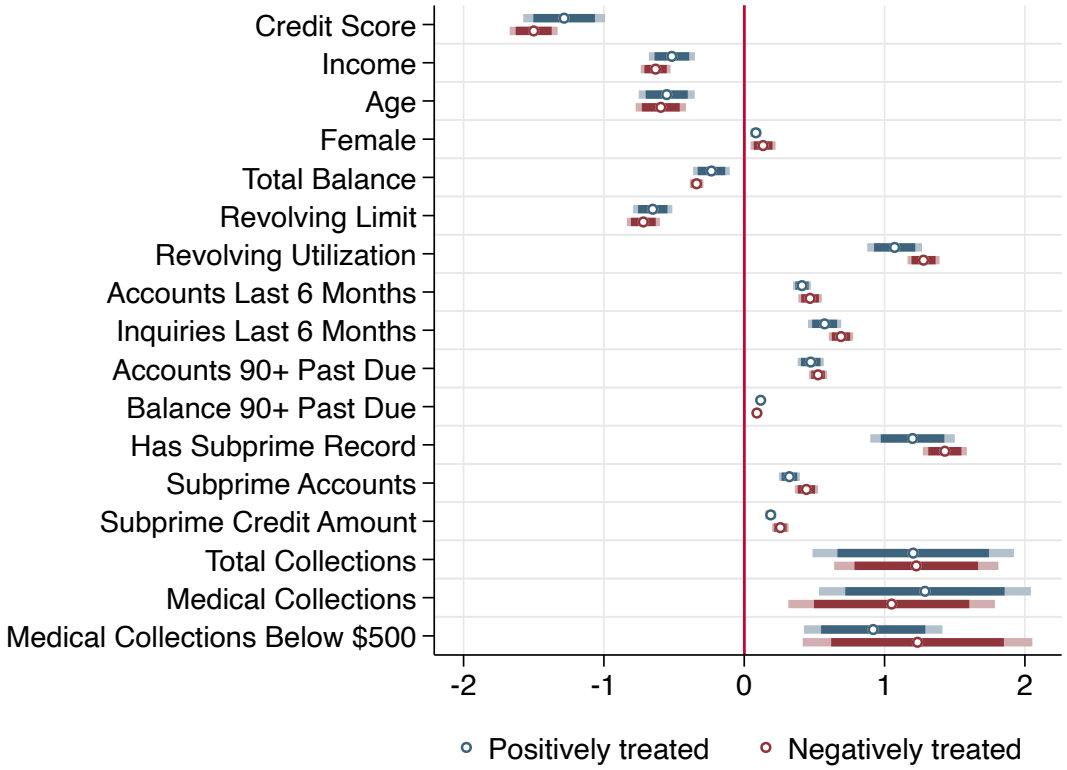


(b) Random Variable



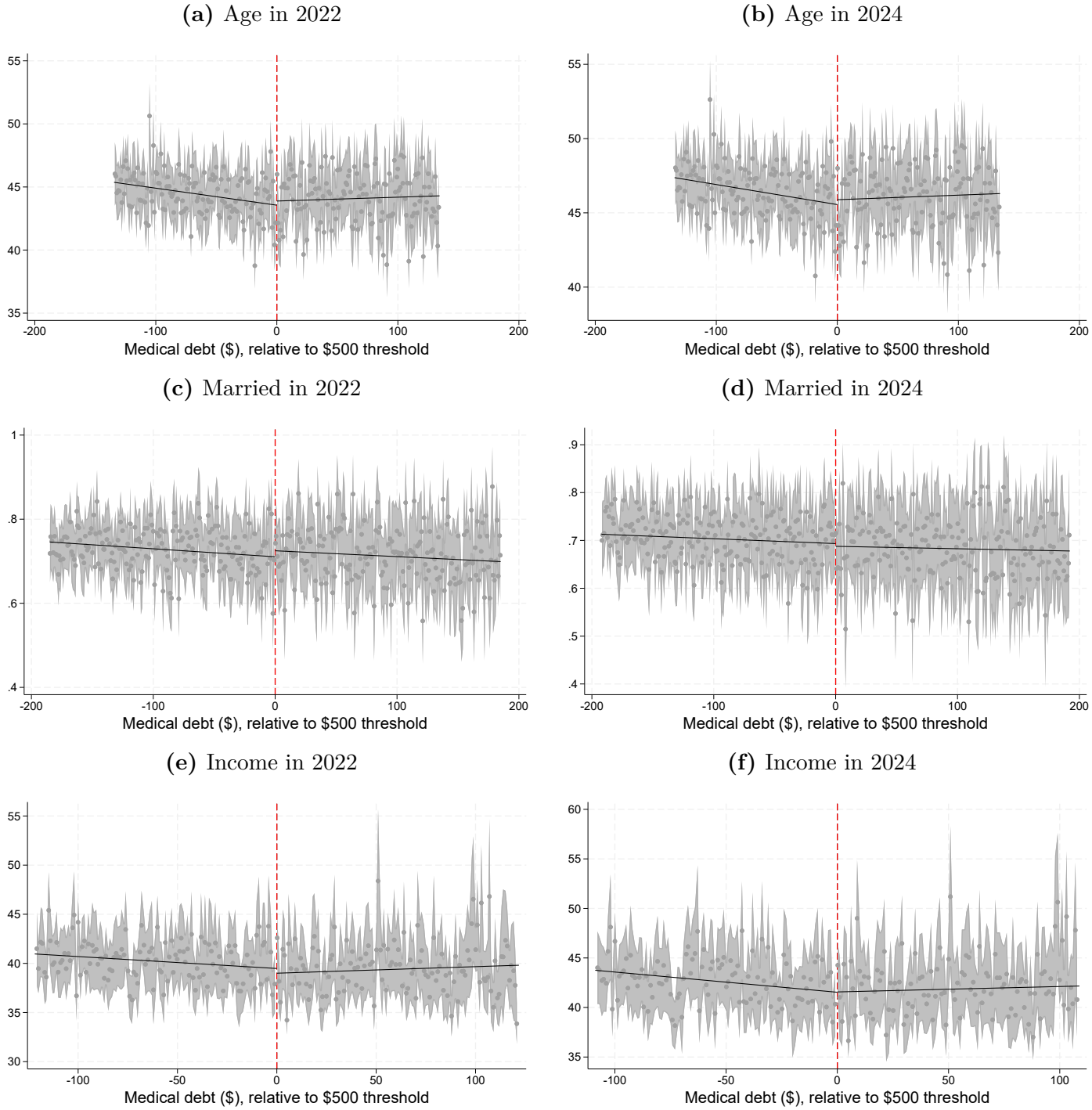
This figure shows coefficients of balancing regressions for a range of covariates, denoted in the y-axis. In each panel, we train credit scoring models with and without a variable and compute the difference in predicted probabilities of default under each model using 2019 data. In panel A, the variable i smedical collections below $500 and, in panel B, the variable is randomly drawn from a uniform distribution. Negatively Treated refers to consumers above the 95th percentile in the distribution of probability differences according to credit scoring models with and without the variable. Positively Treated refers to consumers below the 5th percentile in the distribution of probability differences. The omitted category consists of consumers between the 25th and 75th percentiles. We construct 100 equal-sized bins of the difference in predicted probability of default and cluster standard errors at the bin level.

Figure A.3: Covariate Balance by Changes in Default Probabilities: Logit Model



This figure shows coefficients of balancing regressions for a range of covariates, denoted in the y-axis. It is analogous to Figure 2 but instead of computing predicted probability differences with and without medical collections below $500 according to our baseline credit scoring models, we use logit models instead. Negatively Treated refers to consumers above the 95th percentile in the distribution of probability difference, whose predicted probability of default increases by 2 p.p. or more when medical collections below $500 are removed from our baseline credit scoring model. Positively Treated refers to consumers below the 5th percentile in the distribution of probability difference, whose predicted probability of default decreases by 2 p.p. or more when medical collections below $500 are removed from our baseline credit scoring model. The omitted category consists of consumers between the 25th and 75th percentiles, who experience a change in predicted default probabilities (in absolute value) of 0.0015 p.p. or less. We construct 100 equal-sized bins of the difference in predicted probability of default and cluster standard errors at the bin level.

Figure A.4: Covariates Smoothness



Notes: This figure shows the average Age, Marital Status and Income by medical debt, relative to the $500 threshold. The running variable for medical debt corresponds to the highest debt amount across the consumer's medical collections accounts. The fitted lines are estimated using Equation (4).

Table A.1: Credit Scoring Metrics: Medical vs. Other Variables From Credit Reports

Metric	Baseline	Medical < $500	Medical > $500	All Non-Medical	Record Length	New Credit	90+ Delinquencies
Accuracy	0.9133	0.9132	0.9134	0.9131	0.9131	0.9128	0.9093
Recall	0.4420	0.4418	0.4421	0.4373	0.4407	0.4391	0.4339
Precision	0.7339	0.7330	0.7352	0.7363	0.7328	0.7315	0.7012
F1 Score	0.5517	0.5513	0.5521	0.5487	0.5504	0.5488	0.5361
AUC	0.7100	0.7098	0.7101	0.7079	0.7093	0.7085	0.7043

Notes: This table reports the performance of our two baseline credit scoring models—with and without information of medical collections below $500—and models removing other types of credit information. The first two columns reproduce the results presented in Table 3. The subsequent columns show results when, starting from the full set of 48 features in our baseline model, we alternatively drop: (i) medical collections above $500, (ii) non-medical collection trades and balance, (iii) age of the oldest account and average account age, (iv) age of newest trade and number of inquiries in the past 12 months, (v) 90 days or more past due trades and balances and share of accounts that are 90 days or more past due. Accuracy score is the share of correct predictions. Precision is the proportion of predicted defaults that were correctly classified. Recall is the proportion of actual defaults that were correctly classified. F1 score is a harmonic mean of Precision and Recall. AUC if the area under the Receiver Operating Characteristic curve, which measures the probability that the model will assign a higher default probability to a true defaulter than to a non-defaulter.

Table A.2: Credit Scoring Metrics: Medical vs. a Random Variable

Metric	Baseline	Without Medical < $500	With Random Variable
Accuracy	0.9133	0.9132	0.9131
Recall	0.4420	0.4418	0.4415
Precision	0.7339	0.7330	0.7328
F1 Score	0.5517	0.5513	0.5510
AUC	0.7100	0.7098	0.7097

This table reports the performance of our two baseline credit scoring models—with and without information of medical collections below $500—as well as the performance of a model including a feature randomly drawn from a uniform distribution. Accuracy score is the share of correct predictions. Precision is the proportion of predicted defaults that were correctly classified. Recall is the proportion of actual defaults that were correctly classified. F1 score is a harmonic mean of Precision and Recall. AUC if the area under the Receiver Operating Characteristic curve, which measures the probability that the model will assign a higher default probability to a true defaulter than to a non-defaulter.

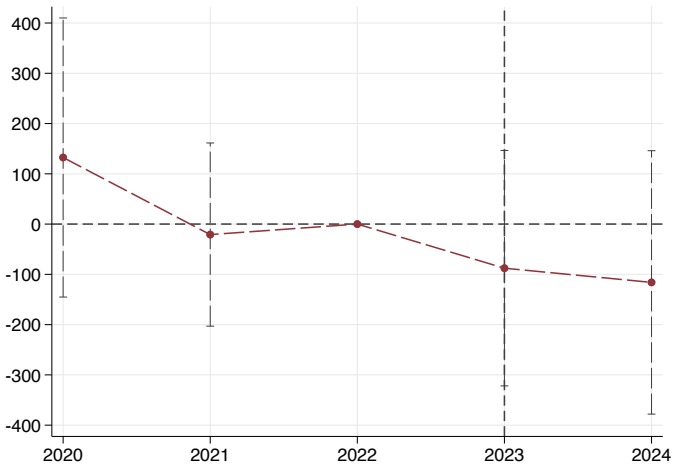
Table A.3: Credit Scoring Metrics: Logit Model

Metric	With Medical < $500	Without Medical < $500
Accuracy	0.9018	0.9019
Recall	0.3430	0.3423
Precision	0.6876	0.6886
F1 Score	0.4577	0.4573
AUC	0.6608	0.6605

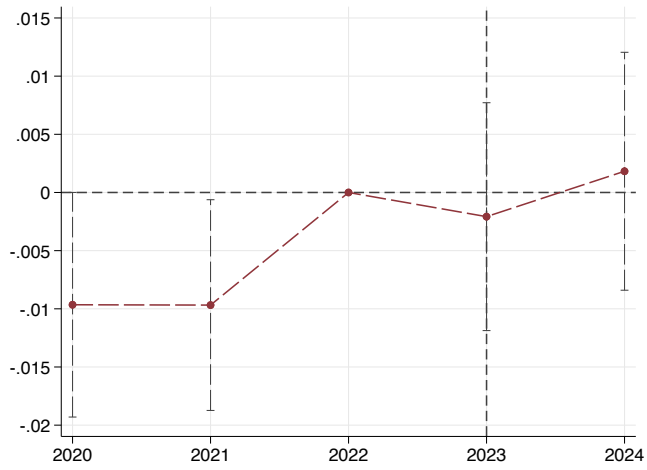
Notes: This table reports the performance of two logit credit scoring models: with and without information of medical collections below $500. Accuracy score is the share of correct predictions. Precision is the proportion of predicted defaults that were correctly classified. Recall is the proportion of actual defaults that were correctly classified. F1 score is a harmonic mean of Precision and Recall. AUC if the area under the Receiver Operating Characteristic curve, which measures the probability that the model will assign a higher default probability to a true defaulter than to a non-defaulter.

Figure A.5: Event Study: Additional Outcomes

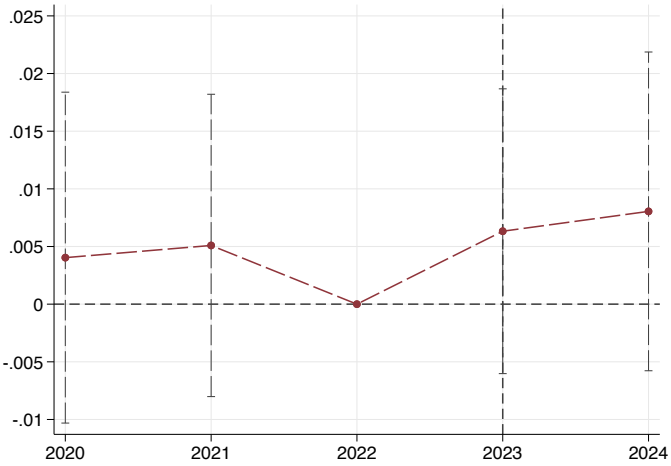
(a) Balance 90 days+ Past Due ($)



(b) Inquiries in Past 6 Months (#)



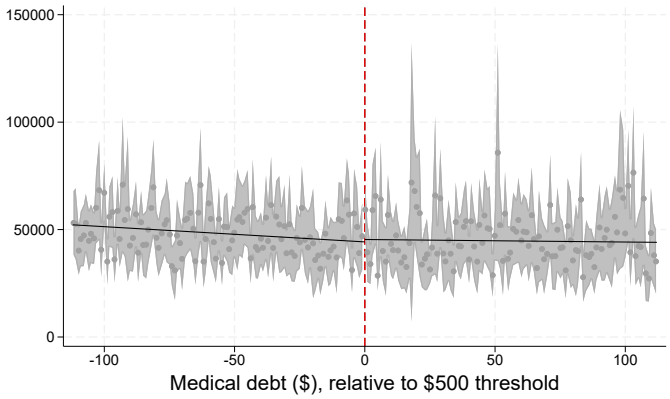
(c) Trades Opened in Past 6 Months (#)



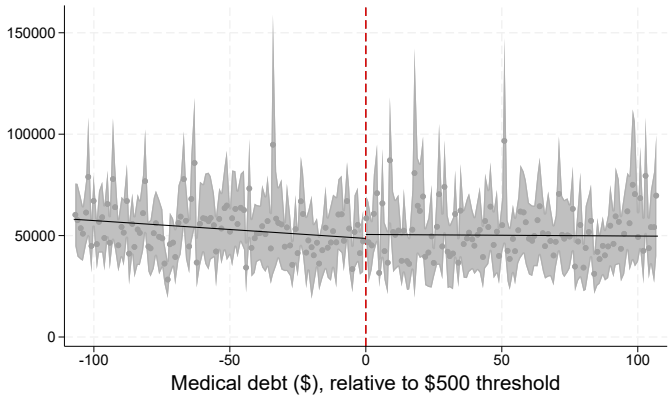
This figure shows coefficient estimates and 95% confidence intervals of Equation (6) for the outcomes denoted in panel captions. Balance 90 days+ Past Due is the balance on trades 90 days or more past due. Inquiries in Past 6 Months is the number of inquiries in the past 6 months. Trades Opened in Past 6 Months is the number of trades opened in the past 6 months. We we create 1,000 equal-sized bins of the difference in predicted probability of default in our full sample and cluster our standard errors at the bin level. Our differences-in-differences sample corresponds to approximately 10% of the full sample, and we are thus left with approximately 100 clusters.

Figure A.6: RD Design: Additional Outcomes

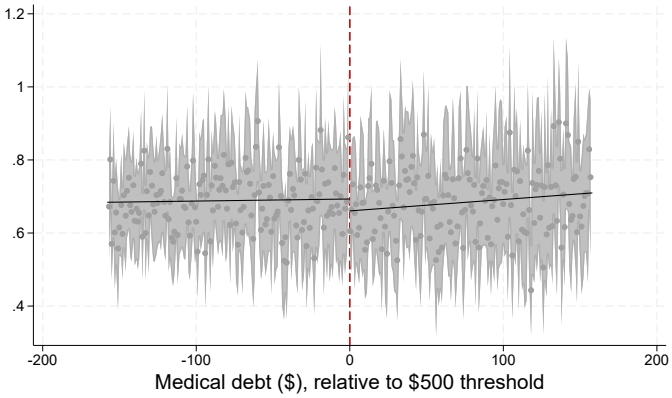
(a) Balance 90 days+ Past Due ($) in 2022



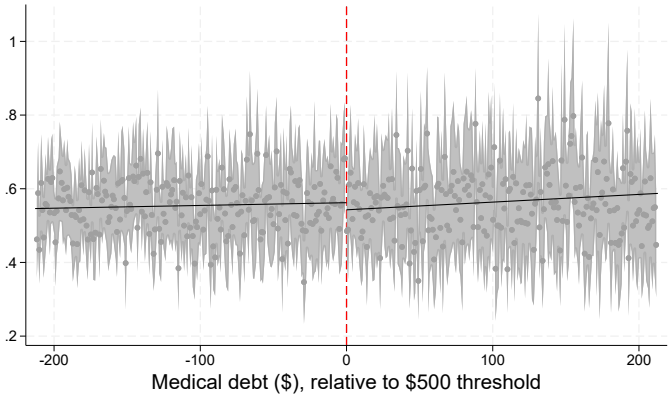
(b) Balance 90 days+ Past Due ($) in 2024



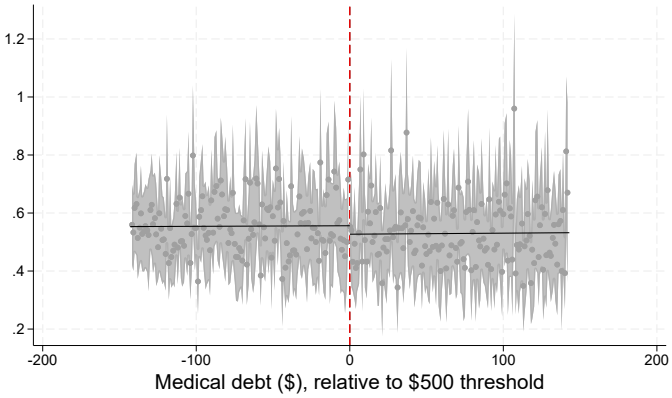
(c) Inquiries in Past 6 Months (#) in 2022



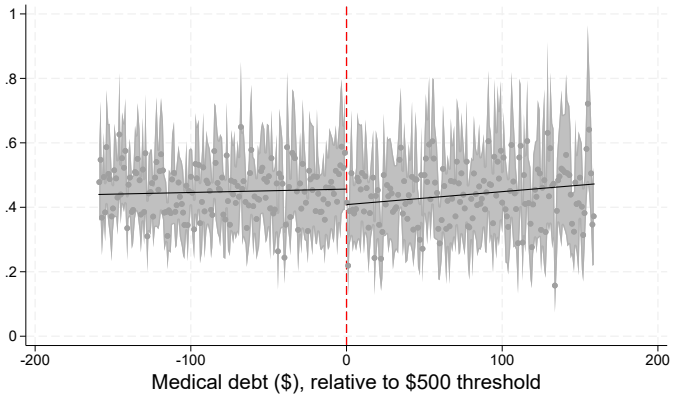
(d) Inquiries in Past 6 Months (#) in 2024



(e) Trades Opened in 6 Months (#) in 2022



(f) Trades Opened in 6 Months (#) in 2024



This figure shows coefficient estimates and 95% confidence intervals of Equation (4) for the outcomes denoted in panel captions. Balance 90 days+ Past Due is the balance on trades 90 days or more past due. Inquiries in Past 6 Months is the number of inquiries in the past 6 months. Trades Opened in Past 6 Months is the number of trades opened in the past 6 months.