NBER WORKING PAPER SERIES

MORTGAGE 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, Revised January 2022

An earlier version of this paper was entitled "High Cost Lenders and the Geographic Concentration of Foreclosures". The authors thank Bob Avery for inspiring the paper, and Morris Davis, Ingrid Ellen, Fernando Ferreira, Kris Gerardi, Andra Ghent, Lauren Lambie-Hansen and additional 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 for providing advice on this research. 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Stephen L. Ross and Yuan Wang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Mortgage Lenders and the Geographic Concentration of Foreclosures
Stephen L. Ross and Yuan Wang
NBER Working Paper No. 28781
May 2021, Revised January 2022
JEL No. D14,G01,G21,R21,R23

ABSTRACT

We use HMDA rate spread loans to identify lenders involved in riskier lending prior to the foreclosure crisis. We develop a shift-share measure of changes in high rate spread share lender representation in housing submarkets across origination years. While half the cross-sectional correlation between foreclosure and high rate spread lender share is explained by borrower observables, we find robust and stable estimates of the within housing submarket relationship between foreclosure and predicted changes in market share. Estimates are not explained by local housing price variation, rather evidence suggests servicer behavior in response to rising local foreclosure rates as a mechanism.

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
wangyuansdu@gmail.com

Mortgage Lenders and the Geographic Concentration of Foreclosures

Stephen L Ross, University of Connecticut and NBER

Yuan Wang, FreddieMac

A central feature of the U.S. foreclosure crisis is that effects were concentrated in specific locations. Subprime lending was concentrated in low income and minority neighborhoods (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 high foreclosure rates (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, neighborhoods experiencing high foreclosure rates also tended to have high rates of subprime lending leading 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]

Subprime and other high risk lending almost certainly contributed to the spatial concentration of foreclosures through either loans to riskier borrowers and/or loans with riskier attributes. However, more limited research exists on whether the spatial concentration of lenders engaging in such lending had effects on foreclosure above and beyond the specific loans originated. Many papers document evidence of neighborhood spillovers from foreclosure (e.g. Campbell et al., 2011; Gerardi et al., 2015; Gupta, 2016; Munroe and Wilse-Samson, 2013). While

[1] This empirical regularity has been established using many different indicators including the Department of Housing and Urban Development (HUD) subprime lender list, non-agency securitized lending, high cost lending based on rate spread loans 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).

those spillover estimates arise from close proximity to a foreclosure (often within 250 feet) and tend to be small in magnitude, a few recent papers suggest that these spillovers could have had larger aggregate impacts (Mian, Sufi and Trebbi 2015; Huang, Nelson and Ross 2018; Gupta 2019; Guren and McQuade 2020).

We examine the effect of geographic lending patterns using a shift-share approach based on the composition of lending during an initial period, similar to Greenstone, Mas and Ngyuen's (2020) study of small business lending. We define mortgage lenders that appear heavily engaged in "higher risk" lending using the Home Mortgage Disclosure Act (HMDA) rate spread variable, i.e. mortgages with a substantial rate spread above the return on comparable maturity treasury bonds (3 points above treasury yields). Both, Mayer and Pence (2008) and Chan, Haughwout, and Tracy (2015) use rate spread mortgages as an indicator of subprime mortgage lending. Further, Bayer, Ferreira and Ross (2018) conclude that lenders with a high share of rate spread loans served a different segment of the market, having unusually high unexplained ex-post foreclosure rates. We categorize lenders as having a high rate spread share, or as a short hand "high cost lenders", if 20 percent or more of their mortgages were rate spread loans. We measure the high cost lender share within housing submarkets using Public Use Microdata Areas (PUMAs), geographically contiguous areas containing at least 100,000 people.[4] As shown below (Table 2), high cost lender share was much more volatile over time than traditional measures of underwriting risk, like loan to value or debt to income ratios.

First, we document the correlation between foreclosures and the geographic concentration of these high cost lenders using home purchase mortgages originated between 2004 and 2007 from

[4] Our data agreement prevents the use of smaller geography like census tract. PUMA's also have the advantage of being at a scale well above the geographic level of the foreclosure spillovers studied by other papers. Further, this geographic scale still captures dramatic variation in foreclosures.

seven large metropolitan/regional sites.[5] We find a strong cross-sectional correlation between foreclosure and the PUMA high cost lender share. However, controlling for borrower credit score and demographics reduces the estimated effect by half. The remaining conditional correlation with foreclosure is reduced by an additional 13% by the inclusion of both loan terms, e.g. loan to value and debt to income ratios, whether an adjustable rate mortgage and use of subordinate debt, and lender fixed effects.[6] After conditioning on PUMA fixed effects and origination year trends, the correlation between foreclosure and high cost lender share is similar to the final cross-sectional correlation, and the inclusion of controls has only modest effects on the estimates. In summary, much of the cross-sectional correlation is explained by borrower observables, and after conditioning on borrower observables only a modest portion of this correlation is explained by loan attributes, lender identity or geographic location.

Next, we examine foreclosure using a Bartik or shift-share prediction of high cost lender share (Autor and Duggen 2003; Brunner, Ross and Washington 2011; Greenstone, Mas and Ngyuen 2020). We use the 2004 initial distribution of loans across lenders in each PUMA, the first year that rate spread is reported in HMDA, and scale those shares by national changes in each lender's market share between 2004 and 2007. We continue to control for PUMA fixed effects and trends to mitigate any correlation between initial shares and location unobservables (Goldsmith-Pinkham, Sorkin and Swift 2020).[7] Further, we also include lender by origination year fixed effects to assure that estimated effects cannot be driven by lender specific differences in foreclosure risk.

[5] See Bayer, Ferreira and Ross (2014, 2016, 2018) for recent analyses of minority borrowers using this data.

[6] We do not observe some terms like loan is no or low documentation, lock period for adjustable rates, information on teaser rates, or prepayment penalties. However, these unobserved features are concentrated in the subprime sector where adjustable rates loans and the use of subordinate debt are much more common and also at subprime lenders. So, our controls for loan terms and lender identity should capture some of the effect of these product attributes.

[7] A cross-sectional shift share model yields results similar to the cross-sectional model using actual share: correlation between foreclosure and predicted high cost lender share that is eroded by 50% using credit score and borrower attributes, and only another 8% with loan terms and lender fixed effects.

A one standard deviation increase in the shift-share proxy implies a 1.46 percentage point change in foreclosure likelihood or approximately a 27% increase in foreclosure rate. Based on this estimate, the difference between the top and bottom terciles on high cost lender share can explain 56% of the across tercile foreclosure rate differences. Similarly, high cost lender share differences by tercile explain 22% of the black-white differences in foreclosure, 15% of Hispanic-white differences and 66% of differences between the top and bottom income quintiles.[8] We conclude that the expansion of higher risk/higher cost lending in the run up to the crisis had larger effects on foreclosure rates in the housing submarkets where the lenders primarily responsible for such lending were concentrated.

Notably, the estimated relationship between foreclosure and our shift-share proxy is very stable in magnitude to the inclusion of borrower attributes and loan terms. Following Oster (2019), we evaluate the conditional exogeneity of our proxy by comparing the change in the estimate to the change in the R-squared as controls are added, and in her words find that the influence of borrower and loan unobservables would have to be over eight times more important than the observables and have effects in the opposite direction to explain our estimates. In comparison, the cross-sectional estimates on share high cost can be explained by borrower and loan unobservables if those unobservables had effects similar to or even smaller than the effects of observables.

These results are also robust to changing the definition of high cost lenders using thresholds of 15 or 25 percent of loans, adjusting the rate spread threshold to address yield curve inversion, selecting a balanced panel of credit report/crisis years, weighting using the inverse of our sampling probability, including contemporaneous measures of non-mortgage delinquencies and controlling

[8] We do not instrument for the share of loans from high cost lenders with the shift share prediction 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, see the Appendix and Chalak (2019).

for a wide array other local mortgage market attributes. Further, Goldsmith-Pinkham, Sorkin and Swift (2020) observe that shift-share models with geographic fixed effects are difference-in-differences analyses. While our panel of origination years is too short for an event study, we address concerns about staggered roll-out by replicating results using pairs of origination years so that our base year is the pre-period and a second year is the treated post-period. This strategy also provides a falsification test showing that origination changes between 2006 and 2007 cannot explain differences between 2004 and 2006 foreclosures. As suggested by Callaway and Sant'Anna (In Press), the PUMA cohort trends address heterogeneous trends over observables when difference-in-differences analyses fail balance over time invariant, initial attributes, see Abadie (2005). Finally, results are robust both to including fixed effects for origination year by PUMA clusters that have very similar initial (2004) shares of loans from high cost lenders and to clustering standard errors over these fixed effects.

These estimated effects are broad based and not just concentrated among observably high risk loans. We estimate similar magnitude effects for our proxy of subprime or high risk lender representation whether the predicted changes over time imply either an increase or a decrease in high cost lender share. Foreclosure effects exist for both rate spread and non-rate spread loans, white and minority loans, regardless of the source of securitization, and whether originated by high cost lenders or not. Notably, predicted high cost lender share effects are significantly larger in absolute terms for loans issued by 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 changes in foreclosure likelihood are similar, 28 and 31 percent, respectively. We also investigate whether our results can be driven by changes in the market share of the lender who originated the loan, and our estimates are robust, even though lender market share also appears to influence foreclosure rates when the lender is high cost.

In terms of mechanism, the difference-in-differences, shift-share identification strategy is intended to break the correlation between the variation in lender activity and unobserved borrower and loan attributes. Given the stability of estimates as controls are added, this strategy would tend to rule out explanations for our findings like Bayer et al.'s (2018) argument that expansions of credit attracted riskier borrowers into the home purchase market or a "race to the bottom" story that the entry of higher cost or riskier lenders influenced the behavior other lenders in the housing submarket.[9] Rather, our findings must be driven by broader spatial effects, possibly aggregate effects of the localized spillovers identified in earlier studies (e.g. Campbell et al., 2011; Gerardi et al., 2015; Gupta, 2016; Munroe and Wilse-Samson, 2013) or broader spillovers that were excluded by the identification strategies used in those papers (Huang, Nelson and Ross 2018).

We also explicitly examine two possible mechanisms for these foreclosure concentrations. First, Mian and Sufi (2009) argue that local expansion of subprime credit increased housing prices leading to greater equity losses and higher foreclosure rates, and Ferreira and Gyourko (2012) document substantial spatial variation in the timing of price increases and the extent of price declines during the crisis.[10] We calculate PUMA housing price indices to measure the level of negative equity. While negative equity explains foreclosure, the inclusion of controls for negative equity had minimal influence on our estimates.[11] Second, we consider whether loan servicers

[9] Such as lax risk assessments (Keys, Mukherjee, Seru, and Vig 2009; Dell'Ariccia, Igan, and Laeven 2012; Bhutta and Keys In Press), low documentation loans (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).

[10] 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, and Palmer (2015) on housing prices and foreclosure in the U.S.

[11] 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).

contributed to this phenomenon. Agarwal, Amromin, Ben-David, Chomsisengphet and Piskorski (2017) document higher foreclosure rates in zip codes where mortgage servicers had less loan modification experience, and Ding (2013) found that mortgage modifications were least likely in the neighborhoods hardest hit by the foreclosure crisis.[12]

While we do not observe servicers, we are able to examine the possible role of servicers indirectly. First, we estimate models of 90-180 day delinquencies, rather than foreclosure filings. Mortgage delinquency represents failure of borrowers to make mortgage payments, while foreclosure represents the combined effects of delinquency and lender servicer decisions on foreclosure and loan modifications. A one standard deviation change in our proxy results in only a 3.5 percent change in the sample delinquency rate and is statistically insignificant. Second, we recognize that foreclosure is purely administrative in California, while courts play a significant role in the other states in our sample creating more opportunities for lender discretion. Splitting our sample between California and the other sites, we find that high cost lender share effects are concentrated almost entirely outside of California. For judicial role states, the standardized estimate is 40 percent of the baseline foreclosure rate, and while the estimates for California are noisy the point estimates imply an effect of only 7 percent.

These findings suggest that loan servicers are responding to the experiences of specific cohorts of loans within housing submarkets. We provide descriptive evidence consistent with such responses by including a control for the previous year's foreclosure experience within a PUMA for each cohort of loans (based on a hold-out sample). The standardized effect of preceding year

[12] A large literature has examined the role of servicers in the foreclosure crisis. For example, Piskorski, Seru and Vig (2010), Agarwal, Amromin, Ben-David, Chomsisengphet and Evanoff (2011a), and Kruger (2018) show that securitization created incentives favoring foreclosure over modification, while Adelino, Gerardi and Willen (2013) and Conklin, Diop, Le and D'Lima (2019) find that information asymmetries were important. Notably, Agarwal, Amromin, Bend-David, Chomsisenphetand, Evanoff (2011b) and Reid, Urban and Collins (2017) document heterogeneity across servicers in whether modifications are offered and in the terms and types of modifications. For an alternative view on remediation and securitization, see Ghent (2011).

foreclosure rates is similar in magnitude to the effect of high cost lender share. In addition, unlike virtually all other controls considered, the inclusion of this control erodes the high cost lender share effect by 25%, and the estimate is no longer statistically significant.

In summary, the likelihood of foreclosure is higher when housing submarkets have higher market representation of lenders involved in riskier mortgage lending, as indicated by a concentration of lenders with a high share of rate spread loans. This relationship is very stable as detailed controls for borrower and loan attributes are added to the model suggesting a limited role for borrower and loan unobservables in explaining these estimates. Therefore, the estimates suggest location effects above and beyond the foreclosure risk of the individual mortgages. Our estimated effects cannot be explained by local changes in housing prices over the cycle, which might be expected if the market representation of these lenders affected local housing prices. On the other hand, our study provides new evidence that servicers may have played an important role in the geographic concentration of foreclosures, responding to higher foreclosure rates among cohorts of loans in specific housing submarkets and thereby contributing to higher future foreclosure rates for all associated loans.

1. Methods

We proxy for the market penetration of lenders engaged in subprime or other risky lending practices as the fraction of loans issued by "high cost" lenders in a housing submarket. We define lenders as high cost (*H*) based on whether the share of their originated mortgages that meet the definition of a rate spread loan (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 initial 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

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

where N_{lnp} is the number of mortgages issued by lender l in submarket n and time p.

Then, 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 high cost lender share, borrower and mortgage attributes (X), and a metropolitan or regional site (s) by credit report/foreclosure year (t) by purchase year (p) fixed effects so that each site (s) has 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 cross-sectional variation by including submarket by credit/ foreclosure year fixed effects and submarket trends over purchase year on location observables. Submarket fixed effects capture the constant (across purchase years) impact of unobservables including any future events that affect foreclosure likelihood in a given year for all mortgages in the submarket. The submarket trends allow the relationship between submarket and foreclosure to vary across origination cohorts based on observables. Specifically, purchase year fixed effects are interacted with time invariant submarket attributes (W). The resulting specification is

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

The geographic fixed effects in equation (2) yield a difference-in-differences analyses (in our case over purchase cohorts). The PUMA trends on observables, $\varphi_p W_n$, follow Callaway and Sant'Anna's (In Press) recommendation of including trends associated with time-invariant, geographic covariates to insulate against failures of parallel trends that arise from heterogeneous trends across subgroups (Abadie 2005).

Finally, we create a quasi-experimental, conditionally exogenous proxy for changes in the high cost lender share for each submarket and purchase year. Specifically, we create a Bartik or shift-share style prediction similar to those used in Autor and Duggen (2003), Brunner, Ross and Washington (2011) and Greenstone, Mas and Ngyuen (2020). 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 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. For all lenders *l* with non-zero market share in a submarket for 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 ($\alpha_{\bar{p}nl}$). 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.

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

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}) capturing any differences in foreclosure risk across individual lenders. The final model specification is

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

Standard errors in all models above are clustered at the submarket level (*n*). Note, we do not instrument for the change in high cost lender share with the shift-share predicted change. Both variables are noisy proxies for the market representation of lenders heavily involved with subprime or other risky lending. Instrumenting for one proxy with another may lead to significant upward bias, see the appendix and Chalak (2019).

Goldsmith-Pinkham, Sorkin and Swift (2020) examine panel applications of shift-share variables in models that control for geographic fixed effects. They observe that strict exogeneity need only hold conditional on controls and argue that identification from change (geographic fixed effects) makes the strict exogeneity assumption much more reasonable. Nonetheless, exposure to treatment based on initial shares may correlate with unobserved changes, not just levels. For example, submarkets with similar initial high cost lender shares, $Z_{n\bar{p}}$, may share unobservables that influence the evolution of mortgage markets over time, e.g.

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

where Ω_n is a set of locations that are similar to submarket *n*. With a small number of purchase cohorts, the submarket fixed effects intended to assure strict exogeneity of the shift-share proxy could introduce an incidental parameters bias from $\mu_{\Omega_n p}$. We attempt to absorb $\mu_{\Omega_n p}$ by dividing submarkets into bins of similar initial high cost lender share

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

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 high cost lending share group by purchase year fixed effects $\tau_{k_n p}$

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

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

This structure also helps 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 initial shares also have similar unobservables creating a correlation across geographies. While their analysis is cross-sectional and concerns should in part be addressed by geographic fixed effects, $\mu_{\Omega_n p}$ could lead to across geography correlations in changes. Conditional on the similar submarket by origination year fixed effects $\tau_{k_n p}$, we can address any general pattern of correlations between these similar submarkets by clustering over those $\tau_{k_n p}$ fixed effects (Abadie et al. 2017). We use two-way clustering to preserve clustering within individual submarkets.

Finally, as noted by Goldsmith-Pinkham et al. (2020), shift-share analyses with fixed effects should be validated using techniques developed for difference-in-differences analyses. Our panel is too short to support event study approaches used to test for parallel pre-trends or to address bias from staggered roll-out of treatment (Sun and Abraham In Press, Callaway and Sant'Anna In Press). Further, these techniques for addressing staggered roll-out primarily focus on discrete treatments. Therefore, we eliminate staggered roll-out by examining pairs of mortgage cohorts p so that treatment takes a before and after form.

$$y_{instTl} = \alpha \widehat{\Delta Z}_{Tn} + \beta X_i + \rho_{Tl} + \delta_{stT} + \theta_{nt} + \varphi_T W_n + \varepsilon_{insT0l}$$

$$y_{inst0l} = \beta X_i + \rho_{0l} + \delta_{st0} + \theta_{nt} + \varphi_0 W_n + \varepsilon_{inst0l}$$

where $p = T$ represents the treatment year and $p = 0$ represents the base year of 2004. Using these pairs, we can also address parallel trends by conducting a falsification test using changes between the treatment year T and a future falsification year F by replacing $\widehat{\Delta Z}_{Tn}$ with $\left(\widehat{\Delta Z}_{Fn} - \widehat{\Delta Z}_{Tn}\right)$: Do predicted future cohort changes in shift-share measure explain foreclosure for earlier cohorts?

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 archival credit reporting data collected by Experian PLC for March 31st 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.[13]

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 loans.[14] 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 in HMDA over the sample period from 2004-2007 using 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 for reporting residential location in individual level data. We calculate the share of loans originated by high cost lenders in each PUMA by purchase/origination year for each year between 2004 and 2007 based on the census tract location of the purchased property again using the population of owner-occupied, home purchase mortgages contained in HMDA.[15]

---

[13] 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's northern and western 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.

[14] 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).

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

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 leads to larger numbers.[16] We calculate PUMA time invariant characteristics using both the 2000 Decennial Census for residents and the 2004 HMDA data for the attributes of home purchasers,[17] and calculate time varying (over origination year) attributes based on HMDA, transaction data and 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, county, and census tract.[18] These mortgages were sampled from May through August and matched using name and address by Experian PLC to the March 31st archival record preceding the mortgage

[16] The threshold for a high cost lender is lowered from 20 to 13 percent to hold total share of loans from high cost lenders fixed during the sample period. 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.

[17] 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.

[18] 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.

transaction and March 31st record for every year following this transaction through 2009.[19] Our panel contains one observation per mortgage 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. Weights are 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.[20]

Table 1 shows the means for our final home purchase sample of post mortgage credit reports by tercile of the share of loans in a PUMA from high cost lenders. The high cost lender share 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. In the highest, high cost lender share tercile, borrowers are less white, lower income, less likely to have a co-borrower and have lower credit scores. The highest tercile 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. Highest tercile PUMA's 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 more than 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

[19] 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.

[20] 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.

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 borrowers with subprime credit score and high loan to value ratios 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. Application denial rates increase throughout the entire period of 37%, and application volume decreases by 41% between 2006 and 2007. 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 high cost lender share by PUMA and origination/purchase year cohorts. Using a loan level sample, we regress whether each loan ever faced foreclosure by the end of our credit profile data (March 2009) and high cost lender share (over entire pre-period) 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 rate spread loans from risks associated with the activity of lenders that tend to issue rate spread 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 high cost lender share 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 the unexplained variation in foreclosure rates in a PUMA and the unexplained portion of high cost lender share. 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 leads to a substantially weaker cross-sectional relationship between foreclosure and high cost lender share, consistent with a relationship that may be driven by omitted borrower and loan attributes.

Figure 2 is based on residuals from the same models except that we 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 high cost lender share. Figure 2 suggests that PUMA fixed effects successfully capture much of

[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.

the bias from omitted borrower and loan attributes observed in Figure 1, consistent with Goldsmith-Pinkham, Sorkin and Swift's (2020) view that strict exogeneity is much more reasonable conditional on geographic fixed effects.

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 rate spread loan and the site by purchase year by credit year fixed effects. The next columns in order add controls for borrower vantage (credit) score, the borrower demographics listed above, the detailed mortgage attributes plus controls for housing unit attributes, and 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. Following Oster (2019), we compare the erosion of the parameter estimate between columns 1 and 4 to the change in R-squared relative to a maximum R-squared that is taken as 1.3 times the R-squared in column 5, following a rule of thumb provided by Oster. The residual variation explained based on the increase in R-squared is only 86% of the erosion in the parameter estimate, suggesting that borrower and loan unobservables could reasonably explain the remaining parameter estimate.[23]

Borrower (rather than loan) attributes explain most of the decline in the estimated effect. 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,

[23] The 0.86 represents Oster's (2019) test statistic, which is the ratio of a. the percent decline in parameter estimate magnitude between a restricted model that omits attributes over which the sample should exhibit balance for estimates to be causal and an unrestricted model that includes those attribute controls, and b. the share of residual variation explained as captured by the change in R-squared between the restricted and unrestricted model divided by the difference between a maximum feasible R-squared and the R-squared in the restricted model.

controls for loan terms such as LTV, income ratios, subordinate debt and adjustable rates loans only reduce the estimated effect by 4 percent of the original estimate, and the inclusion of lender fixed effects only reduces the estimate by 2.5 percent, both of which should correlate with high risk mortgage attributes, such as rate resets or prepayment penalties (Reid, Bocian, Li and Quercia 2016). Based on observables, much of the cross-sectional relationship between foreclosure and the activity of high cost lenders is associated with borrower attributes. These findings appear consistent with conclusions of Bayer, Ferriera and Ross (2018) that lenders with high cost lenders had high unexplained "ex post" foreclosure rates and so appeared to systematically operate in mortgage submarkets involving a priori higher risk lending opportunities.

Table 3 Panel 2 shows the estimates for equation (2). These models include PUMA by credit report/foreclosure 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 attributes with purchase cohort dummies.[24] The estimates in Panel 2 are very similar in magnitude to the smaller estimates in columns 4 and 5 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 modest at 9 percent. PUMA fixed effects appear to eliminate most of the bias associated with omitting our observed borrower and loan attributes. In terms of magnitude, a one standard deviation change in the high cost lender share is approximately 9.7 percentage points, and the standardized effect in column 4 is 1.23 percentage points or 23 percent of the 0.053 sample foreclosure rate. A similar Oster style calculation comparing columns 1 and 4 implies that unobservables would need to be four times larger than observables to explain the estimates.

[24] 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.

4. Quasi-Experimental Results

The last panel of Table 3 presents results from regressions that replace the high cost lender share with the shift-share prediction and also includes lender by year fixed effects. The estimates are very stable in magnitude as additional borrower and loan attributes are added with only a minor increase in the estimate when risk factors are added in column 4. The Oster (2019) statistic implies that the effect of unobservables on the estimates would have to be eight times larger than the effect of observables and operate in the opposite direction. 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 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 panel 2 estimates with a one standard deviation change implying a 1.46 percentage point change in foreclosure, or approximately 27 percent of the baseline foreclosure rate.

We also can evaluate the magnitude of these effects by comparing differences in exposure to elevated PUMA shares of high cost lending. Using the terciles from Table 1, the top tercile on high cost lender share has a share that is 11.5 percentage points higher than the bottom tercile share. Multiplying by our estimated effect implies a difference in foreclosure likelihood of 2.5 percentage points, which explains 56% of the sample foreclosure rate differences between those terciles. Table 1 also shows that black, Hispanic and lower income borrowers are exposed to higher PUMA shares of high cost lending. Those differences can explain 22% of the black-white differences in foreclosure, 15% of Hispanic-white differences and 66% of the differences in foreclosure between the top and bottom income quintiles. The larger share explained for income arises primarily from smaller income differentials in baseline foreclosure rates.

4.1 Robustness Tests and Falsification

Next, we conduct a series of robustness tests. In Table 4 Panel 1, we present estimates for our model from Panel 3 of Table 3 using alternative shift-share predictions for high cost lender share based on thresholds of 15 percent and 25 percent or more of loans being rate spread. Results are shown and are robust for both 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, as discussed above. The threshold share is set to 13% for defining high cost lenders due to the decline in the number of rate spread loans. The third column estimates a regression using the sample weights based on the stratified sampling strategy, see Bayer et al. (2016). Column 4 allows effects to differ by whether the predicted change is negative or positive. The results in Panel 2 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 very similar. The effects for predicted declines are more precisely estimated because declines arise from significant closures of large subprime lenders in 2006 and 2007, unlike predicted increases that arise from more gradual expansions of existing lenders.

In panels 3 and 4, we address recent concerns raised about shift share analyses. We first organize PUMA's into clusters with similar 2004 shares of loans originated by high cost lenders, and then include 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 0.02, 0.03 and 0.04 in size 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 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]

In panel 4, we address the recent concerns that difference-in-differences models with staggered roll-out and two-way fixed effects can be biased. We eliminate staggered roll-out by estimating models with only two cohorts at a time comparing mortgages originated in the base year of 2004 to 2005 mortgages in column 1, to 2006 in column 2 and to 2007 in column 3. While the results for 2005 prior to the national peak in housing prices are small and insignificant, the

[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 0.279 (0.079) in column 1 of panel 3.

results for 2006 and 2007 are sizable and highly significant. The largest estimates are for 2006, which are about 65 percent larger than our estimates in Table 3 Panel 3, and the estimates for 2006 are about 75 percent of the Table 3 Panel 3 estimates. In column 4, we conduct a falsification test comparing foreclosures from the 2004 and 2006 mortgage cohorts replacing the 2006 shift share predicted change with the difference between 2006 and 2007, and find no effects. We now return to our baseline models in Panel 3 of Table 3 as we continue our robustness tests below.[26]

We conduct additional analyses to further rule out borrower and loan unobservables as a source of bias in our estimates. First, we identify information from contemporaneous credits reports on bank card delinquencies and medical collection trades, which were not directly related to mortgage distress and might be driven by unobserved borrower experiences during the crisis. Table 5 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 share and foreclosure.

Further, one might expect borrower or loan term unobservables to be correlated with receiving a rate spread loan, especially given that the rate spread variable was initially created as an indicator for subprime mortgages. In fact, referring back to Table 3, the estimates on the rate spread variable always erodes as controls are added suggesting substantial correlation with

[26] All of the robustness test models presented in Table 4 exhibit stability in the parameter estimate magnitudes as controls are added. We also run the robustness tests shown 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.

borrower and loan attributes. 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 of Table 5 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 high cost lender share does not predict whether a loan is high cost after controlling for PUMA and lender.

Finally, we create a series of competing variables that capture changes in the PUMA composition of originated loans and run horse race regressions between these variables and the shift-share proxy. Panel 1 of Table 6 presents estimates using borrower composition over purchase cohorts for PUMA share of mortgages with reported family income below the poverty line, with either a black or Hispanic borrower[27] and where the borrower has a subprime credit score (Vantage score under 701). Panel 2 presents estimates using loan attribute composition: share of loans that are HMDA rate spread loans, have a high loan to value ratio (above 0.95), and have a conforming loan disqualifying total debt expense to income ratio (above 0.45). The final panel presents estimates including controls from HMDA on the denial rate,[28] the application volume and a Herfindahl measure of market concentration of mortgage lenders. While some PUMA by purchase

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

cohort variables explain foreclosure, their inclusion has at most a modest impact on the estimated effect of predicted high cost lender share and never erodes the magnitude of the estimate.

5. Heterogeneity and Lender Market Share

To understanding what might drive this relationship, we first examine whether the effect of the shift-share prediction is isolated among only a subset of loans. Table 7 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 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 high cost lender share 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.[29] However, Column 4 indicates that the high cost lender share 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. For lenders who are not high cost, the standardized effect falls to 1.06 percentage points, but these loans have a smaller baseline foreclosure rate of only 3.5 percent so the percentage increase is similar to high cost lenders at 31 percent.

Given the larger absolute effect for high cost lenders, perhaps some of the high cost lender share effects arise because the market share of the borrower's lender influences foreclosure outcomes. We include a control for changes in the lender's market share (column 2), or predicted

[29] 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.

changes in the lender's share (column 3) using a shift-share for a single lender's PUMA weighted change in market share. The results are shown in Table 8. Panel 1 shows models that add either lender's actual market share or predicted changes in market share. The estimates on these variables are small and insignificant, and the estimates on the high cos lender share are unchanged.

Panel 2 column 1 of Table 8 presents the estimates from Table 7 Column 4 where predicted high cost lender share is interacted with whether the borrower's lender is high cost. In column 2, the borrower's lender 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,[30] but only when the borrower's lender is a high cost lender. However, controlling for lender market share has at most a modest effect on the predicted high cost lender share coefficients. Specifically, the effect of predicted high cost lender share falls by less than 8 percent for low cost lenders and 10 percent for high cost lenders.

In summary, the effects of our high cost lender proxy are broad based, and are not concentrated among especially vulnerable groups of borrowers or specific types of loans. These effects occur for loans made by both high cost and non-high cost lenders, even after controlling for the market representation of the lender. However, these shift-share prediction effects are substantially larger in absolute (but not relative) terms among loans from high cost lenders.

6. Potential Mechanisms

Given the stability of the estimates as individual borrower and loan controls are added and the broad based nature of the estimates, these findings suggest that foreclosure rates for a cohort of loans increase with share high cost lenders for all loans, above and beyond any direct effects of

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

an increase in the number of mortgages involving higher risk borrowers or loans with riskier terms. One possible explanation is that these estimates could be detecting the aggregate effects of very local foreclosure spillover effects or the effect of spatially broader foreclosure spillovers. We also investigate two additional hypotheses: whether the concentration of these high cost lenders is leads to higher housing prices in the run-up to the crisis and so larger equity losses during the crisis, and second whether the high foreclosure rates arising from the issuance of riskier loans (given the presence of high cost and other lenders making such loans) affects servicer behavior.

6.1 Local Housing Price Variation

High cost lender share may be associated with an expansion of credit that increased demand for housing and drove up local housing prices as these home purchases were occurring and so led to larger price declines and equity losses during the crisis. For example, Mian and Sufi (2009) observe that zip codes with a high share of subprime mortgages experienced greater increases in housing prices followed by higher foreclosure rates. 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 31st archives. Using the price indices, 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]

[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.

We also use the American Community Survey (ACS) to create measures of employment for each credit report year at the county level. Specifically, 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 white, black, Hispanic and Asian subsamples. We interact employment rate with the negative equity variables because 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 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 employment variables, and the third column includes the same interactions using the own-race county employment rate. While not shown, the negative equity variables are strongly associated with foreclosure using either county or PUMA price indices, and those effects weaken as county employment rates rise. However, these controls have at most a modest impact on the estimated effects of predicted high cost lender share. Controls for negative equity using county price indices increases our estimates by 5 to 10 percent, while using PUMA price indices increases our estimates by 15 to 17 percent. Estimates never fall as negative equity controls are added, and the largest high cost share effects arise when we use the more disaggregated PUMA price indices.

6.2 Mortgage Servicers and Foreclosure Decisions

Our second potential mechanism relates to the behavior of loan servicers. Potentially, the entry of high cost lenders into a submarket affects the foreclosure strategy of loan servicers during the crisis. Since each cohort of loans could follow a different time path of foreclosure, changes in lender management of delinquent loans might vary across cohorts. For example, local or regional servicer offices likely track foreclosure rates and may alter foreclosure processing when loans from the same area and originated at a similar time are experiencing high foreclosure rates.

While our data does not contain information on servicers, we pursue two analyses that are 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 because 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 delinquency rate.

A second way to examine servicer discretion is to compare California, a purely administrative foreclosure state, to the other states in our sample where judicial review plays a substantial role in the foreclosure process. In California, loans that enter severe delinquency typically move to auction after 120 days delinquent without requiring any court filings or approval.[32] So, if our effects arise from servicer discretion in foreclosure filings, we should not see these effects in California. We re-estimate our models separately for California, Los Angeles and the San Francisco bay area, and for our other five sites. Due to the size of the California sites, the number of clusters is similar in the two subsamples with over 150 PUMA's in California and about

[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.

180 PUMA's in the other sites. For the other sites, the standardized estimate is 1.83 percentage points or a 40 percent increase in foreclosure likelihood (Table 10 column 3). On the other hand, the point estimate for the California sites in column 4 (while noisily estimated) is only 7 percent of the baseline California foreclosure rate. Notably, having a rate spread loan remains a significant predictor in California, even though high cost lender share is statistically insignificant.

Servicer behavior appears important for explaining geographic patterns of foreclosure. For example, servicers may react to their experiences with specific groups of loans, in this case loans located near each other and originated at about the same time. As an indirect test of this premise, we select a hold-out sample of loans in each PUMA and origination year and use that hold-out sample to calculate foreclosure rates in 2005 through 2008 for PUMA by cohort cells. Then from 2006-2009, we use this foreclosure rate, i.e. experience from the preceding year, as a control in a sample that excludes the hold-out loans. Given the PUMA by crisis year fixed effects, the effect of this foreclosure experience variable is conditional on overall foreclosure rates in the preceding year, and so is only driven by cohort specific foreclosure rates. The estimated coefficient is highly significant and sizable. A one standard deviation change in PUMA by cohort preceding foreclosure rate is associated with a 1.44 percentage point increase in foreclosure, 27 percent of the baseline rate (Table 10 column 6), and shockingly similar to our baseline estimates of 1.46 percentage points for the effect of predicted high cost lender share (column 5). Further, unlike every other control examined, the inclusion of foreclosure experience in a horse race model (column 7) reduces the impact of high cost lender representation by 25 percent, doubles the standard error (leaving the standard error on foreclosure experience unchanged), and the estimate on predicted share is insignificant. Together, the evidence points towards servicers as a key mechanism for the impact of lender composition on the geographic concentrations of foreclosure.

6. Summary and Conclusions

We document a strong within submarket across cohort relationship between foreclosure and the market representation of higher risk lenders using a shift-share prediction of the share of loans by lenders who issue a substantial fraction of their loans as HMDA rate spread loans. After controlling for submarket, 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 initial high cost lender share cluster fixed effects, and the inclusion of a wide variety of time varying mortgage market composition variables. The effects are broad based influencing foreclosure for non-rate spread loans, white loans, loans sold to Government Sponsored Enterprises, and loans made by lenders that are not labelled high cost.

The identification strategy and the broad based effects suggest that foreclosure rates are increasing overall for these cohorts of loans, as opposed to the cohort simply containing a larger share of risky loans. We investigate two potential mechanisms. Our effects cannot be explained by greater equity losses due to housing prices fluctuations. However, our results do not arise when considering mortgage delinquency, which unlike foreclosure, is not directly affected by servicer decisions. Further, our effects are concentrated in the sites outside of California, where foreclosure is purely administrative limiting servicer discretion. Finally, we provide a mechanism for why servicer behavior is influenced by high cost lender representation, i.e. foreclosure experiences with a cohort of loans in a housing submarket explains foreclosure rates in the next year.

In summary, this paper suggests that mortgage servicers played a significant role in explaining the foreclosure concentrations that we document. Earlier work by Agarwal, Amromin, Ben-David, Chomsisengphet and Piskorski (2017) uses zip code variation to demonstrate that modifications lead to lower foreclosure rates, which implies that some locations had substantially worse foreclosure rates due to the servicers who happened to manage those mortgages. Similarly, our findings provide new evidence that servicer behavior, possibly in response to local foreclosure experiences, contributed to high levels of dispersion in foreclosure rates over space. The potentially large aggregate impacts of neighborhood foreclosure spillovers (Mian, Sufi and Trebbi 2015; Huang, Nelson and Ross 2018; Gupta 2019; Guren and McQuade 2020) imply that such geographic concentrations of foreclosure are costly, and large racial and income differences in exposure to subprime and high risk lenders suggests large equity impacts, as well.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens and Jeffrey Wooldridge. 2017. When Should You Adjust Standard Errors for Clustering? Working Paper.

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

Adelino, M., K. Gerardi, and P. S. Willen (2013). Why don't lenders renegotiate more home mortgages? redefaults, self-cures and securitization. *Journal of Monetary Economics* 60 (7), 835–853.

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 D. Evanoff. 2011a. Market-Based Loss Mitigation Practices for Troubled Mortgages Following the Financial Crisis. Federal Reserve Bank of Chicago Working Paper.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet and Douglas Evanoff. 2011b. The role of securitization in mortgage renegotiation. *Journal of Financial Economics*, 102(1), 559-578.

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).

Callaway, Brantly and Pedro H.C. Sant'Anna. In Press. Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment. *Journal of Econometrics*.

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.

Conklin, J. N., M. Diop, T. Le, and W. D'Lima. 2019. The importance of originator servicer affiliation in loan renegotiation. *The Journal of Real Estate Finance and Economics*, 59(1), 56–89.

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.

Ding, Lei. 2013. Servicer and Spatial Heterogeneity of Loss Mitigation Practices in Soft Housing Markets. *Housing Policy Debate*, 23(3), 521-54.

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

Ferreira, Fernando and Joseph Gyourko. 2012. Heterogeneity in neighborhood-level price growth in the United States, 1993-2009. *American Economic Review*, 102, 134-140.

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.

Greenstone, Michael, Alexandre Mas, and Hoai-Luu Nguyen. 2020. Do Credit Market Shocks Affect the Real Economy? Quasi-experimental Evidence from the Great Recession and "Normal" Economic Times. *American Economic Journal: Economic Policy*, 2020, 12(1), 200–225.

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.

Mayer, Christopher, Edward Morrison, Thomas Piskorski, and Arpit Gupta (2014). Mortgage modification and strategic behavior: Evidence from a legal settlement with countrywide. *American Economic Review*, 104 (9), 2830–2857.

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.

Palmer, Christopher. 2015. Why Did So Many Subprime Borrowers Default During the Crisis: Loose Credit or Plummeting Prices? University of California – Berkeley Working Paper.

Piskorski, Thomas, Amit Seru, and Vikrant Vig (2010). Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis. Journal of Financial Economics 97 (3), 369–397.

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.

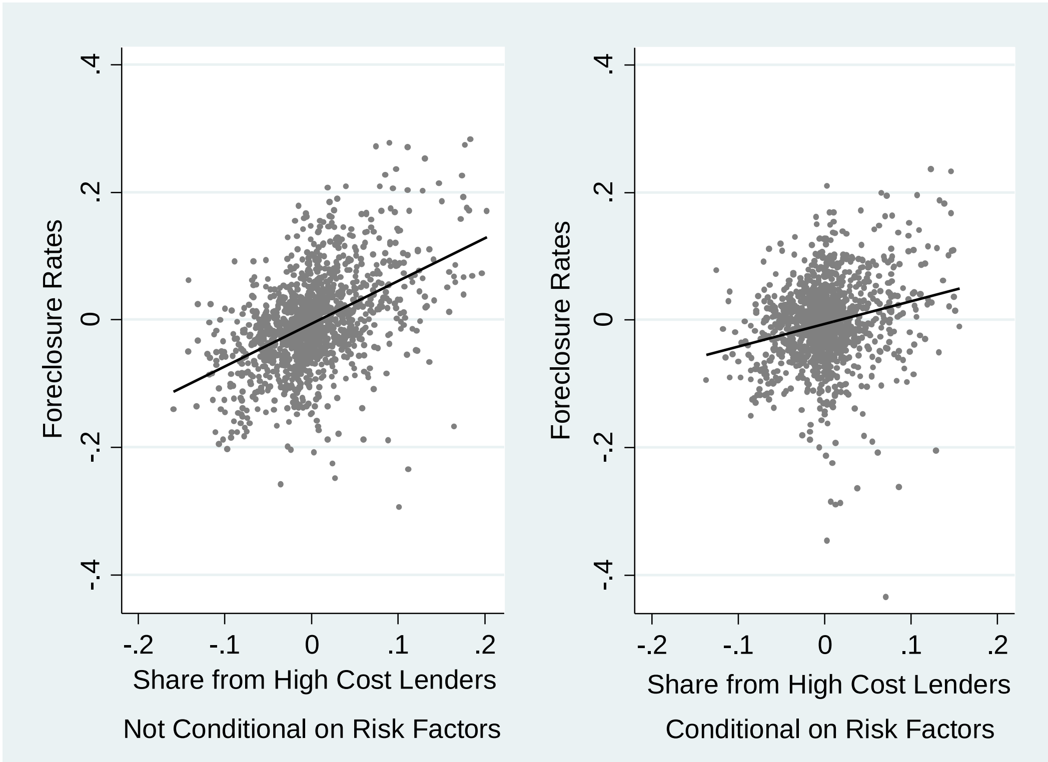
Reid, Carolina and Elizabeth Laderman. 2009. The untold costs of subprime lending: Examining the links among higher-priced lending, foreclosures and race in California. Federal Reserve Bank of San Francisco Working Paper.

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.

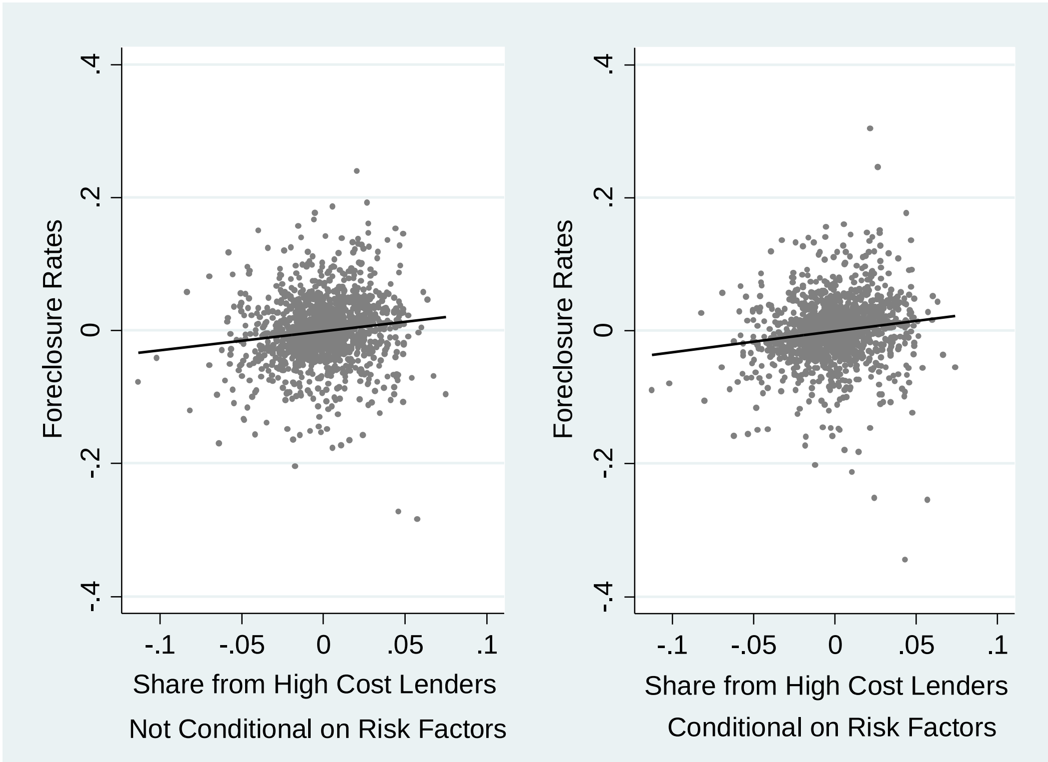
Sun, Liyang and Sarah Abraham. In Press. Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*.

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	0.021	0.143	0.035	0.185	0.070	0.256
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	4.382
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 20 percent or more 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
Share 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

	Rate Spread	Credit Score	Demographic	Risk Factors	Lender FE
Cross-Sectional Variation					
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
Shift Share Proxy 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 designated by *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Robustness Tests for Relationship between 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
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 Rate Spread	Sample Weights	Asymmetric Effects
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 Lenders			
Cluster Size over Share High Cost	0.01	0.02	0.03	0.04
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
	Pairwise Cohort Estimation to Avoid Staggered Roll-out Bias in Difference-in-Differences Models			
Purchase Cohorts in Sample	04 to 05	04 to 06	04 to 07	04 to 06 False 07
Predicted Change in Market Share	0.057	0.361**	0.162**	-0.024
	(0.134)	(0.142)	(0.068)	(0.126)
Observations	179,101	158,478	133,218	158,478
R-squared	0.109	0.149	0.118	0.149

Notes. Table presents models from Panel 3 Table 3. Panel 1 presents column 1 and 4 models for alternative shift share predictions using 15 percent or 25 percent share of rate spread loans for defining high cost lenders. Panel 2 presents the column 4 model 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 total fraction of rate spread loans is constant across years, and column 3 weights the sample based on 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. Panel 4 presents models including only two cohorts of loans where the last column tests whether the changes in the shift share proxy between the 2006 and 2007 cohorts can explain changes in foreclosure between the 2004 and 2006 cohorts. Significance is designated by *** p<0.01, ** p<0.05, * p<0.1.

Table 5: 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 designated by *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Time Varying Measures of Borrower Composition, Composition on Loan 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 designated by *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Heterogeneity in the Effect of the Shift Share Proxy for Lender Composition

	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 designated by *** p<0.01, ** p<0.05, * p<0.1.

Table 8: 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 designated 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: Foreclosure Models Relevant to the Role of Loan Servicers

	Foreclosure vs. Delinquency		Judicial vs. Administrative		Lagged Foreclosure Experience		
	Foreclosure	90 Day Deliquency	Outside California	California	Baseline 06-09	Experience	Horse Race
Rate Spread Loan	0.035***	0.025***	0.033***	0.038***	0.036***	0.036***	0.036***
	(0.002)	(0.002)	(0.003)	(0.005)	(0.002)	(0.004)	(0.004)
Predicted Change in Market Share	0.217***	0.069	0.269***	0.116	0.217***		0.162
	(0.060)	(0.052)	(0.068)	(0.186)	(0.062)		(0.120)
Foreclosure rate by PUMA by Origination						0.377***	0.375***
Year for Preceding Year						(0.046)	(0.046)
Observations	303,015	303,015	212,142	90,872	286191	142,818	142,818
R-squared	0.143	0.092	0.128	0.174	0.144	0.152	0.152

Notes: This table presents estimates of models from the fourth column of Table 3 Panel 3, Columns 1 and 2 present these estimates for foreclosure and for 90-180 day mortgage delinquency as the dependent variable, respectively. Columns 3 and 4 present the foreclosure model for all non-California sites and the two California sites, respectively. Columnd 5-7 present models of foreclosure dropping credit reports in 2005. Column 5 replicates Table 3 Panel 3 column 4 without 2005 foreclosures. Column 6 replaces predicted high cost lender share with the foreclosure rate from a hold-out sample for the same PUMA and origination year in the preceding crisis year, e.g. 2005 experience for credit report data in 2006. Column 7 presents a model containing both the shift share proxy and the previous year foreclosure experience. The hold-out sample of mortgages represents half of all mortgages in each PUMA by cohort cell, and estimations are conducted using the other half of the observations. In Columns 6 and 7, estimates are averaged across a set of 500 randomly sampled hold-out samples, standard errors are bootstrapped by resampling PUMA and resampling PUMA by cohort hold-out samples 500 times, and observation counts and R-squareds are averaged across hold-out samples. Significance is designated by *** p<0.01, ** p<0.05, * p<0.1.

Appendix: Bias in IV Models with Proxy Variables

Consider a model with a proxy (X_1) for the actual variable of interest (X_2), in this paper X_2 might be the market penetration of subprime and other lenders originating higher risk mortgages and X_1 might be either the share of loans originated by high cost lenders or the shift-share predicted change in the share of loans from high cost lenders.

$$y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_{it}$$

where β_1 is zero, X_2 is unobserved, and E[X_1,X_2]>0. Further, assume that E[X_1, ε]=0 so that estimates of β_1 simply proxy for β_1 without raising concerns about endogeneity bias, which will be dealt with later in these notes.

Lets assume that our goal is to estimate the impact of a one standard deviation change in X_2 (labor demand shock) on y (neighborhood voting). Consistent estimates would be

$$\tilde{\beta} = \sigma_2 \beta_2$$

Standard omitted variable bias implies that the standardized coefficient estimate on X_1 using OLS should converge to

$$\hat{\beta}_1^{OLS} = \sigma_1 \left(\frac{\text{cov}[X_1, \beta_2 X_2]}{\text{var}[X_1]} \right) = \sigma_1 \left(\beta_2 \frac{\sigma_2 \rho_{12}}{\sigma_1} \right) = \rho_{12} \tilde{\beta} < \tilde{\beta}$$

The scaling error or attenuation bias that arises is simply the correlation between X_1 and X_2 and since correlations are always less than one the coefficient must be attenuated towards zero. The weaker the correlation between the variable and its proxy is the greater the attenuation bias.

Obviously a second concern is that E[X_1, ε]≠0 or that our proxy is endogenous, i.e. share of loans from high cost lenders. What if a second, truly exogenous proxy (X_3 where E[X_3, ε]=0), such as our shift share prediction, is available, but is less strongly correlated with X_2 than X_1. Then, using X_3 as the proxy would avoid the endogeneity bias, but would exacerbate any attenuation bias arising because the true variable of interest is unobserved. Specifically,

$$\hat{\beta}_3^{OLS} = \rho_{23} \tilde{\beta} < \hat{\beta}_1^{OLS} = \rho_{12} \tilde{\beta} < \tilde{\beta}$$

The question is whether using the second proxy (X_3) as an instrument for the first proxy variable (X_1) can improve our estimates relative to $\hat{\beta}_3^{OLS}$.

Using OLS, the predicted value of X_1 is

$$\hat{X}_{1i} = \hat{\gamma}_3 X_{3i} = \frac{\text{cov}[X_1, X_3]}{\text{var}[X_3]} X_{1i} = \left(\frac{\sigma_1}{\sigma_3} \rho_{13} \right) X_{1i}$$

Next, we can do the same omitted variable calculations based on an OLS regression for the predicted value of X_1.

$$\hat{\beta}_1^{IV} = \sigma_1 \left(\frac{\text{cov}[\hat{\gamma}_3 X_3, \beta_2 X_2]}{\text{var}[\hat{\gamma}_3 X_3]} \right) = \sigma_1 \left(\beta_2 \frac{\sigma_2 \rho_{23}}{\hat{\gamma}_3 \sigma_3} \right) = \frac{\rho_{23}}{\rho_{13}} \tilde{\beta} \begin{cases} > \rho_{23} \tilde{\beta} \\ <> \tilde{\beta} \end{cases}$$

Unfortunately, while the IV analysis mitigates the attenuation arising from X_3 being a worse proxy for X_2, it is impossible to say whether the resulting estimate lies above or below the true value without having a prior about the magnitude of the correlation between the true variable (X_2) and the exogenous proxy (X_3) relative to the correlation between the two proxy variables, X_1 and X_3.

Unlike the case of classical measurement error, estimation using a proxy variable only imposed the requirement that the two variables be positive correlated. In that case, IV estimation cannot provide consistent estimates or even guarantee a lower bound estimate when the true variable of interest is unobserved without imposing an assumption concerning the correlation between the variable of interest and the exogenous proxy. While this assumption may appear innocuous in some contexts, such an assumption is likely to be quite strong when considering correlations that are conditional on high dimensional fixed effects that substantially weaken the correlation between the endogenous and exogenous proxies. In such contexts, the best that we can do is obtain a lower bound of the standardized effect by estimating reduced form models using the exogenous proxy.