Know Thyself: Free Credit Reports and the Retail Mortgage Market*

Amit Kumar†

[This Version: September 2021]

Abstract

Under imprecise creditworthiness information, borrowers may make erroneous credit decisions. Credit reports—which record one's creditworthiness—became free in the entire U.S. in 2005, while these had already been free in seven states. Exploiting this in a difference-in-differences setting, this paper shows that cheaper credit reports to consumers changed the mortgage market outcomes, indicating an improved borrower pool. Specifically, approval ratios and mortgage applications increased, delinquencies decreased, and the lowest-income-quartile borrowers and *ex-ante* high-creditworthy areas experienced larger increase in approvals. Additional findings, including increased interest rates, suggest a demand-driven pool improvement, as consumers receive precise creditworthiness signals from their reports.

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

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

---

*I am profoundly grateful to Utpal Bhattacharya, Emilio Bisetti and John G. F. Nash for their guidance in shaping this paper. I express sincere gratitude to Renée Adams, Daniel Ringo, and Anjan Thakor for providing valuable feedback respectively at the doctoral session of Financial Management Association 2020 annual meeting, the 2021 Junior Scholars Program of American Real Estate and Urban Economics Association (AREUEA), and the 2020 PhD symposium of the McCombs Business School at UT Austin. I am also indebted to Abhiroop Mukherjee, Deniz Okat, Arkodipta Sarkar, and Alminas Zaldokas for various suggestions and to Shashwat Alok, Vimal Balasubramaniam, Francesco D'Acunto, Andreas Fuster, Sebastian Hillenbrand, and John Mondragon for helpful discussions. I also want to sincerely thank the discussants and participants at the following meetings for helpful suggestions: HKUST brownbag seminar, AREUEA 2021 natioanl conference, Asian Bureau of Finance and Economic Research (ABFER) 8th annual conference, 2021 Annual meeting of the Swiss Society for Financial Market Research (SGF 2021), Financial Management Association (FMA) 2020 annual meeting, International Banking, Economics, and Finance Association (IBEFA 2021), 33rd Australasian Finance and Banking Conference (AFBC 2020), 11th Emerging markets conference, Southern Denmark University 5th Finance Workshop, PhD mentoring day at 28th Finance Forum (AEFIN 2020), 15th Annual Conference on Asia Pacific Financial Markets (CAFM), and 2020 Greater China Area Finance Conference.

†Hong Kong University of Science and Technology. All errors are my own. Email: amit.kumar@connect.ust.hk

Having imprecise information of their creditworthiness may lead consumers to make mistakes in their credit-related decisions. Those who *overestimate* their creditworthiness may apply for credit believing that they are creditworthy, while lenders may uncover their true creditworthiness during the application assessment using consumers' credit reports and reject the applications. The phenomenon appears, in particular, in the mortgage applications in the U.S. For example, over 2000 to 2008, about 34% of the mortgage applications were rejected, and among these, the ones rejected for having a bad credit history were twice of those rejected for having a high debt-to-income (DTI) ratio. A lack of perfect information on one's creditworthiness is possibly the main underlying cause, because otherwise such consumers would have opted to not apply and would have avoided the rejection costs, which are severe. In addition to the application costs, which could be $300 to $400, a rejection likely raises the rejection probabilities of all future applications for credit of *all* types and the associated interest rates.

On the other hand, consumers who *underestimate* their creditworthiness, likely due to not having perfect information, may end up not applying for credit despite needing it, because they anticipate that they would be rejected. Referred to as the *discouraged borrowers*, such consumers suffer the opportunity cost of not having the credit. With reference to the mortgage credit, almost 13% of the respondents in the Survey of Consumer Expectations (2013–2020) (SCE) report as being discouraged, whereas with reference to all types of credit, about 15% of the respondents in the Survey of Consumer Finances (1998–2007) (SCF) identify themselves as being discouraged. Moreover, lower usage of credit reports and unawareness of one's credit scores are associated with consumers being discouraged (SCE).

With these frictions in mind, this paper examines the link between consumers' economic cost of accessing their credit reports—an authoritative record of creditworthiness information—and mortgage market outcomes. The key finding is that when the economic costs of the reports are lowered, demand for mortgage and application approval ratios increase. The higher approvals appears to be driven by an improvement in the borrower pool, as the new mortgages flow to more creditworthy areas and to prime consumers, and their delinquency rate declines.

This paper proposes a simple consumer *self-learning* mechanism to explain the findings. Credit reports aid consumers self-assess their creditworthiness more accurately as they contain crucial information on consumers, e.g., their creditworthiness, credit history and borrowing capacity (Figure I), and lenders use these reports in assessing the credit applications. With the help of the reports, consumers can effectively incorporate their self-assessed creditworthiness information into their credit applications. Those whose reports signal high creditworthiness may apply for credit (or, enter the market), while those whose reports signal bad creditworthiness may either search for a suitable (subprime) lender, or do not apply for credit (exit the market). In effect, the pool of borrowers improves due to better self-sorting among consumers and results in higher approval ratio. Whether the credit demand increases depends on the prior distribution of over- and under-estimators of creditworthiness and on the fraction of consumers who are unaware of the role of the reports in credit applications.

Credit reports are not costly. In fact, historically their cost has been just about $8, yet a mere 1.6% of the approximately 1 billion reports generated annually in the U.S. in early 2000's was requested by consumers (Avery, Calem, & Canner, 2004). If we assume that all these reports were requested by the potential mortgage applicants, less than 5% of the total mortgage applications in 2004 could have come from consumers who had checked their credit reports. Such low usage are likely contributed by many factors, such as search costs (where to obtain the reports 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 to use it in credit decisions?).[1] While the reports themselves are trivially cheap, not using them may leave consumers with imprecise creditworthiness information, which in turn may lead to wrong self-assessment of creditworthiness (Perry, 2008), debt underestimation (Brown, Haughwout, Lee, & Van Der Klaauw, 2011), and worse financial outcomes (Courchane, Gailey, & Zorn, 2008).

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 allowed access to three annual free credit reports 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.[2] So, consumers from all states except these seven pre-FACTA states saw a close-to-exogenous reduction in economic costs of the reports from 2005 onward. Now consumers could access reports with just a few

[1] 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 (Federal Trade Commission, January 13, 2000).

[2] The details of enactment of the local laws by the pre-FACTA states are: 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.

clicks, whereas earlier it could take a week for the report to arrive after successfully making a request. The act also potentially raised general awareness about the reports as consumer interest in free credit reports measured using Google Search Interest for the keyphrase *"Free Credit Reports"* heightened in the treated states relative to the control. All in all, the experiment likely reduced the monetary, awareness and search cost of accessing the reports.

A difference-in-differences (DID) research design is set up around the above natural experiment to identify the causal effects. The control group consists of the all the pre-FACTA states except Maine, since its local law and FACTRA were enacted in the same year, and the treatment group consists of 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 event year is 2005, sample period spans 2000 to 2008, and the DID estimator is two-way fixed effects (TWFE) estimator.[3]

This empirical design carefully alleviates the endogeneity in the assignment of the treatment and the control groups. The "treatment" is assigned to all the states by the FACTA, a federal law, whereas the "control" is assigned to the pre-FACTA states by the local laws that had been enacted well before the event took place. Thus the treated states did not opt to become treated, but were mandated to do so. Moreover, the FACTA enactment in 2003 does not appear to be an endogenous response to the prevailing economic conditions. It was In fact not an entirely new law. Its provisions were consolidated from an existing federal law, the *Fair Credit Reporting Act of 1970* (FCRA), which was set to expire in 2003 (via its amendment in 1996). Thus, it is when the FCRA was to expire that the Congress enacted the FACTA, and the motivation behind this was to perpetuate the expiring provisions (Nott & Welborn, 2003), not the prevailing economic conditions. Also, since many of the FACTA provisions except the free credit reports were already in place at the time of the event, they do not create confounding effects.

In a bid to separate the confounding effects of the local economic conditions from the treatment effect, the sample restricts its focus to a narrow geographic area consisting of *only* the counties at the border between the treated and control states, similar in spirit to the strategy used in Huang (2008) and Dube, Lester, and Reich (2010). Furthermore, the key outcome vari-

[3] Since this design is not a staggered treatment design, but 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. More details appear in Section (2), Empirical Research Design.

ables are analyzed at the census tract level, a sub-county micro area that roughly encompasses a population of about 4,000, allowing to flexibly sweep out any regional economic differences.

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. The equivalent dollar amount aggregated across the treated bordering counties is about $5.5 billion due to the former effect and about $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, higher applications indicate that mortgage borrowers on average tend to underestimate their creditworthiness, a finding that is in contrast to the commonly understood belief that consumers tend to overestimate and that is explored in detail later.

Were the investment-seeking borrowers behind the increased demand, as was the case during the financial crisis that unfolded just three years after the event? Tests focusing on the owner-occupied and non-owner-occupied mortgages suggest otherwise. While the non-owner-occupied mortgages saw slight uptick of 1 percentage point as the fraction of total and successful applications, the bulk of increase seems to have come from occupancy-seeking borrowers.

If the increase in mortgage origination and approval ratio was a result of improved borrower pool, the delinquencies would decrease, or at least, would not increase. Over a six-year period since origination, the mortgages originated in the year after the event in the treated areas were indeed *less* likely to become delinquent than those from the control areas; whereas those originated in the pre-event year in the treated areas were just as likely to become delinquent as those originated in the control areas.

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 more 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 in the treated areas vis-à-vis the control areas but the number of applications did not, and the opposite occurred for the higher quartiles. As lower income is associated with overestimation of one's creditworthiness (Perry, 2008), the consumers in the lowest quartile likely corrected this after the event by not applying for mortgages (exiting the market), raising the approval ratios but not the applications. Recall also that the applications increased overall—not to conclude that the mortgage borrowers tend to underestimate their creditworthiness on average, but that the increase in applications was largely driven by the higher-income consumers, who have been shown to be less likely to overestimate.

In support of the self-learning mechanism, several mortgage application and rejection patterns changed indicating an increased mortgage-related cognizance among the borrowers. 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 debt-to-income related denials, pointing to an increased learning among borrowers about their credit history. *Then*, 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. *Finally,* the treated areas saw a 1 percentage point increase in the mortgages by the first-time homebuyers as the fraction of the mortgages originated, indicating new entry by creditworthy borrowers.

It seems unlikely that the increase in the mortgage origination was supply driven, since 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 did not increase.

All in all, imprecise information of creditworthiness among consumers likely play a role in frequent mortgage denials due to credit history and in discouraging potential borrowers from applying for mortgage. Reducing consumers' economic costs of credit reports seems to have improved the outcomes on both these fronts: the borrower pool improved and demand for mortgages increased. As these findings are causal in nature, a policy intervention aimed at educating consumers of their creditworthiness may yield similar results. Also, even though the conclusions in this paper are based on the mortgage market, the findings are relevant to consumer decisions for all credit types.

This paper primarily relates to the literature on effects of information provision on credit market participants. This is the first paper to show that consumers' economic costs of credit reports, when lowered, leads to improved mortgage market outcomes in a way consistent with improved borrower pool. In a field experiment, Homonoff, O'Brien, and Sussman (2019) find that borrowers are less likely to default when provided with information on their FICO® scores. Similarly, bank customers who enroll in free FICO scores program are less likely to default, reduce credit utilization, and increase credit card spending (Mikhed, 2015). Moreover, when financial product disclosure is increased, sophisticated borrowers default less, and when financial product is standardized, unsophisticated borrowers default less (Kulkarni, Truffa, &

Iberti, 2018). Also, consumers increase mortgage borrowing after bankruptcy flags from their credit reports are removed (Dobbie, Goldsmith-Pinkham, Mahoney, & Song, 2016). It also leads to aggregate welfare loss, cheaper credit for poorer defaulters, and expensive for poorer non-defaulters (Liberman, Neilson, Opazo, & Zimmerman, 2018).

This paper also relates to the extensive literature on financial literacy. Low financial literacy leads to detrimental economic outcomes, such as high mortgage delinquency and foreclosure (Gerardi, Goette, & Meier, 2010), poor mortgage choice (Moore, 2003), and large debt (Lusardi & Tufano, 2009; Stango & Zinman, 2009). Further, field experiments reveal that less financially literate distressed borrowers benefit less from loan-modification contracts (Hundtofte, 2017), and educational intervention improves consumers financial product purchases (Balakina, Balasubramaniam, Dimri, & Sane, 2020). This paper shows that providing free credit reports results in increased mortgage demand and lower delinquencies.

The rest of the 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

The FCRA governed consumer credit information-related laws before FACTA. Even under the FCRA, consumers had the right to see the contents of their credit reports, except for the credit score, under specific and restrictive provisions. For example, a consumer could receive a free report if he/she made a request within 60 days after receiving a notice of an *adverse action* taken against him or her on the basis of the information in the credit report (Avery, Calem, Canner, & Bostic, 2003).[4] The 1992 amendment to the FCRA mandated that the cost of disclosure of credit information should be reasonable, while that in 1996 capped the cost at \$8. The latter also provisioned the law to lapse in 2003.

[4] 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. Consumers can receive credit report free of charge once in 12 months if he or she makes a request to the CRA for the credit report and certifies that: (A) she/he is unemployed and intends to apply for employment in the 60 day period beginning on the date on which the certification is made; (B) she/he is a recipient of public welfare assistance; (C) she/he has reason to believe that the file on the consumer at the agency contains inaccurate information due to fraud.

Even though the FCRA allowed free credit reports at the federal level under specific circumstances, consumers rarely proactively requested their credit report for own use. Out of approximately 1 billion credit reports generated annually, only 1.6% were disclosed to consumers (Avery et al., 2004). Of these 1.6%, only 5.25% were proactively requested by consumers, while 94.75% were disclosed to consumers under the FCRA provisions mentioned earlier (Nott & Welborn, 2003).[5] Thus, only 0.084% of all credit reports generated were disclosed to consumers as a result of their own request.

While the FCRA was taking shape at the federal level, many states enacted local laws providing more transparency in credit reporting by allowing residents free credit reports. For example, through the Senate Bill 133 in 1997, Colorado enacted its free credit report law on April 21 of that year. The free credit report provision appears in Section 4, paragraph (E) of this bill, which got added 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.*

Meanwhile, the FCRA was amended in 1996, a sunset clause was added to it mandating its expiry in 2003. In order to make its provisions permanent, the FACTA was enacted on December 4, 2003, with a new key provision added: free annual disclosure of credit reports to consumers by each of the three national credit reporting agencies.

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, the treatment group. Panel (A) of Figure (II) shows these states. The regression sample consists of only the counties lying at the borders of these focal states, shown in Panel (B) of the figure. The event is year 2005, when www.annualcreditreport.com was established to distribute the free credit reports.[6] Sample period is from 2000 to 2008, as extending the sample until 2008 allows for enough post-experiment observations.

[5] Breakdown of the 94.75% credit reports disclosed under FCRA provisions: 84% due to *adverse action*; 11.5% due to fraud claim; 0.4% due to unemployment, 0.1% due to consumer being on public assistance.

[6] The website was rolled-out in four phases over two months, from Dec 2004 to Jan 2005.

Contiguous-county design across state borders provide one of the most compelling identification strategies (Allegretto, Dube, Reich, & Zipperer, 2017), because idiosyncratic trends may not vary widely across neighboring areas and macroeconomic shocks affect them roughly at same time (Dube, Lester, & Reich, 2016). Similar research designs have been used in Huang (2008) and Dube et al. (2010). A synthetic control matching procedure too is a viable approach, though Allegretto et al. (2017) point out that it too may place greater weights on *nearby areas*.

The DID estimator

We utilize the two-way fixed-effects (TWFE) estimator, specified below:

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

where Y_{icsjt} is the outcome variable for a census tract i from a county c lying at the border between treatment state s and control state j. Recall that there are seven control states, thus j ranges from one to seven. t indexes years 2000–2008. $Post_t$ 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 is 1 for those from treatment states s. β_1, the coefficient of interest, captures the change in the dependent variable in the treated counties relative to the control counties after the effect. 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.

All regressions are estimated *with and without* co-variates. Referred to as *Economic Controls* in the equation, the co-variates, when used, include a host of time-varying county- and state-level variables capturing local economic and credit conditions. These are—annual growth rate of county's income per capita, county's aggregate employment and state's gross domestic product (GDP), and number of mortgage lenders (in log) in a census tract.

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 encompassing just about 4,000 population. As census tracts are smaller geographic area than a county or a state, the associated fixed effects account for all state-level differences, including electoral landscape, recourse/non-recourse status etc.

$\gamma_{j,t}$ represents "*Border$\times$Year*" fixed effects, the second of the two-way fixed effects. Here j refers to the border of a control state j. Owing to the interaction with year, these allow for a region-specific time trends that flexibly and robustly account for any time-varying regional shocks affecting bordering states.[7] Thus, together with the time-varying economic controls, these fixed effects are able to reasonably account for the confounding effects of local economic shocks on the outcomes of interest, and allow us to cleanly estimate the desired treatment effect.

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 units (Goodman-Bacon, 2021). Since the current paper does not use a staggered DID design, but rather a *single-treatment* DID design, the issue of heterogeneous treatment effects across time does not arise.[8] The second issue of treatment effect being heterogeneous across treated units is a noteworthy limitation. As the key estimates in the paper are robust in the sub-samples formed by removing one at a time each of the control states (together with respective surrounding treated states), the results remain supportive of the conclusions despite this limitation.

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. This requirement largely holds as the timing and circumstances of the FACTA enactment appear unrelated to the states' actions, as discussed earlier.

In the end, the key assumption the TWFE relies on is parallel-trends: the treated states would have had similar trends as the control states in the absence of the treatment. Though the assumption is unverifiable, Panel (A) of Figure (III) plots the trend of 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 *Approval Ratio* according to the

[7] Consider a 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 (j), thus are grouped as one unit. Thus, the fixed effects only utilize the variation in the outcome variables *within* each such contiguous geographic areas.

[8] The staggered adoption of the local laws by the early states (control group) over 1992–1997 too is not an issue since at least eight years have passed since the last adoption. Had the gap between the adoption by control group and the treatment been small, the dynamic treatment effect on the control (early-adopters) would have co-evolved with that on the later-treated units, compromising the parallel-trends assumption.

following specification:

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

where $Event_k = 1$ *if* $t = T - k$. $Event_k = 0$ *if* $t \neq T - k, k = \{-3, 4\}$. 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) together provide reasonable assurance that the parallel trend assumption is satisfied in the current setting.

Salience of the natural experiment

Whether the natural experiment was a salient event for consumers warrants validation, and some evidence suggest that their interest in free credit reports heightened after the experiment. The examination of the Search Interest data from Google Trends supports this. *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)).[9] *Second,* the plot of differential search interest across the treated and control states using the Interest-by-subregion Google Trends data suggests that consumer interest in free credit reports heightened in the treated states in the year of website's establishment. Panel (B) of Figure (V) shows the mean of the popularity rank for the two groups each year from 2003 to 2008. We see that the keyphrase was equally popular in both the treatment and control states in the pre-event year 2004, but it became more popular in the treatment states in 2005.[10] Also, some anecdotal evidence suggest that the website issued about 52 million credit reports to consumers in the first two years (Wikipedia, n.d.).

A seemingly obvious conjecture for the DID design to work is that higher cost of credit reports leads to its lower usage. Given that a report costs only around \$8, one may wonder

[9] The Google search interest index represents the degree of "search interest" for the 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] For the cross-sectional plot of the search interest, first data spanning one-year intervals were obtained, and then the mean was calculated within each interval for the two sets of states. From this plot, it may appear that control states' popularity after the event decreased. However, this occurs because the popularity measure is essentially a yearly ranking of states, 100 being the most popular; so an increase in rank of one state mechanically decreases rank of other.

whether this is large enough to keep consumers from obtaining their credit reports. Note that the economic costs of accessing credit reports may still be high for consumers, especially for financially less knowledgeable ones, as economic costs include the knowledge whether such reports exist, where to obtain these from, and what role do they play in a mortgage process.

The credit reports usage data do support the above conjecture. The usage were low in general, and were even lower in the treatment states before FACTA was enacted. *First,* recall that only 0.084% of about 1 billion credit reports issued annually are consumer-requested. Assuming each of these reports to be requested by a separate consumer, 0.84 million consumers accessed their credit reports, while the annual mortgage applications around 2004 numbered about 16.8 million. So, even if all these reports were requested by mortgage-seeking borrowers, *less than 5%* of the mortgage applicants had checked their credit reports.[11] *Second,* the data from testimony in the U.S. senate hearing confirms the stark difference in the usage of credit reports in the pre-FACTA states and the rest. Relative to the national average, its usage was 250% higher in GA, 204% higher in MD, 153% higher in CO, 35% higher in NJ, and 25% higher in MA (U.S. Senate. 108th Congress, 2004a).[12] *Third,* the SCE data suggest that even a decade after credit reports became free, the fraction of consumers who report being unaware of their credit score is ~12%, and those who checked their report either never or at least more than two years ago is ~20% (see the Baseline Results on survey evidence for details).

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 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 requesting for reports, and waiting for it to arrive through mail. In fact, in 2000, the three CRA's settled a lawsuit by 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).

[11] This may be why the Federal Reserve Board actively encourage consumers to check their credit reports—*It can be especially helpful to see a copy of your credit report before you apply for, say a car loan, a mortgage, or a credit card*(Federal Reserve Bank of Philadelphia, 2015).

[12] Not only did the pre-FACTA states have higher usage of credit reports, but they 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 (U.S. Senate. 108th Congress, 2004b).

Finally, the effects documented in this paper are conservative measures of the true effect as the treatment here is of intention-to-treat (ITT) nature, since consumers who wanted to get their credit reports could do that before FACTA too, but at higher economic costs. Then, since the DID comparison here is between late- and early-treated units, the effects may also be referred to as the *average treated effects on the late-treated* (ATT-LT).

3 Data and Summary Statistics

This paper primarily draws on three publicly available data: the *Home Mortgage Disclosure Act of 1975* data (HMDA data) for details on mortgage applications; the Federal National Mortgage Agency (Fannie) and the Federal National Home Loan Mortgage Corporation (Freddie) data (GSE data) for details on mortgage performance; and the Call Reports (FFIEC Forms 031/041) for details on financial performance of commercial banks. HMDA data provide 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] GSE data pertain to the 30-year fixed rate single family mortgages, the most popular mortgage type in the U.S. These data record mortgage-level information on interest rate, debt-to-income ratio, credit score, first-time homebuyer status, investment purpose, the first 3 digits of the zip code (zip3) of the mortgaged property etc. Finally, call reports contain detailed financial data of the U.S. commercial banks.

The detailed steps to process each of these data and to link with one another are provided in Appendix (A.), Data Appendix. The regression sample utilizing HMDA data include mortgages taken for all three purposes—home purchase, refinance, and home improvement—and for all three types—conventional loans, loans guaranteed by Veteran Administration (VA) and Farm Service Agency (FSA)/Rural Housing Administration (RHS), and loans insured by Federal Housing Administration (FHA). These application-level data are aggregated to "*Census Tract*$\times$*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*$\times$*Year*" ob-

[13] Until 2003, the census tracts in HMDA are from the Census 1990 definition, while those from the 2004 onward are from 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 Census Bureau (2006). Though necessary, this process is an approximation, and thus introduces some noise in the measurements. Yet the measurement noise 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.

servations. Similarly, the mortgage-level GSE data is aggregated to the zip3-state level, leading to 221 unique zip3-states (91 as control and 130 as treated) and 7,599 "$Zip3-State \times Quarter$" observations.

A few more data sources are used. The data on consumers' credit-related information are taken from Survey of Consumer Expectations (2013–2020) Credit Access Survey, a Federal Reserve Bank of New York rotating panel survey fielded since 2013 over internet every four months. The quarterly data on county subprime population comes from FRBNY and Equifax (n.d.). Data on county-level employment comes from the annual survey of County Business Patterns (CBP) (Census Bureau, 2000–2008). To map the zip code level variables of CBP to census tracts, (Missouri Census Data Center, 2010) is used. Finally, data on state level economic conditions are taken from the Bureau of Economic Analysis, and on population characteristics at census-tract level from Census 2000 (Manson, Schroeder, Van Riper, & Ruggles, 2019).

The key outcome variables of interest are the number of mortgage applications per 1000 adults in a census tract (*scaled applications*) and approval ratio. 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. Other variables of interest are the fraction of total applications denied for credit history or debt-to-income ratio and the fraction of total applications withdrawn by applicants while still under processing.

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 debt-to-income ratio.[14]

Panel (B) of Table (I) shows the comparison of the treatment and control groups in the pre-treatment period using t-test for difference in mean, the p-value 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

[14] We see that the four ratios—the approval ratio, the two denial ratios, and the 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, denial reasons could be other than credit history or debt-to-income ratio. Third, an application might be denied for multiple reasons.

can accommodate pre-existing differences between the treatment and control subjects so long as they satisfy the *parallel-trends* assumption, the concern stands mitigated.

4 Results

This section first shows the findings on credit reports usage and discouraged borrowers. The discussion of baseline results regarding effects of free credit reports on mortgage demand, approval ratio, house prices, and delinquency rates follows. Then, results highlighting the self-learning mechanism and the role of supply- and demand-side factors are discussed.

§A Baseline Results

Survey evidence on the Credit Reports Usage and Discouraged Borrowers

The SCE Credit Access Survey is a representative U.S. consumer survey covering many aspects of credit usage. The rotating panel nature of the survey allows for inferences about the population, aided by "Year-month" level fixed effects, standard error clustering, and sampling weights.

Columns (1) through (3) of Table (II) speak to the usage of credit scores and reports among the U.S. consumers. Column (1) shows the result from regressing the dummy indicator of whether a respondent has never checked or requested credit report on a constant; an estimated 8% of the population falls into this category. Similarly, the indicator variable in column (2) is 1 if respondent has either never checked, or checked it at least more than two years ago (infrequent checkers); a staggering 20% of the population belong to this category. Finally, in column (3), the indicator variable is 1 if a respondent reports that he/she does not know his/her credit score; an estimated 12% of the population does so.

One of the questions in the survey asks, how likely the respondent is to take mortgage and related credit in next 12 months; those who are very/somewhat unlikely to do so, or those who assign less than 10% probability to it, are asked for the reason. Classifying those who respond "I don't think I would get approved" as the *discouraged borrowers*, and then regressing this dummy on a constant gives an estimate of discouraged borrowers as the fraction of those unlikely to apply for mortgage credit. The coefficient in column (4) suggests that the proportion of discouraged borrowers is about 13%. Moving further, whether being discouraged is linked with consumer's usage of credit reports and scores is examined next. The indicator for discouraged borrowers is regressed separately on dummy variable for credit report usage and one's unawareness of credit score. Column (5) and (6) suggest that infrequent checkers and those unaware of their credit score are respectively 3% and 5% *more likely* to be discouraged.

In summary, lack of credit reports/scores usage and unawareness of one's creditworthiness are non-trivially prevalent, and among other things, these have consequences for being discouraged from applying for credit.

Effect of free credit reports on mortgage approval ratio, applications, and house prices

The results on the effects of lower economic costs of credit reports on mortgage approval ratio, number of (scaled) applications and house prices follow next. Each outcome variable is measured at the census tract level. The regression specification used here is from Equation (1); the coefficient of interest is "*Treat*$\times$*Post*", which estimates the change in the outcome variable in the treated areas relative to the control after the event. All specifications include $Census\ Tract$ and "*Border*$\times$*Year*" fixed effects.

Columns (1) and (2) of Table (III) show the regression results for approval ratio; the former is a plain DID without any co-variates, the latter is with controls for local economic conditions included, i.e. 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*$\times$*Post*" suggest that the ratio increased by about 1 percentage point in the treated tracts. It may seem trivial, as the mortgage approval ratios are commonly believed to be high, in the upwards of 80%, but in the sample counties, the average ratio in the pre-event period is just 52%. In real 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 $2.75 billion more successful mortgages, aggregated across the treated bordering counties.[15]

Recall that lower economic costs of accessing credit reports does not affect the information lenders have on borrowers. Their access to the reports saw no change under the FACTA, neither the extent nor the scope of information contained in these reports. In fact, any factor relevant to mortgage decision, i.e. borrower's income, employment, collateralizability of their assets etc., saw no change due to the law. What is most likely to change is borrowers knowledge of their credit history and other information that *lenders* utilize to evaluate them on. This may

[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 $\times$ 0.01 $\times$ 2.7 thousand adults per treated tract), about 18,229 more successful applications across the treated bordering counties (2.6 applications $\times$ 7,011 treated tracts), or a *~$2.75 billion increase in mortgage origination across all bordering treated tracts* (18,229 $\times$ $150,597 average mortgage amount per application).

lead to new entry by creditworthy borrowers and better matching with lenders among existing borrowers (creditworthy borrowers to prime lender and/or less-creditworthy borrowers to sub-prime lenders); all of which predict an increase in approval ratio. The later sections evaluate some of these claims rigorously, especially whether the increase is lender driven.

Columns (3) and (4) show the results for scaled applications. The scaled applications increased in the treated tracts 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 in applications roughly translates to \$37.8 billion increase in mortgages, aggregated across the treated bordering counties.[16] The increase in applications suggests that on average consumers tend to underestimate their creditworthiness when it comes to mortgage borrowing, a finding in contrast to other financial decision-making settings (Perry, 2008).

To ensure that the results are general and are not driven by some specific states, the above regressions are re-estimated over sub-samples formed by removing one at a time each of the control states and their surrounding treated states. 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 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 census tract-level house price index from Bogin, Doerner, and Larson (2016), the index 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.

It is worthwhile to point out a that the post-event sample period of this study includes the years from 2005 to 2008. During this period, the housing market was volatile, and mortgage supply in the U.S. had started to shrink since 2005. Thus it would be injudicious to claim that the effects estimated above are completely uncontaminated with these changes. However, the difference-in-differences design ameliorates this issue to the extent that the market-wide forces

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

evenly affect the neighboring counties across states. Furthermore, even when the post-event sample is restricted to 2006, all the conclusions hold (see Section (§E), Robustness).

It is also understood that one reason for the financial crisis was excessive mortgage borrowing by borrowers without means, often for investment motives vis-à-vis occupancy motives. Whether such borrowers are behind the increased origination in the current setting is important to examine. Table (IV) examines this by focusing on the changes in owner-occupied and non-owner-occupied mortgage category using the same DID specification. The dependent variable in columns (1) and (2) is the number of mortgage applications for the former category, and columns (3) and (4), for the latter. We see 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. Investigating further, columns (5) and (6) examine the composition of non-owner-occupied category as a fraction of total applications, and columns (7) and (8) as a fraction of successful applications. The coefficients in these four columns suggest a modest 1 percentage point increase in non-occupancy mortgages at both the application and origination stage. All in all, the investment-motivated demand does show a slight uptick, but does not appear to be a dominant reason behind the robust 15% increase in the mortgage applications.

Mortgage delinquencies

We saw that lower economic cost of credit reports resulted in the higher mortgage origination. The question 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 and the borrower pool stayed the same as before, the delinquencies would rise.

To examine the patterns in the delinquencies, this paper relies on the GSE data, which are a subset of the HMDA data. First, define a mortgage vintage as a set of mortgages originated in a given area—treated or control—in a given year—2004 (pre-event) or 2005 (post-event), giving us a total of four vintages: the treated vintage in the pre- and post-event year, and the control vintage in the pre- and post-event year. Then define the rate of delinquency of a given vintage as the ratio of the number of mortgages that miss 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. Finally, plotting the rate with the age for different vintages allows one to examine the underlying trends.

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 for the mortgages originated before the event, the delinquency rate of the treated and control vintages follow almost the same trend; whereas the plot on the right reveals that for the mortgages originated in the year after the event, the delinquency rate of the treated vintages is *lower* than that of the control vintage. The same pattern is observed for the 60–89-day delinquency rate, as Panel (B) of the figure shows. 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 as the delinquency rates would stay the same if lenders use the same screening policy after the event as they were using before the event. However, this reasoning assumes that the composition of the pool stays the same, whereas Section (§C.2) shows that 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 it appears to be contributed by an increase in the share of the first-time homebuyers.

§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 on creditworthiness is likely to bind, perhaps useful for policy targeting. This can be achieved by examining the effect heterogeneity across pre-event characteristics; specifically, the borrower creditworthiness and income.

[17] Suppose that lenders deny applicants whose *ex-ante* probability of default falls above some threshold, p^*. Assume that 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.

§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 *more* increase in applications and approvals in areas where 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 following Mian and Sufi (2009), who suggest that such classifications should be done at a time well before the start of the housing boom as creditworthiness of an area endogenously evolves with the housing boom. The earliest year the data on county subprime fraction, (FRBNY & Equifax, n.d.), is publicly available is 1999.

Table (V) shows the results of regressing scaled applications and approval ratio separately using regression Equation (1) for counties with high and low creditworthiness. Columns (1) and (2) show that scaled application increased by 16.8–18.8 (17.4–19.5%), while columns (3) and (4) show that approval ratio increased by 2 percentage points in treated *ex-ante* high-creditworthiness counties vis-à-vis control counties with similar creditworthiness. The contrast emerges in the coefficients in columns (5) through (8), which estimate the effects in *ex-ante* low-creditworthiness treated areas with respect to control areas of similar creditworthiness. The increase is far smaller and barely statistically significant: 8.5–11.6 (8.8%–12%) for scaled application and 1 percentage point for approval ratio. The results support the self-learning mechanism and suggest that creditworthy borrowers are more likely to benefit from easier access to credit reports.

§B.2 Heterogeneous effects by income level of borrowers

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 the subprime lenders is higher for the low-income borrowers vis-à-vis the

[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. 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. 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.

high-income borrowers. This conclusion remains the same in the over- and under-estimation framework. As lower income is associated with a higher likelihood to overestimate one's creditworthiness (Perry, 2008), the downward revision of creditworthiness is more likely to occur for the low-income consumers, leading them to not apply for credit and avoid the rejection costs.

Similarly, the increase in approval ratios across the income groups may differ too. As the marginal propensity to lend to high-income consumers is more (Agarwal, Chomsisengphet, Mahoney, & Stroebel, 2018), the scope of improvement in the borrower pool is smaller for this group than for the low-income group. Thus the ratio is more likely to be high for the low-income consumers.

These predictions are tested next. The cut-offs for the income quartiles are calculated in-sample each year, and then applications in each quartile are aggregated at the census-tract level. 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 it increased significantly for the other three, and the increase was larger for these quartiles. These findings are in line with the prediction. As not applying (exits) are more likely to be a rational choice for low-income consumers, the rise in number of applications would be small for them.

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 only for the lowest income quartile, again consistent with the intuition.

The earlier finding that the demand for the mortgages increased seems to suggest that on average mortgage borrowers tend to underestimate their creditworthiness, whereas a commonly understood belief is that consumers tend to overestimate creditworthiness (overconfidence bias). This contrast appears to be driven by the heterogeneity in the effects across borrower's income. Perry (2008, Table 3) shows that higher income is associated with a *lower* tendency to overestimate, and hence, if the demand is driven largely by the higher-income consumers, we are less likely to see overestimation and more likely to see *underestimation*.

§B.3 Heterogeneous effects for the overestimating borrowers

The theory 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 these predictions 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, there would not be any data on the underestimating type, as borrowers of this type would not have applied for the mortgages in the first place.

Notwithstanding the above limitations, it can be claimed that the borrowers from the areas where the *ex-ante* rejection ratios due to DTI ratio were small but due to credit history were large are more likely to be the over-estimators than the borrowers from the other areas. This is because these rejection patterns fit the borrowers who may have mistakenly overestimated their creditworthiness and ended up applying for a mortgage and were more likely to get rejected for their (bad) credit history than for their repayment inability. Following this reasoning, the next set of regressions separately estimate the effect for scaled applications and approval ratios using Equation (1) for the nine sub-groups formed by sorting the census tracts into *tertiles* using the rejection ratios in 2004 for DTI ratio and credit history.

Table (VII) shows the results of the regressions: Columns (1) through (3) for the scaled applications and Columns (4) through (6) for the approval ratios. Credit-history tertiles vary from top to bottom of the table and DTI tertiles, from left to right. The sub-group of interest corresponds to the third tertile of the credit history and the first tertile of the DTI ratio (the overestimating type). We see that the treated areas saw the smallest increase relative to the control areas for the borrowers likely to be of the overestimating type (6.23 versus 10.67 or 20.32 within the first DTI tertile). This pattern is in line with the intuition that the applications are expected to decline for the overestimating type after the event.

Theoretical prediction for the approval ratio is ambiguous: the ratio should *increase* for both the borrower 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) seems to hold in the data.

§C Mechanism: Consumer Self-learning Channel

The support for the self-learning mechanism is evident from the remarkable patterns that emerged in the treated areas after the event: increase in mortgage-related cognizance among borrowers and entry of new borrowers.

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

The information in credit reports undoubtedly influence lender's assessment of a mortgage application, but just the fact that consumers could access the reports easily after the FACTA would not cause material changes in the market-level outcomes unless reasonably large number of consumers start to access the reports and use the information contained therein in their credit decisions. As the reports contain credit history of consumers, if consumers become more cognizant of this, there should be a decrease in rejection due to credit history, while there should be no change in rejections due to debt-to-income ratio.

Similarly, the increased cognizance would also affect the in-process application withdrawals. Borrowers tend to initiate multiple formal mortgage applications at several lenders because of the uncertainty in approvals and mortgage terms. They incur multiple non-refundable application costs, and in the end, they take the mortgage out with one lender while withdraw from the others (in-process withdrawals).[19] Thus, increase in borrower cognizance of their creditworthiness should result in reduced fraction of in-process withdrawals.

The first prediction can be tested by regressing the fraction of total applications rejected for a given reason. Noting that the value of information in the reports is high when rejection rates are high, these outcomes are estimated for all the census tracts as well as for only those that had *ex-ante* high rejection rates. The latter are the tracts where the rejection rate in the pre-event year 2004 was higher than the *regional mean* (Footnote (18)). Another reason to separately focus on the high rejection areas is that the rejections-for-a-given-reason would not be affected if there are too few rejections in the first place.

[19] The withdrawal ratio was about 12% over the 2000–2008 period. Some anecdotal evidence suggest that consumers tend to withdraw application when they find a better offer from other lenders (Reddit Forum, n.d.). More importantly, credit reporting agencies do not penalize multiple applications if those are made within a short time period. According to Equifax (n.d.), "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."

The second prediction is tested by regressing the fraction of total applications that are formally withdrawn by the borrower before the lender has reached a decision. The specification for these regressions is the same as the one used in the baseline results, the Equation (1).

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 only in the *ex-ante* high rejection rate areas (columns 3 and 4). The coefficients in columns (5) through (8) show that the debt-to-income ratio denials did not decrease statistically significantly. These results, despite having modest statistical significance, indicate that the rejection owing to different reasons changed in a manner consistent with consumers learning more about their credit history.[20] As explained before, borrowers in the *ex-ante* high-rejection areas would value the information in the reports more.

The estimates for withdrawal ratio appear in columns (9) and (10) and imply a reduction of 0.9–0.11 percentage points in the treated tracts vis-à-vis the control.[21] Overall, both the patterns point to an increase in mortgage-related cognizance among borrowers, consistent with the self-learning mechanism.

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

About 15% of households in the SCF survey of early 2000's and 13% of the respondents in the recent SCE surveys report as being *discouraged*, and this tendency is associated with lack of knowledge and usage of credit reports/scores. With access to credit reports becoming easier, the credit applications may increase. Also, the self-learning mechanism predicts that to the extent consumers overestimate rejections and do not apply for credit, there would be entry of creditworthy borrowers into the credit markets. The proportion of first-time homebuyers in the mortgage market allows to examine this.

The information on first-time homebuyer is available in the GSE data, though not in the HMDA data. With reasonable approximations (see Data Appendix, Footnote (28)), the property location information in the GSE data can be mapped to counties. The outcome variable

[20] Another caveat to note regarding rejection reasons is that HMDA does not mandate lenders to report rejection reasons, so if their reporting incentives changed with the event, we would incorrectly attribute lender-induced change to consumer self-learning. However, it seems unlikely as lenders reported a rejection reason in over 70.81% of the rejections. Also, incentives to report rejection reasons would need to change differently in a particular manner between treated and control tracts in the event year to bias the above results, an unlikely scenario.

[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 ~$400 per withdrawn application, this represents ~U.S. $6.6 million saving in upfront mortgage application fees.

of interest is the ratio of the number of mortgages to the first-time homebuyers to the number of (all) originated mortgages with known information on first-time homebuyer status. The regression is now specified at the zip3-state level, different from previous census tract-level regressions:

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

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, similar to that in (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) shows the regression results. Coefficients in columns (1) and (2) show 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 prediction that following the event, more new entry should occur in the treated areas.[23]

§D A Demand- or Supply-side Effect?

The previous results indicate that the treated areas saw increased mortgage demand and origination after the event, and a demand-driven consumer self-learning channel is the likely explanation. However, a supply-driven explanation is plausible as well: in the natural experiment, lenders too were exposed to the knowledge that consumers' access to credit reports became easy and free, and so, they increased the mortgage supply.

Many of the earlier findings favor the demand-driven explanation. Increase in applications and a decrease in in-process application withdrawals indicate a demand-driven mechanism as these decisions are consumer determined and are mostly independent of lenders' influence. Furthermore, under a demand-driven mechanism, the effects should be heterogeneous in relevant consumer characteristics, as was the case for creditworthiness and income. Besides, recall that propensity of lenders to extend credit to low-income borrowers is low (Agarwal et al.,

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

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

2018), yet in the current setting we see that the approval ratio increased significantly for them vis-à-vis the high-income borrowers, making lender-driven effect unlikely.

Notwithstanding, mortgage interest rates and heterogeneous effect by lender's density allow us to further examine whether the supply-side was instrumental in the increased mortgage origination.

§D.1 Interest rates on the GSE-repurchased mortgages

The price at which lenders can sell *conforming* mortgages to the GSEs materially vary only across three dimensions, credit score, loan-to-value ratio (LTV), and loan type (Scharfstein & Sunderam, 2016).[24] Since the GSE sample in this paper include only 30-year fixed rate mortgages, only the first two are relevant. The residuals in a regression of interest rate on the first two attributes roughly reflect lenders pricing schedule independent of borrower risk and mortgage characteristics. Thus, including the "*Treat*×*Post*" in the above regression allows testing whether lenders lowered their pricing in the treated areas vis-à-vis control areas in a bid to increase mortgage origination, in which case the sign on the associated coefficient would be negative. The prediction is tested next using a regression similar to Equation (3) but specified at the loan level i:

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

Columns (3) and (4) of Table (IX) shows the results of the regression. Controls in column (1) are the two relevant pricing variables, credit score and CLTV (combined loan-to-value, it is loan-to-value ratio inclusive of all loans secured by a mortgaged property). In column (2), the following controls are added to make the specification more rigorous: debt-to-income ratio, number of units comprising the mortgaged property, and percentage of mortgage insurance coverage. The coefficient on "*Treat*×*Post*" is 0.009–0.01 percentage points, positive and significant. These estimates suggest that, if anything, lenders increased the risk-adjusted mortgage interest rates in the treated areas rather than lowering it, contradicting the idea of a supply-driven increase.[25]

[24] The pricing schedule published of Fannie Mae is available at https://www.fanniemae.com/content/pricing/llpamatrix.pdf

[25] The magnitude of the increase in the rates is tiny, potentially because of 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.

§D.2 Heterogeneous effects by density of mortgage lenders

If the increase in mortgage origination were driven by lenders, it would be greater in areas where the 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 is more than the *regional mean*, defined in Footnote (18), low otherwise.

Columns (1) through (4) of Table (X) show the results of separately regressing dollar origination volume (in 1000 USD) per adult for the two groups. The regression specification is the same as Equation (1). The estimates are smaller in magnitude and have weaker statistical significance for high-density tracts (columns 2 and 4) vis-à-vis the low-density ones (columns 1 and 3, respectively). Thus, high-lender-density treated areas saw smaller increase than low-lender-density treated areas after controlling for concurrent changes in the control areas with comparable lender density. t-test for the difference in the coefficient of "$Treat \times Post$" in high- and low- lender-density areas ($High - Low$) shows no statistical difference.

Then, columns (4) through (8) of the table repeats the analysis with approval ratio as the outcome variable. 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.

Overall, these findings suggest that the effects were stronger in low-lender density areas (severe low lender density areas are sometimes referred to as banking deserts) and are inconsistent with the explanation that increase in mortgage origination and approval ratio were solely lender driven.

5 Supplementary Discussion

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

An alternative explanation for increased mortgage origination is that higher commissions from private (non-government) securitization led lenders to increase the mortgage supply (Keys, Mukherjee, Seru, & Vig, 2010). If increased approval in the current context were due to private securitization, the fraction of originated mortgages being sold to non-government (private securitization) entities would increase in the treated areas.

Table (XI) examines the above prediction by employing Equation (1). The outcome variables are the fraction of total applications that lenders originated and (1) sold to non-government entities, (2) sold to the four GSEs (Fannie Mae, Freddie Mac, Ginnie Mae, and Farmer Mac), and (3) did not sell. Columns (1) and (2) show that there is no change (increase) in the fraction of mortgages sold to the private entities (non-GSEs); columns (3) and (4) show that the fraction of mortgages sold to the GSEs increased; and column (5) and (6) show that the fraction of unsold mortgages did not change either. Thus, the evidence does not support the conjecture that increase in private securitization in the treated areas could explain increase in origination.

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

It may be argued that increased mortgage origination is due to an increase in the subprime credit (Mian & Sufi, 2009). Employing the comprehensive HMDA data and location-based proxies of creditworthiness, Table (V) already suggest that effect of free credit reports was stronger in the prime counties/census tracts than in the subprime. These proxies are informative and widely used (Di Maggio & Kermani, 2017; Mian & Sufi, 2009), but are imprecise. By restricting ourselves to the GSE sample, we can use precise application-level credit scores.

Table (XII) shows the results of regressing separately the number of prime (credit score ≥ 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. In conclusion, the increased origination did not disproportionately go to subprime consumers. These estimates, however, are not directly comparable to the previous as the observation unit here is zip3-state, not census tracts which was the unit in the previous regressions.

These results come with the same selection issue that applied to the previous results utilizing the GSE data. Same argument as before allays this concern. In addition, before 2007 the GSEs sought to buy more subprime, not prime, mortgages to combat the housing bust (Elul, Gupta, & Musto, 2020), thus their changing incentives appears not to be a critical concern here.

§C Effect on banks

So far we focused on evaluating the effects on borrowers. The effect on banks is evaluated next, though with some caveats. First, commercial banks are not the dominant mortgage originators. Despite being 80% of mortgage lenders by number, banks accounted for just 37% of the mortgage lending in 2005, thus the conclusions drawn from studying banks may not be same for all lenders (Avery, Brevoort, & Canner, 2007). Second, many banks operate across states, so their treatment and control status in this natural experiment is continuous, rather than binary. The treatment intensity is proportional to bank's *ex-ante* mortgage activity in the treated and control states. Nonetheless, we can examine this question by classifying a bank as treated if in the pre-event year 2004 the ratio of mortgage it originated in control states to that in treated *and* control states combined is larger than the cross-sectional mean. The regression equation is:

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

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]

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 median, instead of mean.

It is also useful to understand which financial characteristics allowed the banks 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 *high* share of liquid asset if its ratio of liquid assets to total assets averaged annually was *greater* than the cross-sectional mean in the pre-event year 2004, and as having *low* share of liquid assets otherwise. The total amount of mortgage (in million $) originated by a bank in the control and treated states is regressed using Equation (5) separately for the two groups. Also, since the mortgage information come from the HMDA data, which are available at the annual frequency, the regression is estimated at the *Bank*$\times$*Year* level.

[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.

Panel (B) of Table (XIII) shows the results of the regressions. We see in Columns (1) and (2) 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, among the banks with an *ex-ante* high share of liquid assets, the treated banks on average increased mortgage lending relative to the control by about $111 million in a statistically significant manner.

All in all, with the caveats in mind that were described earlier, 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.

§D An alternative mechanism based on information asymmetry

An alternative mechanism based on asymmetric information is plausible in which borrowers *privately know* their true creditworthiness type, but do not know what lenders know about them. Using free credit reports, borrowers learn that the information on them that lenders have is proportional to their true type. Hence, under the non-trivial search/application cost, bad borrowers self-select out. The borrower pool now improves relative to the situation in which borrowers do not know that 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 self-selecting-out by bad borrowers, but not by self-selecting-in by good borrowers as all borrowers *privately know* their true type. However, under the self-learning mechanism, borrowers themselves have imperfect information of their true type, thus both selecting-in by good borrowers 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 the 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 such as those accumulated 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 borrowers.[27]

§E Robustness

Since the natural experiment utilized in this paper takes place in the year 2005, the sample period is chosen 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. The results are qualitatively and quantitatively similar and are unreported for brevity.

6 Conclusion

A non-trivial proportion of consumers do not check their credit reports regularly, and do not know their credit scores. Various data suggest that consumers err in credit decisions in a manner consistent with them having imperfect information of their creditworthiness. Credit reports contain crucial creditworthiness information, and can aid them in credit-related decisions.

This paper examines the effect of lowering the consumers' economic cost of credit reports on the mortgage market outcomes using a natural experiment in the U.S. The federal *Fair and Accurate Transactions Act of 2003* (FACTA) allowed all consumers to access from an online website three free credit reports annually since 2005, while seven states already had local laws permitting their residents to obtain the reports for free. Thus, the act effectively reduced the economic cost of accessing credit reports in a close-to-exogenous manner, and this paper draws causal conclusions using 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 reducing consumers' economic costs of the reports improves mortgage market outcomes in a way consistent with improvements in the borrower pool, and benefits both consumers and lenders. Specifically, free credit reports resulted in an increase in mortgage demand and approval ratio, more credit to creditworthy borrowers, more first-time homebuyers, fewer delinquencies, and better financial performance of lenders.

[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."

Though the findings pertain to the mortgage decisions of consumers, these broadly hold true for all consumer credit decisions when they have imperfect information of their creditworthiness. Furthermore, the causal nature of these findings implies that a policy intervention aimed at educating consumers of their creditworthiness may yield similar results.

References

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.

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.

Avery, R. B., Calem, P. S., Canner, G. B., & Bostic, R. W. (2003). An Overview of Consumer Data and Credit Reporting. *Federal Reserve Bulletin*, *89*, 47.

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*.

Brown, M., Haughwout, A., Lee, D., & Van Der Klaauw, W. (2011). Do We Know What We Owe? A Comparison of Borrower-and Lender-reported Consumer Debt. *Unpublished*(523).

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

Courchane, M., Gailey, A., & Zorn, P. (2008). Consumer Credit Literacy: What Price Perception? *Journal of Economics and Business*, *60*(1-2), 125–138.

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 Reserve Bank of Philadelphia. (2015). *What Your Credit Report Says About You.* Retrieved 2020-06-10, from https://www.philadelphiafed.org/consumer-resources/publications/what-your-credit-report-says

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

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

Missouri Census Data Center. (2010). *Geocorr 2000: Geographic Correspondence Engine. Version 1.3.3 (August, 2010) with Census 2000 Geography* [Database]. Retrieved from http://mcdc.missouri.edu/applications/geocorr2000.html

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/38k115/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

U.S. Senate. 108th Congress. (2004a). *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. Pages: 375–376)

U.S. Senate. 108th Congress. (2004b). *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. Prepared statement of Joel R. Reidenberg. Pages: 247–251)

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

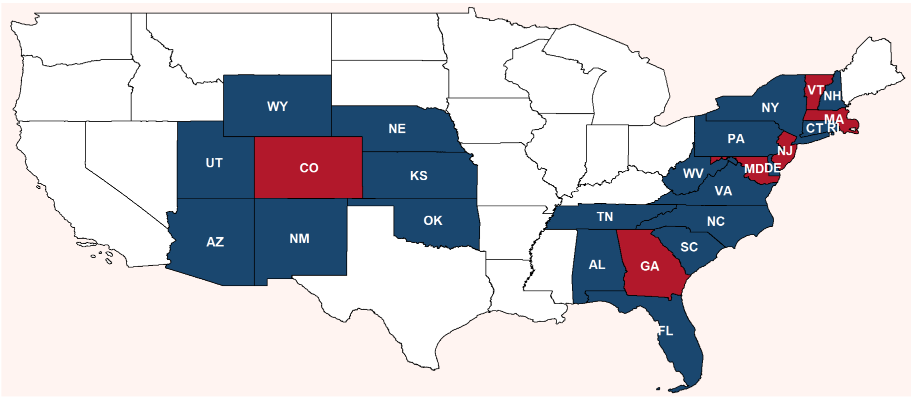
Your credit report includes your Personal Information and, if applicable, Consumer Statements, and could include other items that may affect your credit score and rating.

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, 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.

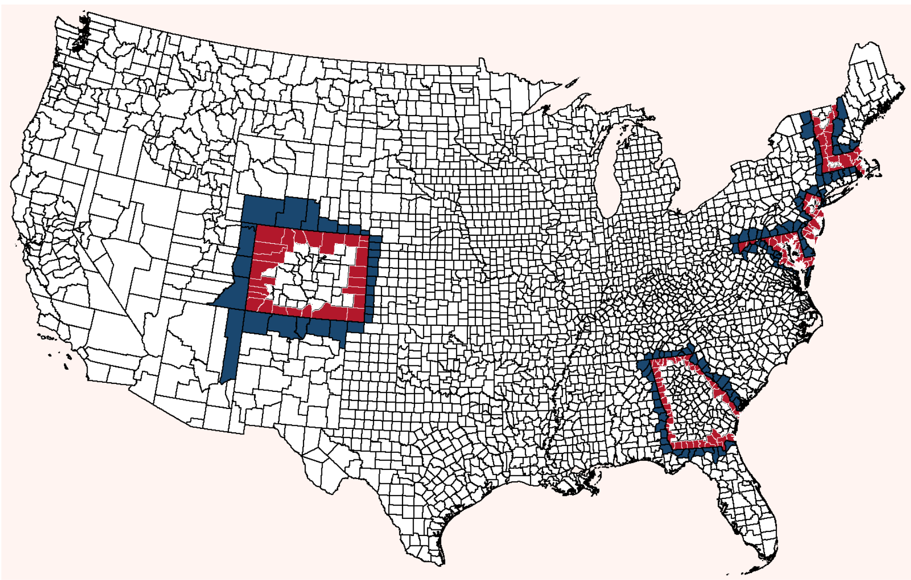
Panel A: Treatment and Control States



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

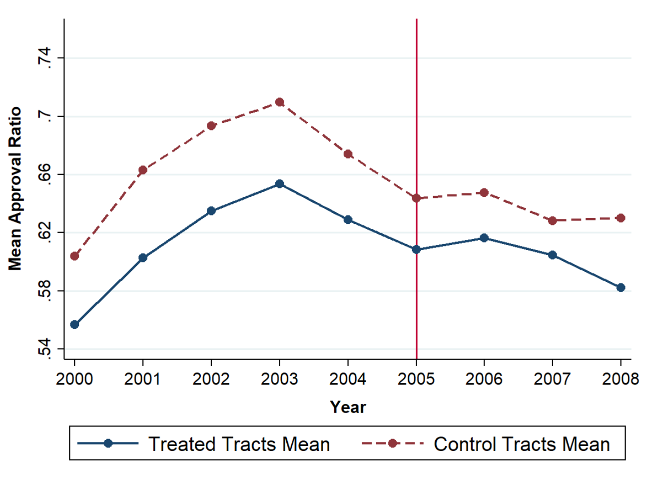
Panel (A) of this figure shows the mean approval ratio in the treated and control census tracts. **Panel (B)** of this figure shows the coefficients β_k from regressing *Approval Ratio* using the specification:

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

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 interval of β_k are also shown. The regression includes "*Border* $\times$ *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

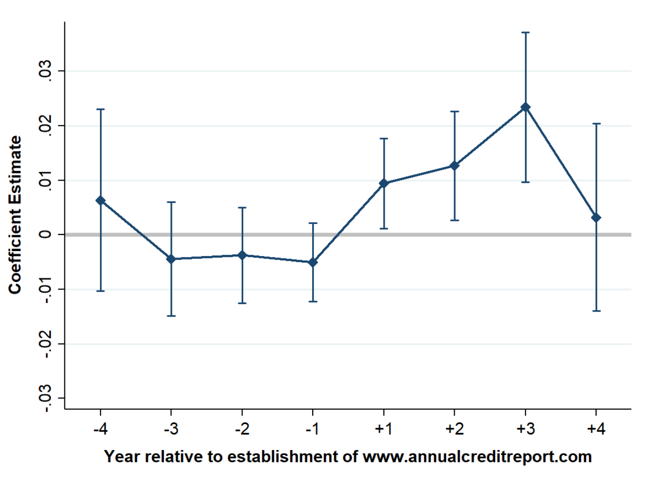
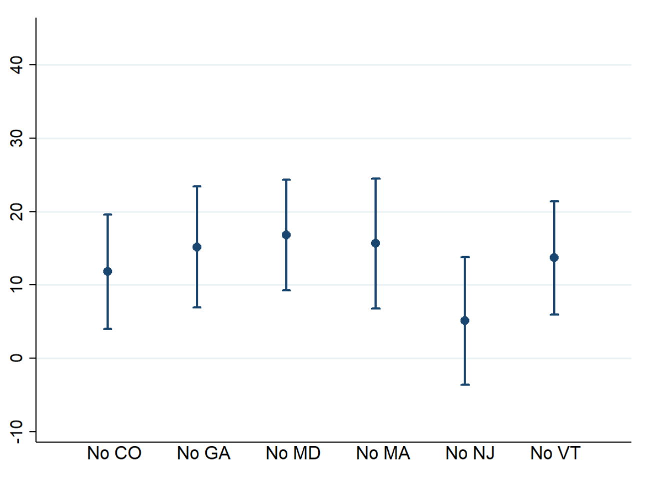


Figure IV: Subsample Analysis

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 were 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

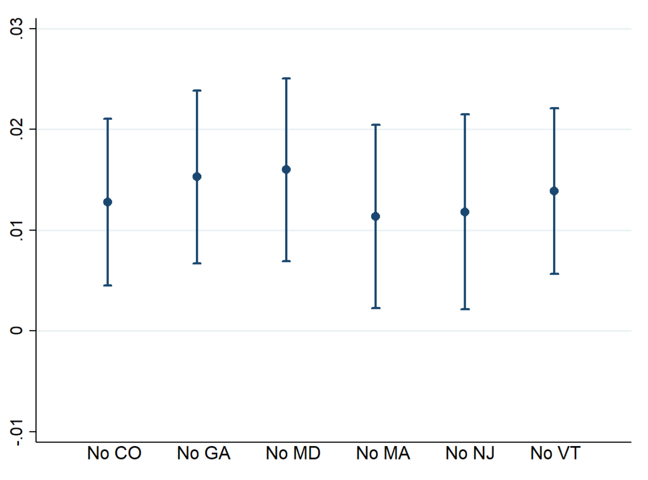
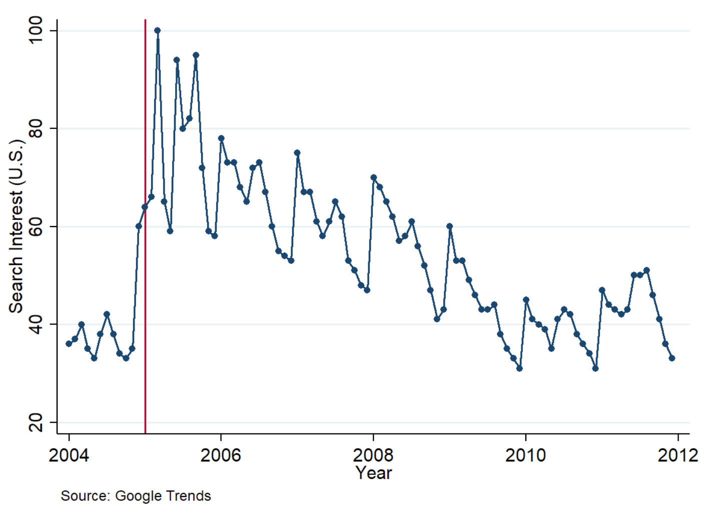


Figure V: Consumer Interest in Free Credit Reports: Google Trends

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 mean popularity rank of treatment and control states for the same keyphrase from 2004 to 2008. The popularity score of each state ranges from 0 to 100 and is re-computed by Google every year. A value of 100 represents the location with the highest popularity of the search term as a fraction of total searches in that location. A value of 50 indicates a location where it is half as popular. For the cross-sectional plot, first data spanning one-year intervals were obtained, and then the mean was calculated within each interval for the two sets of states.

Panel A: Google Search Interest for the Term "Free Credit Report" in the U.S.



Panel B: Relative Popularity of the keyphrase *Free Credit Report*

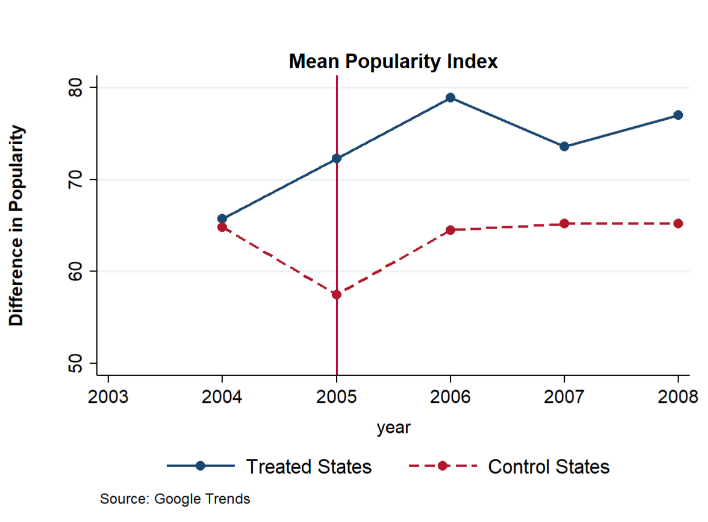
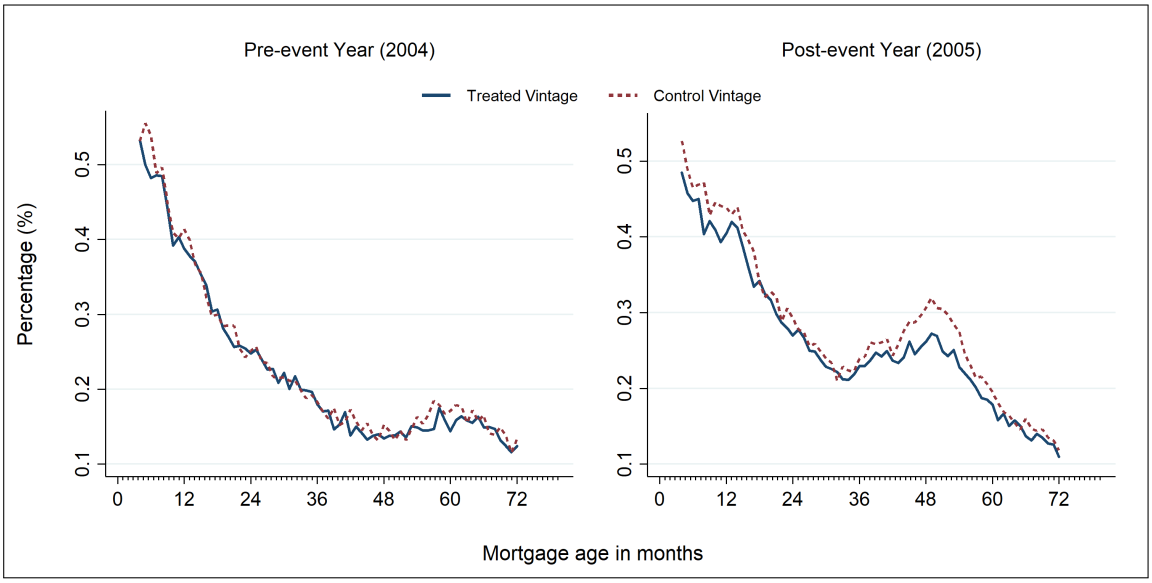


Figure VI: Effect of Free Credit Report on Mortgage Delinquencies

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-days 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-days 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

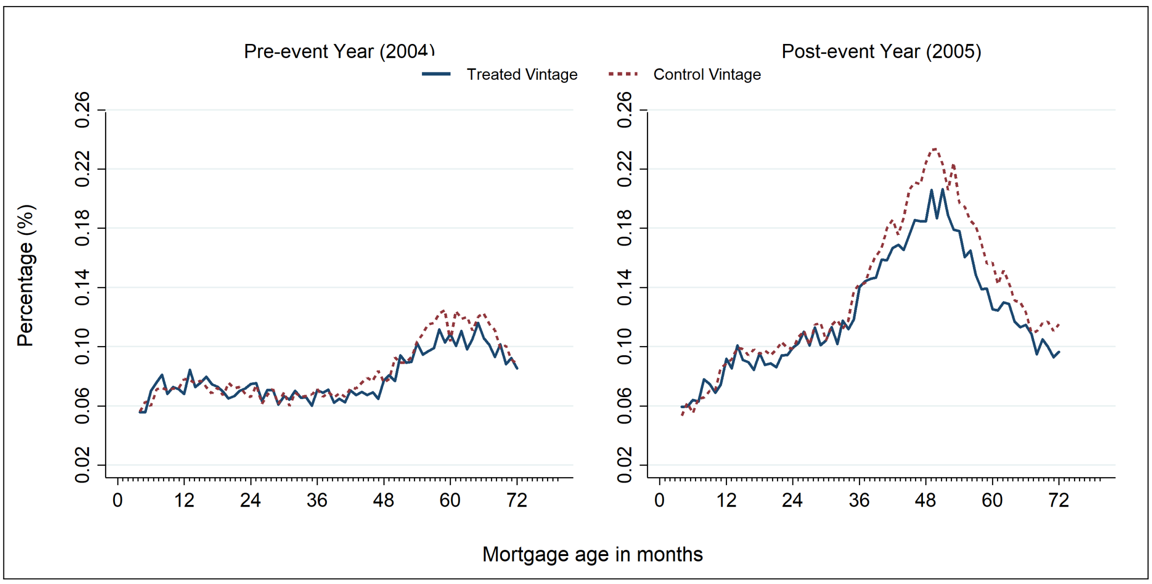


Table I: Summary Statistics

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 the four variables described as follows: (i) *Num. Lenders* (*log*), the number of unique mortgage lenders in a census tracts (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 the 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 one if a respondent has never checked his/her credit score (Q. *N23*). *Infrequently* is one if respondent has never checked it or last checked it more than 2 years ago (Q. *N23*). *Unaware* is one if respondent don't now his/her credit score (Q. *N22*). *Dscrgd* is one if 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 to apply for mortgage or to refinance as less than 10%. All regressions include $Year \times Month$ 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
R^2 (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 mortgages applications, (ii) non-owner-occupied applications, (iii) the fraction of non-owner-occupied mortgages as total applications, and (iv) as a 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):

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

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* 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 are reported in square brackets below the t-statistics. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

		N			Aprv.		
	C. Hist. Tertiles ↓	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 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
R^2 (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 Rate 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 homebuyer is not missing, calculated in at 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, 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); it refers to three more controls *debt-to-income 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 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 4 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 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 *Number of mortgages originated to Prime Borrowers* (credit score≥620) in a given zip3-state area. The dependent variable in column 2 is *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*×*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*×*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 × 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 × 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 *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 *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 *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

A. Data Appendix

The HMDA data contain 190.4 million mortgage applications over the sample period (2000–2008). These application-level data were aggregated to "$CensusTract \times Year$" panel in several steps. First, all observations that have state, county or census tract information missing or "NA", or state Federal Information Processing Standard (FIPS) code as "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: the covered loan 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 the pre-approval requests approved by the financial institutions but not accepted by the applicants, as this data, too, were included in HMDA only from 2004, and its reporting is not mandatory (0.025%). This leaves 150.7 million applications belonging to 77,526 unique census tracts (603,849 "$Census\ Tract \times Year$" observations). Finally, with the help of the county adjacency data from Census Bureau (n.d.), those census tracts were selected that belong to the bordering counties of the treated and control states. This led to the HMDA regression sample: 89,535 "$Census\ Tract \times Year$" 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 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 "$Zip3\text{-}State \times Quarter$" 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 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 a 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 since 2004, so the matching was done for subsequent years using this id, instead of the combination of the agency code and respondent ID.