The Role of Technology in Mortgage Lending

Abstract

We propose to analyze the role of technology in mortgage lending. In the period from 2010 to 2015, technology-based ("FinTech") lenders have increased their market share of U.S. mortgage lending from 2.1% to 7.4%. We present evidence that FinTech lenders process mortgage applications more quickly, even controlling for a large set of loan and borrower observables and fine geographic and time controls. We propose to expand this analysis, and to examine whether FinTech lenders alleviate frictions in access to mortgages. To do so, we study 1) whether FinTech lenders adjust supply more elastically than other lenders in response to mortgage demand shocks, 2) whether FinTech lending is preferred by consumers with a high demand for online services, and 3) whether the presence of FinTech lenders helps reduce inefficiencies in refinancing decisions. Our proposed analysis has broad implications for the effects of technology on lending markets and the evolving role of soft versus hard information.

I. Research Question and Motivation

The residential mortgage industry is experiencing a wave of technological innovation as both startups and existing lenders seek out ways to automate, simplify and speed up each step of the mortgage origination process. For example, *Rocket Mortgage* from Quicken Loans, introduced in 2015, provides a tool to electronically collect documentation about borrower's income, assets and credit history, allowing the lender to make approval decisions based on an online application in as little as eight minutes (Goodman 2015). And in late 2016, Radius Financial Group closed on the first fully paperless mortgage (American Banker 2016). The housing industry magazine *HousingWire* now publishes a widely followed ranking ("HW TECH100") to showcase the top innovative technologies in mortgage lending.

In our proposed research paper, we plan to study whether and how these new technologies are affecting the mortgage market. Our basic approach is to analyze how mortgage lending by a set of financial technology-based (or "FinTech") lenders at the forefront of innovation differs from lending by more traditional mortgage providers. FinTech lenders vary in terms of their business model, but are generally characterized by a complete end-to-end online mortgage application process that is supported by centralized underwriting operations, rather than a network of local brokers or "bricks and mortar" branches.

Between 2010 to 2015, FinTech lenders have experienced year-on-year growth of 26% with total lending of $34bn in 2010 and $111bn in 2016. As a result, the market share of these FinTech lenders in home purchase mortgage originations has increased from less than 1% in 2010 to 5% in 2015 (see Figure 1), and their market share in refinancing has increased from 3% to 10% over the same time period (see Figure 2). Their increase in market share has been particularly pronounced for loans insured by the Federal Housing Administration (FHA), a segment of the market which primarily consists of relatively lower income and wealth borrowers, including first-time home buyers. Fintech lenders have thus emerged as an important source of mortgage credit to U.S. households within only a few years.

Our proposed analysis focuses on four types of mortgage lending outcomes. First, do FinTech lenders process mortgage applications more quickly than traditional lenders? Second, are FinTech lenders better able to accommodate variation in mortgage demand relative to traditional lenders which face short-run capacity constraints due to reliance on physical branches and labor intensive processes? Third, are Fintech lenders preferred by certain consumers with a high demand for online services, including younger and more educated borrowers, borrowers with high-quality Internet access, or borrowers located far from physical bank branches? Fourth, by reducing frictions in the mortgage application process, does the presence of FinTech lenders help reduce inefficiencies in mortgage refinancing by households (as discussed e.g., in Campbell 2006 and Keys, Pope, and Pope 2016)?

The answers to these questions are important in evaluating the impact of technology on the mortgage market. If FinTech lenders do indeed offer a substantially different product to traditional lenders, they may increase the supply of mortgage credit and consumer surplus, at least for certain populations such as younger individuals comfortable with transacting online, and perhaps other underserved populations. The new technology offered by FinTech lenders may also reduce frictions in mortgage lending, leading to fewer refinancing mistakes by borrowers, or ameliorating capacity constraints during periods of high mortgage application volumes, leading to faster passthrough of monetary policy to mortgage rates.[1]

Our analysis also has broader implications beyond mortgage lending. FinTech lending in mortgage markets is arguably the area in which recent financial technology has had the largest economic impact so far. The mortgage market therefore provides an opportunity to learn about how technology affects lending markets more generally, since other loan markets may undergo similar transformations in the future.

We propose to address our research question in several steps. First, we examine the effect of FinTech lending on loan outcomes. In particular, we are interested in processing time and loan risk. Using loan-level variation, we control for loan and borrower observables and examine whether FinTech lenders process mortgages through to origination more quickly than other lenders. As discussed below, our preliminary estimates establish that FinTech lenders reduce processing by time by 5.6 days, corresponding to 10.7% of average processing

[1] Our analysis may instead reveal that lending by FinTech firms is not special on these dimensions, and that such firms offer services that in fact are similar to traditional lenders in terms of processing times and scalability. Under this explanation, there are other economic forces (e.g., regulatory arbitrage or superior marketing) that led to a rise in the market share of lenders we classify as FinTech, but these forces are primarily unrelated to technology.

time.[2] Using both ex-ante observable risk measures and ex-post default data, we also plan to estimate whether FinTech loans are higher risk relative to traditional lenders. This result is important when evaluating other economic factors that may drive FinTech lending, and whether or not the faster processing of mortgage applications associated with the FinTech model comes at the cost of less stringent underwriting standards. Our preliminary evidence suggests that FinTech mortgage lending is if anything associated with *lower* ex-post default. This speaks against a 'lax screening' hypothesis, and may instead imply that FinTech lending technologies are able to attract or screen for less risky borrowers. As discussed in section 4, our proposed analysis will test this hypothesis in more detail, and discuss the welfare implications of any observed 'cream skimming'.

Second, we examine the effect of shocks to mortgage application volume. Applications vary greatly over time and across regions due to variation in long-term interest rates and local variation in refinancing needs. Fuster, Lo, and Willen (2017) show that increases in aggregate application volumes are strongly associated with increases in processing times, and increase the margins that originators charge, thereby attenuating the pass-through of lower interest rates to borrowers. We examine whether changes in application volume have a differential effect on processing times of FinTech lenders relative to traditional lenders. If FinTech lenders have a more scalable business, and face less binding capacity constraints, we expect a smaller change in processing times in response to an increase in application volume than for other lender types. We also test whether there is an increase in FinTech lender's market share at times of spikes in demand. We conduct this analysis both in the time series using high-frequency data and using cross-sectional variation in application volume. We establish plausibly exogenous variation in refinancing volume using changes in long-term interest rates. We expect that this result provides evidence on whether FinTech lenders provide a more elastic supply of mortgage finance than traditional lenders.

Third, we examine geographic variation in the growth in FinTech lending across U.S. census tracts, as a function of age structure, income and other economic and demographic characteristics. We expect that FinTech lenders have higher growth in areas with a higher

[2] Processing time is measured as the time (in days) from the date of submission of a mortgage application until the date of closing.

share of young and more educated people who are more likely to interact online. We also expect that FinTech lenders may have higher growth in areas with better Internet connectivity, or conversely, lower penetration of physical bank branches. Controlling for other observable characteristics, we can characterize the determinants of the market share of FinTech lenders. Using plausibly exogenous variation in pre-determined variables, we can possibly establish some of the causal drivers behind FinTech lending. At a minimum, we can develop a set of careful stylized facts about the determinants of the growth of FinTech mortgage lending.

We also plan to explore in more depth the role of internet access and the digital divide on FinTech mortgage borrowing. There is a widespread concern that unequal access to internet services exacerbates underlying wealth and income inequality across households. Exploiting the staggered roll-out of internet services in selected cities, we plan to evaluate the importance of the digital divide with respect to FinTech lending. As internet service expands, do we find a shift toward FinTech lenders? On the one hand, we may find no effect of the digital divide on FinTech lenders since a large majority of U.S. households can already access internet services. On the other hand, we may find that access to high-speed internet is an important determinant of whether a household uses a FinTech lender. The result of this analysis sheds light on whether a digital divide exists for household access to finance.

Fourth, we propose to examine the role of FinTech lenders on the propensity to refinance. The household finance literature has shown that a large fraction of households refinance suboptimally, and in particular fail to refinance even though it is in their interest. This result is attributed to frictions in mortgage markets such as a lack of financial access or insufficient financial literacy. We examine whether FinTech lenders increase the likelihood of refinancing among household that are likely to benefit from refinancing. Specifically, we test whether geographic variation in the presence of FinTech lenders across locations predicts differential responsiveness of local refinancing propensities to shocks. We expect that this analysis will provide evidence on whether FinTech lenders have a positive impact on overall consumer surplus by improving access to mortgages. More efficient refinancing also potentially affects the pass-through of monetary policy, which is muted if fixed-rate mortgage borrowers do not refinance when it is in their interest to do so. (The "refinancing channel" of monetary policy is studied in recent work by Beraja, Fuster, Hurst, and Vavra

(2016), Di Maggio, Kermani, and Palmer (2016), or Wong (2016).)

This research contributes to a large literature on residential mortgage lending (see Badarinza, Campbell, and Ramadorai (2016) and Campbell (2013) for recent surveys). Little of this work explicitly studies the role of technology, with the exception of very recent research by Buchak, Matvos, Piskorski, and Seru (2017) that studies the growth of FinTech mortgage lending and non-bank lending more generally.[3] To the best of our knowledge, our paper would be the first to study the determinants of mortgage processing times and estimate whether technology can speed up the mortgage origination process and increase the elasticity of mortgage supply. Our work is also closely connected to research on the role of hard versus soft information in lending (Petersen and Rajan (2002); Stein (2002)). FinTech lenders rely heavily on hard information and our paper contributes to understanding how changes in the processing of hard information influence lending supply.

The rest of this proposal proceeds as follows. Section 2 describes the data sources we plan to use. Section 3 describes our approach for identifying FinTech mortgage lenders. Sections 4 through 7 describe the substantive analysis we propose to conduct, and present preliminary evidence to date. Section 8 concludes and presents a timeline for the completion of our analysis.

II. Data Sources

This section describes the data sources that we plan to use for the project. We note whether the dataset is publicly available or restricted. We further note whether we already established data access and whether the data is already used in the preliminary analysis.

[3] We became aware of Buchak et al. after submitting the initial version of this proposal. There is some overlap between our planned analyses and theirs in that both papers study covariates of FinTech market share. We propose to do so at a finer geographic level using additional covariates (including variables derived from proprietary credit bureau data) and mortgage default outcomes (we plan to examine FHA loans whereas Buchak et al. look at loans insured by Fannie Mae and Freddie Mac). Aside from this, however, Buchak et al. focus on the role of regulatory arbitrage and legal risk as drivers of the rise of non-bank lending. In stark contrast, we focus directly on FinTech lending by examining application processing times and supply elasticities, as well as effects on refinancing propensities.

Mortgage characteristics and mortgage application processing times. We draw this information at the loan level from Home Mortgage Disclosure Act (HMDA) data. HMDA data report characteristics of individual residential mortgage applications and originations from most banks and nondepository lenders (only some small lenders are exempt). Data include the identity of the lender, loan amount, borrower income, property location, application date and action date on which the mortgage was either funded or denied. Based on known local conforming loan limits, we can also impute whether a loan has "jumbo" status and thus cannot be securitized by Fannie Mae, Freddie Mac, or Ginnie Mae. We use HMDA data to study the volume and characteristics of mortgages originated by FinTech lenders compared to other types lenders, and the time taken to process each mortgage application (the number of days between the application date and action date). The processing time can only be computed from a restricted version of the dataset to which we have access.[4] All other variables can be computed from the publicly available version of the data. At present, we have access to HMDA data through 2015, but time-permitting we hope to incorporate 2016 in the final analysis when those data are released in late 2017. ***[Status: restricted data, data access established, data is cleaned and used for preliminary analysis]***

Segment-level FHA mortgage default rates by lender. We plan to draw these data for Federal Housing Administration (FHA) mortgages from information reported on the FHA website under their Neighborhood Watch Early Warning System portal. Information on two-year serious delinquency rates is available by lender at the state or county level for different origination cohorts and broken down by mortgage characteristics such as loan size buckets. These data capture a large fraction of all high-credit-risk mortgages originated in the US in recent years, and allow us to use realized loan performance to measure the credit quality of FHA mortgages originated by FinTech lenders compared to other lenders. ***[Status: public data, state level data collected and analyzed, further data collection in progress]***

[4] The restricted-use version of the dataset, which is available to users within the Federal Reserve System, records for each application the dates on which the lender received the application and also the date on which the application was resolved (e.g. origination of the loan or denial or withdrawl of the application). The publicly available HMDA data only contains year indicators.

Agency mortgage delinquency rates by lenders and geography. Similar data to the above (at the loan level) is also available for 30-year fixed-rate mortgages insured by Ginnie Mae, Fannie Mae and Freddie Mac. These data contain a lot of additional borrower and loan characteristics. However, lenders are not in all cases individually identified, as discussed in the next section. ***[Status: public data, data collection in progress]***

Internet Connectivity. There are two data sources for our Internet connectivity data. The earlier period is covered by National Telecommunications and Information Administration (NTIA) data from the National Broadband Map. The latter period is covered by Federal Communications Commission (FCC) data. The two data sets are compatible for our usage of the data and allow us to construct a data set covering the entire period. The NTIA data is reported every 6 months in June and December from December 2010 to June 2014. The data reports the number of services offered by each provider in each census block. Services are differentiated by technology type; Time Warner can report providing both DSL and fiber Internet within a census block, for instance. The FCC data, likewise, is reported every 6 months in June and December from December 2014 to December 2015. The data is similarly structured and also reports the number of services (technologies) provided by each provider by census block. So far we have collected the data for California, Kansas, and Missouri for all time periods. ***[Status: publicly available data, data for California, Kansas, and Missouri are collected and used in preliminary analysis; data collection for other states in progress]***

Age and risk structure based on Federal Reserve Bank of New York / Equifax Consumer Credit Panel (CCP). We compute the risk and age structure of geographic regions using data from the CCP. The CCP is a nationally representative sample of five percent of all individuals with a credit record and a valid Social Security number. The CCP tracks individuals over time at a quarterly frequency and collects data on their debt holdings, payment history, credit scores, age and geographic location (see Lee and van der Klaauw (2010), for more details). We can collapse the CCP by geography in order to summarize the risk and age structure of mortgage borrowers by county or census tract. ***[Status: restricted***

data, data access established]

Demographics and industry composition. We collect data at the census-tract level on local population characteristics such as age structure, adult educational attainment, and population density from the 2010 U.S. Census and the American Community Survey. ***[Status: public data, data partially collected]***

Mortgage servicing data linked to credit records. The Equifax CRISM dataset merges mortgage servicing data (from McDash) with individual-level credit records, which allows to measure for all outstanding loans whether/when they get refinanced and at what terms (including whether the borrower withdraws equity). This is not possible with mortgage servicing data alone, which only measures whether a loan is paid off (which could also be due to a borrower moving). We will thus use these data to measure local refinance propensities and equity withdrawal volumes (as previously done by Beraja, Fuster, Hurst, and Vavra (2016)). ***[Status: restricted data, data access established]***

Bank branch distance. We collect data on bank branch location from the FDIC Summary of Deposits Data. The data set contain GIS coordinates for each branch. We use standard GIS software to compute distance in bank branches across counties and census tract. This data will allow us determine the importance of nearby branches in FinTech lender market share.***[Status: publicly available data, data is cleaned and can be used for analysis]***

Home prices and macro variables. In some of our analyses, we plan to use local home price levels (the dollar price of a median home) and growth rates as control variables. We plan to use publicly available data from Zillow; we find that their county-level data covers about 83% of observations in the HMDA data. We may be able to use finer disaggregations; Zillow now produces census-tract-level indices, although these have not yet been released. We will also use some economy-wide measures of interest rates in some of the analyses; in particular the headline 30-year fixed-rate mortgage rate from the Freddie Mac Primary Mortgage Market Survey. These data can be downloaded from FRED (St. Louis Fed).

[Status: publicly available data, data is cleaned and can be used for analysis]

III. Defining and Measuring Fintech Lenders

While mortgage lenders have adopted new technology to varying degrees in recent years, we have chosen to focus on a distinct subset of lenders that are at the vanguard of technology adoption. To date, we have examined the application process and the mission statement for the largest mortgage lenders in 2015. Table 1 provides the names and market share of the top 25 lenders based on HMDA data.[5]

From this list, we define firms as FinTech lenders if their application process can be conducted entirely or nearly entirely online, such that the applicant is not required to interact with a mortgage salesperson. To verify this, we manually initiated mortgage applications for each of the top 25 mortgage banks by market share. We cross-checked the results of this exercise against industry newsletters and the extent to which each firm emphasized their technology platform as a a differentiating factor to investors and customers. These alternative sources corroborate our classification of which firms offer a comprehensive online application, suggesting this is a good summary indicator of technology focus.

Four lenders qualified as FinTech in our analysis: Quicken, LoanDepot.com, Guaranteed Rate, and Movement Mortgage. While not a comprehensive measure, we believe we have identified the largest lenders with a sophisticated online application process. In addition, there are several relatively new lenders that emerge in 2016 that are not in our current sample but qualify as FinTech lenders by our definition: SoFi, Better (dba Avex Funding), and Lenda. We plan to add these lenders to our sample when 2016 HMDA data become available. We also plan to to conduct a more thorough examination of the lender universe going forward, and in particular to expand our classification to firms outside the top 25 lenders.

Aside from FinTech lenders, our analysis distinguishes between "Deposit Banks" and "Mortgage Banks," where the latter are not depository institutions (i.e., mortgage banks do

[5]We corroborate these numbers with an industry publication *Inside Mortgage Finance*. The two datasets use different methodologies for computing market shares, however we draw similar conclusions.

not take deposits). Mortgage banks (as well as our FinTech lenders) typically rely on a line of credit to fund their mortgages and securitize or sell loans they originate, the latter either to a larger financial institution or directly to Fannie Mae or Freddie Mac. Deposit banks include commercial banks, savings banks and credit unions.

IV. Is Technology-Based Lending Faster?

A. Proposed empirical strategy

Our first research question is whether technology-based lenders are able to process a mortgage application through to origination more quickly than other lenders. Like Fuster, Lo, and Willen (2017) we measure processing time using Federal Reserve confidential-use HMDA data as the number of days between the mortgage application date and the final action date (the origination date for completed applications, or the date on which the application was rejected or withdrawn).

Processing times are likely to vary geographically due to differences in state laws, housing market conditions, and other factors. Thus, we propose to study variation in processing time within region and time by conditioning on location-time fixed effects. We also control for other borrower characteristics such as race and income. Our basic specification is:

$$\text{Processing Time}_{ijct} = \quad \delta_{ct} + \beta \text{FinTech}_j + \gamma Controls_{it} + \epsilon_{ijct} \tag{1}$$

where Processing Time_{ijct} time is measured in days for loan i issued by lender j in location c at time t, FinTech_j is an indicator variable equal to one for FinTech lenders, δ_{ct} is a vector of location-month fixed effects, and $Controls_{it}$ are borrower and loan controls that include log of borrower income, the ratio of the loan amount to borrower income, indicator variables for race and gender, an indicator variable for whether there is a coborrower, and indicator variables for whether the loan is insured by the FHA or Veterans Administration (VA).

This model will be estimated at the loan level using HMDA data from 2010 to 2015 for all loans originated in the U.S. mortgage market. Our baseline specification will include only applications that lead to a mortgage origination. The analysis will be conducted separately for home purchase mortgages and refinancings (the former is a mortgage taken out to finance the purchase of a property, while the latter is not).

The coefficient of interest is β. If FinTech lenders can process mortgage applications faster, we expect $\beta < 0$. As discussed below, we find preliminary evidence that β is indeed negative. Quantitatively, FinTech lenders are found to process mortgage applications through to origination 5.6 days faster than other mortgage lenders.

Bank lenders may face regulatory obstacles that increase processing times, therefore we will consider a sample restricted to mortgage banks (i.e. non-depository lenders). Also, FinTech lender market share, like the market share for mortgage banks, has grown disproportionately for loans insured by the FHA. We plan to explore both processing time and loan risk for this submarket. Furthermore, we plan to study also how different quantiles of processing times differ between FinTech and other lenders (since e.g. average processing times could be shorter simply due to FinTech lenders reducing the incidence of very long processing times — the right tail — rather than the processing time for the median borrower).

One concern is that the endogenous matching of borrowers and lenders may affect measured processing time for FinTech lenders. For instance, if younger borrowers are more likely to apply to FinTech lenders and also tend to submit their paperwork faster, then FinTech lenders will appear to process mortgages faster, even if they do not have an inherent technological advantage. We plan to address this concern as follows:

1. Sensitivity to controls and matching. Our specification partially addresses endogenous matching concerns as we can condition on a large number of observable loan, borrower and geographic controls. As one test, our preliminary analysis finds that the key coefficient of interest is relatively stable to the inclusion or exclusion of these controls, providing some evidence against endogenous matching. Going one step further, we also plan to explore a matched sample analysis studying differences in processing time after matching FinTech and non-FinTech loans along all observable dimensions (location, application month, loan characteristics, borrower characteristics).

2. Placebo tests. We plan to tests two 'placebo' predictions which would hold if shorter processing times of FinTech lenders are a result of endogenous selection:

(a) If FinTech lenders match with 'fast processing' borrowers, the growth in FinTech market share should be larger in geographic areas where processing times were already shortest before the growth in FinTech market share (e.g., 2010).

(b) If non-FinTech lenders lose their faster customers to FinTech lenders, processing times for should have increased disproportionately for borrower/loan types with high FinTech penetration. For example, if FinTech lenders target FHA refinance mortgages but not jumbo purchase mortgages, and if they extract the fastest borrowers from the pool of applicants, then traditional lenders should see increased processing times for FHA loans relative to jumbo loans (conditional on observables).

3. We will also attempt to use plausibly exogenous regional variation in FinTech adoption as a source of variation to test for causal effects of FinTech on processing times. For instance, we are planning to explore within-county, cross-census-tract variation in the expansion of high-speed Internet access (see Section 6) as a potential instrument for the growth in FinTech lending. However, the strength of these potential instruments is yet to be determined.

A.1. Preliminary evidence

We have already estimated preliminary versions of the application processing time regression (3) described above, which compares the processing times of FinTech lenders with those of other mortgage lenders conditional on borrower characteristics, loan type, and fine geographic and time controls (namely county-month fixed effects). Our analysis suggests that FinTech lenders are able to complete the mortgage origination process more quickly than other lenders (both deposit banks and mortgage banks), all else equal. For home purchases, Table 4, and refinancings, Table 5, we find FinTech lenders have faster processing times by 5 to 11 days. This finding is robust to including loan characteristics, demographics controls, and county-month fixed effects (Column 4). In addition, we include a control for Mortgage Banks (Column 5) so that the FinTech coefficient can be interpreted as the difference between FinTech lenders and other non-deposit lenders. The coefficient on FinTech lenders is remarkably stable across specifications. This is suggestive evidence that our results are not an artifact of endogenous matching between firms and lenders.

The findings in Tables 4 and 5 are consistent with the view that FinTech lenders have developed a technological advantage in mortgage processing. Much of this proposal is focused on further investigating the impact of these differences on the mortgage industry and borrowers.

B. *More efficient or just less careful?*

Even if FinTech lenders are found to process mortgage applications more quickly, this may simply reflect less careful screening of borrowers, rather than greater efficiency. We propose to test this screening hypothesis by studying the ex-post performance of FinTech-originated loans compared to similar mortgages from other lenders. We intend to focus on FHA lending, which has been the riskiest segment of the mortgage market in recent years.[6]

We propose to utilize two sources of data on ex-post default for FHA loans, one at the segment level, the other at the loan level. Regarding the first source, data on FHA mortgage default rates by lender are available using a query tool on the FHA Neighborhood Watch Early Warning System data portal. For a given lender, one and two-year default rates are available broken down by origination vintage and geography (county, state and metropolitan statistical area) and by some other loan characteristics (e.g., whether the loan is a refinancing or purchase-money mortgage, and whether the borrower is in a low-income census tract). For a given lender, the Early Warning System query tool reports the lender's scaled default rate relative to all FHA loans within the same 'cell' (that is, the same vintage, geography and set of loan characteristics specified in the data query), as well as other information such

[6]FHA mortgages require a down payment of as little as 3.5% and are generally made to borrowers with low credit scores who do not qualify for a prime conforming loan. FHA loans are government-guaranteed, which limits the credit risk for the lender. However the lender is not fully indemnified agains risk since the FHA can refuse to compensate the lender for credit losses if there is fraud or other defects in mortgage underwriting. FHA lenders have also paid out large legal settlements on FHA loans due to breaches of the False Claims Act and other laws. As a result of these risks, many large bank lenders have withdrawn from FHA lending or wound back their participation in the market (see e.g., Wall Street Journal (2015)). We focus on FHA loans rather than mortgages securitized via Fannie Mae and Freddie Mac because the latter have significantly less credit risk and have experienced very low default rates during the period of our proposed study.

as the total originations by the lender in that cell.

We propose to estimate variations of the following specification:

$$\frac{\text{Default rate}_{lgv}}{\text{Default rate}_{gv}} - 1 = \beta \text{FinTech}_l + \gamma \text{FinTech}_l \times \text{characteristics}_g + \epsilon_{jgv}$$

where default rate is the number of defaults as a share of originations for lender l in geographic area g and mortgage vintage v, FinTech_l is a dummy for FinTech lenders or a vector of dummies for different subsets of FinTech lenders, and characteristics_g includes characteristics of the local geographic region. The regression will be estimated using weighted least squares, weighting by origination volume. Note that this regression does not include geography fixed effects, since the dependent variable is already scaled by aggregate defaults for area g and vintage v.

We will first estimate regressions excluding the interaction terms and examine the sign and magnitude of β. A positive estimated β implies that the set of lenders defined by the Fintech dummy have higher realized default rates than the universe of FHA loans within the same geographic areas, consistent with the 'lax screening' hypothesis. A negative β indicates the reverse, perhaps implying that the automated technologies used by FinTech lenders actually screen borrowers more effectively than the more labor-intensive techniques used by other mortgage lenders. Goodman (2015) argues in favor of this second point of view, but without presenting systematic evidence, as we propose to do.

In additional specifications we will then include interaction terms to test for example whether scaled FinTech default rates are relatively higher in regions where these lenders have grown most quickly. This sheds light on whether rapid observed growth in technology-based mortgage lending has been associated with lower lending standards or less careful screening.

B.1. Preliminary evidence

We have so far collected FHA default data from the FHA portal at the state level for each of the four FinTech lenders identified for this proposal. Results for the specification outlined above are presented in Table 3. The first column of results includes a single constant term for all FinTech lenders. The coefficient of -33.0 means that the default rate for these firms is 33% lower than for FHA loans in the state as a whole. The results in the second column use an alternative measure of performance which may control more precisely for the borrower's observable risk, the supplemental performance metric (SPM) also available from the Early Warning System portal. The SPM is defined as the default rate for the lender relative to a constant benchmark default rate defined by the FHA based on the FICO credit score of the loan (defined based on three bands: FICO<640, FICO between 640 and 680, and FICO>680). We then construct the dependent variable as the percentage difference between the FinTech lender's SPM and the average SPM for the state as a whole. Again the coefficient is negative and statistically significant, albeit somewhat smaller in magnitude (at -21.3% compared to -33% in the first column). The third column separately identifies Quicken (the largest FinTech mortgage lender) vs the other three FinTech firms; our preliminary finding is that Quicken accounts for the lower default rate of FinTech lenders. The fourth column interacts these FinTech dummies with a dummy equal to 1 for the five states where each lender has the highest market share. The comparative FinTech default rate is no higher in these states, suggesting growth in these areas is not occurring through excessive risk-taking or lax screening.

These preliminary results suggest that lenders classified by us as FinTech lenders identified so far as a group have lower realized default rates than FHA lenders as a whole. We propose in the full research paper to repeat this analysis at a finer level of aggregation (by geography and also other characteristics), to ensure these results are not driven by compositional effects. (Note: this finer data collection will take some time because the FHA query tool is somewhat unwieldy). We also plan to augment this segment level default analysis with loan-level performance data from Ginnie Mae, as described below.

B.2. Proposed loan-level FHA analysis

The government agency Ginnie Mae has made loan-level data on the mortgages underlying its MBS available, covering the period since 2013. We also plan to use these data to study the comparative default propensity of FinTech borrowers. The key advantage of these data relative to the FHA Early Warning System data described earlier is that the Ginnie Mae data include a rich set of loan and borrower characteristics (e.g., the borrower's credit score), allowing us to precisely investigate whether FinTech lenders target borrowers who are more or less risky on observable dimensions, and separately whether Fintech borrowers default more or less often *conditional* on observables. For example, if FinTech lenders employ more sophisticated screening algorithms than other lenders, they may be able to attract less risky borrowers among borrowers who look similar based on basic underwriting characteristics.

The key disadvantage of these data is that they include only the *issuer* identity, not the identity of the original lender. Among FinTech lenders only Quicken is a large Ginnie Mae issuer (while the others appear to first sell a significant portion of their loans to other firms before issuance). Thus, these data will only imperfectly identify loans from FinTech lenders (although it will do a good job of identifying mortgages originated by Quicken, the largest and best known of the four FinTech lenders we study).

Operationally, we plan to estimate simple default regressions using a logit model, where the dependent variable is a dummy equal to one if the loan defaults, and the key right-hand-side variable is a dummy equal to one for a FinTech issuer. We would estimate these regressions both unconditionally and conditional on loan covariates, to separately identify the component of credit risk reflecting observable mortgage characteristics versus the residual component that could reflect lender underwriting standards.[7]

B.3. Cream skimming?

The analysis described above may find that FinTech borrowers have lower default rates conditional on observable characteristics. This would suggest that these firms are selecting low-risk borrowers implying that the remaining pool of borrowers is likely to be riskier. Although superior screening can be viewed as an additional advantage of technology-based lending, in some contexts this kind of skimming by lenders may have negative overall welfare

[7]The government-sponsored enterprises Fannie Mae and Freddie Mac have made available loan-level data on mortgages securitized into agency MBS, which identify the mortgage originator of each loan. We may use these data to analyze the relative default probability of agency mortgages originated by FinTech lenders compared to other lender types, using a similar framework. The main drawback of these data are that these prime agency mortgages are significantly less risky than FHA loans, and have experienced very low default rates for recent vintages. Also, not all of our FinTech lenders may be individually included (since for sellers that represent less than one percent of volume within a given acquisition quarter, the seller name is not individually disclosed in these data).

consequences. One reason is that it could shift costs to the government if private and public lenders compete (an argument that has been made in the context of FinTech lenders like SoFi in the student loan market[8]). Another mechanism is that skimming could lead to ex ante credit rationing by weakening the credit quality of the remaining borrower pool. (This mechanism is explored by Mayer, Piskorski, and Tchistyi (2013) in the context of private subprime mortgage lending.)

In the particular context we study, these negative welfare consequences are unlikely to be a significant consideration because essentially all risky mortgages in the U.S. today are government insured at a pre-set price, either by the FHA or other government agencies such as the Veterans Affairs Department. Consequently, skimming of low-risk borrowers by FinTech lenders is unlikely to materially reduce credit access for remaining borrowers, who will still qualify for government insurance. We do however believe it is interesting to establish whether FinTech lenders originate loans which are less risky on unobservables, since this could shed light on the screening effects of technology-based lending more broadly, including contexts where the welfare consequences of skimming may be more significant.

V. Is Technology-Based Lending More Elastic?

A. Proposed empirical strategy

Our second question of interest is whether FinTech lenders are better able to accommodate shocks to the level of demand for new mortgages. Loan application volumes in the U.S. fluctuate enormously over time, primarily due to movements in interest rates that can lead to "refinancing waves." There is also substantial variation in the cross section of locations, due to differential housing market trends. As a consequence of this variation, a key challenge for mortgage lenders is to manage the flow of mortgage applications they are processing. If a lender receives more applications than their underwriting process can handle, their processing times increase and they risk losing money (and future business) due to loans not closing in a timely manner. Figure 3, which is similar to evidence in Fuster, Lo, and Willen (2017),

[8]See e.g. `https://www.bloomberg.com/news/articles/2015-06-10/student-loan-refinancing-boom-could-cost-u-s-taxpayers-billions`

illustrates (i) the large variation in the level of monthly applications (with the peak level being almost three times as high as the trough), and (ii) that looking across all lenders, median processing times are strongly correlated with the total number of loan applications.

FinTech lenders, by automating and standardizing much of the underwriting process, could conceivably improve the short-run elasticity of lending supply with respect to demand shocks. We take several approaches to test the for differences in capacity constraints at FinTech lenders. In our initial specification, we plan to regress a loan's processing time on the application volume received by the lender over the same time period, interacted with whether the lender is a FinTech lender:

$$\text{Processing Time}_{ijt} = \gamma \text{Applications}_{jt} + \beta \text{Applications}_{jt} \times \text{FinTech}_j + \alpha_j + \theta \text{Controls}_{it} + \epsilon_{ijt}$$

where Applications_{jt} is the log of applications received by lender j at time t. The coefficient of interest, β, estimates the differential sensitivity of processing time to loan applications for FinTech lenders. Lender fixed effects, α_j, are included in order to absorb fixed differences in processing times and application volumes across lenders. Note that since this analysis is conducted at the loan level, i, so that we can further control for loan characteristics, $Controls_{it}$, that might influence processing time, including size, location, borrower income and application time fixed effects. Lastly, at the firm level there may be strong trends in applications and significant volatility for small lenders; therefore, this analysis will most likely condition on a set of larger lenders (e.g. top 100 lenders) and may need to be modified to account for firm level trends in applications.

The above regression of lender processing time on lender applications could understate the differences in elasticities across lender types. For example, lenders may solicit applications when processing constraints are slack which would attenuate the relationship between applications and processing time and obfuscate differences across firms. To more cleanly identify the elasticity differences, we plan to exploit demand shocks that vary application volumes independent of firm-specific conditions. The first of such shocks is the aggregate variation in applications which is primarily determined by macroeconomic factors, such as interest rates, and plausibly exogenous to the capacity conditions of any individual lender.

$$\text{Processing Time}_{ijt} = \gamma\text{Applications}_t + \beta\text{Applications}_t \times \text{FinTech}_j + \alpha_j + \theta\text{Controls}_{it} + \epsilon_{ijt} \quad (2)$$

where Applications_t is the log of aggregate mortgage applications, which captures variation in the demand for loans. Here, β is the differential sensitivity of FinTech lenders to aggregate variation in application volumes. A variant on this specification could instrument for the aggregate variation in loan applications with changes in specific macroeconomic factors, like interest rates. Fuster, Lo, and Willen (2017) show that aggregate application volumes are strongly correlated with the difference between the average coupon on outstanding mortgages and the current ten-year Treasury yield; this difference proxies for the average refinancing incentive of current borrowers while being exogenous to current conditions in the mortgage market. The strength of this correlation over our sample period is illustrated in Figure 4.

A drawback of relying on aggregate variation is that there may be unobserved heterogeneity in lender exposure to overall demand fluctuations. To the extent this is based on location differences, we can condition on location-time fixed effects. An alternative would be to construct a Bartik-style estimate of exposure to aggregate demand by taking the weighted sum of location-specific variation in applications where weights vary across lenders according to prior activity in the area. The resulting regression would regress lender processing time on this proxy for exposure to demand variation, $\widehat{\text{Applications}}_{jt}$:

$$\text{Processing Time}_{ijt} = \gamma\widehat{\text{Applications}}_{jt} + \beta\widehat{\text{Applications}}_{jt} \times \text{FinTech}_j + \alpha_j + \theta\text{Controls}_{it} + \epsilon_{ijt}$$

The coefficient of interest, β, estimates the differential sensitivity of processing time to the application activity in its specific geographic areas of operation. Again, one could plausibly instrument for variation in local demand using measures of local refinancing incentives, or local house price appreciation.

One characteristic of FinTech lenders that differentiates them from more traditional mortgage lending models is that they tend to centralize their operations. As a result, it may be that they face similar capacity constraints in response to aggregate changes in loan applications, but are better able to accommodate local demand shocks by sharing resources across different markets. In contrast, a mortgage bank that relies on local resources to process loans is more likely to be constrained by their in-market capacity. To explore this alternative, we test for the correlation of local processing times with *local* demand variation:

$$\text{Processing Time}_{icjt} = \gamma \text{Applications}_{ct} + \beta \text{Applications}_{ct} \times \text{FinTech}_j + \alpha_{jt} + \alpha_c + \theta \text{Controls}_{it} + \epsilon_{icjt}$$

where Applications_{ct} are the log applications volumes in location c at time t. We include location fixed effects, α_c, to ensure we are capturing within location variation in processing times and lender-time fixed effects, α_{jt}, to focus on variation in processing times *within* a firm at a given point in time. In this regression, $\beta < 0$ would suggest that FinTech lenders are less subject to *local* capacity constraints.[9]

Even if FinTech lender processing times exhibit less sensitivity to demand conditions, it may be that they manage this by lending less when demand is high rather than accommodating demand by providing credit. To determine if the the volume of lending is differentially elastic with respect to demand, we plan to examine how the number of loans originated by lenders of different types varies with aggregate or local demand. In levels, this is complicated by the fact that there are trends in application volumes that differ across lender types (and individual firms within a type). Thus, a natural specification regresses changes in monthly applications received by a lender j on changes in monthly applications in the market overall (potentially excluding j) while allowing for differential sensitivity across lender types, and lender-specific fixed effects (meaning different average growth rates over this period):

$$\Delta \text{Applications}_{jt} = \gamma \Delta \text{Applications}_t + \beta \Delta \text{Applications}_t \times \text{FinTech}_j + \alpha_j + \epsilon_{jt}$$

where $\text{Applications}_{(j)t}$ is the logarithm of originated applications (by lender j). Alternative specifications could be run at the loan level (with a FinTech dummy as the dependent variable), or using local variation in demand.

[9]The appropriate definition of location is likely at the county or metropolitan statistical area level.

B. *Preliminary evidence*

FinTech lenders exhibit less variation in processing time. This can be seen visually in Figures 5 and 6 in terms of lower time-series volatility in processing time. In addition, we can see in Table 2 that the standard deviation of processing time in the sample of FinTech lenders is lower than those of the other types of lenders. Furthermore, we have calculated the standard deviation of processing time within firms over time. We do so by regressing processing time for bank-months on firm fixed effects and calculating the standard deviation of the residuals. We find that FinTech banks exhibit a lower standard deviation (28 days) relative to deposit lenders and mortgage banks (35 days). Hence the initial evidence suggests that FinTech lenders are better able to adjust to demand conditions.

More direct evidence on how processing times vary with aggregate demand volume is presented in Figure 7, which provides a visualization of the results from estimating equation (2). This binned scatter plot shows first that FinTech lenders have shorter processing times on average, as already found in the previous section. More importantly, it shows that FinTech lenders' processing times vary less with the level of demand for new mortgages in the economy. While banks' and (to a lesser extent) mortgage banks' processing times increase notably with demand, the same is only very mildly the case for FinTech lenders. The additional analyses described above intends to determine whether these preliminary results are robust, and to further flesh out the implications for loan supply elasticity.

VI. Who Borrows From FinTech Lenders?

A. *Empirical strategy*

This section proposes to examine the cross-sectional determinants of the growth in FinTech lending, which varies substantially across different regions of the U.S. (see Figures 8 and 9). We propose to study the determinants of this cross-sectional variation in FinTech growth at the census tract level using data from 2010 to 2015. We posit that five sets of variables may be important in accounting for FinTech mortgage lending growth: familiarity with using technology, access to technology, diffusion of technology, demand for rapid credit provision, and competition from traditional sources of mortgage finance that rely on a physical local presence. The main empirical challenge in establishing a direct effect of our main variables on FinTech growth is the issue of reverse causality, and the potential that omitted variables may affect both FinTech lending growth and our explanatory variables. We have two broad strategies to address this issue.

The first strategy is to use predetermined local characteristics in predicting FinTech growth. This is useful in our setting because there was little FinTech lending prior to 2010. Hence, it is unlikely that predetermined variables are picking up anticipated FinTech lending growth after 2010. This approach therefore limits the likelihood that our results are driven by reverse causality. Moreover, we can exploit the highly disaggregated nature of our dataset (66,438 census tracts) when examining the effect of predetermined variables. In particular, we can control for region-, state-, and county-specific time trends to control for unobserved local trends. We can also examine the robustness of our main coefficients to adding many control variables, which provides further support for establishing a direct effect. Even though this strategy does not necessarily establish a causal relationship, we believe that this approach will establish a useful set of stylized facts regarding which types of borrowers deliver novel results in terms of which types of borrowers have high our setting.

Our second strategy is to use variation in the availability of Internet access over time. We can combine our highly disaggregated data with detailed data on the changes in Internet availability during our sample period. As discussed in detail below, we are in the process of evaluating the rollout of Internet services across the country during the sample period. We emphasize that the variation in the roll-out is not necessarily exogenous to FinTech lending growth and may (at least partially) be driven by expected demand for Internet services. Yet, to the extent that we can identify variation in the rollout that is driven by technological or logistical considerations, we may be able to use time-series variation for the empirical identification.

B. Do local characteristics predict FinTech growth?

We propose using the following variables (generally measured in year 2010) to measure variation in local characteristics:

1. Familiarity with finance and technology: The literature on Internet usage documents that demographic variables are strong predictors of Internet usage and familiarity. We use three commonly used variables from this literature:

   (a) Age: We measure age as the share of the adult population below the age of 35 or median age within a census tract. We also plan to measure the age structure of mortgage borrowers by tract by constructing this measure using the Federal Reserve / Equifax Consumer Credit Panel just for the subset of mortgage borrowers.

   (b) Education: We measure education as the share of the 25+ adult population with a college degree in a census tract.

   (c) Income: We measure income as the log of median household income. We plan to measure income directly in HMDA, and then aggregate to the census tract level. (Median household income is also available from the American Community Survey). Another closely related measure available from the Consumer Credit Panel is the average FICO score of mortgage borrowers in the Census tract. (Credit scores are highly positively correlated with incomes).

   We expect that younger individuals, more educated individuals, and higher income households are more familiar with technology and therefore face lower costs of adopting technology-based lending. We emphasize that these variables have also been used as proxies for financial sophistication of households. We therefore interpret the variables as broad measures of the level of financial and technological sophistication of households.

2. Access to technology: We use data on computer and Internet usage from the American Community Survey. We note that the data is only available starting in 2013 and only varies at the county-level. As discussed below, we are collecting a separate dataset on

Internet access that is based on the availability of Internet services at a much more disaggregated level. We expect that areas with more Internet usage experience higher FinTech lending growth.

3. Diffusion of technology: Areas with high population density are expected to experience faster diffusion of new technologies because there are more interactions across individuals. We measure population density as the ratio of total population to total area, which we collect directly from Census data. We posit that more densely populated area may have higher FinTech lending growth.

4. Demand for rapid credit provision: Our evidence and proposed analysis discussed earlier suggests that a comparative advantage of FinTech lenders is their ability to process and close mortgage applications more quickly. This will be more valuable in ''hot' markets with rapid housing turnover, and anecdotal evidence suggests that this benefit may explain part of the growth in FinTech lending in such markets.[10]

   We plan to use variation in home price appreciation to identify the strength of local housing markets for this analysis.

5. Bank competition: We measure the availability of bank mortgage financing from traditional sources of bank financing using geographic variation in the location of bank branches. We use ArcGIS to compute the (geographic) midpoint of each census tract and then measure the number of bank branches within a given radius (we currently use a radius of 10 miles but may explore other distances also). We posit that that areas with less competition may experience higher FinTech growth.

[10]For example, a September 2015 The Street article titled 'Online Mortgage Lenders Are Beating Traditional Bank Loans' highlights the shorter closing times of online lenders, and includes the following quote from the CEO of online lender Bank of the Internet "We have very short underwriting term times and that's a plus for our purchase oriented borrowers – we give quick answers," Garrabrants said. "In a really hot market, that's important." Also from the article: Non-traditional, non-bank lenders, such as SoFi as the online lender is commonly called, offer less conventional underwriting for residential mortgages and typically a shorter period to close because of the design of the loans, mortgage experts say. "Your offer is as good as a cash offer in terms of the speed, and that was very important to us as a competitive advantage," Ellis said in terms beating out other buyers in a sellers' market. (see `https://www.thestreet.com/story/13282079/1/online-mortgage-lenders-are-beating-traditional-bank-loans.html`)

We use a standard cross-sectional regression framework to evaluate the effect of predetermined variables on FinTech growth. Specifically, we estimate the following OLS regression:

$$\Delta\text{FinTech Share}_c = \alpha + \beta\text{Technology}_{c,2010} + \gamma\text{Controls}_{c,2010} + \epsilon_c$$

where ΔFinTech Share$_c$ is the change in the market share of FinTech lenders in census tract c from year 2010 to 2015, Technology$_{c,2010}$ is a vector of the predetermined variables discussed above, and Control$_c$ is a vector of control variables. The control variables include observable characteristics that may affect FinTech lending growth directly. Specifically, we include local time trends (e.g., at the state, MSA, and county level), characteristics of the local housing market (e.g., house price levels in 2010) and characteristics of the local mortgage market (e.g., share of jumbo loans in 2010). Depending on availability, we may add additional control variables from Census data and other datasets. The purpose of adding control variables is to understand the robustness of the main variables of interest.

In addition, we may also use loan-level regressions to examine FinTech lending growth during the sample period. The dependent variable in loan-level regression would be an indicator variable equal to one if loan is made by a FinTech lender and zero otherwise. This approach would allow us to control for loan-level observables, and study variation within fine geographic units. To explore the timing of our results, we may also examine year-to-year FinTech growth during our sample period. This approach will be necessary in order to study the 'demand for rapid credit provision' hypothesis described above.

At a minimum, this framework produces stylized facts on the main variables driving demand for FinTech lending. To the extent that the results on age, education, income, Internet speed, and competition are robust to alternative specifications, we believe that it may be reasonable to draw a causal interpretation from these relationships.

C. *Is there a digital divide in FinTech lending?*

There is a widespread concern that the inequalities in access to Internet services is exacerbating underlying wealth and income inequalities across households. One potential channel for such a divide is the access to FinTech lenders. If we observe significant variation in

Internet access, we can use this variation to assess the impact of the digital divide on access to FinTech lenders. On the one hand, we may find no effect of the digital divide on FinTech lenders since the vast majority of U.S. households can already access Internet services at home or elsewhere. On the other hand, we may find that access to high-speed Internet is an important determinant of whether a household uses a FinTech lender. Either way, the result can shed light on the importance of the digital divide in household access to finance.

The main empirical challenge in examining the role of Internet services is the endogeneity of the availability of such services. Internet providers invest in resources where they expect high demand from households, which may also correlate with the likelihood of households in these areas to use FinTech lenders. As discussed above, using Internet access in 2010 alleviates this issue but it may not completely resolve it. As an alternative, we therefore plan to use changes in the availability of Internet services during our sample period. To the extent that changes are driven by logistical or technological consideration, we can evaluate the importance of the digital divide on access to FinTech lenders.

We invested considerable effort in collecting detailed data on Internet access across the U.S. The data covers Internet coverage semi-annually from June 2011 to December 2015 at the census-block and provider-technology level. To get a sense of how we would identify the effect of changes in Internet access, it is best to consider a case study. Below we provide a detailed discussion of the structure, content, and coding of the data.

The case study examines the entry of Google Fiber in Kansas City between 2010 and 2015. Google Fiber is a large-scale initiative by Google to establish a new Internet service provider. The first cities covered by Google are all in the Kansas City area, consisting of Johnson, Leavenworth, and Wyandotte counties in Kansas and Cass, Clay, Jackson, and Platte counties in Missouri. A major factor behind the selection of Kansas City was that households had relatively poor access to high-speed Internet prior to the entry of Google Fiber. Table 6 shows that cable Internet was only available to a small fraction of households over this period (fiber Internet from providers other than Google is likewise limited). Figures 10 and 11 show the availability of Google Fiber as of December 2011 and December 2015, respectively. It is clear from the figure that there was a rapid expansion such that Google Fiber became available broadly across the Kansas City area.

Table 7 provides an overview of the Kansas City coverage by Google Fiber. We find that there was no service in June 2011. In December 2012, Google Service started providing services in 2 census tracts out of 499. The effective population covered was only 0.1%. By June 2014, the service had expanded to 3,191 census blocks and 116 census tracts covering 10.8% of the population. The expansion was considered complete by December 2015. Google Fiber was available in 16,805 census blocks, 359 census tracts, and covered 55.4% of the population of the Kansas City area.

We are in the process of learning more about the determinants of the roll-out of Google Fiber in Kansas City. It is clear from looking at the maps that geographic considerations played an important role in determining the timing of the rollout. Google Fiber effectively started at the center of the city and then expanded outwards (the expansion on the Kansas side was somewhat faster than on the Missouri side). For now, we consider the timing of the staggered roll-out as exogenous relative to the underlying demand for Internet services.

We can use the timing of the staggered rollout (or determinants of timing) to identify the effect of improvements in Internet services on FinTech lending. Specifically, we plan to estimate the following OLS regression:

$$\Delta \text{FinTech Share}_{ct} = \alpha_c + \delta_t + \beta \text{Google Fiber}_{ct} + \gamma \text{Controls}_{c,2010} + \epsilon_c$$

where FinTech Share_{ct} is market share of FinTech lenders in census tract c at time t, Google Fiber_{ct} is the coverage of Google Fiber in census tract c at time t, and Control_c is a vector of control variables, α_c are census tract fixed effects, and δ_t are time fixed effects.

The main coefficient of interest is β, which captures the effect of better Internet service on FinTech lending. As mentioned above, the digital divide would suggest that better access to high-speed Internet improves the take-up of new technologies such as FinTech lending. Alternatively, the improvement in Internet service may not affect FinTech lending if households already have access to sufficiently fast Internet services. We believe that either result is of interest when examining the effect of the digital divide on lending.

We are planning to expend our data coverage to include other cities. As discussed below, we have collected data on three states so far (California, Kansas, and Missouri). We note that many cities do not experience large changes in Internet coverage during our sample period. For comparison, we also examined Los Angeles County, which is the most populous county in the U.S. with a population of about 10 million. Figures 12 and 13 show the availability of the main provider of fiber Internet services, Verizon Fiber, as of December 2011 and December 2015, respectively. There was effectively no change in Verizon Fiber during the sample period (Google Fiber was not effective in California during the sample period except for a small pilot project in the Palo Alto area). As shown in Figures 14 and 15, the result is similar when consider all fiber Internet services in the entire state of California (there is also little change in other high-speed Internet services such as cable). Hence, there is little variation for estimating the effect of changes over time but we can still use the data as a predetermined variable in our cross-sectional analysis.

To summarize, we are planning to use specific service expansion for identification if we can credibly establish that the timing is exogenous to FinTech lending growth. This analysis may focus on a few cities, or even a single city such as Kansas City. We are planning to complement this analysis using variation in 2010 Internet access as one our predetermined variables. We also need to point out that we are still learning about the quality of the underlying data. To the best of our knowledge, we are the first academic paper to use these data and we are still evaluating the data collection process and the proper coding. Appendix A provides additional detail.

VII. FinTech and Mortgage Refinancing

This section examines whether the presence of FinTech lenders has made the mortgage refinancing process more efficient in the aggregate. In a first step, we use the CRISM data to measure monthly refinancing propensities across locations c, and then estimate the following regression:

$$\Delta \text{Refi Propensity}_{ct} = \alpha_t + \beta \text{FinTech share}_{c,t-x} \times \text{AggRefiPropensity}_t + \gamma \text{FinTech share}_{c,t-x} + \epsilon_{ct}$$

where FinTech share$_{c,t-x}$ is the local market share of FinTech lenders measured at an earlier point, and potentially (depending on the findings from the previous section) instrumented for using Internet speed or other variables. The interaction term with $\text{AggRefiPropensity}_t$, the aggregate refinance propensity in the economy, allows to measure whether a stronger presence of FinTech lenders is associated with differential responsiveness, e.g., to rate cuts or other drivers of refinancing, rather than just leading to a level shift in refinancing propensities. Additional controls are as in the previous section and may include location fixed effects in specifications where we use a time-varying lagged measure of FinTech market share. This would allow controlling for the possibility that areas with more FinTech lenders are refinancing differentially simply due to local borrowers' characteristics, rather than the FinTech lenders per se.[11]

An additional important question is whether the presence of FinTech lenders may have affected the propensity of borrowers to make *suboptimal* refinancing decisions, either by not refinancing when they should (errors of omission), or by refinancing when they should not (yet) do so (errors of commission). It is generally difficult to know precisely when any individual borrower should refinance, since this depends on unobservable parameters such as borrowers' discount rate or their mobility expectations. Nevertheless, a number of papers in the existing literature (e.g. Agarwal, Rosen, and Yao 2015, Andersen, Campbell, Nielsen, and Ramadorai 2015, Keys, Pope, and Pope 2016) rely on calibrations of an elegant optimal decision rule devised by Agarwal, Driscoll, and Laibson (2013) to proxy for whether a household should optimally refinance or not. We plan to follow their lead and examine whether a larger presence of FinTech lenders in an area affects local borrowers' propensity for the two errors described above.

It is important to test whether the presence of FinTech lenders has affected refinancing propensities, since this is one channel through which the increased emphasis on technology in lending may have real effects on the economy. Industry evidence indicates that FinTech lenders do indeed exhibit faster prepayment speeds than other lenders (see Goldman Sachs Research 2016), but it is as of yet unclear whether that is simply due to faster-prepaying borrowers selecting into a FinTech loan, without affecting aggregate prepayment speeds. If

[11] A similar analysis could also be conducted at the loan level, where additional control variables could be used (such as an estimate of the current loan-to-value ratio, updated credit scores, etc.).

there is an effect on aggregate refinancing propensities, this could still come not from increased efficiency but rather from some borrowers' refinancing too quickly, which our analysis also tests for.

VIII. Conclusion and Proposed Timeline

To sum up, our proposed research is intended to shed light on how technology is reshaping the mortgage market, by studying how lending by a set of financial technology-based (or "FinTech") lenders at the forefront of innovation differs from lending by more traditional mortgage providers. Since the technologies used by market leaders like Quicken and SoFi are likely to diffuse more broadly through the mortgage lending industry in the years to come, we believe our results will shed light on how mortgage contracting and credit supply in aggregate is likely to evolve in the future. Our results also likely have implications for diffusion of similar technologies in other lending markets.

References

Agarwal, Sumit, John C Driscoll, and David I Laibson, 2013. Optimal mortgage refinancing: A closed-form solution. Journal of Money, Credit and Banking 45, 591–622.

Agarwal, Sumit, Richard J. Rosen, and Vincent Yao, 2015. Why do borrowers make mortgage refinancing mistakes?. Management Science 62, 3494–3509.

American Banker, 2016. What it took to make a fully paperless mortgage. December 19.

Andersen, Steffen, John Y. Campbell, Kasper Meisner Nielsen, and Tarun Ramadorai, 2015. Inattention and inertia in household finance: Evidence from the danish mortgage market. Working Paper 21386 National Bureau of Economic Research.

Badarinza, Cristian, John Y. Campbell, and Tarun Ramadorai, 2016. International comparative household finance. Annual Review of Economics 8, 111 – 144.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra, 2016. Regional heterogeneity and monetary policy. Discussion paper, Federal Reserve Bank of New York Staff Report No. 731.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, 2017. Fintech, regulatory arbitrage, and the rise of shadow banks. Working Paper 23288 National Bureau of Economic Research.

Campbell, John Y, 2006. Household finance. Journal of Finance 61, 1553 – 1604.

——— , 2013. Mortgage market design. Review of Finance 17, 1 – 33.

Di Maggio, Marco, Amir Kermani, and Christopher Palmer, 2016. How quantitative easing works: Evidence on the refinancing channel. Discussion paper, NBER Working Paper No. 22638.

Fuster, Andreas, Stephanie H. Lo, and Paul S. Willen, 2017. The time-varying price of financial intermediation in the mortgage market. Discussion paper, Federal Reserve Bank of New York Staff Report No. 805.

Goldman Sachs Research, 2016. The mortgage analyst: Prepays catching up to quicken? not so quick. http://www.thebondbeat.com/bondbeat/wp-content/uploads/2016/02/gs-TMA.pdf Accessed 2017-03-15.

Goodman, Laurie, 2015. Why rocket mortgage won't start another housing crisis. Urban Institute Urban Wire Blog, February 12. Accessed 2017-03-15.

Keys, Benjamin, Devin Pope, and Jaren Pope, 2016. Failure to refinance. Journal of Financial Economics 122, 482 – 499.

Lee, Donghoon, and Wilbert van der Klaauw, 2010. An introduction to the frbny consumer credit panel. Discussion paper, Federal Reserve Bank of New York Staff Report No. 479.

Mayer, Chris, Tomasz Piskorski, and Alexei Tchistyi, 2013. The inefficiency of refinancing: Why prepayment penalties are good for risky borrowers. Journal of Financial Economics 107, 694 – 714.

Petersen, Mitchell, and Raghuram Rajan, 2002. Does distance still matter? the information revolution in small business lending. The Journal of Finance 57, 2533 – 2570.

Stein, Jeremy, 2002. Information production and capital allocation: Decentralized versus hierarchical firms. The Journal of Finance 57, 1891 – 1921.

Wall Street Journal, 2015. Fha offers olive branch to tentative lenders. Accessed: 2017-03-15.

Wong, Arlene, 2016. Population aging and the transmission of monetary policy to consumption. Discussion paper, Working Paper, Northwestern University.

Figures

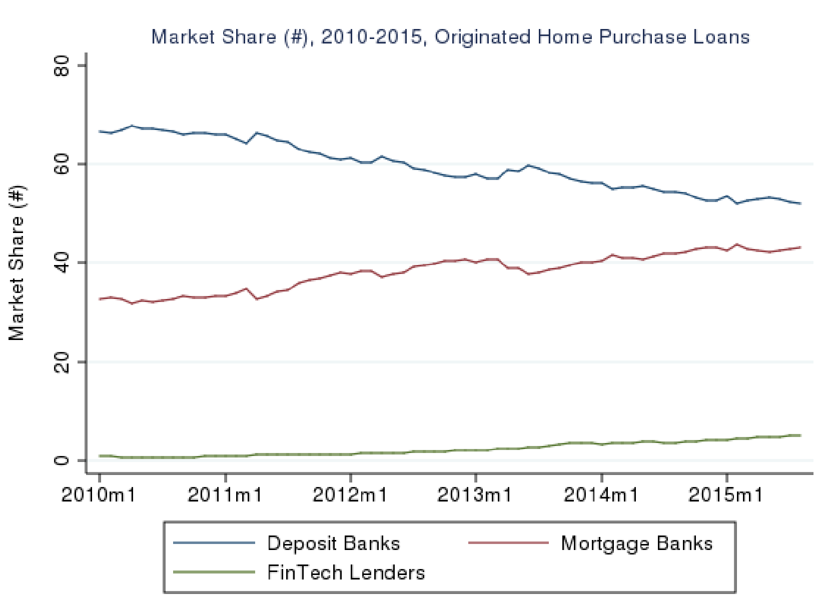

Figure 1: Market share of new mortgage purchases over time

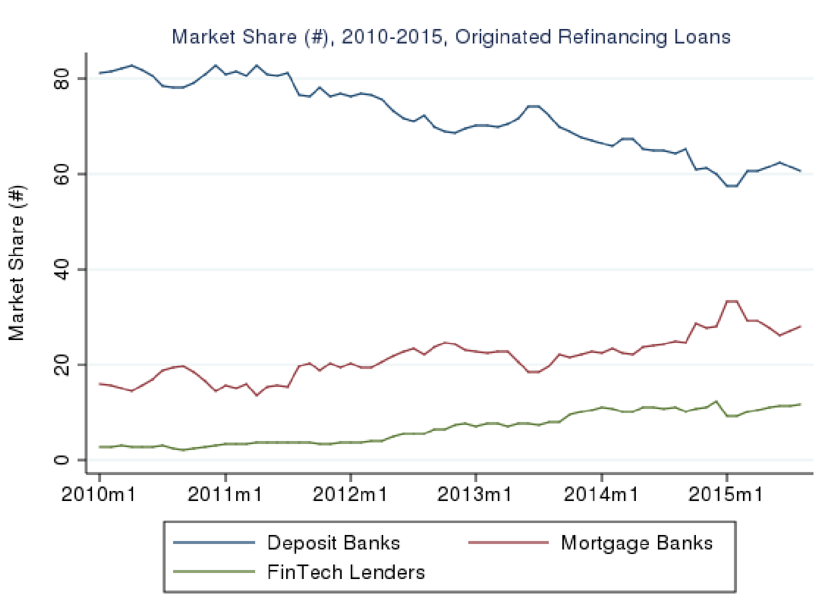

Figure 2: Market share of refinancing mortgages over time

Source: HMDA.

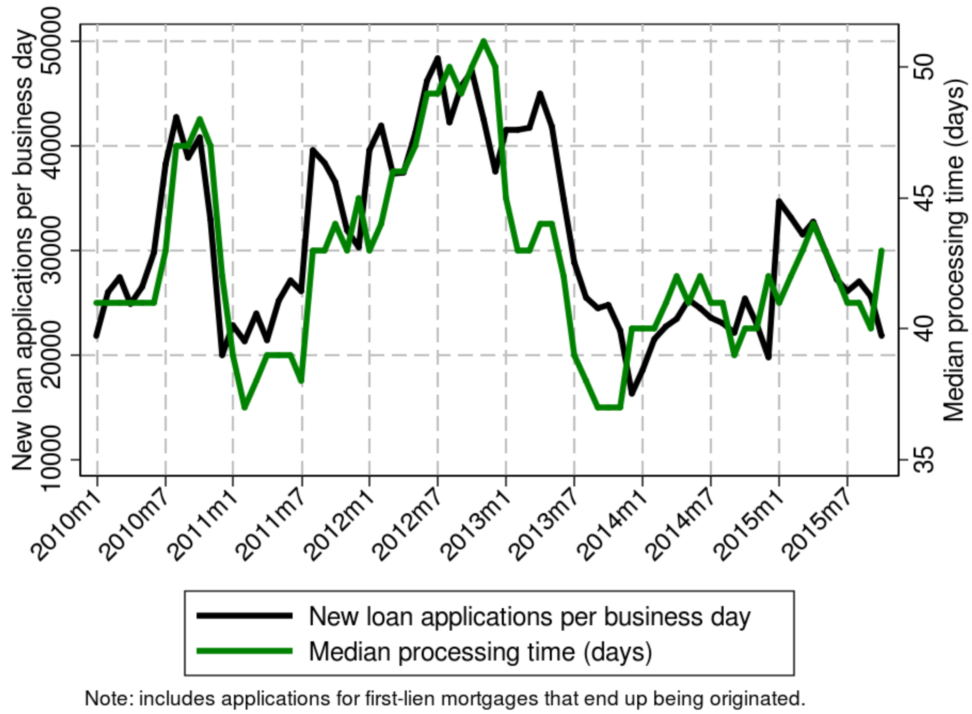

Figure 3: Time series of market-wide loan application volumes and median processing times (source: HMDA)

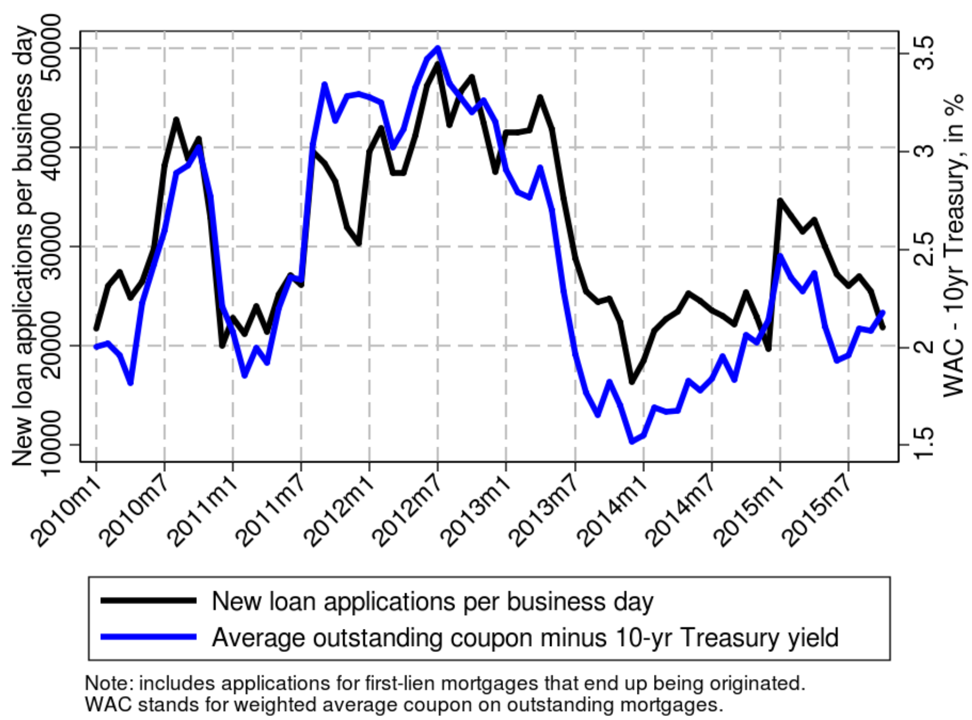

Figure 4: Time series of market-wide loan application volumes and proxy for average refinance incentive (sources: HMDA; Freddie Mac; J.P. Morgan)

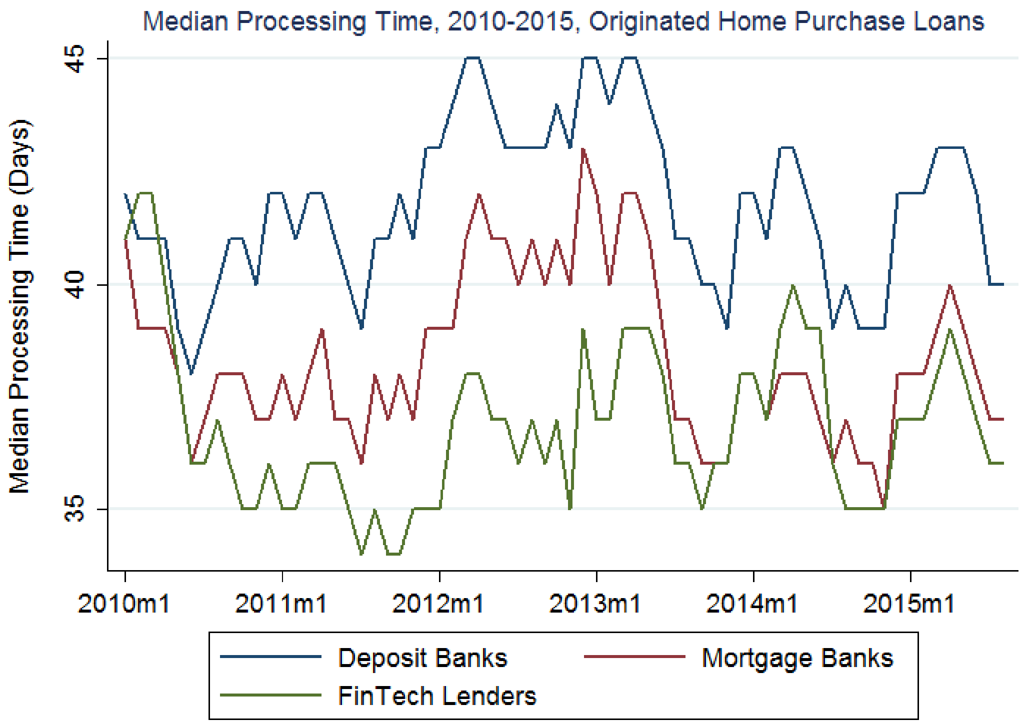

Figure 5: Time series of median processing time by lender type: Home purchase loans

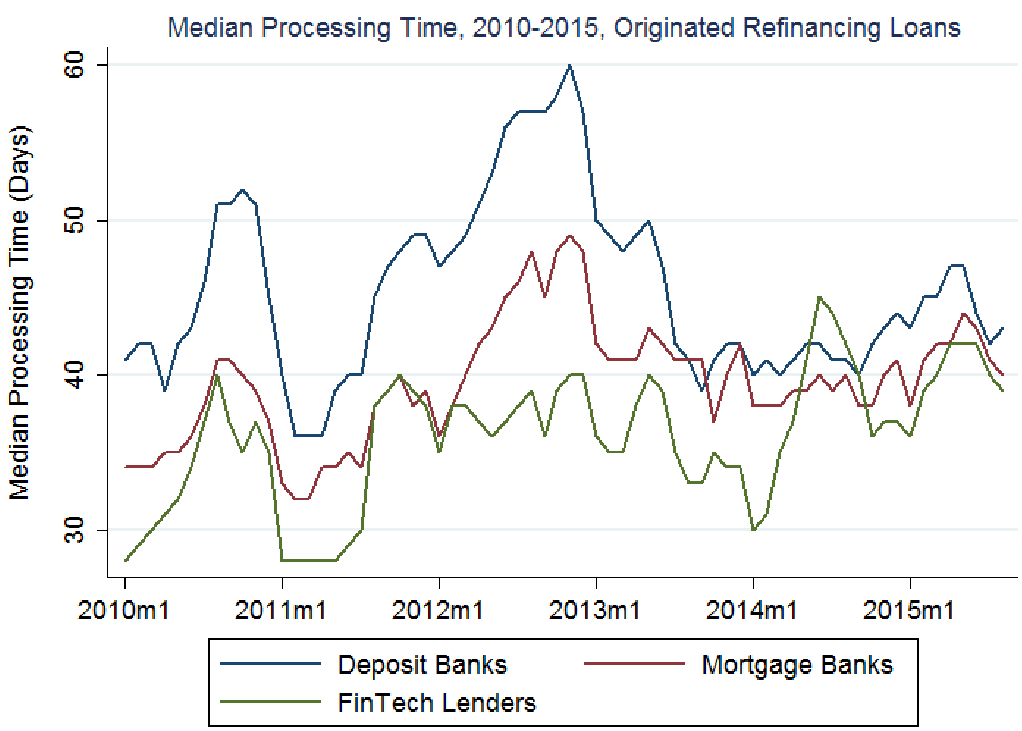

Figure 6: Time series of median processing time by lender type: Refinancings

Source: HMDA.

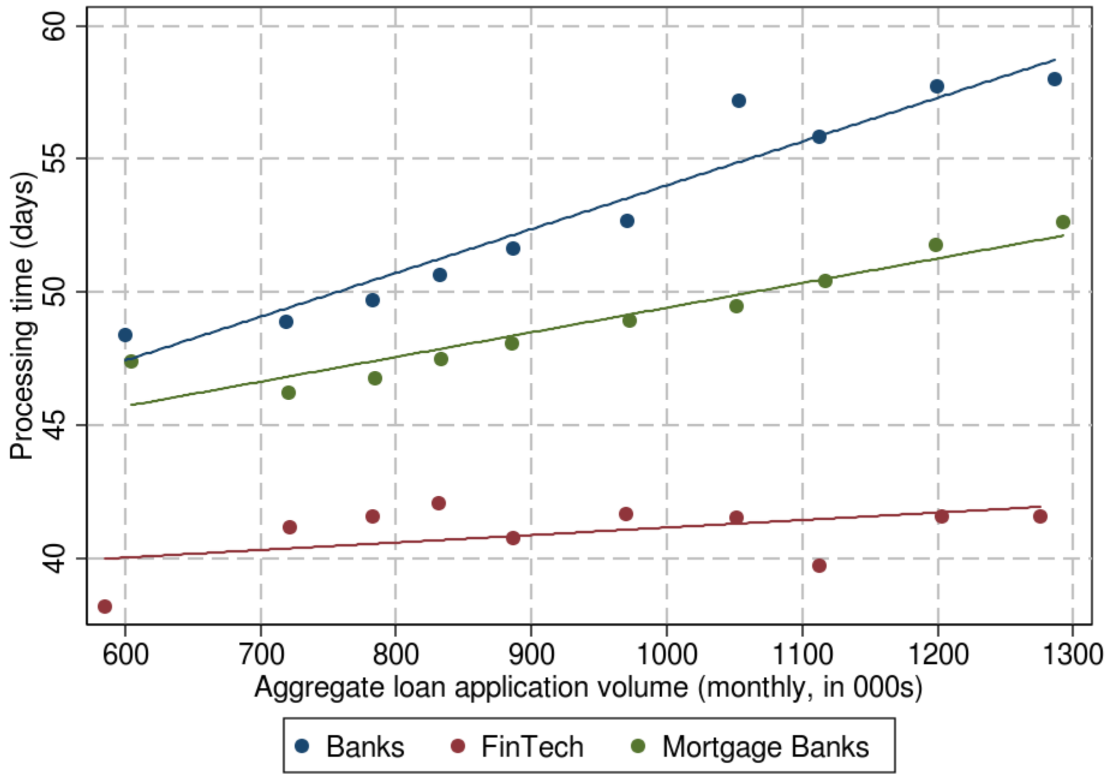

Figure 7: Binned scatterplot of processing times against aggregate level of mortgage applications, by lender type.

Dots represent conditional expectations of processing time for each of 10 deciles of aggregate mortgage application volume, after controlling for county fixed effects, applicant income, loan amount, loan type (FHA/conventional-conforming/jumbo), loan purpose (purchase or refinance), lien status, property type, and applicant race and gender. Data source: HMDA.

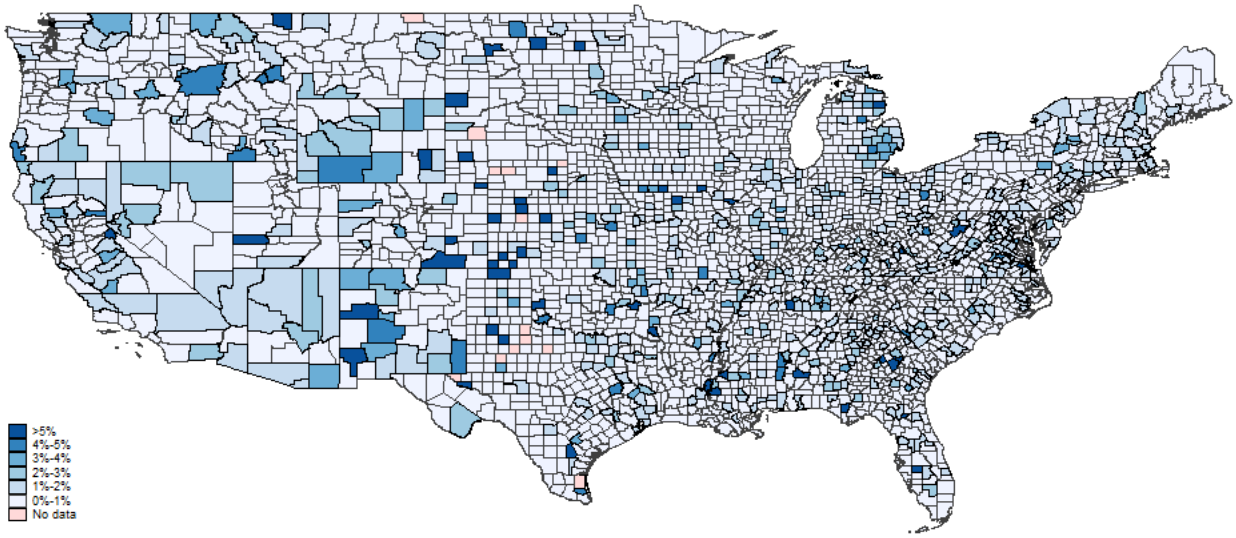

Figure 8: Market Share of FinTech lenders by county in 2010

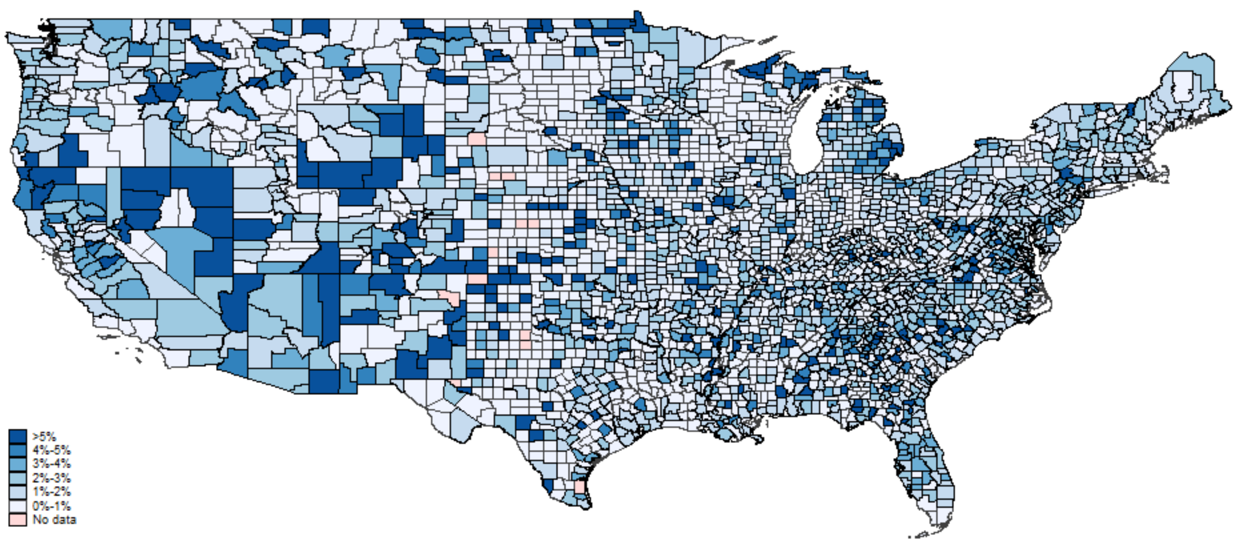

Figure 9: Market Share of FinTech lenders by county in 2015

Market shares are computed as the number of mortgages originated by FinTech lenders relative to the total number of mortgage originations by county and year. Source: HMDA.

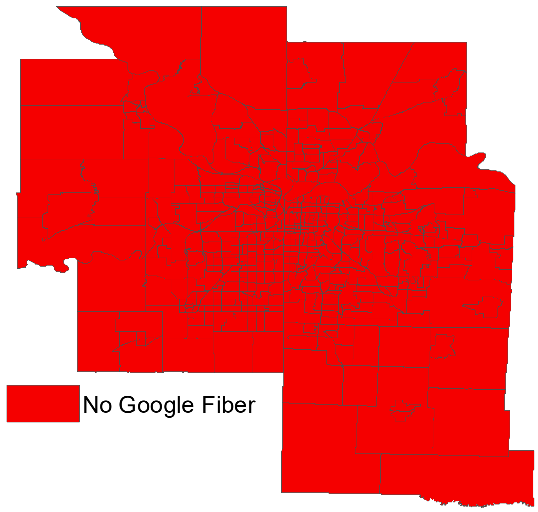

Figure 10: Google Fiber Availability in December 2011

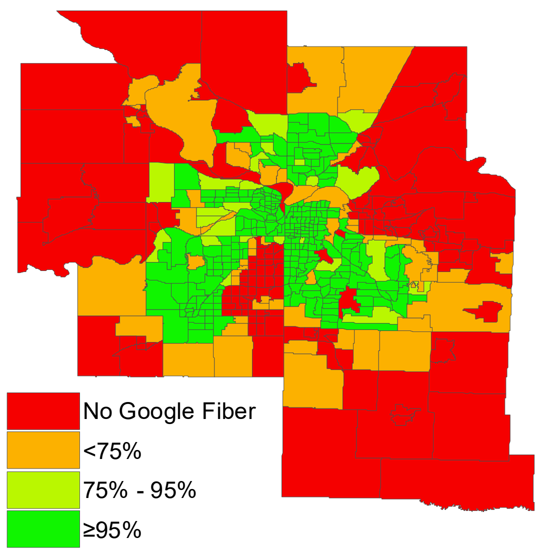

Figure 11: Google Fiber Availability in December 2015

Figure shows the share of the population for each census tract that lives in a census block with Google Fiber in Kansas City. Source: NTIA and FCC data on Internet coverage by census block, provider, and technology in December 2011 and 2015.

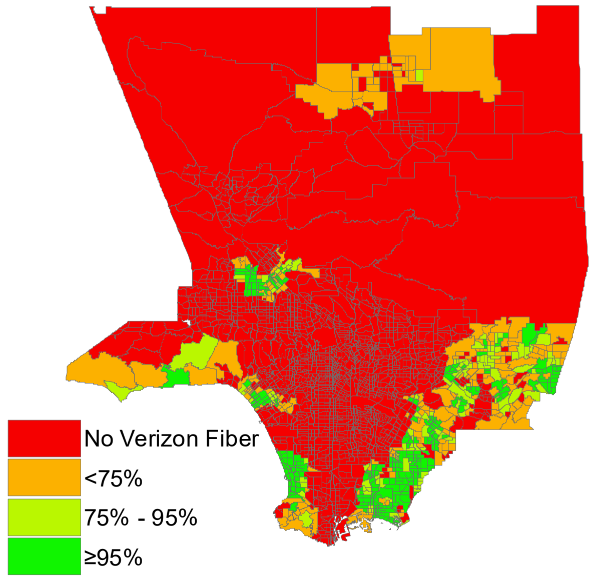

Figure 12: Verizon FIOS Availability in December 2011

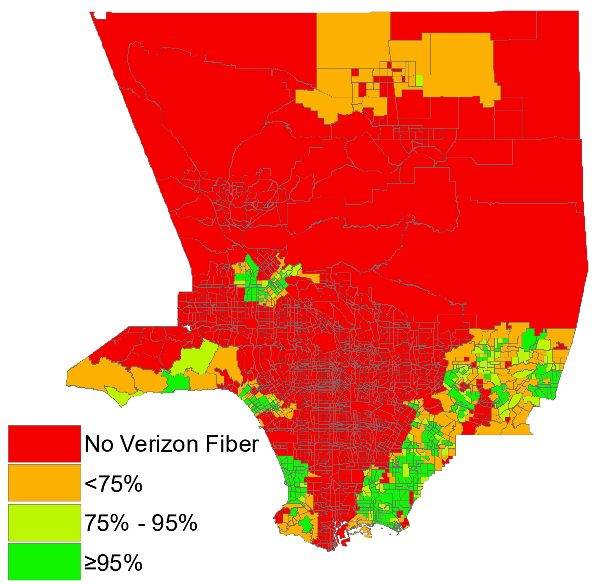

Figure 13: Verizon FIOS Availability in December 2015

Figure shows the share of the population for each census tract that lives in a census block with Verizon FIOS in Los Angeles County. Source: NTIA and FCC on Internet coverage by census block, provider, and technology in December 2011 and 2015.

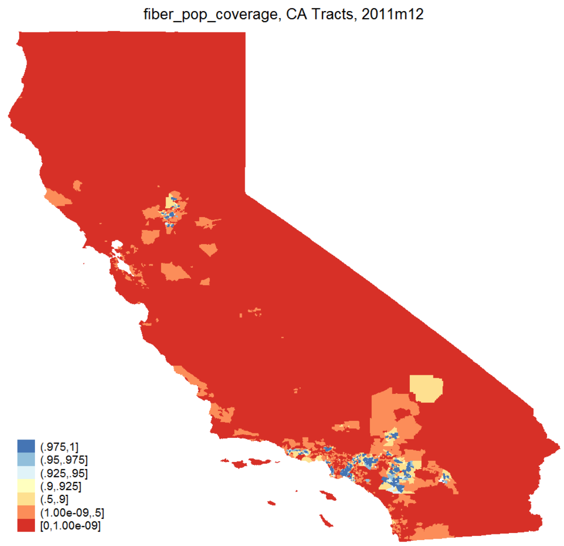

Figure 14: Fiber Availability in December 2011

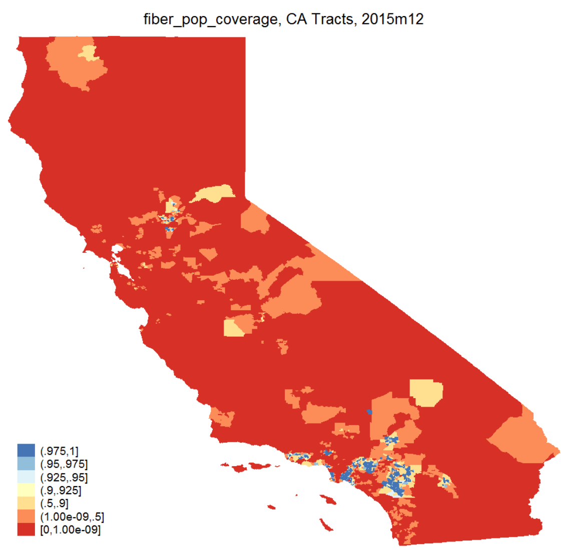

Figure 15: Fiber Availability in December 2015

Figure shows the share of the population for each census tract that lives in a census block with fiber Internet service in California. Source: NTIA and FCC on Internet coverage by census block, provider, and technology in December 2011 and 2015.

Tables

Table 1: Top 25 Mortgage Originators in 2015

Rank	Lender Name	Volume ($mm)	Market Share (%)
1	WELLS FARGO BK NA	109,415	7.28
2	QUICKEN LOANS, INC.	67,993	4.52
3	JPMORGAN CHASE BK NA	56,461	3.75
4	BANK OF AMER NA	47,878	3.18
5	US BK NA	24,490	1.63
6	LOANDEPOT.COM	23,839	1.59
7	FLAGSTAR BK FSB	23,068	1.53
8	CITIBANK NA	21,192	1.41
9	FREEDOM MORTGAGE CORPORATION	19,042	1.27
10	CALIBER HOME LOANS, INC.	16,909	1.12
11	STEARNS LENDING, INC.	14,546	0.97
12	GUARANTEED RATE INC.	12,273	0.82
13	UNITED SHORE FINANCIAL SERVICE	11,907	0.79
14	PRIMELENDING A PLAINSCAPITAL C	11,751	0.78
15	GUILD MORTGAGE COMPANY	11,394	0.76
16	NAVY FCU	11,196	0.74
17	SUNTRUST MTG	10,538	0.70
18	PNC BK NA	10,512	0.70
19	NATIONSTAR MORTGAGE LLC	9,726	0.65
20	FAIRWAY INDP MORTGAGE CORP	9,661	0.64
21	BROKER SOLUTIONS, INC	8,558	0.57
22	BRANCH BKG&TC	8,124	0.54
23	USAA FSB	8,050	0.54
24	PROSPECT MORTGAGE, LLC	7,969	0.53
25	ACADEMY MORTGAGE CORPORATION	7,950	0.53

Constructed using HMDA. Contains all mortgage originations: home purchases and refinancings. FinTech lenders shaded. One additional FinTech lender, Movement Mortgage, had 43bps of market share in 2015.

Table 2: Summary of RHS Variables, Originated Loans Only, Home Purchase and Refinancings

	Deposit Banks			Mortgage Banks			FinTech Lenders		
	Mean	SD	p50	Mean	SD	p50	Mean	SD	p50
Male	0.67	0.47	1	0.69	0.46	1	0.61	0.49	1
Female	0.25	0.44	0	0.27	0.44	0	0.26	0.44	0
No Coapplicant	0.45	0.50	0	0.52	0.50	1	0.49	0.50	0
Owner Occupied	0.88	0.33	1	0.92	0.28	1	0.91	0.28	1
Log(Applicant Income)	4.16	1.33	4.38	4.04	1.34	4.32	4.15	1.24	4.38
LTI	1.99	3.21	1.80	2.49	2.79	2.41	2.39	2.35	2.20
Loan Purpose:									
Home Purchase	0.33	0.47	0	0.53	0.50	1	0.21	0.41	0
Refinancing	0.67	0.47	1	0.47	0.50	0	0.79	0.41	1
Loan Type:									
Conventional	0.85	0.36	1	0.60	0.49	1	0.71	0.45	1
FHA	0.09	0.29	0	0.27	0.45	0	0.20	0.40	0
VA	0.05	0.21	0	0.10	0.30	0	0.09	0.28	0
FSA/RHS	0.01	0.10	0	0.03	0.16	0	0.00	0.06	0
Race:									
White	0.79	0.40	1	0.78	0.41	1	0.71	0.46	1
Black	0.04	0.20	0	0.06	0.23	0	0.05	0.21	0
Asian	0.05	0.22	0	0.07	0.25	0	0.05	0.21	0
Other	0.09	0.29	0	0.09	0.28	0	0.19	0.39	0
Processing Time	52.42	33.68	45	49.05	34.40	39	42.68	25.76	37
Observations	30481086			12781744			2165024		

Table 3: FinTech Loan Performance, FHA defaults

Dependent variable except in column 2 is the percentage difference between the lender's FHA default rate and the aggregate FHA default rate in that state. In column 2 it is the percentage difference between the lender's supplemental performance metric (SPM) and the overall FHA SPM for that state. The SPM measures mortgage default rates relative to a target default rate specified by the FHA for loans within three FICO score bands, <640, 640-680, and >680. Regression weighted by origination volume. Robust standard errors in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% levels respectively.

	(1)	(2)	(3)	(4)
FinTech	-33.0***	-21.3***		
	(5.1)	(5.9)		
Quicken			-57.7***	-58.2***
			(1.8)	(2.0)
Other FinTech			15.9***	13.2*
			(5.5)	(7.5)
Quicken x High Sh				4.2
				(7.2)
Other x High Sh				6.7
				(10.6)
R-squared	0.313	0.121	0.660	0.661
N	199	199	199	199

Table 4: Regression of Processing Time on Lender and Borrower Characteristics: Originated Purchase Loans

	(1)	(2)	(3)	(4)	(5)
FinTech Lender	-6.610**	-7.067**	-6.564**	-6.897**	-5.607**
	(2.801)	(2.790)	(2.828)	(2.839)	(2.743)
LTI		0.846***		0.814***	0.819***
		(0.137)		(0.115)	(0.115)
LTI squared		-0.000353***		-0.000340***	-0.000342***
		(0.0000792)		(0.0000735)	(0.0000737)
Missing Income		17.43***		13.42***	12.34***
		(4.007)		(3.102)	(2.307)
Log(Applicant Income)		2.818***		2.113***	2.000***
		(0.434)		(0.288)	(0.313)
Loan Type: FHA		1.526**		1.512***	1.982*
		(0.682)		(0.578)	(1.056)
Loan Type: VA		2.921**		3.776**	4.109**
		(1.479)		(1.514)	(1.612)
Loan Type: FSA/RHS		9.826***		9.539***	10.05***
		(0.769)		(0.606)	(0.657)
No Coapplicant		-1.568***		-1.779***	-1.754***
		(0.227)		(0.197)	(0.216)
Owner Occupied		5.989***		5.782***	5.817***
		(0.491)		(0.463)	(0.480)
Mortgage Banks					-2.351
					(2.925)
Constant	50.99***	30.16***			
	(1.631)	(1.591)			
Observations	17291648	17291648	17277758	17277758	17277758
Adjusted R^2	0.001	0.014	0.052	0.064	0.065
County-Month FEs	No	No	Yes	Yes	Yes
Demographics	No	Yes	No	Yes	Yes

The dependent variable is mortgage processing time: the time from loan application to closing. FinTech Lender's are a subset of Mortgage Banks that are differentiated by the automation of their mortgage application process. LTI is the reported loan to income of the borrower. Missing income is a dummy that indicates reported income is missing. Demographic controls include indicators for gender and race of the borrower. Standard errors clustered by lender. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Regression of Processing Time on Lender and Borrower Characteristics: Originated Refinancings

	(1)	(2)	(3)	(4)	(5)
FinTech Lender	-9.472***	-10.99***	-9.940***	-11.36***	-5.876**
	(2.911)	(2.929)	(2.659)	(2.851)	(2.801)
LTI		1.102***		0.828***	0.886***
		(0.171)		(0.137)	(0.151)
LTI squared		-0.000264***		-0.000198***	-0.000212***
		(0.0000833)		(0.0000632)	(0.0000689)
Missing Income		19.08***		10.02***	10.47***
		(4.707)		(3.634)	(3.508)
Log(Applicant Income)		5.770***		3.989***	4.124***
		(0.438)		(0.417)	(0.403)
Loan Type: FHA		7.569***		7.882***	10.49***
		(2.007)		(1.918)	(2.085)
Loan Type: VA		4.568*		5.540**	7.434***
		(2.774)		(2.550)	(2.263)
Loan Type: FSA/RHS		15.36***		14.44***	16.76***
		(1.901)		(1.812)	(1.811)
No Coapplicant		0.267		0.133	0.344**
		(0.226)		(0.167)	(0.156)
Owner Occupied		-4.830***		-4.708***	-4.650***
		(0.694)		(0.682)	(0.700)
Mortgage Banks					-7.779***
					(2.528)
Constant	51.70***	26.93***			
	(1.784)	(3.040)			
Observations	28136206	28136206	28126268	28126268	28126268
Adjusted R^2	0.005	0.032	0.080	0.096	0.105
County-Month FEs	No	No	Yes	Yes	Yes
Demographics	No	Yes	No	Yes	Yes

The dependent variable is mortgage processing time: the time from loan application to closing. FinTech Lender's are a subset of Mortgage Banks that are differentiated by the automation of their mortgage application process. LTI is the reported loan to income of the borrower. Missing income is a dummy that indicates reported income is missing. Demographic controls include indicators for gender and race of the borrower. Standard errors clustered by lender. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Cable internet services in Kansas City

	Kansas City Area, Cable			
	June, 2011	December, 2012	June, 2014	December, 2015
Blocks With Service	2,272	839	1,001	5,880
Tracts With Service	189	75	100	457
All Blocks	46,260	46,260	46,260	46,260
All Tracts	499	499	499	499
Share of Blocks With Service	4.9%	1.8%	2.2%	12.7%
Population With Service	1.9%	1.0%	1.1%	9.8%

The table shows the coverage of census blocks and tracts in Kansas City as of June 2011, December 2012, June 2014, and December 2015.

Table 7: Google Fiber in Kansas City

	Kansas City Area, Google Fiber			
	June, 2011	December, 2012	June, 2014	December, 2015
Blocks With Service	0	499	3,191	16,805
Tracts With Service	0	2	116	359
All Blocks	46,260	46,260	46,260	46,260
All Tracts	499	499	499	499
Share of Blocks With Service	0.0%	1.1%	6.9%	36.3%
Population With Service	0.0%	0.1%	10.8%	55.4%

The table shows the coverage of census blocks and tracts in Kansas City as of June 2011, December 2012, June 2014, and December 2015.

A Appendix: Internet data collection

Here is the summary of our work so far and how we are planning to proceed going forward:

1. Geographic coverage: The data is available at the state level and needs to be collected for each state separately. So far, we collected data for three U.S. states (California, Kansas, Missouri). Based on our sample of three states, we note that the structure of the data is similar across states but that there are idiosyncratic differences. We have now developed a protocol that allows us to collect data for other states assuming that the structure is sufficiently similar. We are planning to collect at least the data for the 10 most populous U.S. states, which would cover more than 50% of the U.S. population. Ideally, we will collect the data for the entire U.S.

2. Unit of observation: The data is collected at the census block-technology-provider level. For example, a typical entry would specific census block XYZ, provider Verizon, and technology (fiber, cable, satellite, etc.). We aggregate census block data to census tract data using population weights from the Census. So far, we found that the distribution is bi-modal with most provider-technologies either being active in a census tract or not.

3. Technology: The main technologies reported in both datasets are DSL, cable, and fiber. The FCC data has more disaggregated data by technology and adds some technologies that are not available in the earlier broadband data (wireless, satellite). Our understanding from the literature is that cable and fiber are considered the main technologies for having access to reliable and fast Internet. We therefore use the availability of fiber and cable as our measure of Internet access.

4. Provider coverage: The submission of data to the NTIA was voluntary. The submission of data to the FCC is required by law. We note that the coverage changes over time with more providers reporting to the FCC data. A detailed analysis of the California data shows that the main four providers (AT&T, Time Warner, Comcast, Verizon) as measured by the number of submitted data points, provide data throughout the sample period. Using data in June 2014 (last NTIA data) and December 2014 (first FCC data), we find no abnormal jumps in the data provided by these providers. For now, we therefore focus our definition of access on the four largest providers. Setting aside the issue of consistent reporting, we think this approach is also sensible because it puts the weight on providers that are broadly represented across the state.[12] We

[12] The FCC does not release subscriber data and we take the number of reported data points as a proxy for the number of subscribers.

potentially need to account for M&A activity since ownership of providers and provider-technology can change over time (e.g., Verizon sold FiOS Fiber in California to another company at the end of the sample period). We will need to re-evaluate this issue for each state separately.

5. Internet speed: The NTIA data reports typical and advertised Internet speed for upload and download in categories (e.g., 1-5 MB/sec, 6-50 MB/sec, etc.). Data on typical speed is not well-reported and was therefore disregarded. The FCC data report reports data on advertised Internet speed in numbers. We have tried to construct Internet speed measures over time but Internet speed reporting does not seem consistent even conditional on the technology and company. A possible explanation is that companies seem to have significant leeway on how to report speed and may change their reporting over time. We are therefore not using the speed information when evaluating Internet access but rather rely on the available technology.

6. Time period: The first reporting date is June 2010 and the last reporting date is June 2016. We note that the first reporting data, for June 2010, use a different format and are much less disaggregated than future reporting dates. The data released in December 2010 has the same format as subsequent releases but uses census blocks from the 2000 Census, making it hard to compare to subsequent releases that use census blocks from the 2010 Census. We therefore drop June 2010 and December 2010 from our analysis and use December 2011 as the first reporting data. We further notice that coverage in June 2016 is significantly lower than in previous reporting dates. Our understanding is that the data is still somewhat preliminary and will be updated and completed at the future date. For now, we therefore use December 2015 as our last reporting date. There appears to be consistent reporting from June 2011 to December 2011.