KNOW THYSELF: FREE CREDIT REPORTS AND THE RETAIL MORTGAGE MARKET*

Amit Kumar†

[Job Market Paper, Oct. 2021] [Latest draft available here]

Abstract

Borrowers may overestimate or underestimate their creditworthiness in the absence of precise signals. Credit reports, which contain such signals, became free for consumers since 2005 in the entire U.S., while they had already been free in seven states. A difference-in-differences strategy employing this reduction in the cost of reports shows that mortgage applications and approvals increased, and delinquencies and credit-history related rejections decreased. These findings appear consistent with a mechanism in which high-creditworthiness borrowers enter or stay in the applicant pool whereas low-creditworthiness borrowers exit, leading to an improved pool. A supply-driven interpretation appears unlikely as mortgage interest rates increased.

JEL Codes: D12, D83, G21, G28, L51

Keywords: Credit Reports, Information Provision to Consumers, Household Finance, Mortgages, Regulation of Credit Information

Not knowing what's in your credit report can hurt you... It is especially important to review your credit report before you apply for credit ... For large loans, like home mortgages, lenders sometimes ask borrowers questions about information in their credit reports.

—Federal Reserve Bank of Philadelphia, What Your Credit Report Says About You

Not all consumers possess precise information of their creditworthiness, and thus they may take erroneous credit-related decisions. Those who *overestimate* their creditworthiness may apply for credit believing that they are creditworthy. However, during the application assessment process, lenders may uncover true creditworthiness of the applicants from the information in their credit reports and consequently reject their applications. These consumers end up bearing the rejection costs, which include not only the non-refundable application fees (about USD 300 to 400), but also the increased probability of rejection of all future credit applications and potentially increased interest rates. On the other hand, consumers who *underestimate* their creditworthiness may end up not applying for credit, despite needing it, believing that their application will be rejected. Such consumers suffer the opportunity cost of lack of access to credit and are referred to as the *discouraged borrowers*.

Surprisingly, both the over- and under-estimation frictions appear to be prevalent in the U.S. consumer credit markets, potentially because credit reports usage among credit-using consumers is low, at only about 8.4% in the early 2000's (Nott & Welborn, 2003, p.9). From 2000 to 2008, twice as many mortgage applications were rejected for bad credit history than were rejected for a high debt-to-income (DTI) ratio, which is consistent with creditworthiness overestimation, since such consumers could have avoided the costs of rejection by not applying. Similarly, creditworthiness underestimation may underlie the high fraction of consumers who report being discouraged: about 13% in the Survey of Consumer Expectations (2013–2020) (SCE) for mortgages and about 15% in the Survey of Consumer Finances (1998–2007) (SCF) for all credit categories, including mortgages.

This paper asks the following question: Is there a link between these frictions and economic costs to consumers of accessing their credit reports? Specifically, the paper examines the effect of reducing these costs on mortgage market outcomes. The key finding is that approval ratios and mortgage applications increased after the costs were reduced. The reason appears to be an improvement in the borrower pool, as higher creditworthy areas and prime consumers saw more origination and the mortgage delinquency rates subsequently declined.

A simple consumer *self-learning* mechanism is proposed to explain the findings. As credit reports record crucial information on consumers, e.g., their creditworthiness, credit history and borrowing capacity (Figure I), they may aid consumers to self-assess their creditworthiness more accurately. Those whose reports signal high creditworthiness may apply for credit (or enter the market), while those whose reports signal otherwise may either search for a suitable (subprime) lender, or do not apply for credit (exit the market). This sorting process among consumers results in a better pool of potential borrowers, leading to a higher approval ratio. At the same time, the sorting may increase or decrease the demand for credit depending on the prior distribution of over- and under-estimators and on the fraction of consumers who are unaware of the role of the reports in a credit assessment process.

This paper uncovers the causal link between the economic costs of credit reports and mortgage market outcomes using a natural experiment in the U.S.—the enactment of the federal *Fair and Accurate Transaction Act of 2003* (FACTA). Since 2005, the act has allowed access to three free credit reports annually for all consumers through the website www.annualcreditreport.com, whereas seven states—Colorado, Georgia, Maine, Maryland, Massachusetts, New Jersey, and Vermont—had already been allowing their residents free credit reports prior to the enactment of FACTA.[1] Thus consumers from all states except these seven pre-FACTA states saw a close-to-exogenous reduction in economic costs of accessing the reports from 2005 onward. Now they could access reports with just a few clicks, whereas earlier it could take a week for the report to arrive after successfully making a request. The act also appears to have raised general awareness about the reports as interest in free credit reports, measured using Google Search Interest for the keyphrase *"Free Credit Reports"*, heightened at the time of the enactment. All in all, the experiment likely reduced the monetary, awareness and search cost of accessing the reports.

The empirical research design of the paper utilizes the natural experiment in a difference-in-differences (DID) framework. The control group consists of all the pre-FACTA states except Maine, since its local law and FACTA were enacted in the same year, and the treatment group consists of all the states bordering the six control states. In effect, the *late-treated* states are designated as the treatment group, and the *early-treated*, the control group. Furthermore, the

---

[1] The timing of enactment of the local free credit report law and the underlying legislative bill for each of the pre-FACTA states are as follows: CO in 1997 through Senate Bill (S.B.) 133; GA in 1996 through House Bill (H.B.) 1632; MD in 1992 through S.B. 20; NJ in 1997 through Assembly Bill (A.B.) 2787, enacted as New Jersey Fair Credit Reporting Act; MA in 1995 through S.B. 79; VT in 1992 through S.B. 453; and ME in 2003 through H.B. 419.

event year is 2005, sample period spans 2000 to 2008, and the DID estimator is two-way fixed effects (TWFE) estimator.[2]

The empirical design of this paper carefully alleviates the endogeneity in the assignment of the treatment and the control groups. The "treatment" is effected by FACTA, and it occurred for all 50 states in 2005, whereas the "control" is effected by the local state laws, and it occurred only for the pre-FACTA states over a period of time well before 2005. Thus the treated states did not opt to become treated, they were mandated to do so. Moreover, the FACTA enactment in 2003 does not appear to be an endogenous response to the prevailing economic conditions. An existing federal law related to consumer credit, the *Fair Credit Reporting Act of 1970* (FCRA), was set to expire in 2003 (via its amendment in 1996), and it was then that the Congress enacted FACTA to perpetuate the expiring provisions (Nott & Welborn, 2003). Essentially, FACTA was a repackaged FCRA with a key novelty being the annual free credit reports provision, which also assures that other FACTA provisions do not contaminate the treatment effect.

In a bid to separate the confounding effects of the local economic conditions from the treatment effect, the paper restricts the sample to a narrow geographic area consisting of *only* the counties at the border between the treated and control states, similar in spirit to the empirical strategy of Huang (2008) and Dube, Lester, and Reich (2010). Furthermore, the key outcome variables are analyzed at the census tract level, a sub-county micro area that roughly encompasses a population of only ~4,000, allowing any regional economic differences to be reliably swept out.

The null hypothesis is that cheaper credit reports have no effect on the mortgage market outcomes. The key finding is that the lower economic costs resulted in an increase of 1 percentage point in the approval ratio and 13.8%–16.0% in the number of mortgage applications, and the effect is concentrated in owner-occupied mortgages. The equivalent dollar amount aggregated across the treated bordering counties is about USD 5.5 billion due to the former effect and about USD 38.1 billion due to the latter. The increase in approval ratio is consistent with an improvement in the borrower pool under the self-learning mechanism. Furthermore, the higher number of applications indicates that mortgage borrowers on average tend to underes-

---

[2] Since this design is not based on a staggered treatment, but on a one-shot treatment, and since the early-treated states were treated deep in the past, outside the sample period, the framework of Goodman-Bacon (2021) suggests that the negative-weights issue of TWFE highlighted in Borusyak, Jaravel, and Spiess (2021), De Chaisemartin and d'Haultfoeuille (2020), and Sun and Abraham (2020) may not arise. Further discussion on this issue appears in Section (2), Empirical Research Design.

timate their creditworthiness, a contrasting finding to the common belief that consumers tend to overestimate it, and this is explored in detail later.

Next, the trends in mortgage delinquencies across the two areas are analyzed. If an improved borrower pool underlies the increase in origination, the delinquency rate should decrease, or at least not increase. Indeed, relative to the mortgages from the control areas, those from the treated areas originated in the event-year were *less* likely to become delinquent, but those originated in the pre-event year were just as likely to become delinquent.

Examining the heterogeneity in the effects further helps to uncover the borrowers/areas that are likely to benefit from the free reports. *First*, the areas with an *ex-ante* high creditworthiness saw a greater increase in approval ratios and mortgage applications, in line with the idea that the reports aid consumers in assessing their creditworthiness. *Second*, for the lowest-income-quartile borrowers, the approval ratios increased statistically significantly in the treated areas vis-à-vis the control areas, but the number of applications did not. Since lower income is associated with creditworthiness overestimation (Perry, 2008), the self-learning mechanism predicts that the correction for the overestimation for these consumers will result in fewer applications (market exit), leading to what the regression estimates suggest—a rise in approval ratios but not in number of applications.

Considerable indirect evidence suggests that the proposed self-learning mechanism underlies the effects quantified above. *First*, mortgage-related cognizance among the borrowers seems to increase. For example, the treated areas saw a statistically significant decrease in credit-history related mortgage denials in the *ex-ante* high rejection areas but no decrease in DTI-related denials, pointing to an increased learning among borrowers about their credit history. Also, the fraction of total applications withdrawn while in process dropped , indicating a reduced tendency to formally apply to multiple lenders and thereby saving the costs of multiple applications. *Second*, the fraction of mortgages taken by first-time homebuyers increased in the treated areas, consistent with the new entry prediction under the mechanism.

Finally, a host of supplementary findings describe the response of lenders to the event. A supply-driven explanation for the increase in origination seems unlikely, because in the treated areas (i) mortgage interest rates increased; (ii) the high-lenders-density areas did not see more origination or approvals vis-à-vis the low-lender-density areas; and (iii) private securitization of mortgages did not increase. Furthermore, banks with an *ex-ante* larger mortgage origination

in the treated areas saw higher financial performance ex-post, and those with higher *ex-ante* liquidity originated more mortgages.

This paper concludes that reducing economic costs to consumers of accessing their credit reports brings about changes in the mortgage market that are indicative of an improvement in the borrower pool. To a considerable extent, the findings appear causal, and hence any policy that aims at educating consumers of their creditworthiness may yield similar results, not just for the mortgage markets, but for any consumer credit market.

This paper primarily relates to the literature on effects of information provision on credit market participants. This is the first paper to show that lowered consumers' economic costs of accessing credit reports leads to improved mortgage market outcomes in a manner consistent with an improved borrower pool. A field experiment by Homonoff, O'Brien, and Sussman (2019) reveals that borrowers are less likely to default when provided with information on their FICO® scores. Similarly, Mikhed (2015) shows that borrower participation in a free FICO scores program is associated with lower delinquencies, reduced credit utilization, and increased credit card spending. Using a pair of policy changes in Chile, Kulkarni, Truffa, and Iberti (2018) find that increasing disclosure about financial products leads to lower defaults for sophisticated borrowers, and that standardizing financial products leads to lower defaults for unsophisticated borrowers. Also, bankruptcy flag removal from the credit reports raises mortgage borrowing by consumers (Dobbie, Goldsmith-Pinkham, Mahoney, & Song, 2016), and lowers the cost of credit for poorer defaulters and increases it for poorer non-defaulters (Liberman, Neilson, Opazo, & Zimmerman, 2018).

This paper also speaks to the extensive literature on financial literacy. Low financial literacy leads to mortgage delinquencies and foreclosures (Gerardi, Goette, & Meier, 2010), poor mortgage choice (Moore, 2003), large debt (Lusardi & Tufano, 2009; Stango & Zinman, 2009), and lower ability to benefit from loan-modification contracts when in distress (Hundtofte, 2017). On the other hand, educational intervention improves consumers' financial product purchases (Balakina, Balasubramaniam, Dimri, & Sane, 2020). This paper shows that decreasing the economic costs of accessing credit reports for consumers lowers mortgage delinquencies and raises mortgage application approval ratios and mortgage-related cognizance among borrowers.

The rest of this paper is organized as follows. Section (1) describes the U.S. laws related to consumers' access to credit reports, Section (2) presents the research design, and Section (3) describes the data this paper uses. Section (4) discusses the main results, and Section (5)

contains supplementary results that aid interpretation of the main findings. Finally, Section (6) concludes the paper.

1 U.S. Laws Governing Consumers' Access to Credit Reports

Enacted in 1973, the FCRA was the first legislation regulating the information credit reporting agencies collect and the manner in which consumers could access it. The act provided consumers the right to see the contents of their credit reports, except for the credit score, under specific yet restrictive circumstances. Consumers could receive a free report if they made a request within 60 days after receiving a notice of an *adverse action* taken against them on the basis of the information in the report (Nott & Welborn, 2003).[3] An amendment to the FCRA in 1992 further mandated that the cost of disclosure of credit information should be reasonable, and the next amendment in 1996 capped the cost of the disclosure at USD 8 and provisioned that the FCRA would expire in 2003.[4]

It was when the FCRA was to expire that FACTA was enacted, specifically to perpetuate the FCRA's existing provisions while also adding the new annual free credit report provision. FACTA was signed into law on December 4, 2003, and *inter alia* it allowed for free annual disclosure of credit reports to consumers by each of the three national credit reporting agencies through a centralized source. Subsequently, the website—www.annualcreditreport.com—was established in 2005 to distribute the free credit reports.

Notwithstanding the federal regulations on consumer credit reporting, seven states (CO, GA, MA, MD, NJ, and VT) enacted local state laws over 1992 to 2003 that allowed their residents to access free credit reports (see Footnote (1) for details of the enactments). For example, Colorado enacted its free credit report law on April 21, 1997 through Senate Bill 133. Section

---

[3] An adverse action notice can be sent to a consumer by the *user* of a consumer report (e.g. banks, financial institutions, insurance firms) or a debt collection agency affiliated with the CRA stating that the consumer's credit rating may be or has been adversely affected. Under the FCRA, a consumer can also receive credit report free of charge once in 12 months by making a request to the CRA certifying that she/he either: (A) is unemployed and intends to apply for employment in the 60 day period beginning on the date on which the certification is made; (B) is a recipient of public welfare assistance; or (C) has reason to believe that the file at the agency contains inaccurate information due to fraud.

[4] Even though the cost of credit reports was capped under the FCRA and even though consumers could access free credit reports under specific circumstances, only a small fraction of credit-using consumers did so. Out of approximately 1 billion credit reports generated annually, only 1.6% were disclosed to consumers (Avery, Calem, & Canner, 2004). Only 5.25% of these were proactively requested by consumers, while 94.75% were disclosed following the various FCRA provisions (Nott & Welborn, 2003). Thus, only 0.084% of all credit reports were generated as a result of a consumer request.

6

4, paragraph (E) of this bill added the following to Title 12 Article 14.3-104 of the Colorado Statute:

> *(E): Each consumer reporting agency shall, upon request of a consumer, provide the consumer with one disclosure copy of his or her file per year at no charge whether or not the consumer has made the request in response to the notification required in paragraph (a) of this subsection.*

These state laws were likely an endogenous response to the local environment. For example, Vermont was the first state to enact its law, in 1992, because TRW (a credit reporting firm, now Experian) in 1991 *mistakenly* recorded the tax bills of each resident of the town of Norwich and 650 of the residents of the town of Woodstock as property tax liens, due to which these consumers would have been rejected for credit by every lender had they requested it (Associated Press News, December 24, 1992). In the same year, Maryland enacted a similar law.

2 Empirical Research Design

As discussed, this paper uses a DID setting in which six pre-FACTA states—CO, GA, MA, MD, NJ, and VT—constitute the control group and the states bordering these constitute the treatment group. Panel (A) of Figure (II) shows these states on the map of the contiguous U.S. The regression sample focuses only on those counties that lie at the borders between the treated and control states, and Panel (B) of Figure (II) shows these counties on the map of the contiguous U.S. The event year is 2005, the establishment year of the website. The sample period is chosen to be from 2000 to 2008 to allow for sufficient post-event observations.

This empirical design is influenced, but not compromised, by the fact that the enactments of local laws are endogenous. This is because this design does not utilize the staggered adoption by the six control states, but rather relies on the federal enactment of FACTA after the staggered adoption by those states and the fact that FACTA was binding on all the states.

Differences in state regulations on housing and mortgages—e.g., recourse versus non-recourse mortgages (Ghent & Kudlyak, 2011) and judicial versus non-judicial foreclosures (Gerardi, Lambie-Hanson, & Willen, 2013)—may introduce effects that confound the treatment effects. Additionally, credit-related regulations that are enacted within the sample period, such as the adoption of Anti-predatory Lending laws (APL) by twenty U.S. states over 2000 to 2006 (Di Maggio & Kermani, 2017), may further aggravate the issue.[5]

---

[5] Table (A1) in the Online Appendix lists the treated and control states and their status with respect to these regulations. The distribution of these regulatory differences is as follows: 90% of the treated and all the control states

However, the DID design makes the estimates robust to any state-level differences that do not change over the sample period, such as the recourse versus non-recourse mortgages and judicial and non-judicial foreclosures. The confounding effects of staggered regulatory adoption, too, get averaged out in the estimation when the timing of adoption and the states who adopt them are different from the timing of the natural experiment and the treated and control states. This issue is further alleviated by the fact that removing each of the control and associated treated states one at a time yields similar estimates for key outcome variables (discussed in the Results section).

The contiguous-county design across state borders also alleviates the potential confounding effects of local idiosyncratic trends, since such trends are not likely to vary widely across neighboring areas and macroeconomic shocks affect neighboring areas roughly at the same time (Dube, Lester, & Reich, 2016), making this design one of the most compelling identification strategies (Allegretto, Dube, Reich, & Zipperer, 2017).[6] A similar empirical approach has been used in Huang (2008) and Dube et al. (2010).

The DID estimator

The estimator used in the paper is the two-way fixed-effects (TWFE) estimator, specified as

$$Y_{icsjt} = \beta_0 + \beta_1 \times \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt} , \quad (1)$$

where Y_{icsjt} is the outcome variable measured in year t for a census tract i from a county c lying at the border between treatment state s and control state j. t ranges from 2000 to 2008, and j ranges from one to six, corresponding to each of the six control states. $Post_t$ takes value 0 for year $t < 2005$ and value 1 for year $t \geq 2005$. $Treat_{icsj}$ is 0 for all the census tracts i in counties c from control states j, and it is 1 for all the census tracts in counties from the treatment states s. β_1, the coefficient of interest, captures the treatment effect, which is the change in the dependent variable in the treated counties relative to the control counties occurring in the post-event period

---

have recourse mortgages; 45% of the treated and 83.3% of the control states have judicial foreclosures; and 35% of the treated states and 83% of control states adopted the APL laws within the sample period.

[6] In the context of this paper, another viable approach is to use a synthetic control matching procedure. However, this method, too, may place greater weights on *nearby areas* (Allegretto et al., 2017) as it replicates the unobserved counterfactual by taking a weighted average of the observable units. A benefit of a contiguous-county design is that the treatment and control areas map one-to-one to observable geographic areas, enabling tighter links to the real-world data, while a disadvantage is that economic environments across states are inherently different.

relative to the pre-event period. Standard errors are clustered at the county level to provide for correlation in error terms for the observations from census tracts belonging to the same county.

Economic controls in the equation represent the co-variates. They include a host of time-varying county- and state-level variables capturing local economic and credit conditions. These are the number of mortgage lenders (in log) in a census tract and the annual growth rate of county's income per capita, county's aggregate employment and state's gross domestic product (GDP). To ensure that the treatment effects are not influenced by the co-variates, all regressions are estimated both with and without the co-variates.

α_i represents *Census Tract* fixed effects, the first of the two-way fixed effects. These account for any time-invariant differences across census tracts at a highly granular geographic area that encompasses a population of just about 4,000. As census tracts are smaller geographic areas than a county or state, these fixed effects flexibly and fully account for any state-level time-invariant differences, including recourse and non-recourse mortgages, judicial and non-judicial foreclosures etc.

$\gamma_{j,t}$ represents *Border×Year* fixed effects, the second of the two-way fixed effects. Here j refers to the border of a control state j. These fixed effects are formed by the interaction of the border of each of the control states with year, allowing for each region to have its separate time trend (where a region consists of the counties at the border between a given control state and all contiguous states). Thus any regional shock that may affect the regions across different years are flexibly and robustly accounted for.[7]

All in all, the two fixed effects and the time-varying economic controls are expected to reasonably account for confounding effects of local economic conditions on the outcomes of interest, and thus should allow for cleaner estimation of the treatment effects.

Is the TWFE an appropriate estimator for the current DID design?

A key issue with the TWFE estimator is that in *staggered* DID designs, it may aggregate individual treatment effects by assigning "negative weights" to some of them (Borusyak et al., 2021; De Chaisemartin & d'Haultfoeuille, 2020; Sun & Abraham, 2020). Since the estimator is the variance-weighted average of the treatment effects, the negative weights occur in staggered designs when the treatment effects are heterogeneous across time and/or the treated

---

[7] Consider the control state CO. All census tracts from the counties at the border between CO and the surrounding states—WY, UT, AZ, NM, OK, KS, and NE—take the same value for j, and thus are grouped as one unit (region). Thus the fixed effects only utilize the variation *within* each of the six such regions.

units (Goodman-Bacon, 2021). Here the issue of heterogeneous treatment effects across time does not arise, since the current paper uses a *single-treatment* DID design, not a staggered treatment.[8] The other issue of the treatment effect being heterogeneous across treated units remains a noteworthy limitation. However, the results in the paper remain largely supportive of the conclusions, since the key estimates are robust across the sub-samples formed by removing, one at a time, each of the control states and the respective contiguous treated states.

Time-varying co-variates also potentially introduce bias in the estimator (Goodman-Bacon, 2021), but the conclusions of this paper are robust to this issue, as all the estimates are qualitatively and quantitatively similar, *either with or without* the co-variates. Finally, the TWFE also requires random assignment of the treatment. Though the assignment is not carried out randomly, the timing and circumstances of the FACTA enactment are unrelated to the states' actions, and thus the requirement is expected to be largely satisfied.

In the end, the TWFE relies on the parallel-trends assumption: the treated states would have had trends similar to that of the control states in the absence of the treatment. Though the assumption is unverifiable, Panel (A) of Figure (III) plots the trend of the mean approval ratio across the two groups before the event, and they seem to be parallel. Furthermore, Panel (B) of Figure (III) plots the coefficients (β_k) from regression of the *Approval Ratio* according to the following specification:

$$\text{Y}_{icsjt} = \beta_0 + \sum_{k=T-3}^{T-1} \beta_k \, \text{Treat}_{icsj} \times \text{Event}_k + \sum_{k=T+1}^{T+4} \beta_k \, \text{Treat}_{icsj} \times \text{Event}_k + \alpha_i + \gamma_{j,t} + \varepsilon_{icsjt} \,, \quad (2)$$

where *Event$_k$ = 1 if t = T − k, Event$_k$ = 0 if t ≠ T − k, k = {−3, 4}, and T = Event year 2005.* The coefficients $\beta_k's$ represent the difference in approval ratio for the two groups over the years relative to the pre-event year 2004. We see from the plot in Panel (B) that for the most part no significant difference exists between the treated and control census tracts before the event, but the difference becomes significant afterwards. Overall, the two plots in Figure (III) reasonably assure that the parallel-trends assumption largely holds.

---

[8] The staggered adoption of the local laws by the control states over 1992–1997 is not an issue either, since the post-event outcomes are measured at least eight years after the last adoption by a control state. With smaller gaps, the longer-lasting dynamic treatment effects *may* affect the outcome variable in the control states and the later-treated states at the same time, compromising the parallel-trends assumption.

Salience of the natural experiment and other validating assumptions

As it is the free credit reports that shape the outcomes of interest, it needs to be validated that consumers had interest in accessing their credit reports, or alternatively, whether the natural experiment was salient for consumers. The uptick in interest in free credit reports, measured using Search Interest data from Google Trends, strongly suggests so.[9] *First*, the search interest for the key phrase *Free Credit Report* heightened in Jan 2005, coinciding perfectly with the establishment of the website (Panel (A) of Figure (V)). *Second,* plotting the search interest separately for the treated and control states using the Interest-by-subregion data in Panel (B) of Figure (V) reveals that, although the interest in free credit reports was similar in both sets of states in the pre-event year 2004, the interest was more intense in the treated states relative to the control states after the event.[10] Also, anecdotal evidence suggests that almost 52 million credit reports were issued to consumers through the website in just the first two years (Wikipedia, n.d.).

One may wonder why making credit reports free affects the market outcomes given that the monetary cost of credit reports has been trivially small at just about USD 8 historically. This paper argues that the outcomes respond primarily not to the reduction in the monetary cost of the reports *per se*, but to the reduction in the economic costs of *accessing* them. Despite costing so little, prior to the policy, there existed stark differences in the usage of the reports in the states where they were free (control states) and where they were not. Relative to the national average, the usage of credit reports was 250% higher in GA, 204% higher in MD, 153% higher in CO, 35% higher in NJ, and 25% higher in MA (108th Congress of the U.S. Senate, 2004, statement of Senator Bennett, p. 376).[11] Moreover, a mere 0.084% of the approximately 1 billion reports generated annually in the U.S. (in the early 2000's) were proactively requested

---

[9] The Google Trends data represent the degree of "search interest" for a given keyword at any time relative to the highest point during the period of analysis over a given region (U.S.). In the time series, a value of 100 represents the peak popularity for the term. A value of 50 means that the term is half as popular. In the cross-section, a value of 100 represents the location with the highest popularity of the keyword as a fraction of total searches in that location. A value of 50 indicates a location that is half as popular. A score of 0 means there were not enough data for this term. Google Trends data start from January 2004.

[10] An issue with analyzing cross-sectional trends in the search-interest data (interest-by-sub-region data) is that the data values are normalized by Google within the time interval for which the data are extracted. However, this can be overcome by first extracting the data *separately* for each time interval of interest (one-year intervals, in the current case), and then calculating the mean *separately* within each time-interval for each of the two sets of states. From this plot, it may appear that the popularity in the control states after the event decreased. However, this occurs because the popularity measure is essentially a yearly percentile ranking of states, with 100 being the most popular; so an increase in the rank of one state mechanically decreases the rank of others.

[11] These states also seem to have enjoyed better consumer credit environments. The rate of consumer bankruptcies was the lowest (second lowest) nationally in Vermont (Massachusetts) in 2002, and the interest rate on a conventional mortgage in Vermont and Massachusetts was below the national median (108th Congress of the U.S. Senate, 2004, prepared statement of Joel R. Reidenberg, p. 250).

by consumers (Avery et al., 2004), and the fraction of credit-using consumers who requested their credit reports was just 8.4% (Nott & Welborn, 2003). Even if one were to assume that all these credit-report-using consumers had applied for a mortgage in 2004, their applications would make up less than 5% of the total number of mortgage applications.

It is likely the high economic costs that lead to the lack of usage of the reports among consumers. Many factors potentially contribute to the economic costs, for example, search costs (where can the reports be obtained from?), unawareness costs (what role do the reports play in credit/mortgage applications?), and illiteracy costs (what does the information in the reports mean and how should the information be used in credit decisions?).[12]

It is also worth pointing out that even though the credit reports issued under FACTA do not contain the numerical credit score, consumers are not left in the dark about it. In fact, the website actively asks consumers if they wish to retrieve their scores from any of the three CRAs, and provides a link to the corresponding CRA's website for further steps. It is reasonable to expect that the economic costs of accessing credit reports and credit scores dramatically reduced to just a few clicks, which in the earlier system involved calling a CRA to request a report, and waiting for it to arrive through the mail. In fact, in 2000, the three CRAs settled a lawsuit by the Federal Trade Commission (FTC) for blocking calls of millions of consumers who wanted to discuss the content of their credit reports (Federal Trade Commission, January 13, 2000).

Finally, the effects measured in this paper are conservative estimates of the true treatment effect as the nature of the treatment here is intention-to-treat (ITT)—the consumers who wanted to get their credit reports could do that before FACTA as well, albeit at a higher economic costs. These estimates may also be labeled as the *average treated effects on the late-treated* (ATT-LT), because their computation involves comparing the outcomes in the late- and early-treated areas.

3   Data and Summary Statistics

This paper primarily draws on three publicly available datasets: *Home Mortgage Disclosure Act of 1975* dataset (HMDA data) for information on mortgage applications; Federal National Mortgage Agency (Fannie Mae) and the Federal National Home Loan Mortgage Corporation (Freddie Mac) dataset (GSE data) for information on mortgage delinquency performance; and

---

[12] In 2000, the three consumer reporting agencies (CRAs)—Equifax, Experian, and Transunion—blocked calls of millions of consumers who wanted to discuss the content of their credit reports, and ended up settling a lawsuit for this action (Federal Trade Commission, January 13, 2000).

the Call Reports (FFIEC Forms 031/041) for information on financial performance of banks. HMDA dataset provides application-level details on applicants' race and gender, income, loan amount, the financial institution handling the mortgage application, outcome of the application, and geographic location of the property at the census tract level.[13] The GSE dataset contains only mortgages purchased by Fannie Mae and Freddie Mac, and it covers only the 30-year fixed-rate single-family mortgages, the most popular mortgage type in the U.S. The mortgage-level information in these data include the interest rate, DTI ratio, credit score, first-time homebuyer status, investment purpose, and the first three digits of the zip code (zip3) of the mortgaged property. The mortgage performance information include repayment amount and delinquency status, both updated monthly. Finally, Call Reports contain detailed financial information of the U.S. banks.

The detailed steps to process each of these datasets and to link them with one another are provided in Appendix (6), Data Appendix. Mortgages for all purposes and types in the HMDA dataset are included in the sample. The three purposes are—home purchase, refinance, and home improvement, and the three types are—conventional loans, loans guaranteed by Veterans Administration (VA) and Farm Service Agency (FSA)/Rural Housing Administration (RHS), and loans insured by the Federal Housing Administration (FHA). These application-level data are aggregated to the *Census Tract×Year* panel, resulting in 11,942 census tracts that belong to the bordering counties. There are 7,011 treated census tracts, 4,931 control census tracts, and 89,535 *Census Tract×Year* observations. Similarly, the mortgage-level GSE data are aggregated to the zip3-state level, leading to 221 unique zip3-states (91 as control and 130 as treated) and 7,599 *Zip3−State×Quarter* observations.

Some other data are also collected from public sources. Survey data on consumers' credit usage are taken from the Survey of Consumer Expectations (2013–2020) Credit Access Survey, a Federal Reserve Bank of New York rotating panel survey fielded since 2013 over the internet every four months. The data on county-level subprime population come from FRBNY and Equifax (n.d.), on state-level economic conditions from the Bureau of Economic Analysis, on census-tract level population characteristics from Census 2000 (Manson, Schroeder, Van Riper,

---

[13] Until 2003, the census tracts in the HMDA dataset are from the Census 1990 definition, while those from 2004 onward are from the Census 2000 definition. To facilitate the comparison of the tract-level data pre-2003 with post-2003, the census tract-level variables from 2000 to 2003 were scaled using the ratio of population residing in the 1990 tract definition to that in the 2000 definition using data from the Census Bureau (2006). Even though this process is an approximation and introduces some noise in the measurements, it is necessary. The approximation is limited to just 22% of the tracts across the U.S., since 63% of the 1990 census tracts did not see any change across the two censuses and 15% of the 1990 census tracts were wholly combined into various 2000 tracts.

& Ruggles, 2019), and on county-level employment from the annual survey of County Business Patterns (CBP) (Census Bureau, 2000–2008).

The key outcome variables are scaled applications and approval ratio. The scaled applications is the number of mortgage applications per 1000 adults in a census tract, and the approval ratio is the ratio of the number of successful applications (action type "1" *or* "2" in the HMDA dataset) to the number of total applications in a census tract. Similarly, other variables of interest are defined at the census-tract level: the fraction of total applications withdrawn while still under processing and the fraction of total applications denied for credit history or DTI ratio.

Panel (A) of Table (I) shows the summary statistics for the key variables over the sample period. We see that the treated census tracts have fewer scaled applications, lower mortgage approval ratio, and higher denials related to credit history and DTI ratio.[14]

Panel (B) of Table (I) shows the comparison of the treatment and control groups in the pre-treatment period using t-tests for difference in mean, the p-values for which are also shown. Results from the t-tests suggest that the control and treated census tracts differ in pre-treatment years in terms of mortgage-related variables, but are *similar* in the state- and county-level economic characteristics. The similarities in economic characteristics of treated and control counties support the comparison of outcomes across the two groups, whereas the differences in mortgage-related outcomes raise the concern that these groups may also differ on some unobserved characteristics, potentially causing an endogeneity issue. However, since a DID design can accommodate pre-existing differences between the treatment and control subjects so long as they satisfy the *parallel-trends* assumption, the concern is alleviated.

4 Results

This section first analyzes credit report usage and the discouraged borrowers phenomenon using recent survey data, followed by the causal effects of reducing economic costs of accessing credit reports on the mortgage market outcomes. Then results highlighting the heterogeneous effects and self-learning mechanism are presented.

---

[14] We see that the four ratios—the approval ratio, two denial ratios, and withdrawal ratio—do not sum to one. There are three reasons for this. First, the reporting of the reason for denial is not mandatory under HMDA regulations; hence an application may be recorded as denied without any stated reason (70.81% of denied applications have at least one stated denial reason). Second, denials could be for reasons other than credit history or DTI ratio. Third, an application might be denied for multiple reasons.

§A Baseline Results

Survey evidence on Credit Report Usage and Discouraged Borrowers

A representative survey of U.S. consumers, the SCE Credit Access Survey captures the usage of credit reports and scores among consumers and also their planned credit usage. The rotating panel structure of the survey allows for regression analyses that can accommodate fixed effects and clustering of the standard errors at the *Year×Month* level. Furthermore, sampling weights allow one to make inferences that can be generalized to the population. The data used for this analysis span from 2013 to 2020.

Columns (1) through (3) of Table (II) throw light on the average usage of credit scores and credit reports. To do this, an indicator (dummy) variable for a given characteristic is regressed on a constant. In Column (1) the indicator variable is 1 when a respondent has never checked or requested a credit report and 0 otherwise; in Column (2) it is 1 when the respondent has either never checked his or her credit report or checked it at least more than two years ago (infrequent checkers), and it is 0 otherwise; and in Column (3) it is 1 in when respondent does not know his or her credit score and 0 otherwise. The estimated value of the constant in each case is rather startling: an estimated 8% of the population has *never* checked or requested a credit report (Column (1)), about 20% of the population are infrequent checkers of credit reports (Column (2)), and almost 12% of the population does not know own credit score (Column (3)).

Columns (4) through (6) of Table (II) analyze the phenomenon of discouraged borrowers. One of the questions in the survey asks how likely the respondent is to take out a mortgage and related credit in the next 12 months. Those who are very or somewhat unlikely to do so, or those who assign a less than 10% probability to it are asked for the reason. Defining the indicator variable *discouraged borrowers* as 1 for those who respond "I don't think I would get approved", and then regressing it on a constant provides an estimate of the discouraged borrowers. The regression result in Column (4) suggests that among those not planning to take out mortgages, the fraction of those who are doing so because they are discouraged is about 13%. Furthermore, an indicator for discouraged borrowers is regressed separately on the dummy variable for *infrequent checkers* and for *unawareness* about credit score. Columns (5) and (6) show that the coefficient on both the independent variables is positive and significant, i.e. infrequent checkers and those unaware of their credit score are respectively 3% and 5% *more likely* to be discouraged.

15

Taken together, these findings imply that not using credit reports and scores is non-trivially prevalent among retail consumers, and this tendency also contributes to potential borrowers becoming discouraged from applying for credit.

Credit reports and mortgage market outcomes

Now the effects of lower economic costs of accessing credit reports on the mortgage approval ratio, number of applications per 1000 adults (scaled applications) and house prices are examined. To recall, these outcome variables are measured at the census tract level; the regression specification is from Equation (1); and all specifications include *Census Tract* and *Border×Year* fixed effects. The coefficient of interest is *Treat×Post*, which estimates the change in the outcome variable in the treated areas relative to the control after the event.

Columns (1) and (2) of Table (III) show the regression results for approval ratio. The specification in the first column is without any co-variates, whereas that in the second includes controls for local economic conditions, namely, the number of HMDA lenders (in log) in a census tract, and annual growth rates of county income per capita, county aggregate employment, and state GDP. Coefficients on *Treat×Post* suggest that the ratio increased by about 1 percentage point in the treated tracts relative to the control. In dollar terms, keeping the number of applications in the treated areas at the pre-event level, a 1 percentage point increase in approval ratio corresponds to about USD 2.75 billion more successful mortgages, aggregated across the treated bordering counties.[15] The effect at first may seem trivial, especially since approval ratios are commonly believed to be high, at upwards of 80%, but it is just about 52% in the treated tracts in the pre-event period.

Approval ratio likely improved because borrowers were better informed of their creditworthiness, as the event did not affect other aspects of the mortgage process. Specifically, the event did not target either access to the reports for lenders or the content of the reports. The event also did not have any provisions that would affect any borrower-related characteristics such as their income, employment, or collateralizability of their assets. What may explain the increase in approvals is the self-learning mechanism. Borrowers learn their creditworthiness information from the reports, leading to better matching in the market—new entry of creditworthy borrow-

---

[15] A 1 percentage point increase in approval ratio is equivalent to ~2.6 more successful applications per treated tract (96.27 applications per 1000 adults in the pre-treatment period × 0.01 × 2.7 thousand adults per treated tract), about 18,229 more successful applications across the treated bordering counties (2.6 applications × 7,011 treated tracts), or a ~USD 2.75 billion increase in mortgage origination across *all treated tracts from bordering counties* (18,229 × USD 150,597 average mortgage amount per application).

ers and either exit or matching to subprime lenders of borrowers with bad creditworthiness. The mechanism is evaluated rigorously in later sections, including whether the increase was lender driven.

The regression results for scaled applications appear in Columns (3) and (4) of Table (III). This rose in the treated tracts relative to the control by 13.4–16.6, a 13.9–17.2% increase over the pre-treatment average of 96.3. In real terms, keeping the approval ratio in the treated areas at the pre-event level, the increase roughly translates to a USD 37.8 billion increase in mortgages, aggregated across the treated bordering counties.[16] The increase in applications indicates that on average consumers tend to underestimate their creditworthiness when it comes to mortgage borrowing. This interpretation is in contrast to other financial decision-making settings and masks important heterogeneity that is explored in detail later.

To ensure that these results are not driven by some specific states, the above regressions are re-estimated over sub-samples formed by removing each of the control states and their respective contiguous treated states one at a time. The coefficients from regressions similar to Equation (1) with all controls included are plotted in Panel (A) and (B) of Figure (IV) for scaled applications and approval ratio, respectively. We see that the estimates are mostly similar across all the sub-samples.

Coefficients on $Treat \times Post$ in Columns (5) and (6) quantify the changes in the growth rate of house prices. The regressions use the census tract-level house price index from Bogin, Doerner, and Larson (2016) that starts in 2000. The coefficients suggest that the growth rate of house prices in the treated areas increased relative to the control by 1.7–1.8 percentage points after the event, though the estimates are statistically significant only at the 10% level. This finding is in line with that of Di Maggio and Kermani (2017): in their sample, house price growth rates increased by 3.3 percentage points following a 10% increase in mortgage origination.

A noticeable limitation of the estimates above is that the mortgage supply in the U.S. had started to shrink from 2005, an antecedent of the financial crisis of 2008, and the post-event regression sample includes the years from 2005 to 2008. Hence, while it may be injudicious to claim that the effects estimated above are completely uncontaminated by these changes, the DID design ameliorates the issue to the extent that the market-wide forces evenly affect the

---

[16] The average mortgage size in treated tracts in the pre-treatment period was about USD 150,597. Thus the demand for mortgage credit increased by about USD 2.0 million per 1000 adults per census tract (USD 150,597 × 13.4), by about USD 5.4 million per treated census tract (USD 2 million × 2.7 thousand adults per census tract), or by about USD 37.8 billion across *all treated tracts from bordering counties* (USD 5.4 million × 7,011 treated tracts).

neighboring counties across the state borders. In addition, the effects remain qualitatively and quantitatively similar when estimated in an alternative sample which restricts the post-event period to 2006 (see Section (§E), Robustness).

The financial crisis is also often argued to be a result of excessive mortgages taken by borrowers without means, often for investment purposes rather than for occupancy purposes, and this raises the question of whether the increase in the origination reported above is also driven by such borrowers. Table (IV) examines this assertion using the same DID specification. The outcome variable is scaled applications for the owner-occupied mortgages in Columns (1) and (2) and for the non-owner-occupied category in Columns (3) and (4). The coefficients on *Treat×Post* suggest that the applications increased dramatically and significantly only for the owner-occupied category in the treated areas vis-à-vis the control, but not for the latter category. The assertion is further investigated by examining whether there was a compositional change across the two mortgage categories. Columns (5) and (6) examine the scaled applications in the non-owner-occupied category as a fraction of total applications, and Columns (7) and (8) examine the same as a fraction of successful applications. The coefficients in these four columns indicate a modest 1 percentage point increase in non-occupancy mortgages at both the application and origination stage. By and large, the investment-motivated demand does show a slight uptick, but does not appear to be the dominant reason behind the robust 15% increase in mortgage applications.

Mortgage delinquencies

Previous results imply that lowering the economic costs of accessing credit reports for consumers leads to higher mortgage origination. The question that arises then is, whether mortgage delinquencies too would increase as a result? If the origination increased owing to an improved borrower pool, the delinquencies would fall, or at least not rise. However, if the origination increased due to subprime lending while the borrower pool stayed the same as before, the delinquencies would rise subsequently.

To examine the patterns in the delinquencies, the GSE data, which are a subset of the HMDA mortgages, are used. First, a mortgage vintage is defined as the collection of the mortgages originated in a given area—treated or control—in a given year—2004 (pre-event) or 2005 (post-event), leading to four vintages: the treated vintage in the pre- and post-event year, and the control vintage in the pre- and post-event year. Then, the rate of delinquency of a given vintage

is defined as the ratio of the number of mortgages late on a scheduled payment by n days for the first time at a given age (measured in months since origination) to the total number of mortgages in that vintage. The rates are analyzed for delays of n= 30–89 days and 90–120 days.

Panel (A) of Figure (VI) shows the 30–59-day delinquency rate for the treated and control vintages for the year 2004 on the left-hand side and for the year 2005 on the right-hand side. The plot on the left reveals that, among the mortgages originated before the event, the delinquency rates of the treated and control vintages follow almost the same trend; whereas the plot on the right reveals that, among the mortgages originated in the year of the event (2005), the delinquency rate of the treated vintage is *lower* than that of the control vintage. The same pattern is observed for the 60–89-day delinquency rate, plotted in Panel (B) of Figure (VI). Furthermore, each of the delinquency rates of the treated vintage becomes much lower than that of the control vintage during the financial crisis (48 months after 2004, or 36 months after 2005) than during the earlier period.

These reductions may at first appear puzzling, because if lenders use the same screening policy after the event as they were using before the event, one may predict the delinquency rates to stay the same. However, this prediction implicitly assumes that the composition of the pool stays the same, whereas the self-learning mechanism predicts that the composition of the pool may change—new creditworthy borrowers may enter the market and those with poor credit-worthiness may exit—and Section (§C.2) provides an evidence consistent with such a change—the proportion of the first-time homebuyers in the originated mortgage pool increased.[17]

Overall, the reduction in the delinquency rates after the event suggests that the borrower pool improved, and one reason for this appears to be an increase in the share of first-time home-buyers.

§B   Characterizing the Effect: Who benefits?

Characterization of the consumers and the areas that are more likely to benefit from easier access to credit reports may provide insights about those for whom the information frictions about creditworthiness are likely to be binding, and it may also be useful for policy targeting.

---

[17]Suppose that lenders deny applicants whose *ex-ante* probability of default falls above some threshold, p^*. Assume that the average delinquency rate of originated loans in the pre-event period is p_1, where $0 \leq p_1 \leq p^*$. After the free credit report policy is implemented, suppose that an additional pool of applicants is motivated to request a mortgage, and they are subject to the same upper bound, p^*, but their delinquency rate is p_2, where $0 \leq p_2 \leq p^*$. It is clear that depending on the values of p_1 and p_2, the average delinquency rate after the event may increase or decrease, given that the delinquency rate of the new entrants is different from that of the older pool.

The heterogeneity in the treatment effect across consumer creditworthiness and income is examined next.

§B.1   Heterogeneous effects by creditworthiness of borrowers

Given that creditworthy borrowers are more likely to be granted mortgage credit, if credit reports aid consumers in assessing creditworthiness, easier access to the reports should lead to a *greater* increase in applications and approvals in areas where the fraction of creditworthy consumers was higher before the event. This is because, under the self-learning mechanism, the exit would be larger and the entry smaller in the subprime areas vis-à-vis the prime areas.

To test this, a county is classified as having high creditworthiness if its subprime population fraction is less than the *regional mean* before the event.[18] The year 1999 is chosen as the classification year because Mian and Sufi (2009) suggest that such classification should be done at a time well before the start of a housing boom, as creditworthiness of an area may endogenously evolve with the boom. The earliest year for which data on the county subprime fraction is publicly available is 1999 (FRBNY & Equifax, n.d.).

Table (V) shows the results of regressing scaled applications and approval ratio separately using regression Equation (1) for counties with high and low creditworthiness. Within the *ex-ante* high-creditworthiness counties, vis-à-vis the control counties, the treated counties saw an increase of 16.8–18.8 (17.4–19.5%) in scaled applications (Columns (1) and (2)) and a 2 percentage points increase in approval ratio (Columns (3) and (4)). At the same time, within the *ex-ante* low-creditworthiness counties, vis-à-vis the control counties, the treated counties saw an increase of just 8.59–11.66 (8.8%–12%) in scaled applications (Columns (5) and (6)) and a 1 percentage point increase in approval ratio (Columns (7) and (8)), which is statistically significant only at 10%. Taken together, these estimates support the self-learning mechanism and suggest that creditworthy borrowers are more likely to benefit from easier access to credit reports.

---

[18] The steps to calculate *regional mean* are as follows. A region is defined as the area encompassing a control (pre-FACTA) state and all the surrounding states. Consider the control state Colorado (CO) and all the surrounding treatment states. The regional mean for this region is the average rejection rate for the census tracts in all the counties at the border between CO and WY, UT, AZ, NM, OK, KS and NE. The regional means of rejection rates for all seven control states are calculated in this way, and a census tract is then classified as a "high rejection tract" if its rejection rate is more than the regional mean in 2004.

§B.2 Heterogeneous effects by income level of borrowers

The treatment effect may vary across borrowers of different incomes, because the consequences of a mortgage rejection are more severe for low-income borrowers. Thus, upon learning one's creditworthiness, the likelihood of not applying for credit (exiting the market) or gravitating to subprime lenders is higher for low-income borrowers as opposed to high-income borrowers. Alternatively, the over- and under-estimation tendencies under the self-learning mechanism also lead to the same prediction. As lower income is associated with a higher likelihood to overestimate one's creditworthiness (Perry, 2008, Table III), low-income consumers are more likely to revise their creditworthiness downwards after learning their true creditworthiness from their credit reports, and thus are more likely to avoid rejection costs by not applying for credit. The approval ratio for this sub-group of consumers is thus likely to increase owing to the exit channel of the self-learning mechanism, whereas it is unclear how the ratio would change for high-income consumers.

To test these predictions, first, the cut-offs for the income quartiles are calculated each year within the sample, and then the applications from each quartile are aggregated to the census-tract level and scaled by the population (measured in 1000's). The approval ratio is then calculated within each quartile.

Panel (A) of Table (VI) shows the results of regressing the scaled applications separately for each of the income quartiles using Equation (1). We see that the scaled applications did not increase significantly for the lowest quartile, but increased significantly for the other three, and the increase was larger and statistically significant for these quartiles. The significant increase in applications among the higher quartiles but not among the lowest-quartile consumers is consistent with the prediction that it is the latter for whom exits are more likely.

Panel (B) of Table (VI) shows the results of regressing approval ratios separately for each of the income quartiles using Equation (1). We see that the ratios increased statistically significantly in the treated areas relative to the control areas only for the lowest income quartile borrowers, consistent with the prediction. The high-income borrowers did not see an increase in approvals likely because the marginal propensity of lending to such consumers is larger than that of lending to low-income consumers (Agarwal, Chomsisengphet, Mahoney, & Stroebel, 2018), thereby such consumers are not the ones who are excluded from the credit markets. The absence also re-emphasizes that the value of accessing credit reports is relatively higher for the low-income borrowers.

21

§B.3   Heterogeneous effects for the overestimating borrowers

While the self-learning mechanism predicts that the number of mortgage applications after the event should decline for the overestimating borrower type and increase for the underestimating type, empirically testing them is challenging. First, the two borrower types are not distinguishable in the mortgage application or performance data. Second, to the extent that the predictions are true, the mortgage data would not capture the underestimating type, as this type is predicted to *not* apply for credit (mortgage).

Notwithstanding the above limitations, it may be argued that the borrowers from the areas where the *ex-ante* rejection ratios due to DTI ratio were small but due to the credit history were large are more likely to be the over-estimators, relative to the borrowers from other areas. This is because these rejection patterns fit the borrowers who mistakenly overestimate their creditworthiness, apply for a mortgage, and thus are more likely to be rejected for (bad) credit history than for their repayment inability (high DTI). Following this reasoning, the census tracts in the sample are sorted into *tertiles* of the rejection ratios in the pre-event year 2004 for the DTI ratio and credit history, leading to nine sub-groups.

Table (VII) shows the results of separately regressing the scaled applications and approval ratios using Equation (1) for each of the nine sub-groups. The regression results for the scaled applications appear in Columns (1) through (3) and for the approval ratio in Columns (4) through (6). Credit-history tertiles vary from the top to bottom of the table and DTI tertiles, from left to right. The overestimating sub-group corresponds to the third tertile of the credit history and the first tertile of the DTI ratio. We see that the treated areas saw the smallest increase in scaled applications relative to the control areas for the overestimating borrowers (6.23 versus 10.67 or 20.32 within the first DTI tertile) than the other borrowers. This pattern is in line with the prediction for the overestimating borrowers.

The prediction regarding how the approval ratio would change for the overestimating and underestimating type is ambiguous. The ratio should increase for *both* types, but the size of the increase depends on the *ex-ante* relative proportion of the two types in the applicant pool. In Column (4) of Table (VII) we see that the amount of increase in the approval ratio is indeed similar across the three credit-history tertiles, though it is not significant for the third tertile.

To summarize, the prediction that after the event the number of applications for the overestimating borrower type should decrease (or should increase less than the others) is largely supported in the data.

22

§C Mechanism: Consumer Self-learning Channel

The findings discussed so far document how the mortgage market outcomes changed in response to the reduced economic costs to consumers of accessing their credit reports, and the heterogeneity tests yielded results consistent with the self-learning mechanism. This section investigates the mechanism more directly by examining the changes in mortgage-related cognizance among borrowers and first-time homebuyers.

§C.1 Increase in the mortgage-related cognizance among borrowers

If consumers learn more from their credit reports about their creditworthiness after the event, their decision regarding credit and mortgages will reflect it. In particular, as the reports contain the credit history of consumers, their cognizance of about it should increase. Consumers may thus be able to reduce the likelihood of rejections due to credit history by taking actions such as steps to improve the record before applying for credit or by applying to subprime lenders, who specialize in providing credit to those with poor credit history. At the same time, the likelihood of rejections due to DTI ratio may not change, as it is unlikely that consumers could boost their income strategically before applying for credit.

Similarly, an increase in cognizance would also affect applicants' tendency to withdraw mortgage applications that are still being processed (before the lender has made the decision). It is common for potential applicants to initiate several formal mortgage applications at once at different lenders to hedge against the uncertainty in approvals and mortgage terms. In doing so, they incur multiple non-refundable application costs, but in the end they take out a mortgage with only one lender and withdraw their applications from the others (in-process withdrawals). With an increase in cognizance of their creditworthiness, borrowers' uncertainty over approvals and credit terms decreases, and with that, they are likely to apply to *fewer* lenders at once. Thus the fraction of in-process withdrawals should decrease in the treated areas.[19]

The first prediction is tested by regressing the fraction of total applications rejected for credit history and for DTI ratio. These outcomes are estimated separately for the entire sample and for a sub-sample of only those census tracts where the rate in the pre-event year 2004 was higher

---

[19] The withdrawal ratio over the 2000–2008 period is about 12%, indicating that in-process withdrawals are fairly common. Anecdotal evidence suggests that consumers tend to withdraw applications when they find a better offer from other lenders (Reddit Forum, n.d.). More importantly, credit reporting agencies do not penalize multiple applications if they are made within a short time period, as Equifax (n.d.) explains: "If you're shopping for a new auto or mortgage loan or a new utility provider, the multiple inquiries are generally counted as one for a given period of time. The length of this period may vary depending on the credit scoring model used, but it's typically from 14 to 45 days. This allows you to check at different lenders."

than the *regional mean* (Footnote (18) shows the steps to calculate the mean). The reasons to separately focus on the *ex-ante* high-rejection areas are that the information in the reports are more valuable when the rejection rates are high, and the influence of the event on reason-specific rejection probabilities will be greater in the areas where rejections were frequent before the event. The second prediction is tested by regressing the withdrawal ratio, which is the fraction of total applications that are formally withdrawn by borrowers before lenders could make a decision.

These predictions are tested using the regression specification from Equation (1), and Table (VIII) shows the results. In Columns (1) through (4) we see that the fraction of applications denied due to credit history decreased by 0.3 percentage points in the treated tracts relative to the control, statistically significant in the *ex-ante* high-rejection-rate areas (Columns (3) and (4)). The coefficients in columns (5) through (8) show that the DTI ratio denials did not decrease statistically significantly. Though the estimates carry only modest statistical significance, they indicate that the reason-specific rejection likelihoods changed in a manner consistent with potential borrowers becoming more cognizant of their credit history.[20] The estimates for the withdrawal ratio appear in Columns (9) and (10) and imply that it decreased by 0.9–0.11 percentage points in the treated tracts vis-à-vis the control.[21]

Overall, these findings point to an increase in mortgage-related cognizance among borrowers, consistent with the self-learning mechanism.

§C.2   New entry: Increase in first-time homebuyers

About 15% of households in the SCF survey from 1998 to 2007 and 13% of the respondents in the SCE surveys from 2013 to 2020 report as being *discouraged*, and an analysis of the latter survey in Table (II) reveals that this tendency is associated with a lack of knowledge and usage of credit reports/scores. With the economic costs of accessing credit reports becoming lower, to the extent that consumers underestimate creditworthiness and do not apply for credit when the costs are higher, there would be an *entry* of creditworthy borrowers into the markets.

---

[20] A caveat of this analysis is that HMDA does not mandate lenders to report reasons for rejections, so if the reporting incentives of lenders were also influenced by the event, the estimates reported above would be the result of the changes in borrowers' cognizance and lenders' incentives. However, the incentives to report rejection reasons would need to change in the event year in a particular manner that varies across the treated and control areas, even for lenders that may operate in both areas. Such precise changes in incentives for reporting the reasons for rejections appear unlikely. Moreover, lenders reported reasons for rejection in 70.81% of the rejected applications in the sample.

[21] In economic terms, the drop is equivalent to ~2.34 fewer in-process withdrawals per treated tract or ~16,513 fewer withdrawn applications aggregated over the treated border counties. At an average cost of ~USD 400 per withdrawn application, this represents ~USD 6.6 million saving in upfront mortgage application fees.

The entry of new borrowers can be investigated empirically by examining the share of first-time homebuyers among the mortgages. Specifically, the outcome variable is defined as the ratio of the number of mortgages taken out by first-time homebuyers to the number of all originated mortgages that had known information on first-time homebuyer status. It is important to enumerate two limitations of this analysis. Whether an applicant is a first-time homebuyer is recorded only in the GSE data, not in the HMDA data. Also, since the property location information in the GSE data is only available at the 3-digit zip code level, the properties were mapped to the counties using simplified approximations (see Footnote (28) in Data Appendix). Accordingly, the regression is specified at the zip3-state level as follows:

$$Y_{zsjt} = \beta_0 + \beta_1 \times \text{Treat}_{zsj} \times \text{Post}_t + \delta \times \text{Economic controls} + \alpha_{zs} + \gamma_{jt} + \varepsilon_{zsjt} . \quad (3)$$

Here, z indexes the areas delineated by a 3-digit zip code at the border of treated state s and control state j. α_{zs} is zip3-state fixed effects. γ_{jt} is the *Border×Quarter* fixed effects, and it serves similar function as that of the *Border×Year* fixed effects in Equation (1). The sample is limited to the zip3-state areas that come under the border counties of treated and control states.

Columns (1) and (2) of Table (IX) show the regression results. The coefficients suggest that the percentage of first-time homebuyers increased by 1 percentage point in the treatment areas relative to the control areas.[22] This finding is in line with the prediction that, following the event, more new entries should occur in the treated areas.[23]

§D A Demand- or Supply-side Effect?

The findings so far indicate that the likely explanation for the increase in origination in the treated areas relative to the control after the event is the self-learning mechanism, which is a demand-driven channel. However, a supply-driven explanation is also plausible. Even though the natural experiment reduced the economic costs of accessing credit reports only for the consumers, it also likely updated lenders' knowledge that the consumers' access to the reports had

---

[22] About 6.7% of the observations within the hombebuyer data sample pertaining to the bordering counties do not have information on first-time homebuyer status. The specifications that alternatively define the outcome variable as the ratio of number of first-time homebuyers to *all mortgages* yield similar estimates, and these estimates are left unreported for brevity.

[23] A concern with this estimation is that the mortgage sample consists of only those that were purchased by the GSEs. However, as argued before, this selection would be an issue only if GSEs' incentives to purchase first-time homebuyer mortgages relative to their overall purchase from the treated counties increased relative to the control counties from the event year 2005 onward. Such a time- and location-specific change seems improbable.

become free. With this updated knowledge, lenders could respond by increasing the mortgage supply.

Even though the supply-driven explanation is plausible, many of the earlier findings favor the demand-driven explanation. First, an increase in applications and a decrease in in-process-applications withdrawals are a result of decisions that are taken solely by potential borrowers, and these quantities are mostly independent of lenders' influence. Second, it is the demand-driven mechanism under which the effects would be heterogeneous across consumer characteristics, as it was across creditworthiness and income. Additionally, given that the propensity of lenders to extend credit to low-income borrowers is low (Agarwal et al., 2018), and given that in the current setting we see that the approval ratio increased significantly for such borrowers vis-à-vis the high-income borrowers, the supply-driven explanation appears unlikely.

Notwithstanding the above suggestive evidence favoring the demand-side explanation, two outcomes more directly related to supply-side characteristics are examined next: mortgage interest rates and heterogeneous effect by *ex-ante* density of mortgage lenders.

§D.1 Mortgage interest rates on the GSE-repurchased mortgages

The changes in the mortgage interest rates after the event in treated and control areas can be utilized to examine whether the increase in origination was supply- or demand-driven. If it is the former, the rates would decrease; if the latter, they would increase.

The investigation of the rates needs to account for characteristics of the property and borrower-risk. Scharfstein and Sunderam (2016) argue that the prices (interest rate) at which lenders sell *conforming* mortgages to the GSEs materially vary only across three dimensions: credit score, loan-to-value ratio (LTV), and loan type (adjustable rate, fixed rate etc.).[24] Therefore, the residuals obtained from a regression of the rate on these dimensions approximately measure the lender-specific pricing schedule independent of the characteristics of borrowers and the mortgage. In the context of the current paper, only the first two dimensions are relevant, since the GSE sample only includes one type of mortgage—the 30-year fixed-rate single-family mortgage. Thus, when the rate is regressed on credit score, LTV, and the *Treat×Post* interaction term, the coefficient on the last term captures the change in the pricing schedule of lenders in the treated areas vis-à-vis the control areas after the event.

---

[24]Fannie Mae's mortgage pricing variation across these dimensions can be seen in its pricing schedule here: https://singlefamily.fanniemae.com/media/document/pdf/llpa-matrix-pdf

If the lenders lowered the mortgage interest rates in the treated areas (in a bid to increase mortgage origination), the sign on the coefficient would be negative; if they raised the rates, the sign would be positive.

The regression specification is similar to Equation (3), but it is now specified at the loan level i as follows:

$$\text{Interest Rate}_{izsjt} = \beta_0 + \beta_1 \times \text{Treat}_{izsj} \times \text{Post}_t + \delta \times \text{Controls} + \alpha_{zs} + \gamma_{jt} + \varepsilon_{izsjt} \ . \qquad (4)$$

Columns (3) and (4) of Table (IX) show the results of the regression. In Column (3), *Controls* are the two relevant pricing dimensions, credit score and CLTV (combined loan-to-value: the loan-to-value ratio inclusive of all loans secured by a mortgaged property); and in Column (4), to make the specification more rigorous, *Controls* additionally include DTI ratio, number of units comprising the mortgaged property, and percentage of mortgage insurance. Results show that the coefficients on *Treat×Post* in both the columns are positive and significant, at about 0.009–0.01 percentage points. Thus, if anything, lenders responded to the event by increasing, not decreasing, the risk-adjusted mortgage interest rates in the treated areas relative to the control, contrary to what a supply-driven explanation would predict.[25]

§D.2 Heterogeneous effects by *ex-ante* density of mortgage lenders

If the increase in mortgage origination were driven by lenders, it would be greater in areas where the *ex-ante* density of lenders is high. To examine this, first, census tracts are classified into high and low lender-density groups: high if the number of HMDA mortgage lenders per adult in the pre-event year 2004 in a census tract was greater than the *regional mean* (defined in Footnote (18)), and low otherwise.

Columns (1) through (4) of Table (X) show the results of separately regressing, using Equation (1), the dollar origination volume (in 1000 USD) per adult for the two density groups. The estimates are smaller in magnitude and have weaker statistical significance for high-density tracts (Columns (2) and (4)) vis-à-vis the respective low-density census tracts (Columns (1) and (3), respectively). Thus, among the *ex-ante* high-lender-density census tracts, relative to the control tracts, the treated tracts saw a much *smaller* increase than the treated tracts from the

---

[25]The magnitude of the increase in the rates is tiny, potentially for two reasons. First, interest rates on conforming (GSE-repurchased) loans do not vary across regions or with dimensions other than FICO scores, loan-to-value ratio, and loan type (Hurst, Keys, Seru, & Vavra, 2016). Second, the supply of mortgages in the U.S. is highly elastic because of the large-scale purchases by the GSEs of conforming mortgages in the secondary market.

low-lender-density group relative to the respective control tracts. Moreover, the t-test for the difference in the coefficient of *Treat×Post* in high- and low- lender-density areas (*High − Low*) shows no statistical difference.

Columns (4) through (8) of the table show the above regressions for the approval ratio. The results are similar—there is no statistical difference in the increase in the approval ratio in areas with a high or low lender density.

The finding that the effects were stronger, not in the areas that had *ex-ante* high lender density, but in the areas that had an *ex-ante* low lender density is inconsistent with the supply-driven explanation.

5 Supplementary Discussion

§A   Did origination increase due to rise in private mortgage securitization?

As private securitization of mortgages (selling the mortgage to non-government agencies) offers lenders higher commissions, a rise in the tendency to privately securitize could explain the increase in the mortgage origination of reported in the current paper (Keys, Mukherjee, Seru, & Vig, 2010). If this explanation is true, the fraction of originated mortgages that were sold to non-government entities would increase in the treated areas relative to the control after the event.

Table (XI) shows the result of regressing, using Equation (1), three outcome variables: the fraction of total applications that lenders originated and (i) sold to non-government entities, (ii) sold to the four GSEs (Fannie Mae, Freddie Mac, Ginnie Mae, and Farmer Mac), and (iii) did not sell. The estimates suggest that private securitization did not increase (Columns (1) and (2)); government securitization increased (Columns (3) and (4)); and the fraction of unsold mortgages did not change (Columns (5) and (6)). Overall, private securitization does not appear to a reason behind the increase in origination.

§B   Did origination increase due to subprime lending? Credit score-based evidence

It may be argued that the increase in mortgage origination was due to an increase in the subprime credit (Mian & Sufi, 2009). Using the comprehensive HMDA data and location-based proxies of creditworthiness, Table (V) already suggests that the effect of free credit reports was stronger in the prime counties/census tracts than in the subprime. Even though these proxies

are informative and widely used (e.g., Di Maggio & Kermani, 2017; Mian & Sufi, 2009), they are imprecise. The GSE sample contains the application-level credit scores, and thus it can be used to precisely examine the patterns in prime and subprime origination.

Table (XII) shows the results of regressing separately the number of prime (credit score $\geq$ 620) and subprime *originated* mortgages in zip3-state areas using Equation (3). Columns (1) and (2) show that the number of prime mortgages increased by 308–312 in the treated zip3-state areas relative to similar control zip3-state areas, whereas columns (3) and (4) imply that subprime mortgages increased only by ~10 applications, which is 30 times smaller. Thus the increase in mortgages did not disproportionately go to subprime consumers. Note that these estimates are not directly comparable to the previous regressions, as the observation unit here is zip3-state, not census tracts, and the outcome variable is not scaled by population. Also, these results suffer from the same selection issue that existed with the previous results utilizing the GSE sample, and the same argument that was made before—the implausibility that the incentives of GSEs changed across the contiguous sample counties around the event—alleviates it. In addition, Elul, Gupta, and Musto (2020) show that to combat the onset of the housing bust before 2007, the GSEs sought to buy more subprime, not prime, mortgages. Hence, had the GSEs not changed their buying pattern, the estimates of subprime origination would have been even lower and the contradiction of the subprime hypothesis even stronger.

§C   Effect on banks

The analysis in the paper so far has focused on evaluating the effects on borrowers, but it is the lenders who ultimately evaluate the credit decisions, and hence the effect on banks is evaluated next.

Analyzing the effects on banks has a few notable limitations. First, banks are not the dominant mortgage originators. Despite being 80% of mortgage lenders by number, banks accounted for just 37% of mortgage lending in 2005 (Avery, Brevoort, & Canner, 2007), and thus the conclusions drawn from studying banks may not generalize to all mortgage lenders. Second, since many banks operate across states, their treatment and control status in this natural experiment is not binary, but continuous. To solve this issue, it is assumed that the continuous treatment intensity is proportional to a bank's *ex-ante* mortgage origination in the treated states relative to the activity in treated and control states combined. Following this, a bank is classified as "control" if in the pre-event year 2004 the ratio of mortgage amounts it originated in the control

states to those in the treated and control states combined was larger than the cross-sectional mean in that year across all banks in the sample. The regression equation is

$$Y_{bt} = \beta_0 + \beta_1 \times \text{Treat}_b \times \text{Post}_t + \delta \times \text{Bank controls}_{bt} + \alpha_l + \gamma_t + \varepsilon_{bt} \, , \qquad (5)$$

where Y_{bt} represents the three outcome variables: net interest margin (NIM), return on equity (RoE), and return on assets (RoA); b indexes the banks; $Treat_b$ is 1 if a bank is treated and 0 otherwise; $Post_t$ is 1 if $year \geq 2005$ and 0 otherwise; year t represents year-quarter; α_i is bank fixed effects; γ_t is year-quarter fixed effect; and *Bank controls* include banks' log total assets, share of liquid assets to total assets, and cost of deposit.[26]

The regression results in Table (XIII) show that treated banks saw a 6 basis-points increase in NIM (Columns (1) and (2)), a 0.75–0.76 percentage-points increase in RoE (Columns (3) and (4)), and a 0.07–0.08 percentage-points increase in RoA (Columns (5) and (6)). Also, the results are qualitatively and quantitatively similar when lenders are classified into treatment and control groups using cross-sectional median instead of the mean.

Furthermore, it is also useful to understand which characteristics of the banks allowed them to increase the mortgage supply after the event. To this end, the role of liquid assets on a bank balance sheet is examined. A bank is classified as having a *high* share of liquid assets if its ratio of liquid assets to total assets (computed annually) was *greater* than the cross-sectional mean in the pre-event year 2004, and as having a *low* share of liquid assets otherwise. The total amount of mortgages (in million USD) originated by a bank in the control and treated states is regressed using Equation (5) separately for these two groups. Also, since the mortgage information come from the HMDA data, which are publicly available only at an annual frequency, the regression is estimated at the *Bank×Year* level.

Panel (B) of Table (XIII) shows the results of the regressions. The coefficients in Columns (1) and (2) suggest that, among the banks with an *ex-ante* low share of liquid assets, the change in mortgage lending by treated banks vis-à-vis the control banks is not statistically significant, and the point estimate is negative. At the same time, the coefficients in Columns (3) and (4) suggest that among the banks with an *ex-ante* high share of liquid assets, the treated banks

---

[26] NIM is the ratio of net interest income (sum of RIAD4074 and RIAD4301) to earning assets. I use the definition of earning assets from St. Louis Fred: it is the sum of RCFD0071, RCFD1350, RCFD2122, RCFD3545, RCFD1754, and RCFD1772 (https://fred.stlouisfed.org/series/USNIM). RoE is the ratio of net income (RIAD4340) to book value of equity. RoA is the ratio of net income to book value of total assets. Liquid assets is the sum of RCFD1754, RCFD1773, RCFD3545, RCFD1754, RCFD3545, and RCFD1350. Cost of deposit is the ratio of RIAD4073 to earning assets.

on average increased mortgage lending relative to the control by about USD 111 million, in a statistically significant manner.

All in all, the effect of the event on financial performance of the banks seems to be positive, and the banks with high liquid assets appear to be behind the increase in mortgage origination after the event.

§D   An alternative mechanism based on information asymmetry

An alternative mechanism based on asymmetric information in which borrowers *privately know* their true creditworthiness type, but do not know what lenders know about them, is plausible. Using free credit reports, borrowers learn that the creditworthiness information on them available at lenders are proportional to their true creditworthiness type. Hence, under the non-trivial search/application cost, borrowers with poor creditworthiness (bad type) self-select out. The borrower pool now improves relative to the situation in which borrowers do not know that a lender has information about their true type and optimistically expect that the information is better than what is warranted by their credit reports. Note that the improvement occurs here due to the self-selecting-out by bad type, but not by self-selecting-in by good type, since all borrowers *privately know* their true type. However, since under the self-learning mechanism, borrowers themselves have imperfect information of their true type, both selecting-in by good borrowers and selecting-out by bad borrowers contribute to pool improvement after credit reports become free.

The empirical findings are consistent primarily with the self-learning mechanism. We saw that in the treated areas both the mortgage applications and first-time homebuyers fraction increased, not decreased. Both these findings provide evidence of selecting-in by borrowers, which is plausible only under the *self-learning* mechanism.

Another valid concern is that in assessing mortgage applications, together with the credit reports, lenders use private information that they may accumulate through relationship lending. This attenuates the effects of free credit reports. The concern is partially alleviated by the fact that lenders necessarily look at credit reports and scores when assessing credit applications.[27]

---

[27] Experian (n.d.) explains: "Not all lenders think the same way, and they may have different ways of making their decisions. But all of them will look at some key factors to help them decide. These include: information on your credit report including your credit history and public record data."

§E   Robustness

Since the natural experiment utilized in this paper occurred in the year 2005, the sample period is chosen to be from 2000 to 2008 to allow for enough post-experiment observations. As the experiment is close to the financial crisis of 2008, it is crucial to ensure that the results are not caused by the unique lending environment that existed in 2007–2008. To this end, all the regressions were re-estimated by excluding the observations for the years 2007 and 2008. Mostly the results are qualitatively and quantitatively similar and are left unreported for brevity.

6   Conclusion

Several large-scale surveys in the U.S. indicate that a non-trivial proportion of consumers do not check their credit reports and do not know their credit scores. At the same time, mortgage application data show patterns consistent with consumers making credit decisions under imperfect information of their creditworthiness. Noting that credit reports may aid consumers to self-assess their creditworthiness, this paper uses a natural experiment to examine the effect on mortgage market outcomes of lowering the economic costs to consumers of accessing their reports. The federal *Fair and Accurate Transactions Act of 2003* (FACTA) has made access to credit reports free through a website since 2005 for all consumers, while seven states already had local laws permitting their residents to obtain their reports for free. This effectively reduced the economic cost of accessing credit reports in a close-to-exogenous manner. This paper utilizes this change in a difference-in-differences setting in which the border counties of the early-adopting states constitute the control group and those of the neighboring states, the treatment.

The key finding is that the mortgage market responds to the reduction in consumers' economic costs of accessing their credit reports in a manner consistent with improvements in the borrower pool, and the outcomes appear to benefit both the consumers and lenders. Specifically, approval ratios and mortgage applications increased, more credit was originated to creditworthy borrowers, more first-time homebuyers took out mortgages, and mortgage delinquency reduced slightly, and financial performance of mortgage-lending banks improved.

All in all, there seems to be a causal link between the economic costs of accessing credit reports and mortgage market outcomes, and any policy intervention aimed at educating consumers of their creditworthiness may bring about similar improvements in other consumer credit markets as well.

References

108th Congress of the U.S. Senate. (2004). *The Fair Credit Reporting Act and Issues Presented by Reauthorization of the Expiring Preemption Provisions: Hearings Before the Committee on Banking, Housing, and Urban Affairs United States Senate*. U.S. Government Printing Office. (S. Hrg. 108-579.)

Agarwal, S., Chomsisengphet, S., Mahoney, N., & Stroebel, J. (2018). Do Banks Pass Through Credit Expansions to Consumers Who Want to Borrow? *The Quarterly Journal of Economics*, *133*(1), 129–190.

Allegretto, S., Dube, A., Reich, M., & Zipperer, B. (2017). Credible Research Designs for Minimum Wage Studies: A Response to Neumark, Salas, and Wascher. *ILR Review*, *70*(3), 559–592.

Associated Press News. (December 24, 1992). *TRW Settles Vermont Credit Report Suit*. Retrieved 2020-12-29, from https://apnews.com/article/65342c8b9600af099a3a8dbaa8a4d499

Avery, R. B., Brevoort, K., & Canner, G. (2007). Opportunities and Issues in Using HMDA Data. *Journal of Real Estate Research*, *29*(4), 351–380.

Avery, R. B., Calem, P. S., & Canner, G. B. (2004). Credit Report Accuracy and Access to Credit. *Federal Reserve Bulletin*, *90*, 297.

Balakina, O., Balasubramaniam, V., Dimri, A., & Sane, R. (2020). The Effect of Information Unshrouding on Financial Product Purchase Decision [Working Paper]. *Unpublished*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3519845

Bogin, A., Doerner, W., & Larson, W. (2016). Local House Price Dynamics: New Indices and Stylized Facts [Working Paper]. *Unpublished*. Retrieved from https://www.fhfa.gov/PolicyProgramsResearch/Research/Pages/wp1601.aspx

Borusyak, K., Jaravel, X., & Spiess, J. (2021). Revisiting Event Study Designs: Robust and Efficient Estimation [Working Paper]. *Unpublished*.

Census Bureau. (n.d.). *County Adjacency File* [Database]. Retrieved from https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html

Census Bureau. (2000–2008). *County Business Patterns* (*CBP*) [Database]. Retrieved from https://www.census.gov/programs-surveys/cbp.html

Census Bureau. (2006). *Census of Population and Housing, 2000* [*United States*]: *Census Tract Relationship Files* (*CTRF*) [Database]. Inter-university Consortium for Political and Social Research. Retrieved from https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/13287

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, *110*(9), 2964–96.

Di Maggio, M., & Kermani, A. (2017). Credit-induced Boom and Bust. *The Review of Financial Studies*, *30*(11), 3711–3758.

Dobbie, W., Goldsmith-Pinkham, P., Mahoney, N., & Song, J. (2016). Bad Credit, No Problem? Credit and Labor Market Consequences of Bad Credit Reports. *The Journal of Finance*.

Dube, A., Lester, T. W., & Reich, M. (2010). Minimum Wage Effects across State Borders: Estimates using Contiguous Counties. *The review of economics and statistics*, *92*(4), 945–964.

Dube, A., Lester, T. W., & Reich, M. (2016). Minimum wage shocks, employment flows, and labor market frictions. *Journal of Labor Economics*, *34*(3), 663–704.

Elul, R., Gupta, D., & Musto, D. K. (2020). Concentration in Mortgage Markets: GSE Exposure and Risk-Taking in Uncertain Times [Working Paper]. *Unpublished*. Retrieved from https://www.philadelphiafed.org/-/media/research-and-data/publications/working-papers/2020/wp20-04.pdf

Equifax. (n.d.). *Understanding Hard Inquiries on Your Credit Report.* Retrieved 2020-06-10, from https://www.equifax.com/personal/education/credit/report/understanding-hard-inquiries-on-your-credit-report/

Experian. (n.d.). *What credit score do I need for a mortgage?* Retrieved 2020-06-10, from https://www.experian.co.uk/consumer/mortgages/guides/credit-and-mortgages.html

Federal Trade Commission. (January 13, 2000). *Nation's Big Three Consumer Reporting Agencies Agree To Pay $2.5 Million To Settle FTC Charges of Violating Fair Credit Reporting Act.* Retrieved 2021-06-15, from https://www.ftc.gov/news-events/press-releases/2000/01/nations-big-three-consumer-reporting-agencies-agree-pay-25

FRBNY, & Equifax. (n.d.). *Equifax Subprime Credit Population* [Database]. retrieved from FRED, Federal Reserve Bank of St. Louis.

Gerardi, K., Goette, L., & Meier, S. (2010). Financial Literacy and Subprime Mortgage Delinquency: Evidence from a Survey Matched to Administrative Data [Working Paper]. *Unpublished*. Retrieved from https://www.frbatlanta.org/-/media/documents/research/publications/wp/2010/wp1010.pdf

Gerardi, K., Lambie-Hanson, L., & Willen, P. S. (2013). Do Borrower Rights Improve Borrower Outcomes? Evidence from the Foreclosure Process. *Journal of Urban Economics*, *73*(1), 1–17.

Ghent, A. C., & Kudlyak, M. (2011). Recourse and Residential Mortgage Default: Evidence from US States. *The Review of Financial Studies*, *24*(9), 3139–3186.

Goodman-Bacon, A. (2021). Difference-in-differences with Variation in Treatment Timing. *Journal of Econometrics*.

Homonoff, T., O'Brien, R., & Sussman, A. B. (2019). Does Knowing Your FICO Score Change Financial Behavior? Evidence from a Field Experiment with Student Loan Borrowers. *Review of Economics and Statistics*, 1–45.

Huang, R. R. (2008). Evaluating the real effect of bank branching deregulation: Comparing contiguous counties across us state borders. *Journal of Financial Economics*, *87*(3), 678–705.

Hundtofte, S. (2017). No Such Thing as a Free Option? Offers of Debt Forgiveness Under Imprecise Borrower Beliefs [Working Paper]. *Unpublished*. Retrieved from https://drive.google.com/file/d/1R1ii1F5WuCj6Qr3HhpmuWhmt_PNbPNSL

Hurst, E., Keys, B. J., Seru, A., & Vavra, J. (2016). Regional Redistribution through the US Mortgage Market. *American Economic Review*, *106*(10), 2982–3028.

Keys, B. J., Mukherjee, T., Seru, A., & Vig, V. (2010). Did Securitization Lead to Lax Screening? Evidence from Subprime Loans. *The Quarterly Journal of Economics*, *125*(1), 307–362.

Kulkarni, S., Truffa, S., & Iberti, G. (2018). Removing the Fine Print: Standardization, Disclosure, and Consumer Loan Outcomes [Working Paper]. *Unpublished*. Retrieved from https://static1.squarespace.com/static/58b5e6e15016e1efa0bfd0a5/t/5bd88062f4e1fc38159faefa/1540915300523/informational_frictions_chile.pdf

Liberman, A., Neilson, C., Opazo, L., & Zimmerman, S. (2018, September). The Equilibrium Effects of Information Deletion: Evidence from Consumer Credit Markets [Working Paper]. *Unpublished*. Retrieved from http://www.nber.org/papers/w25097

Lusardi, A., & Tufano, P. (2009, March). Debt Literacy, Financial Experiences, and Overindebtedness [Working Paper]. *Unpublished*. Retrieved from http://www.nber.org/papers/w14808

Manson, S., Schroeder, J., Van Riper, D., & Ruggles, S. (2019). *Ipums national historical geographic information system: Version 14.0* [Database]. IPUMS NHGIS.

Mian, A., & Sufi, A. (2009). The Consequences of Mortgage Credit Expansion: Evidence From the Us Mortgage Default Crisis. *The Quarterly Journal of Economics*, *124*(4), 1449–1496.

Mikhed, V. (2015). Can Credit Cards with Access to Complimentary Credit Score Information Benefit Consumers and Lenders? [Working Paper]. *Unpublished*. Retrieved from https://www.philadelphiafed.org/-/media/frbp/assets/consumer-finance/discussion-papers/dp15-03.pdf

Moore, D. L. (2003). *Survey of Financial Literacy in Washington State: Knowledge, Behavior, Attitudes, and Experiences*. Washington State Department of Financial Institutions.

Nott, L., & Welborn, A. (2003). A Consumers Access to a Free Credit Report: A Legal and Economic Analysis. *Congressional Research Service, Order Code RL32008*.

Office of Policy Development and Research. (n.d.). *HUD USPS Zip Code Crosswalk Files* [Database]. U.S. Department of Housing (HUD). Retrieved from https://www.huduser.gov/portal/datasets/usps_crosswalk.html

Perry, V. G. (2008). Is Ignorance Bliss? Consumer Accuracy in Judgments About Credit Ratings. *Journal of Consumer Affairs*, *42*(2), 189–205.

Reddit Forum. (n.d.). *Withdrawing a Mortgage Application*. Retrieved 2020-06-10, from https://www.reddit.com/r/personalfinance/comments/38k1l5/withdrawing_a_mortgage_application/

Scharfstein, D., & Sunderam, A. (2016). Market Power in Mortgage Lending and the Transmission of Monetary Policy [Working Paper]. *Unpublished*. Retrieved from https://www.hbs.edu/faculty/Pages/item.aspx?num=44239

Stango, V., & Zinman, J. (2009). Exponential Growth Bias and Household Finance. *The Journal of Finance*, *64*(6), 2807–2849.

Sun, L., & Abraham, S. (2020). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*.

Survey of Consumer Expectations. (2013–2020). *Survey of Consumer Expectations Credit Access Survey* [Database]. Federal Reserve Bank of New York. Retrieved from https://www.newyorkfed.org/microeconomics/sce

Survey of Consumer Finances. (1998–2007). *Survey of Consumer Finances* [Database]. Federal Reserve Board. Retrieved from https://www.federalreserve.gov/econres/aboutscf.htm

Wikipedia. (n.d.). *AnnualCreditReport.com*. Retrieved 2020-12-29, from https://en.wikipedia.org/wiki/AnnualCreditReport.com

Figure I: A Sample Credit Report

This figure shows the summary page of a credit report obtained from the website www.annualcreditreport.com for free under the *Fair and Accurate Transaction Act of 2003*. The specific credit history-related details are not shown. The report contains, among other things, the details of the consumer's active accounts, debt-to-credit ratio, and an indication of the available borrowing capacity.

1. Summary

Review this summary for a quick view of key information contained in your Equifax Credit Report.

Report Date	Apr 14, 2020
Credit File Status	No fraud indicator on file
Alert Contacts	0 Records Found
Average Account Age	5 Months
Length of Credit History	8 Months
Accounts with Negative Information	0
Oldest Account	DISCOVER BANK (Opened Aug 29, 2019)
Most Recent Account	AMERICAN EXPRESS (Opened Jan 10, 2020)

Credit Accounts

Your credit report includes information about activity on your credit accounts that may affect your credit score and rating.

Account Type	Open	With Balance	Total Balance	Available	Credit Limit	Debt-to-Credit	Payment
Revolving	2	2	$606	$11,044	$11,650	5.0%	$70
Mortgage							
Installment							
Other							
Total	2	2	$606	$11,044	$11,650	5.0%	$70

Other Items

Your credit report includes your Personal Information and, if applicable, Consumer Statements, and could include other items that may affect your credit score and rating.

Consumer Statements		0 Statements Found
Personal Information		3 Items Found
Inquiries		2 Inquiries Found
Most Recent Inquiry	DISCOVER BANKAug 27, 2019	
Public Records		0 Records Found
Collections		0 Collections Found

Figure II: Empirical Research Design

Panel (A) of this figure shows on the map of the contiguous U.S. the states utilized in the difference-in-differences (DID) setting. Seven U.S. states had enacted free credit report laws prior to the FACTA enactment in 2004: CO (1997), GA (1996), MD (1992), NJ (1997), MA (1995), VT (1992), and ME (2003). All except ME constitute the control group, and the 26 states surrounding the control group are the treatment. **Panel (B)** of this figure shows on the map of the contiguous U.S. the counties included in the estimation sample. These are the counties at the border between the treatment and control states.

Panel A: Treatment and Control States

■ Control: Pre-FACTA States
■ Treatment: States surrounding the pre-FACTA states
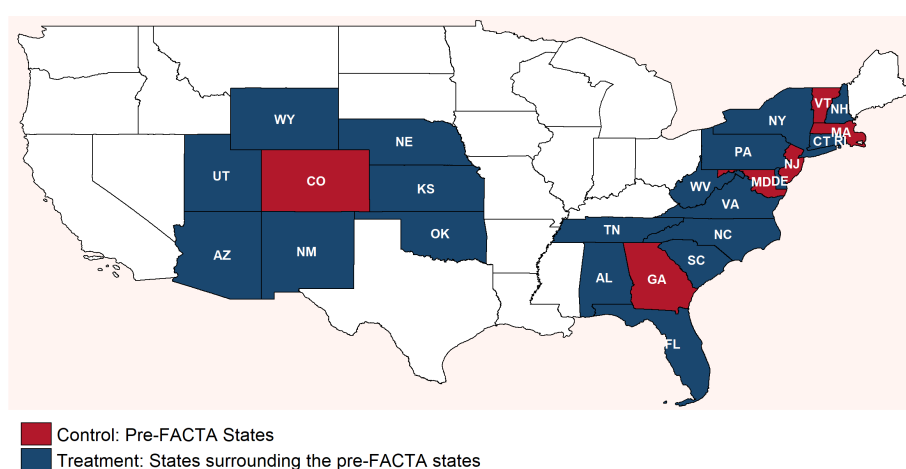
Panel B: Sample Counties from the Treatment and Control States

■ Control: Counties at the border of pre-FACTA States
■ Treatment: Counties which belong to the bordering states and are adjacent to the control counties
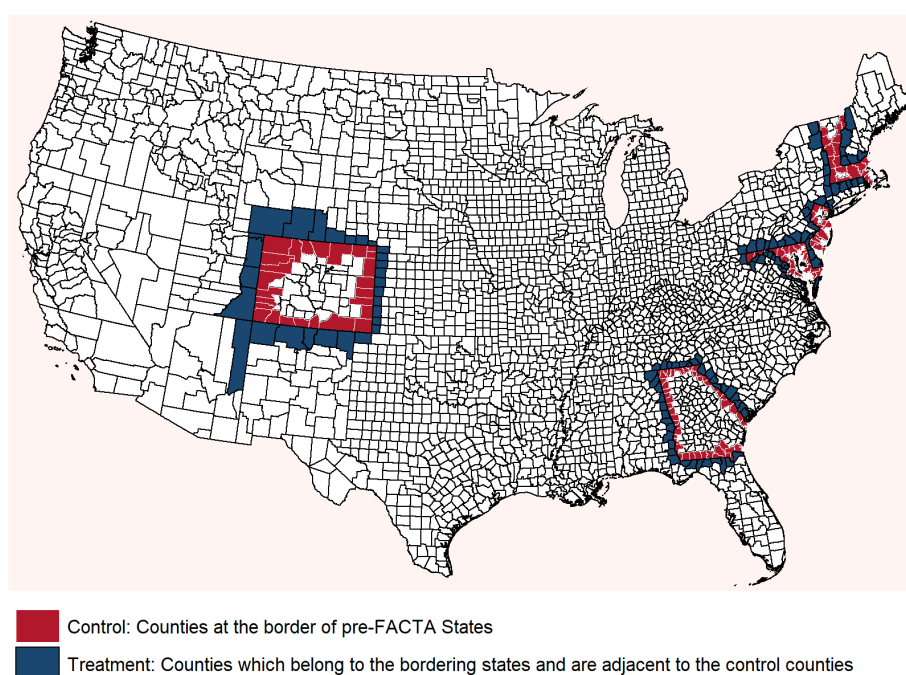
Figure III: Examining the Parallel Trends

Panel (A) of this figure shows the mean approval ratio in the treated and control census tracts.

Panel (B) of this figure shows the coefficients β_k from regressing *Approval Ratio* using the following specification:

$$Y_{icsjt} = \beta_0 + \sum_{k=T-3}^{T-1} \beta_k \, \text{Treatment}_{icsj} \times \text{Event}_k + \sum_{k=T+1}^{T+4} \beta_k \, \text{Treatment}_{icsj} \times \text{Event}_k + \alpha_i + \gamma_{j,t} + \varepsilon_{icsjt} ,$$

where $\text{Event}_k = 1$ if $t = T - k$. $\text{Event}_k = 0$ if $t \neq T - k, k = \{-3, 4\}$. $T = $ Event year 2005. Coefficients are estimated with respect to the base year 2004 ($j = 0$). The *x*-axis shows year relative to the pre-event year 2004; i.e., T = +1 is the first treated year, 2005. The *y*-axis shows the coefficients β_k. The 95% confidence intervals of $\beta_k s$ are also shown. The regression includes *Border×Year* and *Census Tract* fixed effects. Other terms in the equation are the same as those in Equation 1. Standard errors are clustered by county.

Panel A: Mean Approval Ratio in Treated and Control Areas
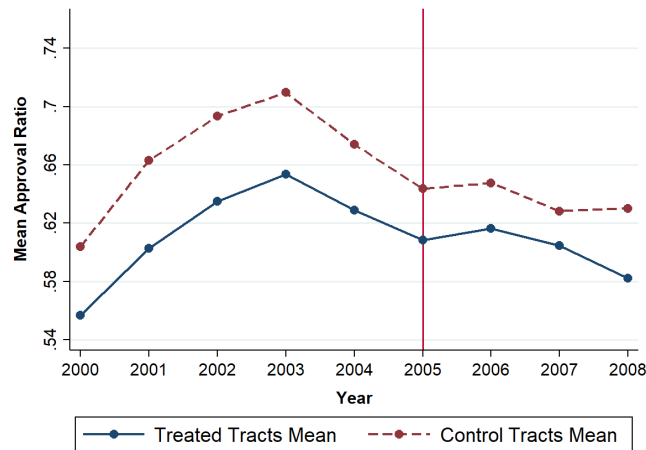
Panel B: Coefficient Estimates of Approval Ratio by Years to Treatment

38
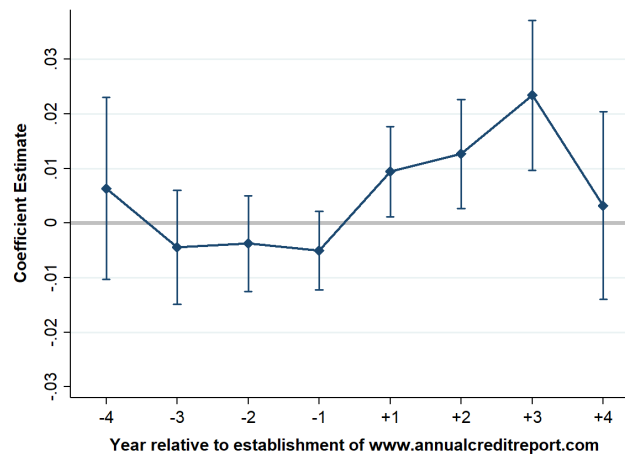
Figure IV: Subsample Analysis

Panel (A) of this figure shows the estimates for changes in number of applications per 1000 adults (scaled applications) when each control state is removed one by one. **Panel (B)** of this figure shows the estimates for changes in approval ratio when each control state is removed one by one. For example, the coefficient corresponding to "No CO" respresents the estimate when Colorado and its surrounding states are removed from the estimation sample. The regressions specifications behind the estimates are the same as those in Table(III). The bands around the estimates show 90% confidence intervals.

Panel A: Number of Applications (per 1000 Adults in a Census Tract)
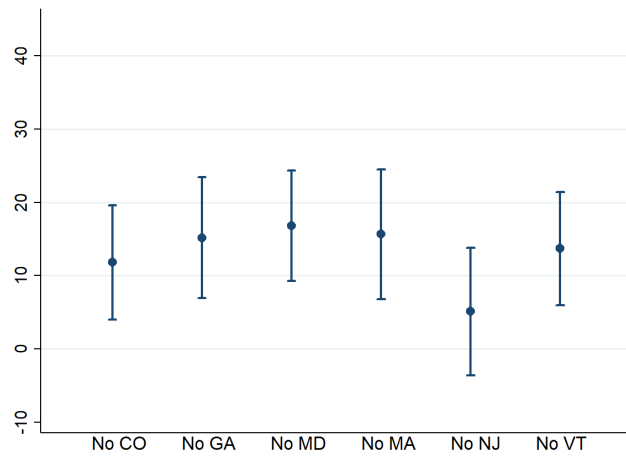
Panel B: Approval Ratio
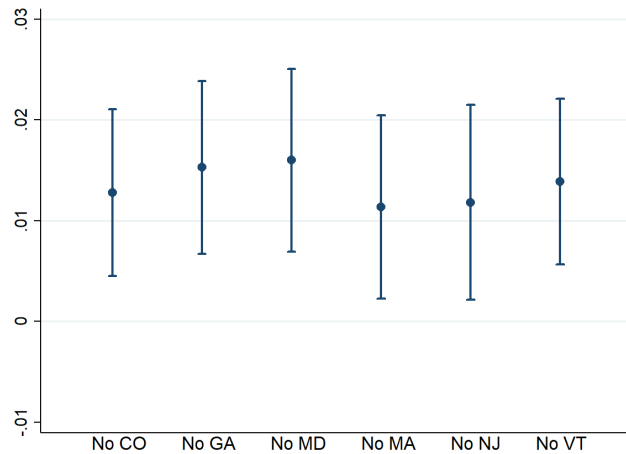
Figure V: Search Interest in Free Credit Reports: Google Trends

This figure plots the search interest in free credit reports using Google Trends data. **Panel (A)** of this figure shows the plot of *Search Interest* for the keyphrase *Free Credit Report* in the US from Jan 1, 2004 till Dec 31, 2011. Numbers on the vertical axis represent search interest relative to the highest point on the chart during this period. A value of 100 (50) represents the peak popularity (half of the peak popularity) for the keyphrase. A value of 0 means there was not enough data. **Panel (B)** of this figure shows the difference in the mean search interest for treated and control states for the same keyphrase from 2004 to 2008 using the interest-by-subregion data from Google. These data are computed within the time period for which the data are extracted from Google. A value of 100 represents the location with the highest popularity of the search term as a fraction of total searches in that location, and a value of 50 indicates a location where it is half as popular. To overcome the issue of data-value normalization by Google, first, the data were *separately* extracted for each one-year interval, and then the means were *separately* calculated within each time interval for each of the two sets of states.

Panel A: Time Series Search Interest in the U.S. for the Terms "Free Credit Report"

Source: Google Trends
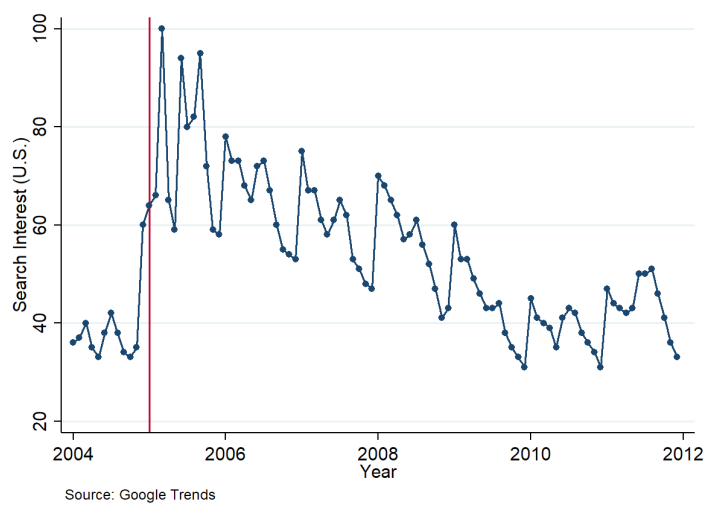
Panel B: Cross-sectional Search Interest for the Terms "Free Credit Report"

Source: Google Trends

40
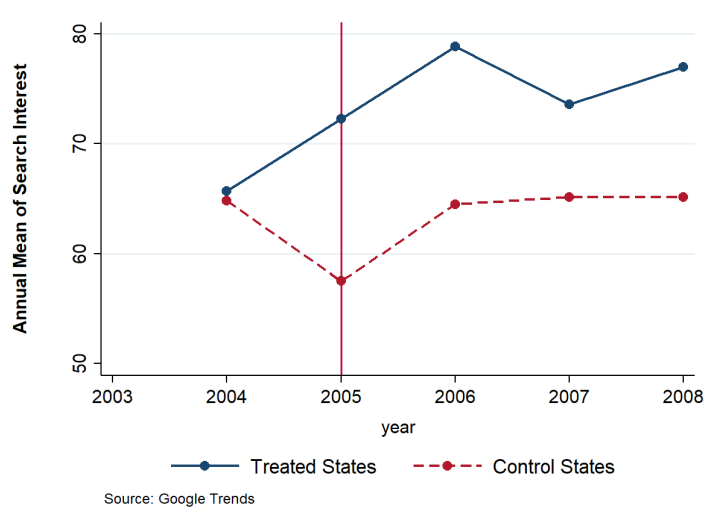
Figure VI: Effect of Free Credit Report on Mortgage Delinquencies

This figure plots the delinquency rates of various mortgage vintages with their age (measured in months). A mortgage vintage is a collection of mortgages originated in a given area—treated or control—in a given year—2004 (pre-event) or 2005 (post-event). Delinquency rate is calculated at each age as the ratio of the number of mortgages becoming delinquent for the first-time to the total number of mortgages in the respective vintage. **Panel (A)** shows 30–59-day delinquency rates separately for treated and control areas for 2004 on the left-hand side and for 2005 on the right-hand side. **Panel (B)** shows these same plots for 60–89-day delinquencies. These plots are based on the 30-year fixed-rate single-family mortgages purchased by Fannie Mae and Freddie Mac.

Panel A: 30–59-day Delinquency Rate
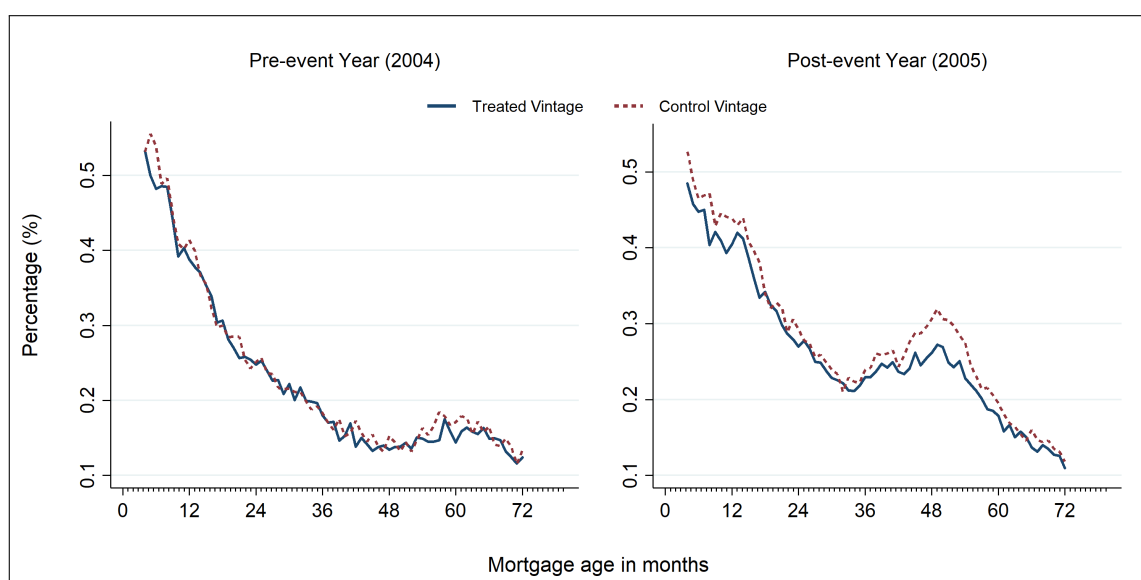
Panel B: 60–89-day Delinquency Rate

41
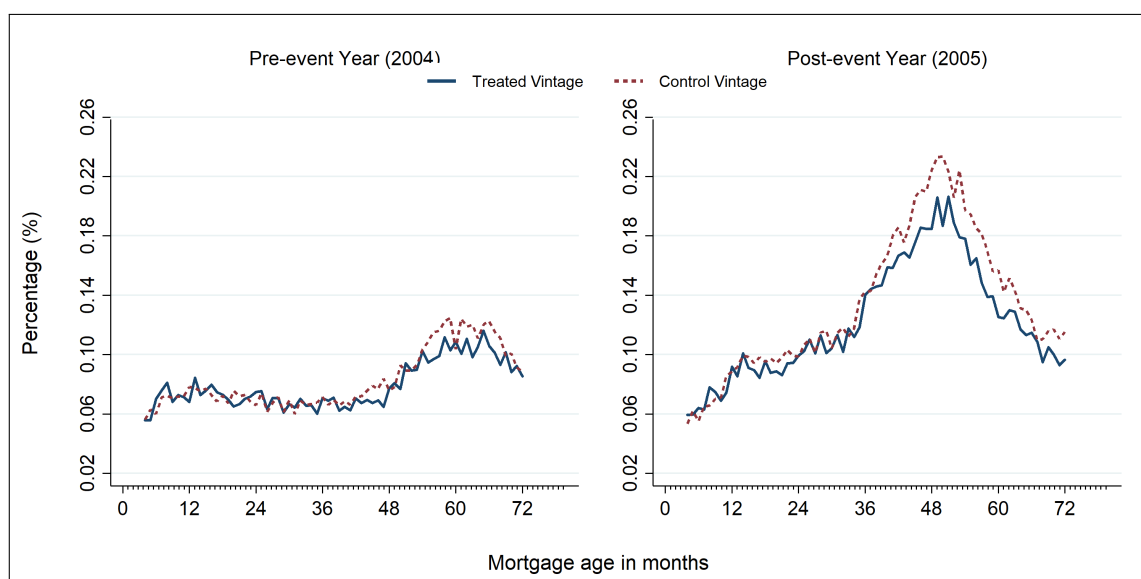
Table I: Summary Statistics

Panel A shows the statistics for the full sample time period (2000–2008). Panel B shows the statistics for the pre-treatment period (2000–2004) and the p-values for the t-test for difference in the control and treatment group. *Scaled applications, N,* is the number of mortgage applications in a census tract scaled by the population aged 18 to 64 years in the tract (scaled applications). *Approval ratio (Aprv.)* is the ratio of the number of successful applications (action type "1" *or* "2" in the HMDA dataset) to the number of total applications in a census tract. *Deny Credit Hist Ratio* and *Deny Debt-to-inc Ratio* are the ratio of applications denied due to credit history and debt-to-income ratio, respectively, to the number of total applications in a census tract. *Withdrawal Ratio* is the ratio of applications expressly withdrawn by the applicant to the number of total applications in the census tract.

Economic Controls consists of four variables described as follows: (i) *Num. Lenders* (*log*), the number of unique mortgage lenders in a census tract (expressed in natural log); (ii) Δ *Inc per capita*, the annual growth rate of income per capita at the county level; (iii) Δ *Emp.*, the annual growth rate of employment by all establishments at the county level; and (iv) Δ *State GDP*, the annual growth rate of the state gross domestic product.

Panel A: Full Sample (2000 – 2008)

	Full Sample				Control Group (C)				Treatment Group (T)			
	N	Mean	SD	Med.	N	Mean	SD	Med.	N	Mean	SD	Med.
Scaled Applications (N)	86017	83.09	74.77	66.04	36002	98.18	77.76	77.44	50015	72.23	70.57	56.22
Approval Ratio (Aprv.)	81914	0.54	0.13	0.55	35386	0.57	0.12	0.58	46528	0.52	0.14	0.53
Deny Credit Hist Ratio	81914	0.06	0.04	0.05	35386	0.05	0.04	0.04	46528	0.06	0.05	0.05
Deny Debt-to-inc Ratio	81914	0.03	0.03	0.03	35386	0.03	0.02	0.03	46528	0.03	0.03	0.03
Withdrawl Ratio	81914	0.12	0.05	0.12	35386	0.12	0.04	0.11	46528	0.12	0.06	0.12
Num. Lenders (log)	82477	3.16	0.78	3.30	33974	3.36	0.60	3.42	48503	3.01	0.85	3.19
Δ Inc per capita	2259	0.04	0.06	0.04	1125	0.04	0.05	0.04	1134	0.05	0.07	0.04
Δ Emp	2262	0.01	0.09	0.01	1120	0.01	0.09	0.01	1142	0.01	0.10	0.01
Δ State GDP	73	0.05	0.03	0.04	29	0.05	0.02	0.04	44	0.05	0.03	0.04

Panel B: Pre - Treatment Sample (2000 – 2004)

|	Full Sample				Control Group (C)				Treatment Group (T)				(C-T)
	N	Mean	SD	Med.	N	Mean	SD	Med.	N	Mean	SD	Med.	p-val
Scaled applications (N)	47923	110.16	83.62	92.99	20015	129.53	86.39	108.37	27908	96.27	78.68	82.84	0.000
Approval Ratio (Aprv.)	46584	0.55	0.14	0.56	19798	0.58	0.13	0.60	26786	0.52	0.14	0.53	0.000
Deny Credit Hist Ratio	46584	0.06	0.04	0.05	19798	0.06	0.04	0.05	26786	0.07	0.05	0.06	0.000
Deny Debt-to-inc Ratio	46584	0.03	0.02	0.03	19798	0.03	0.02	0.03	26786	0.03	0.02	0.03	0.000
Withdrawl Ratio	46584	0.12	0.05	0.11	19798	0.12	0.04	0.11	26786	0.13	0.05	0.12	0.000
Num. Lenders (log)	44383	3.36	0.73	3.48	17987	3.53	0.60	3.59	26396	3.24	0.78	3.39	0.000
Δ Inc per capita	1255	0.04	0.06	0.04	625	0.04	0.05	0.04	630	0.04	0.07	0.04	0.620
Δ Emp	1254	0.01	0.09	0.01	622	0.01	0.09	0.01	632	0.00	0.10	0.01	0.290
Δ State GDP	39	0.05	0.02	0.05	17	0.05	0.02	0.05	22	0.05	0.02	0.06	0.543

Table II: Survey Evidence on the Credit Reports Usage and Discouraged Borrowers

This table reports the regression results from the SCE Credit Access Survey. *Never* is 1 if a respondent has never checked his/her credit score (Q. *N23*). *Infrequently* is 1 if a respondent has never checked it or last checked it more than two years ago (Q. *N23*). *Unaware* is 1 if a respondent does not know his/her credit score (Q. *N22*). *Dscrgd* is 1 if a respondent said "I do not think I would get approved" in *Q. N19*. Note that this question (*Q. N19*) is a conditional question in the survey. Hence the observations in specifications (4–6) include only the responses in which (i) for *Q. N17A*, respondent selected *very unlikely* or *somewhat unlikely* to apply for mortgage/home-based loan, or refinance, or (ii) for *Q. N17B*, mentioned the probability of applying for mortgage or to refinance as less than 10%. All regressions include *Year* × *Month* fixed effects (FE). Standard errors are clustered by survey's Year×Month. p-values are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Check Credit Report		Know Credit Score	Mortgage-discouraged Borrowers		
	(1)	(2)	(3)	(4)	(5)	(6)
	Never	Infrequently	Unaware	Dscrgd	Dscrgd	Dscrgd
Check Infrequently					0.03**	
					(0.05)	
Unaware						0.05*
						(0.06)
Constant	0.08***	0.20***	0.12***	0.13***	0.13***	0.13***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Cluster (Year-Month)	Yes	Yes	Yes	Yes	Yes	Yes
FE (Year-Month)	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.007	0.007	0.007	0.003	0.004	0.005
Observations	19231	19231	20275	9059	9058	9058

Table III: Mortgage Applications, Approval Ratio, and House Prices

This table reports the estimates of the treatment effect of free credit reports on the number of mortgage applications, approval ratio, and growth in house prices. The regression specification is from Equation (1):

$$Y_{icsjt} = \beta_0 + \beta_1 \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic Controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt}.$$

N, *Aprv.*, and ΔHPI are the number of applications per 1000 adults (scaled applications), the approval ratio in a census tract and growth in house prices at census tract level, respectively. *Economic Controls* include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable in the treated census tracts relative to the control census tracts. All regressions include the *Border×Year* fixed effects (FE) and the *Census Tract* FE. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Aprv.	Aprv.	N	N	ΔHPI	ΔHPI
Treat × Post	0.01***	0.01***	13.43***	16.63***	1.83*	2.00*
	(2.80)	(2.84)	(2.95)	(3.79)	(1.88)	(1.94)
Economic Controls	No	Yes	No	Yes	No	Yes
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
Border × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (County)	Yes	Yes	Yes	Yes	Yes	Yes
R² (Adj.)	0.740	0.730	0.806	0.816	0.682	0.693
Observations	81871	76437	86010	80546	24927	23806

Table IV: Owner-occupied and Non-owner-occupied Mortgages

This table examines the changes in (i) owner-occupied mortgage applications, (ii) non-owner-occupied applications, (iii) non-owner-occupied mortgages as the fraction of total applications, and (iv) as fraction of successful applications. The regression specification is from Equation (1):

$$Y_{icsjt} = \beta_0 + \beta_1 \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic Controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt}.$$

The dependent variable in columns (1) through (4) is the number of applications per 1000 adults in a census tract (scaled applications), N. In columns (1) and (2), N measures owner-occupied category mortgage applications only; in columns (3) and (4), non-owner-occupied only. The dependent variable in columns (5) and (6) is the non-owner-occupied mortgage as a fraction of total applications, and in columns (7) and (8), as the fraction of originated applications. *Economic Controls* include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable in the treated census tracts relative to the control census tracts. All regressions include *Border×Year* fixed effects (FE) and *Census Tract* FE. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Owner		Non-owner		Non-owner, % of all appl.		Non-owner, % of succ. appl.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	N	N	N	N	%	%	%	%
Treat × Post	12.92***	15.99***	0.82*	1.00*	0.01**	0.01*	0.01**	0.01*
	(2.90)	(3.75)	(1.67)	(1.80)	(2.02)	(1.74)	(2.05)	(1.91)
Economic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (County)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.808	0.819	0.755	0.757	0.086	0.080	0.085	0.079
Observations	86010	80546	86010	80546	81871	76437	81785	76349

Table V: Characterizing the Effect: Heterogeneity by Consumer Creditworthiness

This table reports the estimates of the treatment effect of free credit reports on the number of mortgage applications per 1000 adults, (scaled applications, N) and the approval ratio (*Aprv.*) in *ex-ante* low- and high-creditworthiness areas. A county is "subprime" if its subprime population fraction is more than the *regional mean* subprime population fraction in 1999. The regression specification is from Equation (1):

$$Y_{icsjt} = \beta_0 + \beta_1 \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic Controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt}.$$

Economic Controls include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable in the treated census tracts relative to the control census tracts. All regressions include *Border×Year* fixed effects (FE) and *Census Tract* FE. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Ex-ante High Creditworthiness (Prime Counties)				*Ex-ante* Low Creditworthiness (Subprime Counties)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	N	N	Aprv.	Aprv.	N	N	Aprv.	Aprv.
Treat × Post	16.82**	18.80***	0.02***	0.02***	8.59	11.66**	0.01*	0.01*
	(2.33)	(2.66)	(3.19)	(3.29)	(1.64)	(2.42)	(1.71)	(1.78)
Economic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (County)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.802	0.822	0.777	0.772	0.825	0.826	0.679	0.672
Observations	39076	35703	38000	34644	46631	44558	43763	41703

Table VI: Characterizing the Effect: Heterogeneity by Income Level of Consumers

This table reports estimates of the treatment effect of free credit reports on the number of mortgage applications (Panel A) and the approval ratio (Panel B) for each of the income quartiles. The regression specification is from Equation (1):

$$Y_{icsjt} = \beta_0 + \beta_1 \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic Controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt}.$$

N and *Aprv.* are the number of applications per 1000 adults (scaled applications) and the approval ratio in a census tract, respectively. Income quartiles are calculated every year for a given census tract. *Economic Controls* include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable in the treated census tracts relative to the control census tracts. All regressions include *Border×Year* fixed effects (FE) and *Census Tract* FE. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Number of Applications per 1000 adults

	Income Quartile 1		Income quartile 2		Income Quartile 3		Income quartile 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	N	N	N	N	N	N	N	N
Treat × Post	0.23	0.49	2.14**	2.38***	2.64**	3.03***	3.97*	4.83***
	(0.16)	(0.35)	(2.54)	(3.11)	(2.35)	(3.19)	(1.90)	(2.61)
Economic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (County)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.760	0.760	0.772	0.773	0.740	0.741	0.659	0.672
Observations	87479	80546	87479	80546	87479	80546	87479	80546

Panel B: Approval Ratio

	Income Quartile 1		Income quartile 2		Income Quartile 3		Income quartile 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Aprv.	Aprv.	Aprv.	Aprv.	Aprv.	Aprv.	Aprv.	Aprv.
Treat × Post	0.01*	0.01**	0.01	0.01	0.00	0.00	-0.00	0.00
	(1.92)	(2.24)	(1.24)	(1.18)	(0.37)	(0.34)	(-0.41)	(0.19)
Economic Controls	No	Yes	No	Yes	No	Yes	No	Yes
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (County)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.316	0.307	0.338	0.326	0.308	0.297	0.169	0.156
Observations	71190	66014	71718	66523	71832	66632	71248	66062

Table VII: Characterizing the Effect: Heterogeneous Effects for the Overestimating Borrowers

This table reports the treatment effect for the number of mortgage applications and approval ratio, estimated separately for the census-tract tertiles created by sorting them independently on the rejection ratios for credit history and DTI. *C. Hist.* and *DTI* respectively represent the ratio of the number of mortgage applications rejected for credit history or DTI to the total number of mortgage applications in a census tract. The tertiles for these two ratios are calculated in the pre-event year 2004. The regression specification is from Equation (1):

$$Y_{icsjt} = \beta_0 + \beta_1 \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic Controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt}.$$

N and *Aprv.* are the number of applications per 1000 adults (scaled applications) and the approval ratio in a census tract, respectively. *Economic Controls* include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable in the treated census tracts relative to the control census tracts. All regressions include *Border×Year* fixed effects (FE) and *Census Tract* FE. Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses, and the number of observations is reported in square brackets below the t-statistics. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	C. Hist. Tertiles ↓	N			Aprv.		
		DTI Tertiles →			DTI Tertiles →		
		1	2	3	1	2	3
		(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	1	20.32***	19.73***	17.41***	0.01**	0.01**	0.02**
t-statistics		(2.65)	(3.38)	(3.13)	(2.11)	(2.54)	(2.20)
Observations		[12873]	[7533]	[4518]	[12199]	[7434]	[4298]
Treat × Post	2	10.67**	9.83**	12.27**	0.01**	0.02***	0.01*
t-statistics		(2.07)	(2.44)	(2.29)	(2.20)	(3.12)	(1.77)
Observations		[7053]	[9927]	[7865]	[6956]	[9886]	[7746]
Treat × Post	3	6.23	10.47**	5.70	0.01	0.01	0.02**
t-statistics		(1.48)	(2.29)	(1.38)	(1.15)	(1.43)	(2.18)
Observations		[4955]	[7180]	[13452]	[4701]	[7145]	[13175]

Table VIII: Increase in Mortgage-related Cognizance among Borrowers

This table reports the estimates of the treatment effect on the fraction of mortgage applications denied for credit history and debt-to-income ratio, and the in-process withdrawal ratio. The regression specification is from Equation (1):

$$Y_{icsjt} = \beta_0 + \beta_1 \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic Controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt}.$$

The outcome variables are: *%C.Hist*, *%DTI*, and *%WDR*. *%C.Hist (%DTI)* is the ratio of the number of denied applications due to credit history (debt-to-income ratio) to the total number of mortgage applications in a census tract. *% WDR* is the ratio of number of borrower-withdrawn applications before the lender reached a decision. *High-Denial Areas* are the census tracts where denial per capita in the pre-event year 2004 was more than the regional mean of denials across the census tracts. *Economic Controls* include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat×Post* interaction term captures the change in the fraction of mortgage applications denied due to a given reason in the treated census tracts relative to the control census tracts. All regressions include *Border×Year* fixed effects (FE) and *Census Tract* FE. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	All Areas		High Denial Areas		All Areas		High Denial Areas		All Areas	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	% C.Hist	% C.Hist	% C.Hist	% C.Hist	% DTI	% DTI	% DTI	% DTI	%WDR	%WDR
Treat × Post	-0.003	-0.003	-0.003**	-0.003*	-0.002	-0.002	-0.002	-0.002	-0.009***	-0.010***
	(-1.47)	(-1.52)	(-2.01)	(-1.80)	(-1.03)	(-1.17)	(-1.43)	(-1.35)	(-2.92)	(-3.95)
Economic Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (County)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R² (Adj.)	0.541	0.532	0.575	0.568	0.267	0.264	0.319	0.320	0.340	0.338
Observations	81871	76437	38737	36447	81871	76437	38737	36447	81871	76437

Table IX: First-time Homebuyers and Mortgage Interest Rates in the GSE Data

This table reports the estimates of the treatment effect on the fraction of first-time homebuyers and interest rate using the GSE data. In column (1) and (2), the dependent variable is the ratio of the number of mortgages taken by first-time homebuyers to total number of mortgages for which the information on first-time homebuyers is not missing, calculated at the zip3-state area level. The regression specification is from Equation (3):

$$Y_{zsjt} = \beta_0 + \beta_1 \text{Treat}_{zsj} \times \text{Post}_t + \delta \times \text{Economic controls} + \alpha_{zs} + \gamma_{jt} + \varepsilon_{zsjt}.$$

In columns (3) and (4), the dependent variable is interest rate on the GSE mortgages (in percentages), and the regression specification is from Equation (4):

$$\text{Interest Rate}_{izsjt} = \beta_0 + \beta_1 \text{Treat}_{izsj} \times \text{Post}_t + \delta \times \text{Controls} + \alpha_{zs} + \gamma_{jt} + \varepsilon_{izsjt}.$$

Economic Controls include annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). *Mortgage Controls* refer to *credit score* and *combined loan-to-value* (*CLTV*) in column (3), and to three additional controls, *DTI ratio*, *number of units in the property*, and *mortgage insurance percentage*, in column (4). The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable in the treated zip3-state areas vis-a-vis the control. All regressions include *Zip3−State* fixed effects (FE) and *Border×Quarter* FE. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	First-time Borrower Ratio		Interet Rate (%)	
	(1)	(2)	(3)	(4)
Treat × Post	0.011***	0.010**	0.009***	0.010***
	(2.70)	(2.47)	(13.36)	(12.06)
Economic Controls	No	Yes	-	-
Mortgage Controls	-	-	Yes	Yes
Zip3-State FE	Yes	Yes	Yes	Yes
Border × Qtr FE	Yes	Yes	Yes	Yes
Cluster Zip3-State	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.694	0.695	0.731	0.758
Observations	7593	7593	7579052	3512619
Reg. Unit	Zip3-state Aggregate	Zip3-state Aggregate	Individual Mortgage	Individual Mortgage

Table X: Effect Heterogeneity by Lenders Density

This table reports the estimates of the treatment effect on the origination volume (in 1000 USD) per adult and the approval ratio, estimated separately for census tracts having a high and low density of mortgage lenders per capita in 2004. The regression specification is from Equation (1):

$$Y_{icsjt} = \beta_0 + \beta_1 \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic Controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt}.$$

Low (*High*) identifies a census tract having a lower (higher) number of HMDA lenders than the *regional mean* number of HMDA lenders (per census tract) within the bordering counties between the given control state and all the treatment states surrounding it in 2004 (See Footnote 18). *Difference [High - Low]* shows the result of the t-test for the difference in coefficients of *Treat×Post* in specifications *High* and *Low*. The dependent variable in Columns (1) through (4) is volume of mortgages originated (in 1000 USD) per adult in a census tract. The dependent variable in Columns (5) through (8) is the approval ratio of mortgage applications at census tract level. *Economic Controls* include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable in the treated census tracts relative to the control census tracts. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Volume (in 1000 USD) per Adult				Approval Ratio			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low	High	Low	High	Low	High	Low	High
Treat × Post	0.002**	0.001	0.003***	0.002*	0.015***	0.010*	0.016***	0.009*
	(2.21)	(1.16)	(3.06)	(1.66)	(3.07)	(1.92)	(3.15)	(1.73)
Difference [High - Low]		-0.001		-0.001		-0.006		-0.007
p-value		(0.592)		(0.498)		(0.474)		(0.413)
Economic Controls	No	No	Yes	Yes	No	No	Yes	Yes
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (County)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.643	0.571	0.636	0.616	0.751	0.716	0.740	0.709
Observations	60210	25497	56188	24073	57134	24629	53135	23212

Table XI: Did Origination Increase due to Rise in Private Mortgage Securitization?

This table reports the estimates of the treatment effect on the approval ratio estimated separately for mortgages sold to non-GSEs, sold to GSEs, and not sold. The regression specification is from Equation (1):

$$Y_{icsjt} = \beta_0 + \beta_1 \text{Treat}_{icsj} \times \text{Post}_t + \delta \times \text{Economic Controls} + \alpha_i + \gamma_{jt} + \varepsilon_{icsjt}.$$

The dependent variables are the fraction of total mortgage applications originated and sold to the non-GSEs (Columns 1 and 2); originated and sold to the GSEs (Columns 3 and 4); approved and not sold by the lending institution (Columns 5 and 6). All the dependent variables are calculated at the census tract level. *Economic Controls* include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable in the treated census tracts relative to the control census tracts. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Sold to Non-GSE		Sold to GSE		Not Sold	
	(1)	(2)	(3)	(4)	(5)	(6)
	Fraction	Fraction	Fraction	Fraction	Fraction	Fraction
Treat × Post	-0.004	0.001	0.048**	0.047***	0.001	0.002
	(-0.28)	(0.05)	(2.49)	(2.78)	(0.11)	(0.52)
Economic Controls	No	Yes	No	Yes	No	Yes
Census Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
Border × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (County)	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.008	-0.003	0.003	-0.003	0.055	0.028
Observations	81871	76437	81871	76437	81871	76437

Table XII: Did Origination Increase due to Subprime Lending? Credit Score-based Evidence

This table reports the estimates of the treatment effect on the number of mortgages originated to prime and subprime borrowers by government sponsored enterprises (GSEs) Fannie Mae and Freddie Mac. The regression specification is from Equation (3):

$$Y_{zsjt} = \beta_0 + \beta_1 \text{Treat}_{zsj} \times \text{Post}_t + \delta \times \text{Economic controls} + \alpha_{zs} + \gamma_{jt} + \varepsilon_{zsjt}.$$

The dependent variable in Column (1) is *N-Prime*, the number of mortgages originated to prime borrowers (credit score$\geq$620) in a given zip3-state area. The dependent variable in Column (2) is *N-Subprime*, the number of applications to subprime borrowers (credit score$<$620) in a given zip3-state area. *Economic Controls* include the number of mortgage lenders in a census tract and annual growth rate of county income per capita, county aggregate employment, and state gross domestic product (GDP). The coefficient associated with the *Treat$\times$Post* interaction term captures the change in the dependent variable in the treated zip3-state areas relative to the control zip3-state areas. All regressions include *Zip3$-$State* fixed effects (FE) and *Border$\times$Quarter* FE. All variables are defined in Table (I). Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
	N-Prime	N-Prime	N-Subprime	N-Subprime
Treat $\times$ Post	325.63***	325.87***	11.28**	11.40**
	(3.54)	(3.44)	(2.24)	(2.23)
Economic Controls	No	Yes	No	Yes
Zip3-State FE	Yes	Yes	Yes	Yes
Border $\times$ Qtr FE	Yes	Yes	Yes	Yes
Cluster Zip3-State	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.761	0.762	0.795	0.796
Observations	7599	7599	7599	7599

Table XIII: Effect of Free Credit Reports on Banks

Panel (A) of this table reports the estimates of the treatment effect on financial performance of banks. Panel (B) reports the estimates of the originated mortgage amounts (in million $) separately estimated for the banks with an *ex-ante* low and high share of liquid assets. The regression specification is from Equation (5):

$$Y_{bt} = \beta_0 + \beta_1 \text{Treat}_b \times \text{Post}_t + \delta \times \text{Bank Controls}_{bt} + \alpha_l + \gamma_t + \varepsilon_{bt}.$$

t refers to *Year−Quarter* in Panel (A) and *Year* in Panel (B) *NIM* (Net Interest Margin) is the ratio of net interest income to earning assets (in percentages); *RoE* (Return on Equity) is the ratio of net income to book value of equity (in percentages); and *RoA* (Return on Assest) is the ratio of net income to book value of total assets (in percentages). A bank is classified as having an *Ex-ante High Liq. Share* if its share of liquid assets (of their total assets) in the pre-event year 2004 was *greater* than the cross-sectional sample mean, and as having an *Ex-ante Low Liq. Share* otherwise. Bank Controls in Panel (A) include: natural log of the total assets (in $1000); cost of deposit (ratio of total interest expense to total earning assets, expressed in percentages); and share of liquid assets in total assets (in percentages). Bank Controls in Panel (B) include only the first two of these variables. The coefficient associated with the *Treat×Post* interaction term captures the change in the dependent variable for the treated banks relative to the control banks. All regressions in Panel (A) include *Year−Quarter* fixed effects (FE) and *Bank* FE; and those in Panel (B) include *Year* FE and *Bank* FE. Standard errors are clustered by county. t-statistics are reported below the coefficients in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Financial Performance of Banks

	(1)	(2)	(3)	(4)	(5)	(6)
	NIM (%)	NIM (%)	RoE (%)	RoE (%)	RoA (%)	RoA (%)
Treat × Post	0.06***	0.06***	0.75***	0.76***	0.07***	0.08***
	(5.49)	(5.96)	(5.13)	(5.32)	(5.18)	(5.53)
Bank Controls	No	Yes	No	Yes	No	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Qtr FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (Bank)	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.807	0.814	0.586	0.597	0.556	0.573
Observations	85847	85847	85847	85847	85847	85847

Panel B: Mortgage Origination by *Ex-ante* Share of Liquid Assest of Banks

	Ex-ante Low Liq. Share		*Ex-ante* High Liq. Share	
	(1)	(2)	(3)	(4)
	Amt	Amt	Amt	Amt
Treat × Post	-159.18	-157.24	111.33*	111.35*
	(-1.29)	(-1.28)	(1.67)	(1.69)
Bank Controls	No	Yes	No	Yes
Bank FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster (Bank)	Yes	Yes	Yes	Yes
R^2 (Adj.)	0.569	0.570	0.792	0.792
Observations	10940	10940	7748	7748

Appendix

Data Appendix

The HMDA data contain 190.4 million mortgage applications over the sample period (2000–2008). These application-level data were aggregated to the *Census Tract×Year* panel in several steps. First, all observations that had state, county or census tract information missing or "NA", or state Federal Information Processing Standard (FIPS) code of "0", "00" or "0 " were dropped (2.5% of the observations), leaving 185.6 million mortgages with an identifiable county. Then, observations on three action types were removed: covered loans purchased by the financial institutions from other institutions (18.80%), as these are not borrower initiated; pre-approval requests denied by financial institutions (0.01%), as these data were included in HMDA reporting only from 2004; and pre-approval requests approved by the financial institutions but not accepted by the applicants, as these data, too, were included in the HMDA reporting only from 2004, and this reporting is not mandatory (0.025%). This leaves 150.7 million applications belonging to 77,526 unique census tracts (603,849 *Census Tract×Year* observations). Finally, with the help of the county adjacency data from the Census Bureau (n.d.), those census tracts that belong to the bordering counties of the treated and control states were selected. This led to the HMDA regression sample: 89,535 *Census Tract×Year* observations consisting of 11,942 unique census tracts of which 7,011 are treated and 4,931 are control.

The GSE data contain 33 million observations over the sample period. The property locations in this data do not contain the census tract information, but only the first 3 digits of the zip code (zip3) and state. Hence, to identify the mortgages from the zip3-states that lie within the bordering counties of the sample, the zip code-to-county crosswalk file provided by the U.S. Department of Housing was used.[28] Then, aggregating the individual mortgages to the zip3-state level and restricting the sample to only those zip3-states that lie within the sample border counties yielded 221 unique zip3-states (91 control and 130 treated) and 7,599 *Zip3-State×Quarter* observations.

Finally, the mortgage lenders in the HMDA data were matched with the commercial banks in the Call Reports (FFIEC Forms 031/041) data using lenders' Federal Deposit Insurance Cor-

---

[28] Areas delimited by 3-digit zip codes do not align with the county borders. Hence, to identify the 3-digit zip codes that lie along the county borders, first, a crosswalk file of 5-digit zip codes to county is obtained from the Office of Policy Development and Research (n.d., 2010 Q1 version). Then all such 3-digit zip codes are filtered out from the sample for which none of the underlying 5-digit zip codes lie within the bordering counties.

poration (FDIC) certificate ID, or Office of the Comptroller of the Currency (OCC) charter number (henceforth, the identifiers). Call Reports contain information on banks' identifiers and also a unique id called RSSD ID. At the same time, HMDA data contain a lender's agency code (lender's regulator) and a respondent ID. A respondent ID equals the FDIC Certificate ID if the lender's regulator is the FDIC; and it equals the OCC charter number if the regulator is the OCC.

Some HMDA mortgage lenders are the affiliates of the commercial banks, but are not banks themselves. Such lenders were matched using their parent entities (available in the HMDA Ultimate Panel data). If both an HMDA reporter and its parent entity had a successful match in the call reports, the parent's match was kept. Finally, the RSSD ID began to be directly available in the HMDA data from 2004, so the matching was done for subsequent years using this ID, instead of the combination of the agency code and respondent ID.

Know Thyself: Free Credit Reports and The Retail Mortgage Market

Online Appendix

Table A1: Sample States and Status of Selected Regulations

This table lists all the states included in the sample and their treatment and control status. For each state, it also indicates whether mortgages are recourse or non-recourse (Ghent & Kudlyak, 2011), whether foreclosures are judicial or non-judicial (Gerardi et al., 2013), and whether and when the state adopted Anti-predatory Lending laws (APL) (Di Maggio & Kermani, 2017).

State	Treatment (T) / Control (C)	Recourse (R)/ Non-Recourse (NR)	Judicial (J) / Non-judicial (NJ)	APL (Adoption Month, Year)/ Non-APL (NAPL)
Alabama	T	R	NJ	NAPL
Arizona	T	NR	NJ	NAPL
Colorado	C	R	NJ	APL (Jul, 2003)
Connecticut	T	R	J	APL (Jan, 2002)
Delaware	T	R	J	NAPL
Florida	T	R	NJ	NAPL
Georgia	C	R	J	APL (Mar, 2003)
Kansas	T	R	J	NAPL
Maryland	C	R	J	APL (Oct, 2002)
Massachusetts	C	R	J	APL (Nov, 2004)
Nebraska	T	R	J	NAPL
New Hampshire	T	R	NJ	NAPL
New Jersey	C	R	J	APL (Nov, 2003)
New Mexico	T	R	J	APL (Jan, 2004)
New York	T	R	J	APL (Apr, 2003)
North Carolina	T	NR	NJ	APL (Jul, 2000)
Oklahoma	T	R	NJ	NAPL
Pennsylvania	T	R	J	NAPL
Rhode Island	T	R	NJ	APL (Dec, 2006)
South Carolina	T	R	J	APL (Jan, 2004)
Tennessee	T	R	NJ	NAPL
Utah	T	R	NJ	NAPL
Vermont	C	R	J	NAPL
Virginia	T	R	NJ	NAPL
West Virginia	T	R	J	APL (Jun, 2000)
Wyoming	T	R	NJ	NAPL