NBER WORKING PAPER SERIES

HIGH COST LENDERS AND THE GEOGRAPHIC CONCENTRATION OF FORECLOSURES

Stephen L. Ross
Yuan Wang

Working Paper 28781
http://www.nber.org/papers/w28781

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2021

The authors thank Fernando Ferreira for providing advise on this research and conference and seminar participants at the Econometric Society meetings, the Federal Reserve Banks of Boston, Kansas City and New York, the Furman Center at New York University, and the Housing-Urban-Labor-Macro Conference. The authors thank Gordon MacDonald, Kyle Mangum, Yuan Wang, Ailing Zhang, and Samantha Minieri for outstanding research assistance. The analyses presented in this paper use information provided by one of the major credit reporting agencies. The credit agency has the legal right to view the research prior to its public release. However, the substantive content of the paper is the responsibility of the authors and does not reflect the specific views of any credit reporting agencies, nor does it reflect the views of Freddie Mac or its board of directors. This work was supported by the Ford Foundation, the Research Sponsors Program of the Zell/Lurie Real Estate Center at Wharton, and the Center for Real Estate and Urban Economic Studies at the University of Connecticut. Professor Ross has worked in the past as a consultant on fair housing and fair lending issues, but he has not done any consulting in this area within the last three years. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

High Cost Lenders and the Geographic Concentration of Foreclosures
Stephen L. Ross and Yuan Wang
NBER Working Paper No. 28781
May 2021
JEL No. D14,G01,G21,R21,R23

ABSTRACT

We define high cost lenders as lenders that issue a disproportionate number of high cost loans. We develop a shift-share measure to capture the market representation of these high cost lenders in housing submarkets. After conditioning on housing submarket fixed effects, origination year fixed effects and trends over origination years based on housing submarket attributes, the magnitude of the estimated relationship is very stable as detailed controls for borrower attributes, credit score and loan terms are added. The relationship between the representation of high cost lenders and foreclosure is broad based across borrowers and types of loans, but is strongest for loans originated by high cost lenders whether or not the loans themselves are high cost. We investigate three potential mechanisms: reverse causality where high cost lenders respond to an increase in demand from higher risk borrowers, the types of mortgages issued when high cost lenders increase their market presence, and the behavior of loan servicers when a cohort of loans contains a large number of loans issued by high cost lenders. While we do not have direct information on loan servicers, our evidence points towards foreclosure decisions during the crisis as the primary mechanism behind our findings.

Stephen L. Ross
Department of Economics
University of Connecticut
341 Mansfield Road, Unit 1063
Storrs, CT 06269-1063
and NBER
stephen.l.ross@uconn.edu

Yuan Wang
FreddieMac
McLean, VA
USA
wangyuansdu@gmail.com

High Cost Lenders and the Geographic Concentration of Foreclosures

Stephen L Ross, University of Connecticut and NBER

Yuan Wang, FreddieMac

A central feature of the foreclosure crisis in the U.S. is that the effects were concentrated in specific locations. Subprime lending was concentrated in low income and minority neighborhoods in the run up to the crisis (Calem, Gillen and Wachter 2004; Calem, Hershaff, Wachter 2004; Mayer and Pence 2009; Ghent, Hernández-Murillo and Owyang 2014; Bayer, Ferreira and Ross 2014, 2018).[1] Minority and low income neighborhoods also experienced especially severe foreclosure rates during the crisis (Bayer, Ross and Ferreira 2016; Ghent, Hernández-Murillo and Owyang 2014; Chan, Gedal, Been and Haughwout 2013; Geradi and Willen 2009; Fisher, Lambie-Hanson and Willen 2011; Edminston 2009).[2] Finally, several studies document that the same neighborhoods experiencing high foreclosure rates during the crisis also had especially high rates of subprime or high cost lending in the run up to the crisis (Reid and Laderman 2009; Mian and Sufi 2009; LaCour-Little, Calhoun and Yu 2011; Reid, Bocian, Li and Quercia 2016).[3]

In this study, we carefully document for home purchase mortgages both the cross-sectional and across time correlation between foreclosures and the geographic concentration of the activity

[1] This empirical regularity has been established using many different indicators for subprime or high risk lending including the Department of Housing and Urban Development (HUD) subprime lender list, non-agency securitized lending, high cost lending based on APR rate spread in the Home Mortgage Disclosure Act data, and measures of ex-post foreclosure rates.
[2] Ghent, Hernández-Murillo and Owyang (2014) find higher rates of foreclosure in low income neighborhoods, but find lower rates of foreclosure in minority neighborhoods in their sample of privately securitized mortgages.
[3] For recent reviews, see Foote and Willen (2018) and Chan, Haughwout and Tracy (2015). For a more general discussion of redlining see Ross and Yinger (2002).

of lenders that we describe as “high cost” lenders.[4] As shown below in the data section (Table 2), volatility in the share of loans from “high cost” lenders during the housing boom is much larger than volatility over more traditional measures of mortgage underwriting risk, like loan to value or debt to income ratios. The paper’s focus on “high cost” lenders is also driven in part by Bayer, Ferreira and Ross (2018) who find that lenders who issue a large share of high cost loans appear to serve a different segment of the mortgage market in that they tend to have unusually high ex-post foreclosure rates, after controlling for borrower and mortgage attributes. We identify high cost lenders as lenders for whom a substantial share (20 percent or higher) of their mortgages were high cost, i.e. defined in the Home Mortgage Disclosure Act (HMDA) as mortgages with a substantial rate spread above the return on comparable maturity treasury bonds, annualized percentage rate 3 points above treasury yields. We define our housing submarkets as Public Use Microdata Areas (PUMA), which are defined by the U.S. Census Bureau as geographically contiguous areas containing at least 100,000 people.[5] We proxy for the market penetration or representation of high cost lenders first using share of loans issued by high cost lenders in a PUMA and origination year and then using a Bartik or shift-share prediction (Autor and Duggen 2003; Brunner, Ross and Washington 2011) of changes in the PUMA representation of high cost lenders.

Using a sample of home purchase mortgages that were originated between 2004 and 2007 from seven large metropolitan/regional sites across the U.S.,[6] we find a strong cross-sectional

---

[4] A related literature (Campbell et al., 2011; Gerardi et al., 2015; Gupta, 2016; Munroe and Wilse-Samson, 2013) examines neighborhood spillovers from foreclosures. While those studies tend to find modest effects that are very locally concentrated, often within several hundred feet, Mian, Sufi and Trebbi (2015), Huang, Nelson and Ross (2018) and Guren and McQuade (2020) demonstrate that these effects may have had larger effects at a broader geographic level as the crisis worsened and spillover effects multiplied.

[5] While our data contain census tract attributes, we were restricted in our data agreement to not include census tract specific identifiers. This larger housing submarket geography also has the advantage of being at a scale well above the geographic level of foreclosure spillovers discussed earlier. Further, as we will show below, this geographic region captures dramatic variation in the distribution of high cost loans, loans by high cost lenders and foreclosures.

[6] See Bayer, Ferreira and Ross (2014, 2016, 2018) for recent analyses of racial differences in foreclosures and high cost lending using this data.

correlation between foreclosure and the share of loans in a PUMA and year originated by high cost lenders. However, controlling for borrower credit score and demographics reduces the magnitude of the estimated effect by half. The remaining conditional correlation between foreclosure and the share of mortgages from high cost lenders is at most modestly reduced (about 13 percent) by the inclusion of detailed loan terms, such as loan to value and debt to income ratios, whether an adjustable rate mortgage and use of subordinate debt, and the inclusion of lender fixed effects.[7] After conditioning on PUMA fixed effects, the correlation between foreclosure and share of loans from high cost lenders is very similar to the cross-sectional correlation after including borrower, loan and lender controls, and generally robust to the inclusion of credit score, borrower demographics, loan terms, and lender fixed effects. After including these controls, a one standard deviation increase in share of loans from high cost lenders increases foreclosure rates by 1.23 percentage points or about 23 percent of the sample foreclosure rate of 5.3 percent. In summary, a substantial portion of the cross-sectional correlation between foreclosure and the activity of high cost lenders can be explained by borrower observables, and after conditioning on borrower observables only a small portion of this correlation can be explained by loan attributes, lender identity or time invariant attributes of housing submarkets.

Further, we estimate PUMA fixed effect models using a Bartik or shift-share prediction of changes in the PUMA share of loans from high cost lenders based on the initial distribution of loans across lenders in each PUMA and national changes in lender market share. We view this measure as a more exogenous proxy for changes in the market penetration or representation of

[7] Unfortunately, we do not observe some key loan terms like whether the loan is no or low documentation, the lock period for adjustable rate loans, information on teaser rates, or prepayment penalities. However, many of these unobserved features are concentrated in the subprime sector where adjustable rates loans and the use of subordinate debit are much more common for home purchase loans, and so many of our actual controls for loan terms should capture at least some of the effect of these product attributes. Further, these attributes also tend to be concentrated among subprime loans that tended to be issued by specific lenders, which in part is captured by lender fixed effects.

high cost lenders in a PUMA. To assure that the results are not driven by national changes in the market share of specific lenders, we also include lender by origination year fixed effects in all models. The estimated relationship between foreclosure and our shift-share proxy is again very stable in magnitude to the inclusion of borrower attributes, loan terms and lender fixed effects. The standardized effect from the shift-share prediction is similar, only about 19% higher, than the effect of observed lender share with a one standard deviation change implying a 1.46 percentage point change in foreclosure or approximately 27 percent of the sample foreclosure rate.[8]

These results are robust to changing the definition of high cost lenders to be based on thresholds of 15 or 25 percent of loans. Results are also robust to using an alternative measure of high cost loans that adjusts the rate spread threshold, selecting a balanced panel of credit report years, or weighting the sample based on the inverse probability of selection. Finally, results are robust to controlling for detailed fixed effects based on clusters of PUMA's that have very similar initial (2004) shares of loans from high cost lenders and were originated in the same year and to clustering standard errors over these fixed effects, to address concerns raised by Goldsmith-Pinkham, Sorkin and Swift (2020) and Adão, Kolesár and Morales (2019), respectively.

These estimated effects are broad based. We estimate similar magnitude effects of high cost lender representation whether the predicted changes over time imply either an increase or a decrease in share of high cost lenders. The geographic foreclosure effects exist for both rate spread and non-rate spread loans, white and minority loans, loans regardless of the source of securitization, and loans originated by high cost and non-high cost lenders. Notably, however, these geographic foreclosure effects are significantly larger in absolute terms for loans issued by

[8] We do not instrument for the share of loans from high cost lenders with the Bartik predicted change because both variables are proxies for the variable of interest and instrumenting for one proxy with another when measurement error is non-classical may lead to significant bias (Chalak 2019).

high cost lenders, a standardized effect of 3.07 percentage points relative to 1.06 percentage points for non-high cost lenders. However, given the higher baseline foreclosure rates, the percent change in foreclosure is 28 and 31 percent for high cost and non-high cost lenders, respectively. Given the larger absolute effects for loans originated by high cost lenders, we investigate whether our results can be driven by changes in the market share of the lender who originated the loan. While the lender level shift-share analysis implies higher foreclosure rates for loans made by high cost lenders when they increase their PUMA market share, the effect of the market representation of foreclosure is robust, falling by only about 8-10 percent with the inclusion of these lender market share controls.

We next consider three possible mechanisms behind these results. First, we consider the possibility that this geographic concentration of foreclosures may be driven by a kind of reverse causality where increased demand for credit among riskier borrowers leads to the entry of high cost lenders in order to meet this demand. First, the very stable estimates for the shift-share prediction as controls for credit score and borrower demographics are added to the model are inconsistent with bias from borrower unobservables, because it is very unlikely that unobservable attributes will erode the estimates if important observable attributes like credit score or race/ethnicity have no impact on the estimates (Altonji, Elder and Tabor 2005; Oster 2019). To further investigate possible bias from borrower unobservables, we show that the results are not eroded by including contemporaneous information on bank card and health expense delinquencies, which might correlate with unobserved borrower risk factors. Further, we show that our proxy for high cost lender representation has no effect on the likelihood of a loan being high cost, which might be expected if the share of high risk borrowers was increasing. Finally, Bayer, Ferriera and Ross (2016) suggests that high risk borrowers will rationally enter credit markets as access to credit increases. Therefore, we develop measures of changes in the composition of borrowers at the PUMA level on credit score, race/ethnicity and income. Controlling for these changes does not erode the foreclosure effects of high cost lender representation in a PUMA, even though changes in borrower composition over race/ethnicity and income explain foreclosure. Therefore, we find no evidence to support borrower unobservables as a mechanism behind our results.

The next mechanism that we consider is that the estimated effects arise because subprime lenders are concentrated in these submarkets and those lenders issue mortgages with minimal assessment of risk (Keys, Mukherjee, Seru, and Vig 2009; Dell'Ariccia, Igan, and Laeven 2012; Bhutta and Keys In Press), little documentation (Jiang, Nelson and Vytlacil 2014a; LaCour-Little and Yang 2013) and/or high risk and possibly predatory terms (Reid and Laderman 2009; Jiang, Nelson and Vytlacil 2014b; Agarwal, Amromin, Ben-David, Chomsisengphet, and Evanoff 2014, 2020; Reid, Bocian, Lie and Quecia 2016) leading to high levels of foreclosure in those locations during the financial crisis. As with borrower unobservables, the stability of our estimates to controlling for loan terms and lender identity, which both almost certainly correlate with subprime mortgage attributes, works against subprime loan features as an explanation for these geographic effects. Next, we include controls for PUMA wide changes in key loan attributes (rate spread loans, loan to value ratio and debt to income ratio) that might indicate a shift in the types of loans being originated within the PUMA. Again, these controls do not have any impact on our location estimates. Finally, we consider Mian and Sufi's (2009) argument that the expansion of subprime credit in a location increases housing prices in that housing submarket leading to greater equity losses and higher foreclosure rates during the following housing market downturn.[9] We use our transaction data to calculate price indices and measure the level of negative equity faced by

[9] Also see Carozzi, Hilber and Yu (2020) for evidence of local price effects in England from a policy that expanded access and affordability of mortgage credit.

borrowers. While our measure of negative equity explains foreclosure, the inclusion of negative equity as a control had minimal influence on our estimated location effects.[10]

Finally, we consider whether the geographic clustering of foreclosures associated with the activity of high cost lenders can be explained by differential mortgage servicing decisions during the crisis. For example, Kruger (2018) shows that servicers have broad discretion concerning loan modifications, but securitization creates incentives favoring foreclosure.[11] Agarwal, Amromin, Ben-David, Chomsisengphet and Piskorski (2017) and Reid, Urban and Collins (2017) document significant heterogeneity in the terms and types of modifications offered or even whether a modification is offered at all. Unfortunately, we do not observed data on servicers.

However, we are able to conduct two final analyses that are suggestive that foreclosure decisions during the crisis may have driven our estimated effects. First, we estimate the same models using 90-180 day delinquencies as the outcome, rather than foreclosure filings. Mortgage delinquency represents failure of the borrower to make their mortgage payments, while foreclosure represents the combined effects of a serious mortgage delinquency and lender servicer decisions concerning foreclosure filings and loan modifications. The effect of our proxy for high cost lender market penetration on delinquency represents only a 3.5 percent change in the sample delinquency rate and is statistically insignificant. Second, we recognize that foreclosure is purely administrative in the state of California, while courts play a significant role in all other states represented within our sample creating significantly more opportunities for lender discretion in foreclosure. Splitting our sample between the two California sites and the other sites within our sample, we find that our geographic foreclosure effects are concentrated almost entirely outside of California. For states

[10] These measures of current negative equity are also interacted with the current employment rates at the county level because default in response to negative equity is often triggered by income shocks (Bayer, Ferreira and Ross 2016; Gerardi, Ross and Willen 2011).

[11] For an alternative view on remediation and securitization, see Ghent (2011).

with a judicial role in foreclosure, the standardized effect of our estimate is about 40 percent of the baseline foreclosure rate, and while the estimates are noisy in our California subsample the point estimates imply an effect that is only 7 percent of the baseline rate.

In summary, we identify a strong correlation between subsequent loan foreclosures in a housing submarket and the activity of lenders that tend to issue a disproportionate number of high cost loans. We find this correlation with foreclosure risk both in the cross-sectional variation across places, as well as in changes within place across purchase cohorts of loans based on the year of origination. The cross-sectional correlation is very sensitive to the inclusion of our limited number of borrower controls, and yet is relatively unaffected by controls for detailed loan attributes or by the inclusion of lender fixed effects. The across origination cohort correlation is not sensitive to either borrower or loan attributes whether based on the observed share of loans from high cost lenders or a Bartik style prediction of this share. These across cohort effects are broad based, but larger among loans originated by high cost lenders. In terms of mechanisms, we do not find evidence that these effects are driven by riskier borrowers attracting high cost lenders into the housing submarket, nor evidence that these effects are driven by the expansion of risky mortgage lending when the representation of high cost lenders increases. Rather, while indirect, our evidence appears consistent with these effects being driven by more aggressive foreclosure activity of loan servicers for cohorts of loans that were originated when high cost lenders were heavily represented in the housing submarket.

1. Methods

We start by creating a proxy for the market penetration or representation of high cost lenders in a housing submarket within a broad metropolitan or regional market. In order to do this, we define lenders as high cost (H) based on whether the share of their originated mortgages that meet the definition of a rate spread or high cost loan in the Home Mortgage Disclosure Act data (α_l) exceeds some pre-specified threshold ($\bar{\alpha}$).

$$l = \begin{cases} H & if\ \alpha_l \geq \bar{\alpha} \\ L & if\ \ \alpha_l < \bar{\alpha} \end{cases}$$

Our proxy for market penetration is the fraction of home purchase mortgages originated by high cost lenders in a submarket (*n*) during a given year (*p*) defined as

$$Z_{np} = \frac{\sum_{l \in H} N_{lnp}}{\sum_{l} N_{lnp}}$$

where N_{lnp} is the number of mortgages issued by lender l during the submarket and time period considered. More details on the definition of rate spread loans is provided in the data section below.

Then, we document the conditional correlation between foreclosure and our first proxy for market penetration, the share of loans in a housing submarket originated by high cost lenders. We follow Bayer, Ferreira and Ross (2016) and estimate foreclosure models using an annual panel of foreclosure notices from a sample of home purchase mortgages using a linear probability model. The model controls for share of loans from a high cost lender (*Z*) in the submarket (*n*) and purchase year (*p*), a host of borrower and mortgage attributes (*X*), and site (*s*) by credit report or foreclosure year (*t*) by purchase year fixed effects to allow each metropolitan or regional market (*s*) to have its own time path of foreclosures for each cohort of loans

$$y_{instp} = \alpha Z_{np} + \beta X_i + \delta_{stp} + \varepsilon_{instp} \tag{1}$$

Our next model removes the cross-sectional variation by including housing submarket by credit/foreclosure year fixed effects and submarket trends over purchase year based on location observables. The housing submarket fixed effects capture the constant (across purchase cohorts) impact of time invariant submarket unobservables, while the submarket trends allow the relationship between submarket and foreclosure to vary across origination cohorts based on time invariant observables. Specifically, the submarket trends are captured by adding purchase year fixed effects interacted with time invariant submarket attributes (*W*) allowing for the purchase year cohorts of loans in each submarket to follow a non-linear trend over purchase year. The resulting specification is

$$y_{instp} = \alpha Z_{np} + \beta X_i + \delta_{stp} + \theta_{nt} + \varphi_p W_n + \varepsilon_{instp} \qquad (2)$$

Finally, we attempt to capture the causal relationship between the market penetration of high cost lenders and foreclosure by creating an exogenous proxy for or prediction of changes in the representation of high cost lenders for each housing submarket and purchase year. Specifically, we create a Bartik or shift-share style prediction similar to those used in Autor and Duggen (2003) and Brunner, Ross and Washington (2011). We measure the fraction of loans in a housing submarket for a base year ($\bar{p}$) from each lender ($\alpha_{\bar{p}nl}$) to capture the baseline share and also measure the national market share for each of those lenders for every purchase year (μ_{pl}) to capture the shift over time leading up to the crisis. For all lenders *l* with non-zero market share in a housing submarket in the base year, we calculate a predicted growth in the housing submarket volume of both high cost lenders (*H*) and all lenders as a weighted average of the percentage changes in each lender's overall market share $\left(\mu_{pl} - \mu_{\bar{p}l}\right)$ where the weights are the base year housing submarket share for each lender ($\alpha_{n\bar{p}l}$). The predicted percentage change for high cost lenders is added to the original high cost lender share in the base year, divided by one plus the predicted percentage change for all lenders, and finally the base year share is subtracted to obtain a predicted change.

$$\widehat{\Delta Z}_{np} = \frac{Z_{n\bar{p}} + \sum_{l \in H} \alpha_{n\bar{p}l}\left(\mu_{pl} - \mu_{\bar{p}l}\right)}{1 + \sum_{l} \alpha_{n\bar{p}l}\left(\mu_{pl} - \mu_{\bar{p}l}\right)} - Z_{n\bar{p}} \qquad (3)$$

where $Z_{n\bar{p}}$ is the PUMA market share of high cost lenders in the base year $\bar{p}$ so that $\widehat{\Delta Z}_{n\bar{p}}$ is always zero in $\bar{p}$. We also add lender by purchase year fixed effects (ρ_{pl}) so that our model will not be identified by national trends in the market share of large high cost lenders. The final model specification is

$$y_{instpl} = \alpha\widehat{\Delta Z}_{pn} + \beta X_i + \rho_{pl} + \delta_{stp} + \theta_{nt} + \varphi_p W_n + \varepsilon_{instpl} \qquad (4)$$

Standard errors in all models above are clustered at the housing submarket level (*n*).

Several recent papers have examined the application of Bartik or shift-share variables of this type. Goldsmith-Pinkham, Sorkin and Swift (2020) examine panel applications of shift-share variables in models that control for geographic fixed effects. They conclude that a sufficient condition for consistency is strict exogeneity of the initial shares that are used to scale the time varying changes or shifts. Importantly, they observe that strict exogeneity need only be established conditional on controls and argue that identification based on changes (based geographic fixed effects), rather than levels, makes the assumption of strict exogeneity much more reasonable.

Nonetheless, areas with high versus low exposure to treatment based on initial shares may have unobservables that correlate with the predicted changes (again drawing on the language of Goldsmith-Pinkham et al. 2020). For example, perhaps submarkets with similar levels of initial shares of loans from high cost lenders, $Z_{n\bar{p}}$, share unobservables that influence the evolution of mortgage cohorts in the run-up to the financial crisis. In this case, we might rewrite the unobservable as

$$\varepsilon_{instpl} = \mu_{\Omega_n p} + \tilde{\varepsilon}_{instpl} \quad n \in \Omega_n$$

where Ω_n is a set of locations that are similar to submarket *n*. With a small number of purchase cohorts and a limited number of geographic submarkets, the submarket fixed effects could suffer from an incidental parameters bias due to a failure of strict exogeneity. Accordingly, we attempt to absorb $\mu_{\Omega_n p}$ by dividing submarkets into bins of similar initial share of high cost lenders using

$$n \in \Omega_k \; if \; \bar{Z}_k \le Z_{n\bar{p}} < \bar{Z}_{k+1}$$

where $\bar{Z}_k$ represents the bottom threshold of high cost lender share for the kth group. As a robustness test, we then estimate a revised model where we add initial share group by purchase year fixed effects

$$y_{instpl} = \alpha\widehat{\Delta Z}_{pn} + \beta X_i + \rho_{pl} + \delta_{stp} + \theta_{nt} + \varphi_p W_n + \tau_{k_n p} + \varepsilon_{instpl}$$

where k_n represents the share group to which submarket *n* belongs.

This structure also helps us address a second concern raised by Adão, Kolesár and Morales (2019). They demonstrate that inference in shift-share analyses could be biased if places with similar share structure also share similar unobservables creating a correlation across geographies. Notably, their analysis is based on cross-sectional use of shift share variables, and so their concerns should in part be addressed by geographic or submarket fixed effects in our application. Nonetheless, submarkets that are similar on initial shares could also have similar purchase cohort time trends in the unobservables. After including fixed effects for loans belonging to the same cohort in similar submarkets, we can address any general pattern of correlations for the same cohort between these similar submarkets by clustering at the purchase cohort by submarket share of loan fixed effect level. We will use two way clustering in this robustness test so that clustering at the individual submarket level is preserved, as well.

2. Mortgage Foreclosure Sample

Our data set is based on public Home Mortgage Disclosure Act (HMDA) data from between 2004 and 2007. We begin with a convenience sample of seven metropolitan/regional housing markets based on the counties comprising these areas: Chicago IL CMSA, Cleveland OH MSA, Denver CO MSA, Los-Angeles CA CMSA, Miami-Palm Beach Corridor, San Francisco CA CMSA, and Washington DC-Baltimore MD suburban Corridor. We matched the HMDA mortgage originations to housing transaction data purchased from Dataquick Inc., and then selected a stratified random sample of mortgages to match to credit reporting data collected by Experian PLC annually from the year of origination to end of our data in 2009, see Bayer, Ferriera and Ross (2014, 2016, 2018) for earlier applications of this data related to the outcomes of black and Hispanic borrowers.[12]

Beginning in 2004, HMDA data began reporting information for whether the Adjusted Percentage Rate (APR) of each loan exceeds the yield on treasury bonds of comparable maturity by at least 3 percentage points, and these loans are referred to as rate spread or high cost loans.[13] HMDA also identifies lenders using a respondent identification number. For all loans originated in HMDA during our sample period and in our seven sites, we calculate the cross-sectional fraction of rate spread loans originated during the entire period by each lender. We then define "high cost lenders" as any lender that had at least 20 percent of their loans classified as rate spread loans in HMDA over the sample period from 2004-2007 including all loans originated in the seven sites. Then, we define the housing submarkets for each major/metropolitan market as Public Use Microdata Area's (PUMA). PUMA's are defined by the U.S. Census as geographically contiguous areas containing at least 100,000 people, and the U.S. Census uses the PUMA definitions in order to provide residential location information at the individual level for the Decennial Censuses and American Community Surveys. We calculate the share of loans originated by high cost lenders in each PUMA by purchase year for each year between 2004 and 2007 based on the census tract

[12] Miami-Palm Beach Corridor contains Miami-Dade, Broward and Palm Beach counties. The Washington DC-Baltimore MD suburban Corridor contains all counties in the state of Maryland, but the population of transactions is dominated by Baltimore and Annapolis, their suburbs, and Washington suburbs. The mortgage data set assembled originations from 2004 to 2008, but this study does not include originations made in 2008 due to the on-going financial crisis during that year.

[13] The Annual Percentage Rate (APR) estimates cost of credit including interest rate and closing costs. These high cost or rate spread loans are sometimes referred to as subprime loans (Mayer and Pence 2009; Chan, Haughwout and Tracy 2015), but other authors study the subprime market based on a list of top subprime lenders, e.g. Ferreira and Gyourko (2015), based on borrowers who have a low credit score, e.g. Mian and Sufi (2009), or private label securitized loans, e.g. Ghent, Hernández-Murillo and Owyang (2014).

location of the purchased property again using the population of owner-occupied, home purchase mortgages contained in HMDA.[14]

We calculate the market share of high cost lenders in a PUMA and purchase year for several additional definitions of high cost lender. First, we redefine high cost lender based on the lender having at least 15 or 25 percent of their loans classified as rate spread. Next, we recognize that the share of rate spread loans is sensitive to the yield curve over bond maturities because APR is compared to treasury rates of comparable maturity to mortgage terms, but mortgages are often pre-paid (Avery, Brevoort and Canner 2007). Therefore, we redefine the rate spread variable adjusting the high cost loan threshold by year in order to keep the share of high cost loans constant over time, anchored to 2004 which had the lowest share. This revised rate spread variable is then used to identify a new set of high cost lenders that is not affected by lenders issuing large numbers of "high cost" loans in years when the yield curve lead to larger numbers of rate spread loans.[15] We also calculate time invariant characteristics of PUMA's using both the 2000 Decennial Census for residents and the 2004 HMDA data for the attributes of home purchasers at the beginning of our sample,[16] and calculate time varying (over origination year) PUMA attributes based on HMDA, our housing transaction data and our matched credit history data.

The home purchase sample is constructed as a sample of owner-occupied, 1-4 family properties drawn from HMDA and merged to both proprietary housing transaction/lien and assessor's databases purchased from Dataquick based on year, loan amount, name of lender, state,

[14] The Census Bureau provides detailed cross-reference files mapping census tracts into PUMA's.

[15] The threshold for a high cost lender is lowered from 20 to 13 percent in order to hold the total share of loans from high cost lenders fixed during the sample period. Note that HMDA reports the actual APR for all loans where the APR exceeds the 3 percentage point threshold. Therefore, by adjusting the threshold upwards from 3 percent in years with a higher share of rate spread loans, we can set the share of rate spread loans to the same percentage for each year.

[16] These controls include share of residents black, Hispanic, Asian, 65 years old or older, or married, share of households in poverty, and median household income; and from the 2004 HMDA data share of borrowers black, Hispanic, Asian and in poverty, and median family income of borrowers.

county, and census tract.[17] These mortgages were sampled from May through August and then matched using name and address by Experian PLC to the March 31st archival record preceding the mortgage transaction and March 31st record for every year that follows this transaction through 2009.[18] Our panel contains one observations for every year following the year of origination through 2009, and the foreclosure variable is set to 1 if one or more foreclosure reports are present in the credit record in that year and zero otherwise. The sample includes weights calculated at the loan level based on the probability of selection where each site receives equal weight in the pooled sample, and we use these weights as a robustness test.[19] See Bayer, Ferriera and Ross (2016) for more details on the data.[20]

Table 1 shows the means for our final home purchase sample of post mortgage credit reports by tercile of the share of loans from high cost lenders in a PUMA. The share of loans from high cost lenders is strongly correlated with many loan attributes. Both the foreclosure rate and the likelihood that an individual loan is a high cost or rate spread loan increases dramatically across the terciles. Borrowers are less white, lower income, less likely to have a co-borrower and have lower credit scores in the highest tercile by share of loans from high cost lenders. The mortgages also have higher loan to value ratios, higher expense to income ratios, are more likely to have subordinate debt and are more likely to have adjustable interest rates in the highest tercile. The

---

[17] In the Dataquick sample, we eliminate non-arm's length transactions, transactions where the name field contains the name of a church, trust, or where the first name is missing, and transactions where the address could not be matched to a 2000 Census tract or the zip code was missing. Data was provided by DataQuick Information Systems, Inc. www.dataquick.com.

[18] The sample includes oversamples of mortgages to minority borrowers, mortgages to white borrowers in minority or low-income neighborhoods, and high cost mortgages as designated in HMDA as high rate spread loans. The match rate for the pre-mortgage archive is 81.4 for the home purchase sample. For years following the mortgage origination, the match rate rises by 4 to 5 percentage points.

[19] The sampling is explicitly based on 8 strata for each site: black borrowers, Hispanic borrowers, white borrowers in minority or low-income neighborhoods, and all other borrowers divided into rate spread and non-rate spread loans. All loans from the same strata and year receive equal weight.

[20] The sample composition is quite stable except for a moderate decline in share white and moderate increase in loan amount between columns 1 and 2 associated with the difficulty of matching lender names between HMDA and the Dataquick provided assessor files.

PUMA's that are identified as having a higher share of loans from high cost lenders also have a higher share of black and Hispanic residents, and a lower median family income.

Table 2 shows the means of PUMA variables measured by origination or purchase year and weighted for our sample of home purchase mortgages. The table contains four columns, one for each purchase year between 2004 and 2007. Depending upon the threshold selected, the share of loans from high cost lenders doubles or even triples between 2004 and 2006, and falls by between 60 and 80 percent in 2007. This volatility in share of loans from high cost lenders is larger than any other observed volatility in the sample. The share of rate spread or high cost loans in a PUMA exhibits a similar, but less dramatic, pattern doubling between 2004 and 2006 and falling by half in 2007. All other borrower and loan attributes exhibit notably less variation over the period. Among traditional credit risk variables, only the PUMA composition over loan to value ratio increases substantially by about 25%, and similarly only share subprime credit score and high loan to value ratio exhibit large changes in 2007 falling by about 25%. The largest demographic composition change is in share of black borrowers increasing by 35% between 2004 and 2006 and falling by 12% in 2007. Market wide, there is a substantial increase in the application denial rate throughout the entire period of 37%, and a similar decrease in application volume of 41% between 2006 and 2007. The county level price indices rise by 29 percent between 2004 and 2006, but only fell by 4 to 5 percent in 2007.

2. Descriptive Results

We begin our analysis by creating some simple scatter plots of unexplained foreclosure rates versus the share of loans originated by high cost lenders by PUMA and origination/purchase year cohorts. We begin with our loan level sample of home purchase mortgages originated between 2004 and 2007. In our simplest model, we regress whether each loan ever faced foreclosure by the end of our credit profile data in March of 2009 and the share of loans from high cost lenders on whether the loan is a rate spread loan and on purchase year by site fixed effects. We condition on whether the loan is a rate spread loan to separate the risk associated with high cost loans from risks associated with the activity of lenders that tend to issue high cost loans. We also include purchase year by site fixed effects because Table 2 illustrates large changes in the volume of high cost loans and in the activity of high cost lenders over time. The timing and magnitude of these changes also vary significantly across our seven sites. These residuals for the ever foreclosed variable and for share loans from high cost lenders are then collapsed into purchase year by PUMA cells.

Figure 1A on the left hand side of the figure presents the scatter plot for the cell means of residuals from the regression above with a linear regression line plotted for the PUMA by purchase year data. The scatter plot and the regression line indicates a strong positive cross-sectional correlation between unexplained variation in foreclosure rates in a PUMA and the share of loans from high cost lenders in the PUMA. Figure 1B on the right hand side uses cell means from residuals after also controlling for detailed borrower and loan attributes, including race and ethnicity, family income deciles, borrower gender, presence of a co-borrower, Vantage credit score in 20 point bins, bins for loan to value and expense to income ratios,[21] dummies for whether the loan is adjustable rate, has a mortgage amount above the conforming loan limit and whether the purchase involves the use of subordinate debt (a second lien), whether held in portfolio or if not the source of securitization, and structural attributes of the housing unit.[22] As shown by the much flatter regression line, the inclusion of these controls in the regression model leads to a substantially

[21] Loan to value ratio bins are based on thresholds of 0.6, 0.8, 0.85, 0.9, 0.95, 1.0 and 1.05. Mortgage payment and debt payment to income ratio bins are based on 0.02 and 0.03 increments, respectively, with larger bins on the edges of the distribution.

[22] The housing attributes include number of bedrooms, number of bathrooms, square feet of living space, lot size, whether the units is single or multi-family, condominium or mobile home, and the number of stories and number of units for multi-family structures.

weaker cross-sectional relationship between foreclosure rates and the share of loans from high cost lenders, consistent with a relationship that is driven heavily by omitted variable bias.

Figure 2 is based on residuals from the same models except that the models also include PUMA fixed effects so that the scatter plot residuals are based on changes over time within PUMA's. Again, Figure 2B also conditions on borrower and loan attributes. Figures 2A and 2B are virtually identical. The inclusion of borrower and loan attributes in the regression models yield residuals in Figure 2B that imply a very similar relationship between changes in foreclosure rates and changes in the share of high cost lenders across purchase cohorts of loans. Figure 2 suggests that PUMA fixed effects successfully capture much of the bias from omitted borrower and loan attributes that was observed in Figure 1.

Table 3 Panel 1 shows the cross-sectional estimates from equation (1) and is comparable to the scatterplot in Figure 1. The first column only includes the PUMA share of loans from high cost lenders, whether the borrower has a high cost loan and the site by purchase year by credit year fixed effects. The next columns in order add controls for borrower vantage (credit) score in 20 point bins, the borrower demographics listed above, the detailed mortgage attributes plus the controls for the physical structure of the housing unit, and finally lender fixed effects. The first column shows the strong conditional correlation between the likelihood of foreclosure and the share of loans from high cost lenders. However, as in Figure 1B, the inclusion of controls substantially erode the magnitude of the estimates, and the estimate in column 4 is approximately half the size of the estimate in column 1.

It is notable that it is primarily borrower, rather than loan attributes, that erode the estimated effect of lender share. The inclusion of credit score reduces the initial estimate by 20 percent, and the inclusion of demographics further reduces the initial estimate by another 26 percent. On the other hand, controls for loan terms such as LTV, income ratios, subordinate debt and adjustable rates loans, which are expected to correlate strongly with subprime lending activity, only reduce the estimated effect by 4 percent, and the inclusion of lender fixed effects only reduces the original estimate by 2.5 percent. Based on observables, much of the cross-sectional relationship between foreclosure and the activity of high cost lenders is associated with borrower attributes, rather than mortgage attributes like high loan to value ratio, adjustable interest rate or subordinate debt that might correlate with high risk mortgage attributes, such as rate resets or prepayment penalties as described in Reid, Bocian, Li and Quercia (2016). These findings appear consistent with conclusions of Bayer, Ferriera and Ross (2018) that lenders with high “ex post” foreclosure rates systematically operated in a segment of the mortgage market that involved a priori higher risk lending opportunities.

Table 3 Panel 2 shows the estimates for equation (2). These models include PUMA by credit report year fixed effects and so are comparable to the scatterplots in Figures 2A and 2B. The model also includes the interaction of time invariant PUMA observables from the 2000 Decennial Census and from the 2004 HMDA data.[23] The estimates in Panel 2 are very similar in magnitude to the smaller estimates in columns 3 and 4 of Panel 1. Only the controls for borrower demographics have an appreciable effect on the estimates for share of loans from high cost lenders, and the reduction is quite modest at only 9 percent. The inclusion of PUMA fixed effects appears to have eliminated most of the bias associated with omitted borrower and loan attributes that was captured by the controls in Panel 1. In terms of magnitude, a one standard deviation change in the share of loans from high cost lenders is approximately 9.7 percentage points, and so the

[23] The controls include the share of residents who are black, Hispanic, Asian, 65 years old or older or married, the share of families in poverty and the median family income in the PUMA based on the 2000 Decennial Census, as well as the share of homebuyers who are black, Hispanic, Asian, or in poverty, and the median family income of homebuyers from 2004 HMDA data.

standardized effect in panel 2 column 4 is 1.23 percentage points or 23 percent of the 0.053 foreclosure rate in the sample.

4. Quasi-experimental Results

The last panel of Table 3 presents results from regressions that replace the share of loans from high cost lenders with the Bartik or shift-share style prediction and also includes lender by year fixed effects to assure that results are not driven by national trends in lender market share. The estimates are very stable in magnitude as additional borrower and loan attributes are added to the model.[24] Further, while not shown, the inclusion of lender by year fixed effects have minimal effects on our estimates. In order to calculate the standardized effect, we add the 2004 level back into the shift-share predicted change to include the cross-sectional variation. The standard deviation of this prediction is smaller than the deviation for the actual share at 0.067, while the point estimate is about double the point estimate from panel 2. As a result, the standardized effect from panel 3 column 4 is 19% higher than the estimated effect in panel 2 with a one standard deviation change implying a 1.46 percentage point change in foreclosure, or approximately 27 percent of the baseline foreclosure rate.

Next, we conduct a series of robustness tests. In Table 4 Panel 1, we present the estimates for our model from Panel 3 of Table 3 using two alternative shift-share predictions for the change in the share of loans from high cost lender based on thresholds of 15 percent and 25 percent or more of high cost loans for defining whether a lender is high cost. Both of these results are shown

---

[24] Oster (2019) argues for evaluating parameter stability by comparing the change in the parameter estimate to the change in the residual variation based on the R-squared. Assuming that 50 percent of the variation in foreclosure can be explained by observed and unobserved factors at the time of mortgage origination (a conservative assumption relative to Oster's recommendations for setting the maximum R-squared), a comparison of column 1 to column 4 in panel 3 implies that unobservables would have to bias estimates in the opposite direction of observables and be 2.6 times as influential as the observables, which in our case contain all critical underwriting variables, in order to eliminate our estimated effects.

for the baseline model specification in column 1 of Panel 3 in Table 3 and for the specification including all borrower and loan controls (column 4).

Panel 2 of Table 4 presents a series of additional robustness tests using the Table 3 column 4 specification that includes all controls. The first column presents results from a balanced panel of credit reports where foreclosure outcomes are only included for the 2008 and 2009 credit report years that are responsible for the vast majority of foreclosure filings. The second column assigns lenders as high cost using an alternative rate spread definition that is designed to keep the total share of rate spread loans constant over purchase year during our sample period. Adjustable Percentage Rates (APR) are based on the full term of the mortgage, and expected rates of pre-payment may affect the spread between APR and treasury bonds, which cannot be pre-paid. By holding the total fraction of high cost loans in the sample constant across cohorts, we eliminate or at least mitigate these interest rate environment effects on the definition of a rate spread loan. The threshold is set to 13% for high cost lenders due to the substantial decline in the number of high cost loans given this approach. The third column estimates a regression using the sample weights based on the stratified sampling strategy used for collecting the sample of mortgages, see Bayer et al. (2016) for more discussion of the weights. Column 4 allows the effects to differ based on whether the predicted change is negative or positive. The results in Panel are robust with the standardized foreclosure effects ranging between 1.2 and 1.5 percentage points. The estimated effects for negative and positive changes are virtually identical. The effects for predicted declines are more precisely estimated because, unlike predicted increases in high cost lender share that arise from gradual expansions, predicted declines include some significant closures of large subprime lenders in 2006 and 2007.

Finally, as discussed above, we address recent concerns raised about shift share analyses by organizing PUMA's into clusters with similar 2004 shares of loans originated by high cost lenders, and then including purchase year by initial share cluster fixed effects in the model from Table 3 Panel 3 Column 4. Panel 3 of Table 4 shows these results for several initial share cluster definitions. Column 1 presents results based on a separate cluster for 0.01 intervals on the fraction of loans in 2004 originated by high cost lenders leading to 124 purchase year by initial share cluster cells. Columns 2, 3 and 4 present similar results except the intervals are based on 0.02, 0.03 and 0.04 with the number of cells falling to 60, 48, and 36, respectively. The magnitude of the estimates on the predicted change in market share rise from 0.217 in Table 3 to 0.279 in column 1 for a 0.01 interval, falling as the interval is enlarged to 0.244 and 0.182 for columns 2 and 3, and finally rising back to 0.204 in column 4 with an interval of 0.04. While these estimates are not as stable as the rest of the estimates in Tables 3 and 4, they are always sizable and strongly significant, and never separated from the estimates in Table 3 by more than a standard deviation. Further, in terms of inference, after a modest increase in the standard errors with the addition of purchase cohort by cluster fixed effect, the standard error estimates are quite stable as we change the interval even when the number of clusters falls to only 36.[25]

All of the robustness tests in Table 4 exhibit similar stability in the parameter estimates as controls are added. We also run the robustness tests in Panels 1 and 2 for the non-shift share models from Panels 1 and 2 of Table 3, and the qualitative and quantitative results in Table 3 continue to be robust. We return to our baseline models in Panel 3 of Table 3 for all follow-up analyses below.

[25] We can also change the cluster definition by altering where the intervals start. In column 1, for example, the intervals always start on a round value of 0.01 or 1 percentage point. Alternatively, we can always start intervals at points involving 0.005 with the interval ending on the next 0.005. This alternative clustering yields very similar estimates and standards errors, 0.277 (0.079) as compared to 0.0279 (0.079) in column 1 of panel 3.

4.1 Heterogeneity and Lender Market Share

Our first step in understanding what might drive this relationship is to examine whether the effect of the shift-share prediction is isolated among only a subset of loans. Table 5 estimates a model where the effect of PUMA share of loans from high cost lenders is allowed to be heterogeneous across borrowers, loans or lenders. The first column presents results where effects are allowed to vary by whether the loan itself is a high cost or rate spread loan. The second column allows effects to vary by the race or ethnicity of the borrower. The third column allows for differential effects based on whether the loan was securitized by one of the government sponsored enterprises, held in portfolio, or privately securitized. The final column interacts the predicted change in PUMA share of loans from high cost lenders with whether the borrower's lender was a high cost lender. We do not find any evidence that effects are larger for high cost loans, minority borrowers or for privately securitized loans.[26] However, Column 4 indicates that the effect of PUMA share of loans from high cost lenders on foreclosures is substantially larger for borrowers who obtained loans from high cost lenders. The total standardized effect for these high cost lenders using the shift-share prediction is 3.07 percentage points or 28 percent of the higher average foreclosure rate of 10.7 percent. Notably, we still observe a sizable effect of this proxy for market penetration of high cost lenders for borrowers who did not receive their loan from a high cost lender, the resulting standardized effect falls to 1.06 percentage points, but these loans have a smaller baseline foreclosure rate of only 3.5 percent so the relative standardized increase is sizable at 31 percent.

Given the larger absolute effect for high cost lenders themselves, perhaps some of the estimated effects of PUMA high cost lender representation arise because a lender's own market

[26] The coefficient on the interaction with whether the borrower is Hispanic is significant at the 10% level, but the estimate is negative suggesting that these effects are less concentrated among Hispanic borrowers.

share influences foreclosure outcomes for the loans that this lender originated. Therefore, we include a control for changes in a lender's market share (column 2), or predicted changes in the lender's share (column 3) using the same Bartik/shift-share formula described in equation (3) except for just a single lender's PUMA weighted change in market share in the numerator. The results are shown in Table 6. Panel 1 shows models that just include either lender's actual market share in each purchase year or predicted changes in market share for each year. The estimates on these two variables are small and insignificant, and the estimates on the share of loans from high cost lenders are unchanged.

Panel 2 column 1 of Table 3 presents the estimates from Table 5 Column 4 where predicted PUMA share is interacted with whether the borrower's lender is a high cost lender. In column 2, the lender's actual market share is also interacted with the high cost lender dummy, and in column 3 the predicted change in lender market share is interacted with high cost lender. An increase in predicted lender's market share over time leads to higher foreclosure rates, a standardized effect of 2.1 percentage points,[27] but only when the borrower's lender is a high cost lender. However, controlling lender market share has at most a modest effect on the predicted high cost lender share coefficients. Specifically, the standardized effect of predicted change in market share for low cost lenders falls by less than 8 percent and the standardized effect of predicted share for high cost lender falls by only 10 percent.

In summary, the effects of high cost lender market penetration or representation are broad based, and are not concentrated among especially vulnerable groups of borrowers. These effects occur for loans made by both high cost and non-high cost lenders. However, these shift-share

[27] The standard deviation of lenders' PUMA market share is approximately 0.01.

prediction effects are substantially larger in absolute (but not relative) terms among loans from high cost lenders, even after controlling for the market representation of the individual lenders.

5. Potential Mechanisms

As discussed in the introduction, we investigate three broad mechanisms that might explain the geographic correlation between the market penetration or representation of high cost lenders and the foreclosure rates observed during the crisis. First, these correlations may represent a type of reverse causality where high cost lenders increase their representation within housing submarkets in response to changes in the demand for mortgage credit from borrowers who are relatively risky based on their unobservables. Second, high cost lenders themselves may play a role in the concentration of foreclosures by issuing riskier mortgages that have features like very high combined loan to value ratios, resetting variable rates and pre-payment penalties. Finally, a third possibility is that cohorts of loans are serviced differently in housing submarkets and origination years where high cost lenders issued a substantial fraction of loans.

5.1 Unobserved Borrower Heterogeneity

Our evidence from Table 3 Panels 2 and 3 show that the inclusion of observed borrower attributes has little effect on the conditional correlation between our proxies for the market penetration of high cost lenders and foreclosure rates. As a further test on the importance of borrower attributes, we identified two pieces of information from contemporaneous borrower credits reports, bank card delinquencies and medical collection trades, which were not directly related to mortgage distress and might be driven by unobserved borrower attributes and experiences during the crisis. Table 7 presents the estimates from models that include these controls. The first column presents the results after including a control for the number of bank card delinquencies, the second column presents results conditional on the number of medical collection trades and the aggregate dollar amount of those trades, and the third column presents results for a model including both the bank card and medical collection controls. While bank card delinquencies have a strong correlation with foreclosure, the inclusion of these controls does nothing to erode the relationship between high cost lender market representation at origination and foreclosure during the crisis.

Next, we recognize that borrower unobservables may affect the likelihood of a borrower obtaining a high cost or rate spread loan. Therefore, we estimate models similar to our foreclosure models except that we move whether the loan was a rate spread loan to the left hand side of the regression equation. We estimate this model using our sample of mortgages (as opposed to mortgages by credit report year) and condition on PUMA fixed effects, the trends based on PUMA observables over origination year, site by origination year fixed effects and lender by origination year fixed effects. The fourth column presents these results with no additional controls other than predicted share, and the fifth column presents results after including all borrower, loan and housing unit controls. The resulting estimates are small and statistically insignificant. Unlike like foreclosure, the predicted share of loans from high cost lenders does not predict whether a loan is high cost after controlling for PUMA and lender.

Finally, Bayer et al. (2016) find that the foreclosure risk of minority borrowers is substantially larger for loans originated near the peak of the housing market, even after including controls for negative equity. They conclude that as credit constraints relax borrowers who are higher risk on unobservables will choose to enter the mortgage market, contributing to the higher foreclosure rates for mortgages originated at the peak of the market. One way to capture changes in the composition of borrowers is to control for trends over purchase years in borrower attributes at the PUMA level. We create three PUMA by purchase year level variables: share of mortgages with reported family income below the poverty line, share of mortgages with either a black or Hispanic borrower[28] and share of mortgages where the borrower has a subprime credit score (Vantage score under 701). While individual level controls capture the direct effect of correlated borrower attributes, these PUMA trends can capture broad shifts in the borrowing population that might correlate with population changes on unobservable attributes. Note that family income and minority borrower variables are based on the full HMDA sample, while credit score must be based on our specific matched sample. We then include each of those variables one at a time in order to test whether their inclusion erodes our estimated relationship between foreclosure and the predicted share loans from high cost lenders. These estimates are shown in Table 8 Panel 1. While foreclosure rates increase as the share of minority borrowers in the PUMA increases and surprisingly decrease with the share of loans to borrowers in poverty, the inclusion of these controls has no impact on the relationship between the expected activity of high cost lenders in a location and future foreclosure rates.

5.3 Unobserved aspects of Subprime Lending Activity

As noted earlier, controls for observed loan terms and attributes and inclusion of lender fixed effects had no impact on our estimated geographic effects, even though the use of adjustable rate products and subordinate debt, as well as lender identity, strongly correlate with subprime lending activity. Further, in the second Panel of Table 8, we develop additional controls at the PUMA level that could capture submarket level changes in the types of loan products being offered and the types of loans being made in a PUMA. Specifically, we present estimates that include controls for the share of loans in a PUMA and purchase year that are high cost based on the HMDA rate spread variable, have a high loan to value ratio (above 0.95), and have a conforming loan

[28] Similar results arise including share black loans and share Hispanic loans as separate control variables.

disqualifying total debt expense to income ratio (above 0.45). As above, the high debt expense variable must be calculated using only our matched sample. The final panel presents estimates from models that include controls from HMDA on the market environment in the PUMA and purchase year including the denial rate,[29] the application volume and a Herfindahl measure of market concentration of mortgage lenders using the HMDA respondent identification number. While some of these variables are statistically significant, their inclusion in the model has virtually no impact on the relationship between the expected activity of high cost lenders in a location and future foreclosure rates.[30]

Another potential way in which high cost lending may have influenced foreclosures is through an expansion of credit that increased demand for housing and drove up local housing prices. Specifically, an increase in the representation of high cost lenders in a PUMA may have pushed up prices in the PUMA and led to larger price declines during the crisis. For example, Mian and Sufi (2009) observe that zip codes with high levels of subprime mortgage market activity experienced greater increases in housing prices leading up to the crisis and higher foreclosure rates during the crisis. We calculate housing price indices by county by quarter by year and by PUMA by quarter by year using the full sample of Dataquick housing transactions. The price level for a purchase cohort of loans is based on the average of the second and third quarter price index in each year because we have a sample of matched transactions between May and August, and the price level in each credit report year is based on the average of the first and second quarter price indices because our credit profile data is based on March 31 archives for each year. Using the price indices,

---

[29] The denial rate is calculated as the number of loan applications denied divided by the sum of the number of loan applications denied and the number of loans applications originated.

[30] For example, an increase in the share of loans with high loan to value ratios is associated with lower foreclosure rates overall after conditioning on the actual loan to value ratio of the individual loans, perhaps because lenders become more restrictive in lending as the share of low down payment loans increase.

the purchase year and the initial loan to value ratio, we create dummy variables for whether the mortgage was near negative equity (above 0.9 or above 0.95), in negative equity, or has negative equity levels 10-30%, 30-50% and above 50%.[31]

We also use the American Community Survey (ACS) to create measures of employment for each credit report year at the county level. Specifically, for each ACS survey year and county in our sample identified in the ACS, we calculate the fraction of all prime aged males from age 26-55 (whether in the labor market or not) who report being employed. We also calculate a race specific employment rate for each year and county using the white, black, Hispanic and Asian subsamples of the ACS. We interact these employment rate controls with the negative equity dummy variables to capture the well-established phenomenon that households in negative equity often do not enter foreclosure without some trigger event, such as loss of a job.

Table 9 presents our estimates after including additional controls for negative equity based on county (Panel 1) and PUMA (Panel 2) housing price indices. The first column simply includes the negative equity dummy variables. The second column includes the interaction of negative equity with the county level employment variables, and the third column includes the same interactions with the own-race county employment rate. While not shown, the negative equity dummy variables are strongly associated with foreclosure using either county or PUMA price indices, but those effects weaken substantially as county employment rates rise. However, these controls have at most a modest impact on the estimated effects of the market representation of high cost lenders. Adding controls for negative equity based on county price indices and county employment rates increases our estimate on share of loans from high cost lenders by 5 to 10

[31] Given the broad negative equity bins and the short period of time between origination and the foreclosure crisis, we do not consider amortization of the mortgage balance.

percent, while using PUMA price indices increases our estimates by between 15 and 17 percent, and the estimated effect never falls in magnitude for any of our models in Table 9.

5.5 Mortgage Servicers and Foreclosure Decisions

A final potential explanation relates to the foreclosure behavior of loan servicers. Potentially, the entry of these high cost lenders into a housing submarket affects the foreclosure strategy of the major loan servicers during the crisis. Since each cohort of loans in a PUMA could follow a different time path of foreclosure during the crisis, changes in lender management of delinquent loans might vary across cohorts in ways that are related to the composition of each cohort of mortgages.

While our data does not contain information on servicers, we are able to pursue two analyses that are at least suggestive of a role for servicers in the geographic concentration of foreclosures. First, we estimate models to explain whether the loan experienced a 90 to 180 day mortgage delinquency during a given credit year because, while foreclosure is to some extent a lender decision, a mortgage delinquency represents the borrower's failure to make their mortgage payments independent of any lender actions or decisions. Column 1 of Table 10 presents the foreclosure results from Table 3 Panel 3 column 4, while column 4 presents the comparable 90 day delinquencies estimates. While the estimated standardized effect for 90 day delinquency is positive, the estimates are statistically insignificant, and the standardized effect is only 1/3rd of the size of the estimated foreclosure effect, which implies only a 3.5 percent change in the sample average 90-180 day delinquency rate.

We also re-estimate the models from Table 6 where both predicted high cost lender and individual lender market shares can vary by whether the lender is high cost or not. Columns 2 and 3 repeat the foreclosure results from Table 6, and columns 5 and 6 present the results for 90 day delinquency for comparable models. The market share level and high cost lender interaction estimates are both statistically insignificant in column 5. The level point estimate is a little over half the magnitude of the equivalent foreclosure estimate in column 2. The interaction estimate between high cost lender and our shift-share prediction is negative, the opposite sign of the estimate in column 2. In column 6, the level market share estimate is significant in the lender market share model and larger in magnitude, but this result occurs because of the negative, statistically insignificant estimate on the interaction of high cost lender market representation with loan originated by high cost lender, which again is the opposite sign of the effect in our foreclosure models. All other coefficients in the model are insignificant, and in summary the empirical patterns that we have documented for foreclosure are simply not present for mortgage delinquency. Notably, the effect of having a high cost loan on mortgage delinquency is consistently positive and statistically significant.

One last way to examine the role of discretion in foreclosure is to compare California, a purely administrative foreclosure state, to the other states in the sample where the judicial branch plays a substantial role in the foreclosure process. Specifically, in California, loans that entered severe delinquency typically move to auction after 120 days delinquent without requiring any court filings or court approval.[32] So, if our effects arise from servicer discretion in when to file foreclosures in court, we should not expect to see these effects in California. We re-estimate our baseline models separately for our California subsample of mortgages from Los Angeles and the San Francisco bay area and for our other five sites. Due to the large size of the California sites, the

[32]During our sample period, delinquent loans in California were regularly issued a Notice of Default when the loan became 90 days delinquent followed by a Notice of Trustee Sale when the loan became 120 days delinquent. The Notice of Trustee Sale should typically lead to a foreclosure report to the credit reporting agencies. While Maryland and Colorado officially have administrative foreclosure, their foreclosure process is court supervised and still allows for substantial lender discretion in the timing of mortgage delinquency.

number of clusters are similar in the two subsamples with over 150 PUMA's in California and about 180 PUMA's in the five other sites. The results for the other five sites are shown in column 1 of Table 12, and the standardized estimate is 1.83 percentage points or a 40 percent increase in foreclosure likelihood. On the other hand, the point estimated for the two California sites (while noisily estimated) is only 7 percent of the baseline California foreclosure rate. Notably, the variable for having a rate spread loan remains a significant predictor in the California subsample, even though share of loans from high cost lenders is statistically insignificant.

6. Summary and Conclusions

In this paper, we first estimate the cross-sectional correlation between foreclosure and the share of loans in a housing submarket originated by lenders who tend to issue a large number of high cost loans. The magnitude of these estimates is very sensitive to the inclusion of borrower controls. Controlling for credit score, borrower race and ethnicity, age and co-borrower status reduces the estimated relationship between the share of high cost lenders by almost half. On the other hand, the inclusion of detailed controls for standard underwriting risk variables and lender fixed effects had only a modest effect on the estimated relationship. These findings are consistent with borrower unobservables being an important factor in explaining the cross-sectional concentration of foreclosure in housing submarkets.

We also document a strong within submarket across cohort relationship between the market share of high cost lenders and foreclosure using both the share of loans from high cost lenders and a Bartik or shift-share prediction of the share of loans. After controlling for submarket or PUMA fixed effects, the estimates are robust to a wide variety of specifications including detailed controls for borrower and loan attributes, lender fixed effects, alternative definitions of high cost lenders, purchase year by cluster of similar PUMA fixed effects, and the inclusion of a wide variety of alternative mortgage market composition variables for each PUMA and purchase year. While these effects are larger when the loans were originated by high cost lenders, the effects are broad based influencing foreclosure for non-rate spread loans, white loans, loans originated to be sold to the Government Sponsored Enterprises, and loans made by lenders that are not labelled high cost.

We investigate several potential mechanisms behind this second phenomenon. We find no evidence that our effects can be explained by contemporaneous shocks to borrowers, as captured by credit report outcomes on bank card delinquencies and medical collections, or captured by trends in the composition of PUMA borrowers. We also find no evidence that our effects can be explained by changes in observable mortgage attributes, changes in the patterns of lending in the PUMA or changes in housing prices that might have been driven by expansions in mortgage credit. However, our results do not arise when considering mortgage delinquency, which unlike foreclosure, is primarily driven by borrower circumstances. For 90 day delinquency, our estimated effects of the predicted share of loans from high cost lenders is one-third the size of the effects on foreclosure and statistically insignificant. Further, we find that our effects are concentrated in the sites that are located outside of California. In California, foreclosure is purely administrative, while in all other states in our sample courts play a significant role and as a result lenders may exercise more discretion in terms of when to enter the foreclosure process.

While we began this study with the supposition that lender activities in the run up to the crisis might have contributed to concentrated foreclosure rates in specific housing submarkets, the evidence in this paper suggests that borrower attributes can explain a substantial share of the cross-sectional correlation between foreclosure and our geographic submarkets and that lender servicers likely play a significant role in explaining the geographic concentrations that we document across cohorts of loans within submarkets.

References

Adão, Rodrigo, Michal Kolesár and Eduardo Morales. 2019. Shift-share designs: Theory and inference. *Quarterly Journal of Economics*, 134(4), 1949-2010.

Altonji, Joseph G., Todd E. Elder and Christopher R. Tabor. 2005. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1), 151-84.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet and Douglas Evanoff. 2014. Predatory Lending and the Subprime Crisis? *Journal of Financial Economics*, 113(1), 29-52.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet and Douglas Evanoff. 2020. Financial Education versus Costly Counseling: How to Dissuade Borrowers from Choosing Risky Mortgages?" American Economic Journal: Economic Policy, 12(1), 1-32.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet and Thomas Piskorski. 2017. Policy intervention in debt renegotiation: Evidence from the home affordable modification program. Journal of Political Economy, 125(3), 654-712.

Autor, David H. and Mark G. Duggan. 2003. The Rise in the disability rolls and the decline in unemployment. *Quarterly Journal of Economics*, 118(1), 157–205.

Bayer, Patrick., Fernando Ferreira and Stephen L. Ross. 2014. Race, ethnicity and high-cost mortgage lending. NBER Working Paper #20762.

Bayer, Patrick., Fernando Ferreira and Stephen L. Ross. 2016. The vulnerability of minority homeowners in the housing boom and bust. *American Economic Journal: Economic Policy*, 8, 1-27.

Bayer, Patrick, Fernando Ferreira and Stephen L. Ross. 2018. What drives racial and ethnic differences in high cost mortgages? The Role of High Risk Lenders. *Review of Financial Studies*, 31(1), 175-205.

Bhutta, Neil and Benjamin Keys. In Press. Moral hazard during the housing boom: Evidence from private mortgage insurance. *Review of Financial Studies*.

Brunner, Eric, Stephen L. Ross and Ebonya Washington. 2011. Economics and ideology: Causal evidence of the impact of income on support for redistribution and other proposals. *Review of Economics and Statistics*, 93, 888-906.

Calem, Paul S., Kevin Gillen and Susan Wachter. 2004. The neighborhood distribution of subprime mortgage lending. *Journal of Real Estate Finance and Economics*, 29, 393-410.

Calem, Paul S., Jonathan E. Hershaff and Susan M. Wachter. 2004. Neighborhood patterns of subprime lending: Evidence from disparate cities. *Housing Policy Debate*, 15(3).

Campbell, John Y., Stefano Giglio and Parag Pathak. 2011. Forced sales and house prices. *American Economic Review*, 101(5), 2108-2131.

Carozzi, Filipe, Christian Hilber and Xiaolun Yu. 2020. On the economic impacts of mortgage credit: evidence from help to buy. CEP Discussion Paper No 1681.

Chalak, Karim. 2019. Identification of average effects under magnitude and sign restrictions on Confounding. *Quantitative Economics*, 10, 1619–1657.

Chan, Sewin, Michael Gedal, Vicki Been and Andrew Haughwout. 2013. The role of neighborhood characteristics in mortgage default risk: Evidence from New York City. *Journal of Housing Economics*, 22(2), 100-118.

Chan, Sewin, Andrew Haughwout and Joseph Tracy. 2015. How mortgage finance affects the urban landscape. In the *Handbook of Urban and Regional Economics Vol. 5A*, Eds. G. Duranton, V. Henderson and W. Strange, 987-1045. Amsterdam: North Holland.

Dell'Ariccia, Giovanni, Deniz Igan and Luc Laeven. 2012. Credit booms and lending standards: Evidence from the subprime mortgage market. *Journal of Money, Credit and Banking*, 44(2-3), 367-384.

Edmiston, Kelly D. 2009. Characteristics of high-foreclosure neighborhoods in the tenth district. Economic Review (2nd Quarter), 51-75.

Fisher, Lynn, Lauren Lambie-Hanson and Paul S. Willen. 2011. A profile of the mortgage crisis in low and moderate income neighborhoods. In *The American Mortgage System: Crisis and Reform (The City in the Twenty-First Century)*, Eds. S. Wachter and M. Smith, 137-158. Philadelphia: University of Pennsylvania Press.

Foote, Christopher L. and Paul S. Willen. 2018. Mortgage-default research and the recent foreclosure crisis. *Annual Review of Financial Economics*, 10, 59-100.

Gerardi, Kristopher, Eric Rosenblatt, Paul S. Willen and Vincent W. Yao. 2015. Foreclosure externalities: New evidence. *Journal of Urban Economics*, 87, 42-56.

Gerardi, Kristopher S., Stephen L. Ross and Paul Willen. 2011. Understanding the foreclosure crisis. *Journal of Policy, Analysis and Management: Point-Counterpoint*, 30, 382-388.

Gerardi, Kristopher S. and Paul Willen. 2009. Subprime mortgages, foreclosures, and urban neighborhoods. *BE Journal of Economic Analysis and Policy, 9 (3), Symposium Article 12*.

Ghent, Andra C. 2011. Securitization and Mortgage Renegotiation: Evidence from the Great

Depression. *Review of Financial Studies*, 24(6), 1814-47.

Ghent, Andra, Ruben Hernández-Murillo and Michael Owyang. 2014. Differences in subprime loan pricing across races and neighborhoods. *Regional Science and Urban Economics*, 48, 199–215.

Goldsmith-Pinkham, Paul, Isaac Sorkin and Henry Swift. 2020. Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8), 2586-2624.

Gupta, Arpit. 2019. Foreclosure contagion and the neighborhood spillover effects of mortgage defaults. *Journal of Finance*, 74(5), 2249-2301.

Guren, Adam M. and Tim J. McQuade. 2020. How Do Foreclosures Exacerbate Housing Downturns? *Review of Economic Studies*, 87(3), 1331-1364.

Huang, Weiran, Ashlyn Nelson and Stephen L. Ross. 2018. Foreclosure Spillovers within Broad Neighborhoods. HCEO Working Paper 2018-096.

Jiang, Wei, Ashlyn Nelson and Edward Vytlacil. 2014a. Liar's loan? Effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics*, 96(1), 1-18.

Jiang, Wei, Ashlyn Nelson and Edward Vytlacil. 2014b. Mortgage securitization and loan performance: A contrast of ex ante and ex post relations. *Review of Financial Studies*, 27(2), 454-483.

Keys, Benjamin, Tanmoy Mukherjee, Amit Seru and Vikrant Vig. 2009. Did securitization lead to lax screening? Evidence from subprime loans. *Quarterly Journal of Economics*, 125, 307–362.

Kruger, Samuel. 2018. The effect of mortgage securitization on foreclosure and modification. *Journal of Financial Economics*, 129(3), 586-607.

LaCour-Little, Michael, Charles A.Calhoun and Wei Yu. 2011. What role did piggyback lending play in the housing bubble and mortgage collapse? *Journal of Housing Economics*, 20(2), 81-100.

LaCour-Little, Michael and Jing Yang. 2013. Taking the Lie Out of Liar Loans: The Effect of Reduced Documentation on the Performance and Pricing of Alt-A and Subprime Mortgages. *Journal of Real Estate Research, 35(4), 507-554.*

Mayer, Christopher and Karen Pence. 2009. Subprime mortgages: What, where, and to whom? In *Housing Markets and the Economy: Risk, Regulation and Policy*, Eds. E. Glaeser and J. Quigley, 149-197. Cambridge, MA: Lincoln Land Institute of Land Policy.

Mian, Atif and Amir Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *Quarterly Journal of Economics*, 124(4), 1449–1496.

Mian, Atif, Amir Sufi and Francesco Trebbi. 2015. Foreclosures, House Prices, and the Real Economy. *Journal of Finance*, 70, 2587-2634.

Munroe, D. J. and L. Wilse-Samson. 2013. Foreclosure contagion: measurement and mechanisms. Working Paper.

Oster, Emily. 2019. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business and Economic Statistics*, 37(2), 187-204.

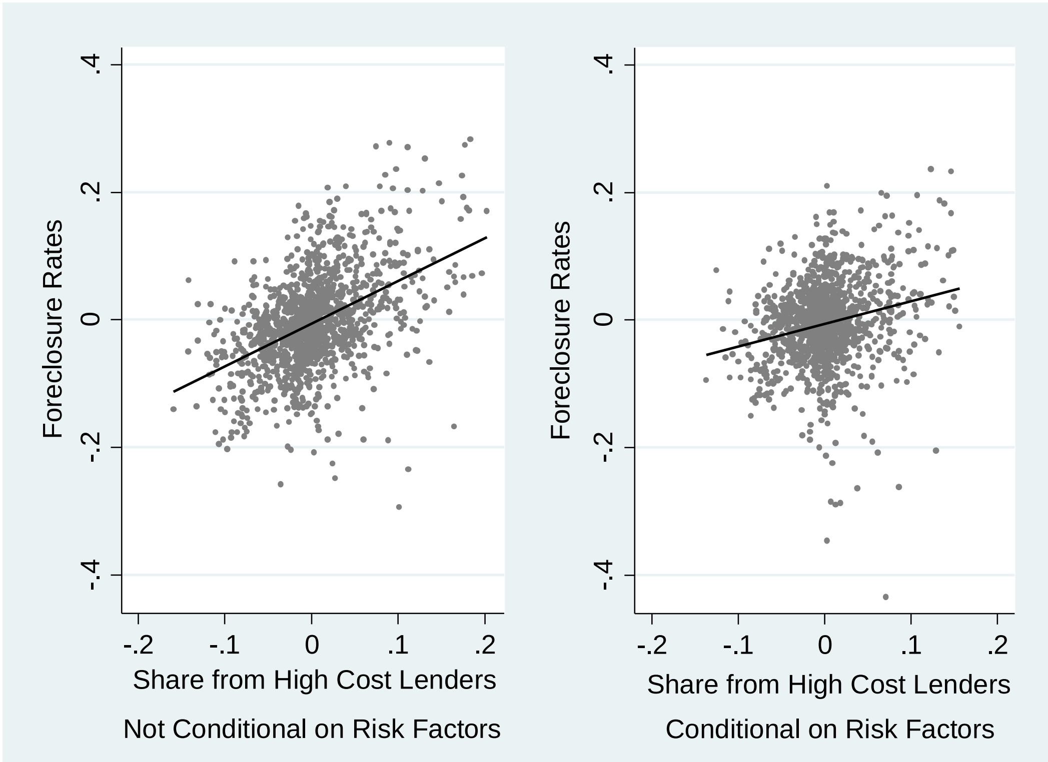
Reid, Carolina, Debbie Bocian, Wei Li and Roberto G. Quercia. 2016. Revisiting the subprime crisis: The dual mortgage market and mortgage defaults by race and ethnicity. *Journal of Urban Affairs*, 39(4), 1-19.

Reid, Carolina and Elizabeth Laderman. 2009. The untold costs of subprime lending: Examining the links among higher-priced lending, foreclosures and race in California. Federal Reserve Bank of San Francisco Working Paper.

Reid, Carolina, Carly Urban and Michael Collins. 2017. Rolling the dice on foreclosure prevention: Differences across mortgage servicers in loan modifications and loan cure rates. *Housing Policy Debate*, 27(1), 1-27.

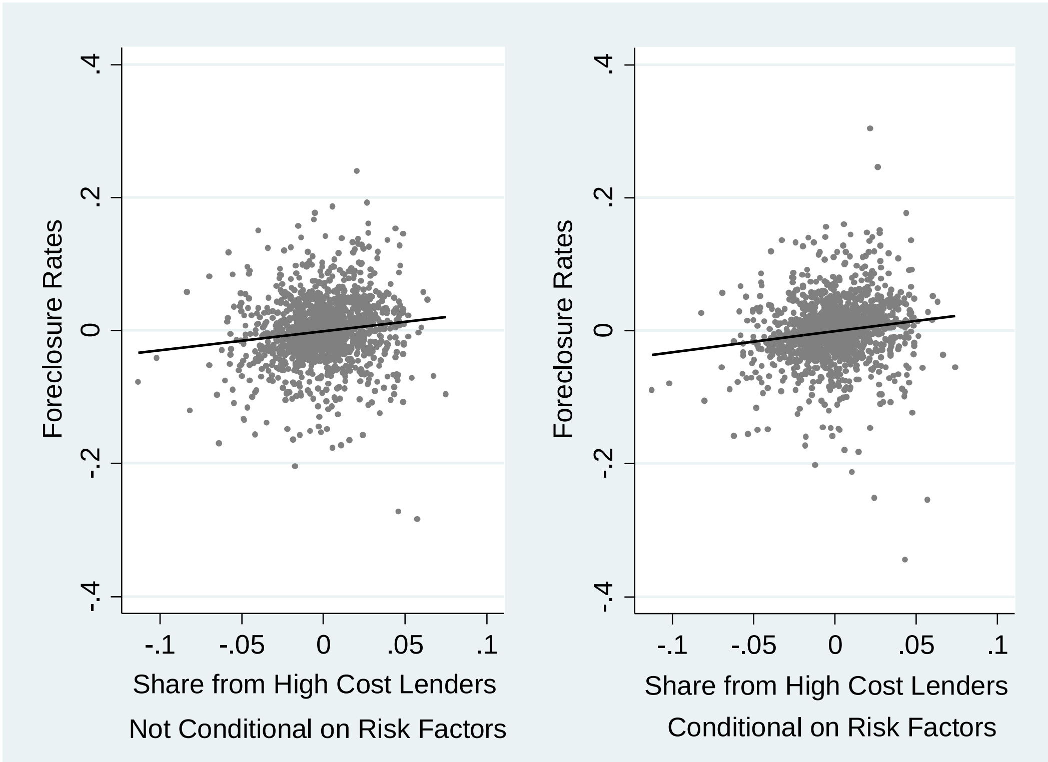
Ross, Stephen L. and John Yinger. 2002. *The Color of Credit: Mortgage Lending Discrimination, Research Methodology, and Fair-Lending Enforcement.* Cambridge, MA: MIT Press.

Figure 1: Mean Residuals by PUMA and Purchase Year



Notes: Residuals for foreclosure rates are based on regressing either foreclosure or share of loans from high cost lenders on rate spread loan and MSA by purchase year fixed effects. These residuals are then collapsed by PUMA and purchase year to get the means for these variables. The graph on the right hand side of the figure also includes additional risk factor controls, including individual level information on credit score and demographics, as well as loan and home characteristics. For a complete list of these controls, see Table 1.

Figure 2: Mean Residuals Conditional on PUMA



Notes: Residuals for foreclosure rates are based on regressing either foreclosure or share of loans from high cost lenders on rate spread loan, MSA by purchase year fixed effects, and PUMA fixed effects. These residuals are then collapsed by PUMA and purchase year to get the means for these variables. The graph on the right hand side of the figure also includes additional risk factor controls, including individual level information on credit score and demographics, as well as loan and home characteristics. For a complete list of these controls, see Table 1.

Table 1: Sample Descriptive Statistics

PUMA Shr from Lndrs >0.20	Lowest		Medium		Highest	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Foreclosure	0.027	0.218	0.047	0.299	0.096	0.435
Rate Spread Loan	0.047	0.211	0.134	0.341	0.317	0.465
American Indian	0.003	0.052	0.004	0.061	0.002	0.048
Asian	0.014	0.348	83.000	0.277	0.055	0.228
Black	0.036	0.186	0.081	0.274	0.167	0.373
White	0.398	0.459	0.678	0.467	0.429	0.495
Hispanic	0.112	0.328	0.154	0.361	0.347	0.476
Male	0.681	0.466	0.648	0.478	0.596	0.491
Female	0.316	0.465	0.350	0.477	0.402	0.490
Loan Amount (in 1000s)	340.834	240.105	269.724	207.197	239.664	155.362
Applicant Income (in 1000s)	119.413	111.695	103.532	105.169	92.002	96.516
Borrower Age	28.071	23.045	28.126	23.224	24.464	23.344
Coborrower Present	0.417	0.493	0.382	0.486	0.280	0.449
Jumbo Loan	0.410	0.492	0.246	0.430	0.182	3.249
Adjustable Interest Rate	0.543	0.498	0.481	0.499	0.575	0.494
Subordinate Debt	0.436	0.496	0.411	0.492	0.484	0.500
Loan to Value Ratio	0.867	0.249	0.882	0.242	0.912	0.216
Vantage Score	801.688	100.245	784.211	104.151	748.994	101.423
Mortgage Payment to Income Ratio	0.246	0.309	0.249	0.283	0.268	0.235
Debt Payment to Income Ratio	0.307	0.379	0.318	0.334	0.337	0.286
Condo	0.206	0.404	0.199	0.399	0.247	0.431
Mobile	0.001	0.036	0.001	0.026	0.002	0.044
Single Family	0.788	0.409	0.796	0.403	0.745	0.436
Lot Size (sf in 1000s)	14,086.73	136,283.00	14,018.82	63,752.21	7,360.16	24,508.92
Unit square feet (in 1000s)	2,102.00	42,079.06	1,812.32	12,607.39	1,420.24	852.42
Number of Bathrooms	2.048	1.297	2.170	11.348	1.707	1.063
Number of Bedrooms	2.283	1.605	2.304	11.694	1.907	1.548
Number of Stories	1.253	2.001	1.221	1.597	1.006	0.894
Units in Building	1.584	10.797	1.538	19.484	1.019	13.187
PUMA Share Residents Black	0.060	0.074	0.077	0.098	0.144	0.169
PUMA Share Residents Hispanic	0.083	0.094	0.092	0.122	0.212	0.227
PUMA Median Family Income (1,000s)	65.02	16.75	59.60	14.21	50.69	10.68
PUMA Homebuyers Share Black	0.030	0.044	0.062	0.099	0.131	0.145
PUMA Homebuyers Share Hispanic	0.074	0.087	0.089	0.117	0.214	0.198
PUMA Homebuyer Med Inc (1,000s)	88.38	25.17	74.82	18.15	62.84	12.51
Number 90 day bankcard	0.040	0.322	0.057	0.391	0.095	0.515
Number medical collection	0.001	0.034	0.001	0.036	0.001	0.046
Agg Med coll ($1000's)	0.590	54.839	0.279	15.374	0.504	36.746
# 90 Day Deliquncies	0.069	0.669	0.111	0.856	0.223	1.215
Current Loan to Value (County Prices)	1.003	0.617	1.036	0.762	1.139	0.774
Sample size	**103,018**		**102,200**		**97,802**	

Notes. This table presents the means and standard deviations of variables by terciles defined by PUMA by origination year share of loans issued by high cost lenders where high cost lenders are defined as lenders for whom more that 20 percent of their loans in our seven sites qualified as rate spread loans, i.e. APR 300 basis points over treasury rates of comparable maturity.

Table 2: PUMA Attributes in Home Purchase Sample

	Purchase/Origination Year			
	2004	2005	2006	2007
Share Loans from Lenders >0.15	0.140	0.248	0.284	0.119
Share Loans from Lenders >0.20	0.106	0.208	0.230	0.065
Share Loans from Lenders >0.25	0.064	0.188	0.193	0.040
Share Loans w/ Subprime Credit Score	0.237	0.241	0.251	0.190
Share High LTV Loans	0.338	0.370	0.427	0.329
Share High DTI Loans	0.450	0.457	0.477	0.504
High Cost/Rate Spread Loans	0.059	0.115	0.123	0.064
Share Black Loans	0.080	0.088	0.108	0.095
Share Hispanic Loans	0.140	0.160	0.166	0.141
Share Low Income Borrowers	0.319	0.325	0.342	0.342
Denial Rate	0.251	0.271	0.302	0.343
Number of Applications	9,713.103	9,973.553	8,505.152	5,742.987
Herfindahl	0.266	0.267	0.263	0.263
Employment Rate	0.944	0.944	0.940	0.933
PUMA Housing Price Index	158.739	189.079	204.180	193.084
Sample Size	83,894	95,210	74,588	49,328

Notes. Table presents sample means by year of origination of variables measured at the PUMA level.

Table 3: Foreclosure Notice in Credit Report on Share of Loans in PUMA Originated by High Cost Lenders

	Cross-Sectional Variation				
	Rate Spread	Credit Score	Demographic	Risk Factors	Lender FE
Rate Spread Loan	0.093***	0.079***	0.066***	0.051***	0.038***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PUMA Shr from Lndrs >0.2	0.256***	0.204***	0.138***	0.128***	0.122***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.018)
Observations	303,020	303,020	303,020	303,020	303,019
R-squared	0.091	0.095	0.115	0.126	0.132
	Changes in Share Loans from High Cost Lenders				
	Rate Spread	Credit Score	Demographic	Risk Factors	Lender FE
Rate Spread Loan	0.092***	0.077***	0.065***	0.049***	0.037***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PUMA Shr from Lndrs >0.2	0.135***	0.136***	0.122***	0.121***	0.127***
	(0.044)	(0.045)	(0.044)	(0.043)	(0.043)
Observations	303,020	303,020	303,020	303,020	303,019
R-squared	0.097	0.101	0.119	0.131	0.136
	Instrument conditional on Lender by Purchase Year Fixed Effects				
	Rate Spread	Credit Score	Demographic	Risk Factors	
Rate Spread Loan	0.055***	0.047***	0.041***	0.035***	
	(0.002)	(0.003)	(0.002)	(0.002)	
Predicted Change in Market Share	0.211***	0.211***	0.212***	0.217***	
	(0.060)	(0.061)	(0.061)	(0.060)	
Observations	303,015	303,015	303,015	303,015	
R-squared	0.116	0.118	0.133	0.143	

Notes. Table presents estimates from regressions of whether the credit report contained a report of foreclosure in a given year based on a sample of annual credit reports following the mortgage origination controlling for whether the loan was a rate spread loan, the PUMA share of loans from high cost lenders in the year of purchase/origination, and purchase year by credit report year by metropolitan/regional site. Panel 1 presents these regresssion with column 2 adding controls for 20 point vantage score bins; column 3 additionally adding demographic controls for race, ethnicity, gender, coborrower status and age; column 4 adding loan terms like loan to value ratio bins, mortgage payment to income ratio bins, debt payment to income ratio bins, whether adjustable rate, whether a jumbo loan amount and whether the purchase included a subordinate lien; and finally column 5 adds lender fixed effects. Panel 2 presents the same models except that the models also include PUMA by credit report year fixed effects. Panel 3 presents similar models except that the share of loans from high cost lenders is replaced by a shift-share style prediction in the change in expected change in share of loans from high cost lenders and all models include both PUMA by credit report year and lender by purchase year fixed effects. Standard errors are clustered at the PUMA level, and significance on two tailed t-test is designed by *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Robustness Tests for Relationship betwee High Cost Lender Market Representation and Foreclosure

	Robustness to Definition of High Cost Lender			
	Market Share > 15		Market Share > 25	
	Rate Spread	Risk Factors	Rate Spread	Risk Factors
Rate Spread Loan	0.055***	0.035***	0.055***	0.035***
	(0.002)	(0.002)	(0.002)	(0.002)
Predicted Change in Market Share	0.180***	0.197***	0.188***	0.213***
	(0.059)	(0.057)	(0.062)	(0.061)
Observations	303,015	303,015	303,015	303,015
R-squared	0.116	0.143	0.116	0.143
	Additional Robustness Tests with Risk Factor Controls			
	Balanced Panel 04-07, 08-09	Constant Share of Rate Spread Loans	Sample Weights	Asymmetric Effects
Rate Spread Loan	0.043***	0.035***	0.037***	0.035***
	(0.003)	(0.002)	-0.003	(0.002)
Predicted Change in Market Share	0.222***	0.190***	0.202***	
	(0.074)	(0.039)	(0.067)	
Positive Predicted Change				0.203
				(0123)
Negative Predicted Change				0.227***
				(0.065)
Observations	179,994	303,015	303,015	303,015
R-squared	0.169	0.143	0.138	0.143
	Conditional and Clustered Standard Errors by Mortage Year by PUMA Initial Share High Cost Loans			
Cluster Size over Share High Cost	0.01	0.02	0.03	0.04
Rate Spread Loan	0.035***	0.035***	0.035***	0.035***
	(0.003)	(0.004)	(0.005)	(0.005)
Predicted Change in Market Share	0.279***	0.244***	0.182**	0.204***
	(0.079)	(0.070)	(0.074)	(0.073)
Number of Clusters	124	60	48	36
Observations	302,959	302,959	302,959	302,959
R-squared	0.116	0.118	0.133	0.143

Notes. Table presents models based off of Panel 3 Table 3. Panel 1 presents the equivalent of columns 1 and 4 for alternative shift share predictions based on either a 15 percent or 25 percent share of rate spread loans for defining high cost lenders. Panel 2 presents the equivalent of column 4 for three different samples/models. Column 1 is a balanced panel only retaining credit report years of 2008 and 2009. Column 2 adjusts the rate spread variable so that the total fraction of rate spread loans is constant across years, and column 3 weights the sample based on weights based on the sampling probability. Column 4 allows the estimated effect to vary based on whether the predicted change is positive or negative. Standard errors for panels 1 and 2 are clustered at the PUMA level. Panel 3 presents the same models as column 4 of Panel 3 Table 3 with the addition of fixed effects for mortgate origination year by PUMA initial (2004) share of loans from high cost lender bins where the bins are based on 0.01, 0.02, 0.03 and 0.04 increments of initial share for columns 1-4, respectively. Standard errors in Panel 3 use two way clustering based on PUMA and on mortgage year by PUMA initial share of loans from high cost lender bins. Significance is designed by *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Heterogeneity in the Effect of the Market Representation of High Cost Lenders

	Lender by Purchase Year by Rate Spread Loan FE	Lender by Purchase Year by Race/Ethnicity FE	Lender by Purchase Year by Type of Loan Purchase FE	Lender by Purchase Year FE
Predicted Change in Market Share	0.194***	0.309***	0.169**	0.165**
	(0.064)	(0.073)	(0.072)	(0.063)
Predicted Change in Market Share*Rate Spread Loan	0.074			
	(0.094)			
Predicted Change in Market Share*Black		-0.116		
		(0.080)		
Predicted Change in Market Share*Hispanic		-0.208*		
		(0.107)		
Predicted Change in Market Share*Held in Portfolio			-0.019	
			(0.083)	
Predicted Change in Market Share*Non-agency Securitization			0.102	
			(0.072)	
Predicted Change in Market Share*High Cost Lender (>0.20)				0.261**
				(0.116)
Observations	303,011	302,999	303,014	303,015
R-squared	0.147	0.154	0.145	0.143

Notes. This table presents estimates for the shift share prediction of changes in share of loans from high cost lenders based on the model from Table 3 Panel 3 column 4, but adds interactions of this predicted change variable. Column 1 includes the interaction with whether the loan is a rate spread or high cost loan, column 2 includes interactions with both whether the borrower is black and whether the borrower is Hispance, column 3 includes interactions with whether the loan was held in portfolio or was securitized outside of traditonal government sponsored channels, and column 4 includes an interaction with whether the lender that originated the loan is defined as a high cost lender. Standard errors are clustered at the PUMA level, and significance is designed by *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Controlling for the Effect of Originating Lender's Market Share

Baseline Model			
	Baseline	Share	Predicted Share
Rate Spread Loan	0.035***	0.035***	0.035***
	(0.002)	(0.002)	(0.002)
Predicted Change in Market Share	0.217***	0.216***	0.218***
	(0.060)	(0.060)	(0.067)
Change in lender share		-0.070	
		(0.100)	
Predicted change in lender share			0.024
			(0.052)
Observations	303,015	303,015	303,015
R-squared	0.143	0.143	0.143
Differential Effects for Origination by High Cost Lender			
	Baseline	Share	Predicted Share
Rate Spread Loan	0.035***	0.035***	0.035***
	(0.002)	(0.002)	(0.002)
Predicted Change in Market Share	0.165**	0.163**	0.153**
	(0.063)	(0.063)	(0.070)
Predicted Change*High cost lender	0.261**	0.263**	0.230***
	(0.116)	(0.117)	(0.116)
Change in lender share		-0.020	
		(0.073)	
Change*High cost lender		-0.358	
		(0.456)	
Predicted change in lender share			-0.061
			(0.055)
Predicted change lender*high cost			1.178**
			(0.478)
Observations	303,015	303,015	303,015
R-squared	0.143	0.143	0.143

Notes. This table presents models that add controls for either the change in lender market share by purchase year or the shift-share predicted change in share. Panel 1 presents baseline models based on Table 3 Panel 3 column 4, and Panel 2 presents models including interactions with whether the originating lender was a high cost lender based on the model in Table 5 column 4. Column 1 repeats models without information on lender share, column 2 includes actual lender share and column 3 includes the shift share based prediction. Standard errors are clustered at the PUMA level, and significance is designed by *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Additional Tests for Whether Borrower Unobservables Matter

	Contemporaneous Risk Factors			Whether Rate Spread Loan	
	Bankcard Trades	Medical Collection Trades	Both	No Additional Controls	All Controls
Rate Spread Loan	0.034***	0.035***	0.034***		
	(0.002)	(0.002)	(0.002)		
Predicted Change in Market Share	0.210***	0.217***	0.210***	-0.025	-0.005
	(0.058)	(0.060)	(0.058)	(0.090)	(0.089)
# of Bankcard Trades	0.064***		0.064***		
	(0.002)		(0.002)		
Acl190 Log(amount+1)		0.000008	0.000008		
		(0.000)	(0.000)		
Acl200 Log(#trades+1)		-0.007	-0.006		
		(0.011)	(0.010)		
Observations	303,015	303,015	303,015	93,771	93,771
R-squared	0.158	0.143	0.158	0.640	0.670

Notes: The first three columns of this table present estimates of models from the fourth column of Table 3 Panel 3. The first column contains of estimates based on including a control for the logarithm of 1 plus the number of bank card accounts that are 90 days past due or longer, the next column includes controls for number of medical collection trades and the aggregate amount of medical collection trades, again using the transformation of the logarithm of 1 plus variable, and the third column includes both the bankcard and medical collection information. The last two columns estimate very similar models where whether the loan is rate spread or high cost is moved to the left hand side of the model. These models models are estimates using a sample of mortgages (as opposed to mortgages by credit report year) and conditional on the purchase/originiation year trends, and PUMA, lender by origination year and metro area by origination year fixed effects. The first of these columns contains no additional controls (equivalent to column 1 Table 3 Panel 3), and the last column includes all controls as in column 4. Standard errors are clustered at the PUMA level, and significance is designed by *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Measures of Trends in Mortgage Borrowers, Attributes and Market Conditions

	PUMA Borrower Attributes		
	Share low income	Share Black & Hispanic	Share subprime
Rate Spread Loan	0.035***	0.035***	0.035***
	(0.002)	(0.002)	(0.002)
Predicted Change in Market Share	0.222***	0.235***	0.221***
	(0.057)	(0.060)	(0.059)
PUMA share	-0.143***	0.124***	-0.023
	(0.039)	(0.042)	(0.015)
Observations	303,015	303,015	303,015
R-squared	0.143	0.143	0.143
	PUMA Loan Attributes		
	Share High Cost Loans	Share high ltv	Share high dti
Rate Spread Loan	0.035***	0.035***	0.035***
	(0.002)	(0.002)	(0.002)
Predicted Change in Market Share	0.212***	0.216***	0.216***
	(0.060)	(0.059)	(0.060)
PUMA share	0.122	-0.022**	-0.005
	(0.098)	(0.009)	(0.012)
Observations	303,015	303,015	303,015
R-squared	0.143	0.143	0.143
	PUMA Mortgage Attributes		
	Denial Rate	No. Applications (10,000)	Herfindahl
Rate Spread Loan	0.035***	0.035***	0.035***
	(0.002)	(0.002)	(0.002)
Predicted Change in Market Share	0.236***	0.245***	0.227***
	(0.059)	(0.060)	(0.063)
PUMA attribute	0.067	-0.000001**	-0.118**
	(0.044)	(0.000)	(0.058)
Observations	303,015	303,015	303,015
R-squared	0.143	0.143	0.143

Notes: This table presents estimates of models from the fourth column of Table 3 Panel 3 adding additional controls for PUMA variables that vary by purchase year one at a time. Panel 1 adds borrower controls: share borrowers with family income below the federal poverty line, share of borrowers who are black or Hispanic and share of borrowers with subprime credit scores (below 701) by PUMA and origination year for columns 1, 2 and 3, respectively. Panel 2 adds loan attributes: share of high cost loans, share of loans with an LTV over 0.95, and share of loans with a debt to income ration of 0.45. Finally, Panel 3 adds common market descriptors from HMDA: the denial rate, number of applications and a Herfindahl of market concentration among lenders. Standard errors are clustered at the PUMA level, and significance is designed by *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Negative Equity and Employment Rate Controls

Equity based on County Price Indices			
	Negative Equity	Negative Equity and Employment	Negative Equity and Employment by Race
Rate Spread Loan	0.035***	0.035***	0.035***
	(0.002)	(0.002)	(0.002)
Predicted Change in Market Share	0.228***	0.226***	0.239***
	(0.058)	(0.058)	(0.059)
Observations	303,015	303,015	302,252
R-squared	0.146	0.146	0.147
Equity based on PUMA Price Index			
	Negative Equity	Negative Equity and Employment	Negative Equity and Employment by Race
Rate Spread Loan	0.035***	0.035***	0.035212***
	(0.002)	(0.002)	(0.002)
PUMA Shr from Lndrs >0.2	0.249***	0.251***	0.254***
	(0.056)	(0.057)	(0.056)
Observations	303,015	303,015	302,252
R-squared	0.144	0.145	0.145

Notes: This table presents estimates of models from the fourth column of Table 3 Panel 3 adding controls for whether the individual is in negative equity. Panel 1 presents results based on county price indices, and panel 2 presents results based on PUMA price indices. Column 1 includes only the negative equity variables, column 2 interacts negative equity with county level by credit report year employment rates from the American Community Survey, and column 3 includes the same interactions with a county by credit report employment rate for the individual's race or ethnicity. Standard errors are clustered at the PUMA level, and significance is designed by *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Estimates for 90-180 Day Mortgage Delinquency

		Foreclosure			90 Day Deliquency	
	Baseline	Interaction	Lender Share	Baseline	Interaction	Lender Share
Rate Spread Loan	0.035***	0.035***	0.035***	0.025***	0.025***	0.025***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Predicted Change in Market Share	0.217***	0.160**	0.142**	0.069	0.094	0.124**
	(0.060)	(0.062)	(0.068)	(0.052)	(0.057)	(0.062)
Predicted Change*High cost lender		0.297**	0.220*		-0.128	-0.166
		(0.118)	(0.113)		(0.103)	(0.120)
Predicted change in lender share			-0.082			-0.045
			(0.057)			(0.064)
Predicted change lender*high cost			1.196**			0.417
			(0.490)			(0.408)
Observations	303,015	303,015	279,565	303,015	303,015	279,565
R-squared	0.143	0.143	0.138	0.092	0.092	0.090

Notes: This table presents estimates of models from the fourth column of Table 3 Panel 3, the fourth column of Table 5 and the third column of Table 6. Columns 1 through 3 present these estimates, and columns 4 through 6 present estimates for the same models using 90-180 mortgage delinquency as the dependent variable. Standard errors are clustered at the PUMA level, and significance is designed by *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Judicial versus Non-Judicial (California) States

	Judicial Role	California
Rate Spread Loan	0.033***	0.038***
	(0.003)	(0.005)
Predicted Change in Market Share	0.269***	0.116
	(0.068)	(0.186)
Observations	212,142	90,872
R-squared	0.128	0.174

Panel 3. Seperately, for a subsample of the five sites outside of California in column 1 and the subsample of the two sites in California in column 2. Standard errors are clustered at the PUMA level, and significance is designed by *** p<0.01, ** p<0.05, * p<0.1.