THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Essays in Financial Economics

MARCO PELOSI

A thesis submitted to the Department of Finance of the London School of Economics and Political Science for the degree of *Doctor of Philosophy*

June 2022

To my parents

Declaration of Authorship

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

As a disclaimer, I declare that the views expressed in this thesis are those of the authors and should not be attributed to the Bank of Italy or the Eurosystem.

Statement of conjoint work

I confirm that Chapter 2 is jointly co-authored with Angelo D'Andrea and Enrico Sette. I contributed 33% of the work for Chapter 2.

I confirm that Chapter 3 is jointly co-authored with Francesco Nicolai and Simona Risteska. I contributed 33% of the work for Chapter 3.

I declare that my thesis consists of 35,995 words.

Acknowledgements

I am forever indebted to my advisors Dirk Jenter and Daniel Paravisini for their support and for being a constant source of inspiration. While not every interaction I had with them was followed by joy and enthusiasm, I can safely say I could not possibly have attempted this journey (and made it to the end) without their guidance and encouragement.

I am also thankful to Vicente Cuñat for all the fruitful and insightful discussions we had, often over a brief break from his evening teaching while eating a bad Pret's sandwich, and to Michela Verardo for her continuous and unconditional support and life advice. I would also like to thank Ashwini Agrawal, Mike Burkart, David de Meza, Paula Lopes-Cocco, Martin Oehmke, Ulf Axelson, and many others for the many immensely instructive discussions over my work.

I would have not survived my PhD journey at LSE without the amazing help of Mary Comben and Simon Tuck, as well as the whole staff at the Department of Finance.

Completing this thesis would have been incredibly harder without the help and understanding of Andrea Linarello, Francesca Lotti, Enrico Sette, and all my colleagues at the Bank of Italy, together with Angelo D'Andrea, a coauthor soon to become a colleague.

Every Tuesday, for more than four years, I have been counselling with Giovanni Felice Pace. Those three hours (including a not-always-enjoyable commuting) always made the rest of the week lighter. I am not sure he would agree, but his help has been crucial throughout and beyond this endeavour.

I consider myself very lucky to have met a group of such amazing people at LSE. Zhongchen, Bruce, Fil, Yue, James, Dimitris, Una, Arthur, Su, and Xiang made my long days in the FMG open space (and later at the George's) brighter. Agnese, Jesús, Lukas, Lorenzo, and Bernardo have been the company of (way too) many coffee breaks and are to blame for the excess of caffeine in my veins. A special mention goes to Simona (Simonce) and Francesco for their friendship, for putting

up with me as a coauthor, and for the many barbecues and fun times in Winchester St.

On this side of the Channel (and to some extent, of the world), Rome has welcomed me with warmer weather, cheaper food, and a group of new and old friends who made the move much easier. Thanks to my new friends Massimo, Lavinia and Antonio (and Bart) for the many evenings spent eating, all the while educating me on how to navigate Rome and Romans; to Stefano (Pipa), for a long-lasting friendship and for offering many silent comedy shows at his expense; to Samuel for his continuous support, understanding and a shared passion for mountains.

I am also grateful to my friends in my hometown for all the good times during holidays and short weekends. Thanks to Manu, for being one of my oldest friends and still treating me as he always had; to Vincenzo, Inve, Matteo, and Enrico (aka *Numeri Uno*, although none of us really believe that) for all the silly jokes that can still make me laugh; to Roberta, for her optimism.

I suspect the submission of this thesis would have been quicker if Guidogiorgio (GG) had not added me to a chat group in December 2015 with a few friends and a bunch of strangers. Six years later, I have to thank him and Ale (Pres), Decio, Niccolò (Niccam), Ricca (Internet Explorer), Stefan, Suri, Tondi, Tri, and Vincenzo (Vincy) for all the laughs (a few memorable ones), all the "POPOLO" and "PHD BELLISSIMO" moments, and for teaching me how to spot the strongest trolls.

London has been home for five wonderful years. If I still cannot accept to have left it, the blame goes to all the friends I shared most of my time (typically over a beer or a meal) with. In particular, I am grateful to Viola, for the friendship, for letting me meet Massimo, and without whose help I would have never passed Micro during my first year of PhD; to Lucio, for all the cooking (mostly baking) sessions, and a lot of good laughs; to Alice and Iacopo, for the love of mountains, the many barbecues at 252 Crondall St, and for bringing a bit of London to Rome; to Alessandro, for all the conversations about life and having the accurate Boris citation for all circumstances; to Agnese, for always (literally always) being up for a pint, a climbing session, or a never-ending meal.[1]

[1] Agnese would also be late at all these things.

A special mention is for my flatmates, colleagues, and friends Fabrizio and Alberto. I met Fabrizio during my first year in London, and since then I have tried to take his example of work-life balance. I owe him wise life advice, many vibrant discussions, and a great deal of fun in and out of our place.[2] For the last twelve years, since the first year of our Bachelor's, I had Alberto to rely on. He is one of the smartest, kindest, and funniest person I know, and my best friend. I shared with Fabrizio and Alberto the *Corpellosi*'s flat in 46 Borough Road, which not only has changed my life in so many ways, but felt like home for three incredible years.

My deepest gratitude goes to my family for all the love and understanding in pursuing this project. Thanks to Mimmo (a Colonnella) for sending many *pacchi da giù*, for sharing many of the funniest stories about my father, and for doing his best to step in for him in the last seven years; to my uncle Franco and my aunt Giusy, for setting the table every time I visit, and for always being there for my mother; to my grandmothers Licia and Sisina, who helped raising me, but cannot see me now. To Nunzia, for the shared love of good food, for showing me that things are often (she would say always) less complicated than I imagine, for constantly reminding me I am not a dog (despite all the barking), and for being the better version of the best life partner I could possibly imagine. To my mother, Ornella, for always being my biggest fan, for her endless tolerance, loving encouragement, and for raising me as the person I am today (often at her expense). To my father, Giampiero. Grumpy at times, he was a fun man and a loving father, and I miss him every single day. I cannot share this achievement with him, but I am sure he would be proud of me more than anyone else.

[2] Next flatmates beware, he cannot stand the sight of anchovies and mushrooms.

Abstract

This dissertation consists of three chapters. In the first, I document advantageous selection in loan amount in the largest peer-to-peer (P2P) lender in the United States. By exploiting a natural experiment within the platform, I show that borrowers who select larger loans are less likely to default. This selection is driven by households who live in states with bankruptcy-friendly laws, where borrowers' default costs are lower. Standard models where borrowers maximize their utility cannot rationalize my results and make the opposite prediction. In a simple model of household borrowing, I show that my results can be explained by the fact that borrowers facing higher loan prices search more intensively for cheaper loans. This effect is stronger for the safest borrowers, as they enjoy the greatest benefits from the switching.

In the second chapter, co-authored with Angelo D'Andrea and Enrico Sette, studies the effect of access to broadband internet on bank credit supply to non-financial firms. We find that banks with branches in municipalities reached by fast internet increase loan supply, both at the extensive and the intensive margin. We document that the expansion of credit goes through two main channels: internal efficiency and competition. To increase lending, fast internet also leads banks to expand their geographical markets and to make internal credit reallocation. Finally, while broadband connection moves credit away from smaller municipalities, it still benefits local economic growth as firms obtain more credit from branches located in larger municipalities.

The third chapter, co-authored with Francesco Nicolai and Simona Risteska, provides evidence of the disparity in the incidence of property taxes levied at different points in time. Housing demand is significantly less elastic with respect to taxes deferred to the future relative to taxes levied at the moment of the purchase. We attribute this difference to the lack of salience of future taxes at the moment

of purchase. We provide directions on the optimal tax mix between salient and nonsalient taxes with the help of a model.

Contents

List of Tables

List of Figures

1. Advantageous Selection in Fintech Loans

MARCO PELOSI[1]

1.1 Introduction

The aftermath of the 2007-08 financial crisis saw the widespread birth and fast rise of peer-to-peer (P2P) and marketplace lenders. As of 2018, fintech loans in the United States accounted for 38% of outstanding balance of unsecured personal loans.[2] These platforms allow demand and supply of loans to meet directly, cutting off traditional banks from their role as intermediaries.

Whether and when online marketplaces will disrupt conventional lenders' business model has been an open question since the emergence of credit marketplaces in the early 2000s. Both anecdotal evidence and reports argue that these lenders are attractive as they charge lower interest rates than traditional banks (Adams, 2018). Such discount is possibly due to the near absence of physical and monitoring costs.

Indeed, cheaper loans are probably one of the main drivers of the surging popularity of these platforms. For example, about 85% of loans issued through Lending Club (henceforth, LC), the largest online marketplace lender in the United States, are used to refinance other debt. Despite this popularity, we still know little about individuals who switch from traditional credit markets and borrow through P2P platforms. In particular, understanding the riskiness of people turning to online lenders is crucial to learn how these compete with conventional banks.

Adverse selection is often assumed to be pervasive in credit markets, as shown theoretically by Jaffee and Russell (1976) and Stiglitz and Weiss (1981) and docu-

[1] I benefited from helpful comments by Dirk Jenter, Daniel Paravisini, Mike Burkart, Fabrizio Core, Vicente Cuñat, David De-Meza, Zhongchen Hu, Francesco Nicolai, Martin Oehmke, Alberto Pellicioli, Nunzia Saggiomo, Enrico Sette, Lorenzo Trimarchi, and seminar participants at LSE for comments and suggestions. The opinions expressed do not necessarily reflect those of the Bank of Italy, the Eurosystem or their staffs. All errors are my own.

[2] https://newsroom.transunion.com/fintechs-continue-to-drive-personal-loans-to-record-levels/

mented by a large body of empirical work.[3] Indeed, there is evidence that banks set cyclical credit limits to reduce excessive borrowing from high-risk individuals (Jambulapati and Stavins, 2014). This phenomenon is likely to be exacerbated in markets with a debtor-friendly bankruptcy code, where it is harder for lenders to seize borrowers' assets. As a result, lenders in such markets will have to charge higher interest rates to break even, and high interest rates discourage borrowing by the safest borrowers.

Despite this evidence, in this paper, I show that advantageous selection in loan amount may occur when a low-cost borrowing competitor enters a market characterized by adverse selection. More specifically, I document that safe borrowers take large loans in markets with better borrowers' assets protection by bankruptcy law. As this finding contrasts with the above theoretical predictions, I present a model where advantageous selection arises because of the existing credit equilibrium conditions. Consumers that pay higher interest rates on their existing loans put more effort into finding a cheaper lender. However, this is optimal only for safer borrowers, as those who are more likely to default do not want to bear the search and switching costs.

Measuring selection on loan size is empirically challenging. Using the universe of LC loans, I exploit the roll-out of larger loans by the platform as a quasi-experiment to disentangle such selection. Analyzing the repayment of borrowers who choose loans of different size might seem sufficient to measure selection on loan amount. Yet, Karlan and Zinman (2009) and Hertzberg et al. (2018) argue that to isolate empirically selection from the causal effect of contract terms, the lender (or the econometrician) must compare the loan performance of different groups of borrowers who choose an identical contract. Therefore, this especially requires comparing the behavior of borrowers choosing loans of the same size. LC loan menus consist of bundles of price, maturity, and amount. Thus, keeping the first two fixed, comparing the behavior of borrowers who sort into the same size bucket at different times allows me to isolate the effect of amount selection from that of other contract characteristics.

[3] For a literature review in the context of household finance see Zinman (2015)

I apply this approach using the loan limit increase announced by LC in March 2016, from $35,000 to $40,000. I measure selection by comparing the default probability of borrowers choosing $35,000 before and after the availability of $40,000 loans (i.e., before and after the borrowers were selected on loan amount). Taking the difference between borrowers who select identical contracts allows me to rule out that different contractual terms (e.g., installments) cause differences in repayment behavior. If borrowers' composition was not time-varying, I would be able to isolate borrowers selection from this simple difference. Yet, it is hard to argue that this is the case, for example loans take-up could be seasonal.

Therefore, to account for changes over time in the composition of LC borrowers I estimate a difference-in-differences exploiting the increase in the loan limit from $35,000 to $40,000 in March 2016, using loan amounts between $20,000 and just below $35,000 as counterfactual. In a nutshell, this test compares the default probability of borrowers choosing $35,000 before and after new and larger loans become available, relative to the change in default rate for loans just below $35,000.

I find that selected borrowers are 1.9 percentage points more likely to default than loans issued before the limit extension. This difference implies that borrowers choosing loans larger than $35,000 after March 2016 are safer, hence the advantageous selection. It is worth mentioning that since my sample does not include such loans - as I do not have a counterfactual for larger loans - my estimate sets a lower bound for the selection effect.

To study selection heterogeneity with respect to the prevailing interest rate in traditional credit markets, I then focus on homestead exemption laws. These legislations represent a form of insurance for borrowers who file for bankruptcy under Chapter 7, and protect the equity in their primary residence up to the level of exemption (Auclert et al., 2019; Indarte, 2019). Because these limits are highly heterogeneous across states (e.g., the protection is unlimited in Florida and none in Pennsylvania) and levels hardly change over time (Hynes et al., 2004), they provide a credible proxy for interest rates charged by banks. To this end, Severino and Brown (2017) show that since banks bear higher costs when a borrower defaults in states with greater protection, they also charge higher prices. However, this is not the case for LC, as I show that the platform does not price the state of residency.

Thus, I can analyze selection heterogeneity by running the same regression outlined above in samples with increasing levels of exemptions. I do not find any evidence of selection in states with exemption level below the median. Instead, the whole differential change in default probability comes from borrowers who live in states with higher than median exemptions, where traditional banks are likely to charge higher interest rates. In these states selected borrowers are 3 percentage points more likely to default than the control group. These results are robust to the sample considered. In another test I divide states into four groups, and I still observe advantageous selection in the two groups with higher protection.

This evidence is counter-intuitive. Standard models of households borrowing predict that when the default cost is low, as in states with high (or even unlimited) homestead exemptions, selection on loan amount is adverse (de Meza et al., 2019). Intuitively, if borrowers do not have to pay anything when they default (for example, because their homes are not seized), the riskiest type finds it optimal to borrow as much as possible. Yet, the above findings contrast this apparently simple prediction, as risky borrowers in LC tend to ask for smaller loans precisely where default costs are low.

To rationalize these results, in the last part of the paper I present a simple model where households search for cheaper loans, and borrow to repay external debt and minimize future cash flows. First, I show that borrowers switch to online lenders as long as their interest rate is lower than traditional banks' one, accounting for switching fees. Then, I prove that the heterogeneity across states arises because of search and switching costs, and credit conditions in conventional lending markets. In fact, paying higher interest rates on outstanding loans leads borrowers living in states with higher homestead exemptions to exert more effort in searching for cheaper lenders. On top of that, those who have the greatest incentives to look for lower prices are the safest individuals, as they enjoy the greatest benefits from switching.

I present many tests that corroborate the predictions of the model. For example, I show that the existing debt of LC borrowers is significantly lower in states with higher than median exemption, all else equal. While I cannot argue that this difference is causal, it hints that the alternative interest rate these households face is higher. Most importantly, I show that households' private information lies in their default probability. LC data allows me to observe the time series of the FICO score for borrowers with active loans. I show that, *ceteris paribus*, borrowers choosing $35,000 loans in high exemption states after March 2016 face a bigger drop in their credit score than loans in the control group, hinting that worse income shocks hit them.

To date, there is little evidence on selection in online credit marketplaces. In contrast with my results, Hertzberg et al. (2018) find adverse selection in longer-term contracts using a similar menu extension.[4] This finding may lead to the conclusion that these markets ultimately do not constitute creative destruction. Instead, they could be another threat to financial stability. The evidence I present sheds new light on the mechanism underlying the move from traditional credit markets of safe and reliable borrowers. My results are also relevant from an investor's perspective, as they shed light on a potential source of alpha in consumer loan portfolio investments.

The critical identification assumption to isolating selection from the menu extension is that any change in loan demand for borrowers with the same characteristics does not differentially influence households borrowing $35,000 and other amounts at the time of introduction of new menus (for example, due to shocks to economic conditions or to credit supply of other lenders). If this was the case, any difference in default probability would be due to shifts in the composition of borrowers caused by reasons other than selection.

I back this assumption by showing that the switch from $35,000 occurs exactly at the time of the introduction of new loans. Moreover, there is no evidence of changes in loan demand around March 2016 that could specifically affect borrowers at $35,000 compared to others. Finally, I include a rich set of borrowers' characteristics and contract fixed effects in all regressions to compare observationally equivalent loans. These controls include the 4-points FICO score range, debt-to-

[4] They use the introduction of 60-month maturity for $10,000-16,000 loans

income ratio, state, loan amount bucket, and maturity. In most specifications, I also control for borrowers' default probability. To predict it, I estimate a random forest algorithm that uses all relevant information on borrowers and contracts.

Related Literature

This paper mainly contributes to two strands of literature. First, it is related to the growing literature on online unsecured consumer credit. Tang (2019) also uses LC data to pose the question of whether P2P lenders are substitutes or complements to traditional banks. With the same data, Hertzberg et al. (2018) study how lenders can screen borrowers using maturity. Di Maggio and Yao (2020) use both online and credit bureau data to show that Fintech borrowers gain market share by lending to risky borrowers first, and then turns to safer borrowers. Cespedes (2019) analyzes interest rate sensitivity of LC borrowers and suboptimal behaviors. From investors' perspective, Vallee and Zeng (2019) argue that P2P platforms maximize loan volume by trying to limit information asymmetries between sophisticated and unsophisticated investors. Marketplace lenders have also attracted attention from policy makers, as the work by Adams et al. (2017) on US consumer awareness indicates. This work brings novel evidence on the intersection between traditional credit markets and online platforms. While many of the previous studies focus on selection on observable characteristics, I analyze borrowers' loan selection based on their unobserved default risk. Finally, I also show that online lenders are able to attract the best borrowers where banks' ex-ante losses given default are higher. To this extent, it is possible to conclude that new platforms constitute a complement to traditional banks.

Finally, this paper also contributes to a large literature on credit rationing and selection. In their pioneering study, Jaffee and Russell (1976) and Stiglitz and Weiss (1981) are the first to show how adverse selection leads to credit rationing. Since then, a large literature has documented the impact of asymmetric information on credit markets. Among these, De Meza and Webb (1987) show that asymmetric information can instead lead to overinvestment. Moreover, their work on advantageous selection in insurance market (De Meza and Webb, 2001) is seminal, as they show that selection can reverse once risk aversion is introduced. Einav and Finkelstein (2011) provide a great simple framework to think intuitively about both adverse and advantageous selection in insurance market. While there a rich body of empirical work on adverse selection (Ausubel, 1991; Dobbie and Skiba, 2013; Edelberg, 2003, 2004; Dobbie and Skiba, 2013; Stroebel, 2016), very few papers detect advantageous selection in credit markets. With a structural model of loan demand Einav et al. (2012) find advantageous selection into large loans in the auto loans market. Yet, their findings are specific to secured loans. While this literature is rich, my results not only constitute the first evidence of advantageous selection in unsecured credit markets. They also show it is possible that the selection created by incentives to look for better credit conditions prevails that induced by moral hazard (for example, because of lower default costs).

The rest of the paper proceeds as follows. Section 2 describes the LC platform and data, and homestead exemptions. Section 3 first delineates the main empirical strategy to measure selection, then it shows main empirical results. Section 4 studies selection heterogeneity based on homestead exemptions. Section 5 presents the theoretical model and tests its predictions. Section 6 concludes.

1.2 Setting

1.2.1 Lending Club

Lending Club is the largest peer-to-peer lender in the United States. Since its registration with the Security and Exchange Commission in 2008, it has funded over 4 million loans for a total value of $53.7 billion. As a comparison, the second biggest P2P lender (Prosper) helped just below 1 million people borrowing $16 billion. Households in all States (except Iowa) can apply for a loan, while all US residents (except for Ohio's) can invest either in the primary or the secondary market.

To apply for a loan, a prospective borrower has to insert the wished amount and the loan purpose in LC's homepage, as shown in Figure A.1. She then gives sufficient personal information (including the Social Security Number) to allow LC to pull credit information from credit score agencies. LC applies a few eligibility criteria to publish the loan on the platform, including a FICO score larger than 660, satisfactory debt-to-income ratio, a credit history of at least 36 months, and a limited number of inquiries in the most recent months.[5]

Once a person is eligible for a loan, LC assigns her one of the 25 possible initial scores (from A1 to E5) according to a proprietary algorithm. After that, she observes a menu of possible loan amounts with two different maturities: 36 and 60 months. The amount-maturity bundle chosen by the borrower puts her in one of the 35 final rating scores (A1 to G5). To each of these final scores corresponds a unique interest rate. Except for the smallest loans, the final price is a non-decreasing function of amount and maturity. Finally, the loan is published on the platform, where both peer households or big investors can participate in the loan funding. During the sample period I use for this study, LC claimed that over 99% of loans are funded. Indeed, in my data, all but two loans are fully funded. For this reason, the whole analysis focuses on loan demand, thus ignoring the supply side of funds.

Most of LC earnings come from two sources. First, an origination fee is paid by borrowers upfront. This means that once the loan is funded, applicants receive the desired amount net of the fee. According to the platform's website, such fee ranges between 1-5% of the loan amount, depending on the above final rating. Second, a service fee paid by investors who receive monthly payments from borrowers, of approximately 1%. On top of these two standard fees, investors pay between 30-40% of any amount collected when a borrower stops repaying and the loan becomes *charged off*. In this case, LC "makes a reasonable effort to recover the money owed to investors".[6] In my sample, LC raises on average 11% of the residual debt.

1.2.2 The Increase in Maximum Loan

Before March 2016, it was possible to borrow any amount between \$1,000 and \$35,000. In March 2016, without any previous announcement, LC increased the

[5] http://www.snl.com/Cache/c33047201.html
[6] https://www.lendingclub.com/investing/investor-education/interest-rates-and-fees

maximum amount lent to $40,000 and published a blog post on its website. Borrowers can ask for any amount in $25 intervals. It is clear from Figure A.2 that the share of $35,000 loans issued goes down in response to the increase in the loan limit. While the share peaks at 16% of issued loans in February 2016, it drops to just above 11% by the end of the sample period.

1.2.3 Bankruptcy in the United States

Bankruptcy helps borrowers discharge or make a plan to repay their debts. Individuals in the United States can choose to file under three chapters of the Bankruptcy Code: Chapter 7, Chapter 13, and Chapter 11. The first is the most bankruptcy-friendly type, as it allows filers to keep some exempt property. Exemptions vary state by state and can include real estate, various sources of income (e.g., wages, pensions), and insurance. Households who file for bankruptcy under Chapter 13 must repay their lenders at least the amount they would reimburse under Chapter 7. To do so, they must propose a repayment plan that can last from three to five years. Compared to Chapter 7, this plan allows borrowers to stop foreclosure proceedings and to keep their properties. Since in bankruptcy borrowers tend to have a low level of non-exempt properties, 61 percent of filers opted for Chapter 7 in 2016. In 2016 alone, 458 thousand households filed for bankruptcy under Chapter 7, discharging a total of $138 billion.[7]

The Homestead Exemption

Of all debt relief tools, the homestead exemption is the most important, as it allows borrowers to keep some of their equity in their primary residence. Also, it induces the greatest cross-state variation in potential debt relief Auclert et al. (2019). Such exemption is non-existent in New Jersey and Pennsylvania and unlimited in eight states, including Texas and Florida. Table A.1 lists homestead exemptions for all States. When a house owner files for bankruptcy under Chapter 7, her seizable

[7] https://www.uscourts.gov/statistics-reports/bapcpa-report-2016

assets can be determined according to the following rule:

$$Seizable\ Assets = Max\{Home\ Equity - Homestead\ Exemption, 0\}$$

Goodman (1993) and Skeel (2001) argue that states set the initial levels of exemptions in the nineteenth century for reasons that are likely to be unrelated to their current state of the credit market and personal bankruptcy. To support this view, Figure A.3 maps the different levels of exemptions across all States. There is neither a geographical nor an economic pattern in the variation of debt relief generosity. Also, Auclert et al. (2019) and Hynes et al. (2004) note that the state variation has been stable over time and that cross-state differences in exemption levels are persistent. States mainly update their exemption levels based on inflation.

1.2.4 Data

The two main datasets I use come from LC's website. The first contains information on newly issued loans. Crucially for this study, each borrower LC publishes the loan repayment status, terms - such as the loan amount, price, and maturity - and some personal information. Such information includes the location (at state and three-digit ZIP code levels), the employment status and tenure, the annual salary, homeownership, plus a rich set of variables from the credit pull. Not only can I observe the FICO score range, the debt-to-income ratio, but also the debt outstanding on various lines of credit (e.g. revolving loans, installment debt), the number of credit inquiries, and the number of delinquencies in the previous two years.

The second dataset consists of the whole history of payments for each loan. Because of the focus on charged-off loans, this dataset allows me to observe the month of the default, the outstanding principal at the time of default, and the post-charge off recovery.

I conduct the analysis using data downloaded in April 2019. I restrict the sample to loans issued between October 2015 and July 2016. This time interval contains the date at which LC increased the maximum loan amount (March 2016), but also limits the length of the sample period to address the concerns raised by Bertrand et al. (2004). Also, I only consider loans larger or equal to $20,000 to make a meaningful comparison between loans of similar size. The final sample is made of 119,747 loans.

Figure A.4 plots the distributions of loan amounts before and after the increase in the loan limit. There are two interesting patterns. First, borrowers tend to choose loans in round numbers, despite Cespedes (2019) shows that is a sub-optimal behavior because of the excessive interest paid. Second, people bunch strongly at every multiple of $5,000. This effect is particularly evident at $35,000, where LC set the initial loan limit. While such bunching persists after March 2016, it is clear that there is a significant pool of borrowers applying for larger loans.

Table A.2 presents the summary statistics of loan and borrowers characteristics of the 63,330 loans issued before March 2016. Panel A shows the characteristics of loans. When borrowers have only loans up to $35,000 available, the average loan amount is $26,143. The average APR is 12.96%, with a corresponding installment of $740. At the take-up stage, borrowers self-report the purpose of the loan. Table A.3 shows that in my sample, 86% of borrowers use their funds to refinance another debt, either in a credit card or other lines of credit. Only a smaller portion takes a loan to make home improvements or to finance small businesses. Half of the sample borrows at 60 months, which is the longest maturity available. Finally, roughly 18% of loans in this period are in default - defined as being delinquent for more than 120 days.

In Panel B, I present statistics about borrowers' characteristics. LC only admits prime borrowers to its platform. Indeed, the average FICO score is 700, and the minimum score in the sample is 660. A share 19.45% of borrowers' monthly income is spent on repaying current debt. Their outstanding revolving balance is $29,413, which represents 56.43% of the available line of credit. Almost three-quarters are either on a mortgage or own a house.

1.3 Measuring Advantageous Selection

1.3.1 Empirical Strategy

To document advantageous selection in loan amount, I exploit the increase in the maximum loan size that occurred in March 2016. Figure A.5 helps visualizing the research design. Given the bunching at the previous limit of $35,000, two groups of people borrowed such amount before the extension. A first group actually wanted to borrow $35,000, and therefore was not rationed to start with. Together with these, some borrowers would have preferred to ask for a higher amount but were capped by the loan limit.

The empirical strategy compares the default rate of observationally equivalent households who borrow $35,000 before and after the increase in the loan limit (that is, before and after they are selected on loan amount). After March 2016, borrowers who would have been rationed without the limit extension can take up larger loans. Because of that, I expect that those who still decide to borrow $35,000 would have asked for the same amount under the old cap. Then, the difference in the default probability around the extension captures the shift in the unobserved quality of borrowers who choose $35,000 loans.

Given the very simple structure of Lending Club loans, the natural experiment is surprisingly clean. Three possible outcomes can arise. Suppose risky borrowers now choose larger loans and safe individuals keep asking for $35,000. In that case, I will observe a decrease in the default rate at this threshold, therefore concluding that the selection is adverse. Instead, the default rate will increase if risky borrowers stay at the old limit and safe borrowers ask for larger loans, thus finding advantageous selection. Finally, the no selection case arises if the default probability does not change, as risky and safe borrowers pick large loans once they are available.

Crucially, nothing else moves when introducing new menus, neither the credit model LC uses nor the interest rate for any rating. Also, given that nearly every loan posted on the platform is funded I am able to rule out any supply effect, and conclude that the demand for loans drives all the variation in the quality of borrowers. I consider a loan as in default if LC categorizes it as in *Default* or *Charged Off*. According to LC's website, a loan is in *Default* if a payment has been past due for more than 120 days. It becomes *Charged Off* when investors should stop expecting additional payments.

However, differences in the default probability of loans of the same size could also be driven by time-varying unobserved factors. The composition of borrowers could change over time due to external factors. If that was the case, disentangling the effect of amount choice from other drivers would be infeasible. Therefore, I use a difference-in-differences strategy that compares the default probabilities of other loans on top of the above variation. The rationale for doing this is two-fold. First, this variation captures the change in the quality of loans of similar size. Second, it computes the difference in the default probability across borrowers who do not apply for the largest possible loan. Therefore, these consumers picked their menus without any constraint both before and after the extension.

Table A.4 summarizes observable characteristics of borrowers asking for $35,000 before and after the limit increase. Panel A already hints that the ex-post default rate increases by two percentage points, suggesting that the selection may be advantageous. Borrowers characteristics in Panel B are similar in all respects, except for the revolving balance in other accounts. Thus, it is possible to conclude that some borrowers who asked for $35,000 before March 2016 would have asked more to repay outside loans had the limit been higher.

The underlying identification assumption is that unobserved trends in loan demand (e.g. economic conditions, meaningful events in LC competitors) do not affect differentially borrowers who choose $35,000 and other bundles in March 2016. Within this assumption, comparing the ex-post performance of borrowers in the treated and control amounts allows me to disentangle the effect of the selection on the amount that followed the increase in the loan limit.

A few factors are supporting this assumption. First, there is no reason to believe that any shift in loan demand should affect differently borrowers choosing $35,000 and bundles just below the threshold in March 2016. Also, none of LC's competitors made meaningful moves that shifted the composition of borrowers from one platform to another (in line with LC, Prosper had the loan limit at $35,000 in March 2016 and increased it to $40,000 only in May 2018). LC did not change any of the eligibility criteria, prices charged to borrowers, or its credit scoring algorithm in March 2016. They made the news about newly available loans via a blog post at the time of the introduction. It is impossible to find hints that the platform was about to increase the limit through web searches, nor that users were demanding such change.

Most importantly, I run an event study to provide more convincing evidence that the shift in borrowers' composition occurred precisely when LC introduced new loans. Finally, using a rich set of time-varying fixed effects - including location, FICO score among others - allows me to account for unobserved trends common to borrowers facing the same economic conditions and with equal creditworthiness at the time of take-up.

1.3.2 Take-up of Larger Loans

First, Figure A.2 shows a substantial decline in the take up of loans at the old loan limit. While before March 2016 the average frequency was between 15% and 16% of all loans in my sample, once larger amounts become available, it settles around 12%. The same trend can be seen in Figure A.4. Red bars represent the distribution of loans before the limit increase. Borrowers prefer round numbers, and in particular in multiples of $5,000. Most importantly, the loan limit at $35,000 leads to a strong bunching effect, as it includes people who would like to borrow more. This effect is more clear once the platform introduces larger loans. Blue bars depict the distribution of loans after March 2016. While the first observation remains true, many borrowers choose newly available loans instead of those at the old limit, thus making the bunching effect weaker.

Nevertheless, while the plots above suggest a weaker take-up of $35,000 loans, take-up at other size bundles changes as well. Therefore, it is useful to verify that the increase in loan limit corresponds a drop at the old cap, all else equal. To do that, I have to check whether similar borrowers are less likely to take \$35,000 loans once \$40,000 loans are also offered in the menu. Therefore, I run the following regression:

$$D35_{ilt} = \alpha + \beta POST_t + \gamma Rate_l + \Theta_{il} + \varepsilon_{ilt} \tag{1.1}$$

where $D35_{ilt}$ is a dummy that equals one if the loan amount is \$35,000, $POST_t$ is a time-dummy that turns on after March 2016, $Rate_l$ is the interest rate on the loan. Finally, Θ_{il} is a rich set of fixed effects that includes maturity, state, 4-digit range FICO score, employment status, home ownership, income bin, number of delinquencies, debt-to-income ratio, and default probability fixed effects.

To construct the latter, I perform a random forest algorithm that produces a default probability for each loan. Starting from October 2015, I estimate the probability of default using as training set all loans issued in the previous 12 months (i.e., starting in October 2014). Thereafter, and until the end of the sample (July 2016), I also include all loans issued in the previous month. Finally, I categorize loans in 20 buckets based on their default probability. The advantage of using machine learning is dual. First of all, it allows me to control for non-linear determinants of the default probability and to select the main variables that influence it. Second, it is the best attempt to match what LC does to assign borrowers' risk grades.

Results of the above regression are shown in Table A.5. The only difference across the three columns is the sample used. In the regression in the first column, I include all loans between \$20,000 and \$35,000. Individuals taking loans after March 2016 are 3.69 percentage points less likely to ask for \$35,000 compared to equivalent borrowers in months before. To better understand the effect, since 15% of people in my sample borrow \$35,000, it represents a drop of roughly 25%. In the second and third columns I report the same regression results using different samples, \$25,000-\$35,000 and \$30,000-\$35,000, respectively. They show similar patterns, with a drop of loans at the old limit always larger than 17%. These effects survive the inclusion of a rich set of fixed effects, including the default probability.

These conclusions allow strengthening the interpretation of the above graphs. Some borrowers were rationed when the old limit was in place and would have borrowed more in the presence of larger loans available. Now it remains to establish whether safe or risky individuals prefer larger loans.

1.3.3 Evidence of Advantageous Selection

The most basic question of this paper is: does riskiness matter when choosing the amount of a loan? If it does, which borrowers choose large loans? To document such an effect, I run a difference-in-differences model to understand the unobserved quality of borrowers who ask for \$40,000. I define a loan as *Treated* if its amount is exactly \$35,000, then I estimate the following regression:

$$y_{lit} = \alpha + \beta DID_{lt} + \Theta_{lit} + \varepsilon_{lit} \tag{1.2}$$

where y_{lit} is the outcome variable, DID_{lt} is a dummy equal to one if loan l is *Treated* after March 2016. The main coefficient of interest is β, the difference-in-differences estimator. In the main specification, the outcome variable is a default dummy. Therefore, β measures the change in the default rate of treated loans relative to loans in the control group. I include a granular set of fixed effects to compare borrowers who can choose similar menus, have the same credit risk, and face similar business cycles. As in the above regression, Θ_{lit} includes amount bin, default probability bin, maturity, state, employment status, home ownership, 4-digits range FICO score, income bin, debt-to-income ratio bin, and number of delinquencies (all except amount bin interacted with month) fixed effects. I remove the loan interest rate as that results from borrowers' choice, thus probably endogenous to their privately known default probability.

1.3.4 Empirical Results

I report the main results of regression (1.2) using default as outcome variable in Table A.6. The first column reports a difference-in-differences estimator of 2.33 percentage points. That is, selected borrowers are 2.33 percentage points relatively more likely to default than those who pick smaller loans. The mechanism is described above and depicted in Figure A.5. Once new loans are available, borrowers who take up loans larger than the old limit - but would have borrowed $35,000 had new amounts not been offered - are less likely to default. This finding is evidence of advantageous selection. The point estimate is statistically significant at the 1% level, and it is robust across samples including different amount limits. Such effect survives the inclusion of a granular set of fixed effects. These absorb any variation in default common to all loans with the same maturity, of borrowers living in the same state, with the same employment status, etc. In other words, one should interpret this result keeping in mind that I am comparing loans with the same maturity of observably equivalent borrowers.

The second column exhibits the same regression results after I include twenty time-varying default probability bins in the set of controls. When added to the aforementioned set of fixed effects, they allow me to study any difference in default probability not captured by a random forest algorithm. While slightly smaller, the advantageous selection remains sizeable and strongly statistically significant. The DID estimator suggests that new loans at the old loan limit are 1.9 percentage points more likely to default when compared with other loans in the sample.

Lending Club data allows me to analyze these results more deeply. In fact, not only can I study the effect of the extension on the extensive margin (i.e., default probability), but also on the intensive margin. It could be the case that even though the default probability increases, eventually the principal repaid does not change. I test this hypothesis by using the share of initial loan eventually repaid as the outcome variable of regression (1.2). Results are displayed in columns 3 and 4 of Table A.6. The third column shows a DID estimator of -0.0139, meaning that selected borrowers pay 1.39 percentage points less of their loan than other borrowers. This effect is again strongly statistically significant. In column 4, where I add default probability bin fixed effects, the point estimate drops (in absolute value) to -0.0122, but it keeps its statistical significance.

These estimates are economically sizeable. Taking into account that the mass of people who choose larger loans equals 25% of borrowers, the DID estimator implies that individuals selecting large loans are 7.6 (1.9/25) percentage points less likely to default than the average $35,000 borrower in my sample, who has an ex-post default probability of 19%. Thus, if the difference in default probability comes only from borrowers choosing large loans, it follows that these are 40% less likely to default.

Similarly, the estimate in column 4 implies that the share repaid by borrowers choosing large loans is 4.9 (1.22/25) percentage points larger than those selecting $35,000 loans. Since the average $35,000 borrower in my sample pays back 78% of the initial loan, it means that larger loans pay 6.2% dollars more of their initial debt. These calculations suggest that LC made a win-win choice. Not only it introduced larger loans that led to more fees for the platforms, it also improved the pool of borrowers sorting into these loans, arguably improving its reputation and increasing ex-post returns for investors.

It is worth stressing that the change in loan performance found is due to equivalent borrowers who would have most likely chosen $35,000, absent any change in the platform rules. If the better performance of new loans was only driven by the *extensive margin* - that is, borrowers who join the platform only once larger loans become available - then I would not observe any differential change in the default rate of already existing menus, as the average type who choose $35,000 would not differ. Instead, comparing borrowers who choose always-available bundles before and after the extension allows me to isolate the effect of selection into new bundles of borrowers who would have been on the platform even without the limit increase.

Finally, a possible explanation to the mechanism described above is that the selection of borrowers changes across the whole distribution of loan amounts. Since there is significant bunching at multiples of $5,000, it could be the case that once new loans become available the composition of borrowers change at those bundles as well. In order to rule out this possibility, I run regression in Equation (1.2) in the sample of loans strictly smaller than $35,000, using $30,000 as the treated group. If the above mechanism holds, the DID estimator should be statistically significant, as the unobserved type of borrowers who choose that threshold after March 2016 changes relative to smaller loans.

I present evidence for this test in Table A.7. None of the two specifications shows a significant DID estimator for the \$30,000 loan bundle, meaning that the ex-post performance of loans strictly below \$35,000 does not differentially change. In turns, the selection effect I highlight is specific to loans at the old loan limit, and due to new borrowers choosing new loans.

Supporting Identification Assumption

The above analysis can measure the selection on loan amount as long as other determinants of borrowers' creditworthiness do not affect differentially borrowers taking \$35,000 and those taking other amounts in March 2016. For example, it could be the case that economic conditions (loan demand, meaningful changes at LC alternatives) drive the choice of borrowers in the same way that the menu expansion does. If this was true, the DID estimator might be capturing effects not due to the selection of borrowers into new loans. In order to rule out this possibility, I run the regression in Equation (1.1) breaking $POST$ into month dummies:

$$D35_{ilt} = \alpha + \sum_{t=-5}^{5} \beta_t EXP_t + \gamma Rate_l + \Theta_{il} + \varepsilon_{ilt}$$

In the alternative story, conditional on a rich set of fixed effects, I should observe a change in the probability of \$35,000 take up before March 2016, and the above selection would merely pick up a change in the composition of borrowers due to external phenomena. Instead, under the assumption that the effects I show in the previous section are due to the increase in the loan limit, I should observe a shift from the \$35,000 bundle precisely in March 2016. Each dot in Figure A.6 represents the point estimate of the dummy of the correspondent month in the above regression, together with its 95% confidence interval. I drop the February 2016 dummy to make the comparison more meaningful.

Once I control for characteristics that lead borrowers to make their optimal choice, there is no pre-trend in the unobserved probability of picking \$35,000 loans. Instead, there is a sizeable drop just after new loans become available, in March 2016. This plot supports the assumption that the increase in the maximum loan limit induced some borrowers to apply for new loans, not other unobserved events.

1.4 Understanding the Source of Selection

It is well known that debt relief generosity affects credit supply (Severino and Brown, 2017) and bankruptcy behavior (Dobbie and Song, 2015; Indarte, 2019), but their influence over the selection of borrowers has received limited attention from scholars. Such selection is driven by the interest rate charged by traditional lenders. In particular, Severino and Brown (2017) show that banks set higher interest rates where homestead exemptions are larger, as they bear most of the default cost.

My analysis aims at studying whether (and how) the pool of borrowers switching to LC is affected by the price they face in conventional credit markets. I subdivide states into groups of equal size based on their homestead exemption level to shed light on this issue. For example, when I form four groups, the least bankruptcy-friendly group includes states with exemptions lower than $20,000. Instead, the thirteen most generous states have an exemption at least as large as $250,000 (and eight do not have any limit). As there is significant heterogeneity across groups, estimating the regression in Equation (1.2) in these different samples allows me to see how debt relief generosity leads borrowers to make different choices. As already pointed out, homestead exemption levels were set in the nineteenth century (possibly not randomly) for reasons unlikely related to nowadays credit markets.

First, Table A.8 reports some relevant information about borrowers in two groups of states defined as *Low Exemption* and *High Exemption*. The first comprises the 24 states with exemptions lower than the median $75,000, while all the remaining 27 are in the second (the two groups are not equally sized as there are five states with median exemption). Before March 2016, individuals in both groups borrowed just above $26,000 at a rate roughly equal to 13%. Consumers in *High Exemption* states prefer shorter loans, which translates into a higher monthly installment. Most importantly, the default rate is 18% in both states. The average borrower in both groups has a 699 FICO score and a debt-to-income ratio just below 20%. Residents of high exemption states earn $5,000 larger incomes and are less indebted by $3,000. Also, they are as likely to have been delinquent in the previous two years, and more likely to rent their home rather than being on a mortgage.

First of all, I check whether borrowers in the two groups of states differ in their probability of choosing newly available loans. To test this, I estimate the following difference-in-differences regression:

$$D35_{ilst} = \alpha + \beta DID_{st} + \gamma Rate_l + \Theta_{lit} + \varepsilon_{ilst} \tag{1.3}$$

where $D35_{ilst}$ is a dummy that equals one for \$35,000 loans, while the dummy DID_{st} turns on for loans in *High Exemption* states after March 2016. Remaining variables are the usual controls: $Rate_l$ is the price of the loan, and Θ_{lit} is the above-defined set of fixed-effect. The coefficient of interest is β, as it measures the differential take-up of \$35,000 loans in the two groups of states after March 2016.

Results are reported in Table A.9. As in previous cases, the only difference between the two columns is the inclusion of the default probability bin fixed effects. The difference-in-differences estimator is not statistically significant in both columns. There is no differential take-up of larger loans after March 2016 between the two groups of states. If this was the case, any heterogeneous selection across states could have been related to different preferences for loan amount. Instead, any selection in the analysis that follows has to be linked to the unobserved qualities of borrowers.

As anticipated, the main test of this section consists in estimating the regression in Equation (1.2) in different samples. First, I only consider two groups called *Low Exemption* and *High Exemption* states. Second, I further split every sample in two, forming four groups. Main results are reported in Table A.10 and Table A.11.

In the analysis reported in Table A.10, I split the sample into states whose homestead exemption is below and above the median \$75,000. The first column summarizes results in the *Low Exemption* group. Interestingly, the point estimate is very close to zero and not statistically significant. This result means that in states more hostile to bankruptcy filers the relative quality of selected borrowers does not change. Thus, the average quality of borrowers choosing \$35,000 and \$40,000 after March 2016 is the same.

The second column reports the opposite result. Here, the relative increase in default probability amounts to 3 percentage points, suggesting that selected borrowers are of poorer quality, as safer individuals prefer larger loans. The coefficient is strongly statistically significant despite the richest set of fixed effects (which comprises default probability bins, interacted with months). Taking the two columns together, it is clear that the effect seen above in Table A.6 is entirely driven by the selection in more generous states.

Using the same back of the envelope calculations I used in the previous paragraph, these estimates imply that, in states with high exemptions, borrowers choosing large loans are 12.5 (3/25) percentage points less likely to default, compared to $35,000 loans (whose average default probability is 19%). That is, in these states switching borrowers who choose large loans are 65% less likely to default.

The previous results are not only true in the extensive margin, but also in the intensive margin. In the third column, there is no evidence of a significant difference between the share of the initial loan repaid by selected borrowers in states with lower-than-median exemption. Instead, in the last column I show that the share repaid by selected borrowers in *High Exemption* is 1.78 percentage points lower than other borrowers.

Finally, to deliver more convincing evidence, I split the above two samples further. Table A.11 reports the results of the same regression, this time with four samples with increasing levels of debt relief generosity. Column one reports the results in states with the lowest level of exemptions, while the last sample considers the most generous states. Conclusions do not change. Borrowers' quality in states more hostile to bankruptcy filers does not show any differential change. Indeed, difference-in-differences estimators in the first two columns are both not statistically significant. Instead, estimators in the third and fourth columns are both strongly significant and sizeable, as they both show a 3 percentage points increase in bankruptcy probability.

The last test I run is possibly the more convincing. As I have established that there is no selection in states whose exemption is below the median, I can add states to the sample one level of exemption at a time and observe any pattern in the DID point estimate. I plot the observed point estimates, together with their 95%

confidence intervals, in Figure A.7.

The first coefficient on the left is the same shown in the first column of Table A.10. Indeed, it is very close to zero and not statistically significant. As I add states to the sample, one level at a time, the DID estimator starts increasing while still being not statistically significant. It shows any relevance when I add states with homestead exemption at least equal to $165,500. And as I enrich the sample, it keeps being significant and with a higher magnitude. This trend is the second interesting fact. Advantageous selection is almost monotonic in the debt relief generosity. That is, the vast majority of safe borrowers who switch to LC live in bankruptcy-friendly states and are, therefore, likely to face higher banks' interest rates.

1.5 A Simple Model of Debt Refinancing

The above empirical evidence is counter-intuitive. Banks tend to lower credit limits avoid excessive borrowing by high-risk individuals (Jambulapati and Stavins, 2014). Similarly, standard models predict that low default costs (such as high homestead exemptions) lead banks to set higher interest rates, which lead the way to adverse selection. To reconcile my empirical findings with a theoretical predictions, my model has to consider i. the current equilibrium in credit markets and ii. the reason why individuals borrow from LC. To this extent, I have already shown in Table A.3 that more than 85% of LC borrowers use their loans to repay existing debt. Interestingly, in Table A.12 I run the usual difference-in-differences regression in Equation (1.2) in the sub-sample of borrowers who use LC loans for purposes other than refinancing, and I do not find any evidence of selection. This is true both when I consider the whole sample (column 1) and when I split the sample in lower-than-median and higher-than-median exemption states (columns 2 and 3). Therefore, in my model, borrowers make an effort to search for lenders charging lower interest rates. Once they find it, they decide how much to borrow to minimize future cash flows.

Each household lives for one period. At the beginning of the period, she faces uncertainty over her disposable income (w). For simplicity, I assume this is either positive (w) or equal to zero. Borrowers' heterogeneity lies in their exogenous probability of default (p_i). She has an outstanding debt (k) she wishes to refinance. Such debt expires in the next period and carries an interest rate r_B. Each borrower chooses how much effort (t) to exert to find a new lender. Given the effort, she finds a cheaper lender with probability $\alpha(t)$, an increasing and concave function of the effort. Yet, searching costs $c(t)$, an increasing and convex function of the effort exerted.

If she finds a lender (for example, LC), she chooses how much to borrow (d). This debt has the same maturity as external debt, but she pays an interest rate r_{LC}. After a loan is funded, she pays an origination fee (δ) upfront. That is, this is deducted from the money received from the lender. Therefore, each borrower actually gets $d(1-\delta)$. In the bad state of the world, she defaults on her LC and external debt. The default cost is fixed (Θ), representing stigma, the impossibility of future borrowing, etc. If she defaults, she only repays a share θ of the outstanding debt with the new lender.[8] Finally, households own a house of value H and, depending on the states a household resides in, she is subject to a homestead exemption (e).

Household i maximizes her expected utility in the payment period. Denoting with U^{LC} and U^B households' utilities if they switch to LC or stick with traditional bank, respectively, they are defined as follows:

$$\begin{aligned} \mathbb{E}[U^{LC}] &= (1-p_i)u(c_N^{LC}) + p_i\left(u(c_D^{LC}) - \Theta\right) \\ \mathbb{E}[U^{B}] &= (1-p_i)u(c_N^{B}) + p_i\left(u(c_D^{B}) - \Theta\right) \end{aligned} \tag{1.4}$$

where $u(c)$ is an increasing and concave function of consumption, and subscripts N and D denote consumption in the non-default and default states, respectively. Therefore, the final maximization problem looks as follows:

$$\max_{d,t}\{\alpha(t)\mathbb{E}[U^{LC}] + (1-\alpha(t))\mathbb{E}[U^B] - c(t)\} \tag{1.5}$$

[8] This is motivated by the fact that in case of default LC "makes a reasonable effort to recover the money owed to investors". While it is true that traditional lenders could do the same, the predictions of the model are not affected by the settlement borrowers negotiate with them.

subject to the following budget constraints:

$$\begin{aligned}
c_N^{LC} &= w - d(1 + r_{LC}) - (k - d(1-\delta))\,(1 + r_B) + H \\
&= w + d\left[(1-\delta)(1+r_B) - (1 + r_{LC})\right] - k(1 + r_B) + H \\
c_D^{LC} &= \min\{e, H\} - \theta d \\
c_N^{B} &= w - k(1 + r_B) + H \\
c_D^{B} &= \min\{e, H\}
\end{aligned} \tag{1.6}$$

When they do not default, borrowers consume their income w after repaying all their debts to LC and traditional banks. Also, they consume their housing wealth. In default, they enjoy their housing wealth only up to the exemption limit.

The first order condition with respect to d is then:

$$(1 - p_i)u'(c_N^{LC})\left[(1-\delta)(1+r_B) - (1 + r_{LC})\right] = p_i u'(c_D^{LC})\theta \tag{1.7}$$

The interpretation of this first order condition is very intuitive. Conditional on finding LC and refinancing, borrowers have to equate the marginal benefit (the savings on interests paid on their outstanding loans) with the marginal cost (the marginal settlement in case of default). Moreover, the necessary condition for an internal solution is that the interest rate savings must be positive, accounting for entry fees. That is:

$$(1-\delta)(1+r_B) - (1 + r_{LC}) > 0$$

The first order condition with respect to t is:

$$\alpha'(t)(\mathbb{E}[U^{LC}] - \mathbb{E}[U^B]) = c'(t) \tag{1.8}$$

That is, the marginal gains from exerting effort (gain in expected utility) must equal marginal costs. Moreover, the gain in expected utility over the *status quo* is a necessary condition for exerting any effort.

It is this last FOC that allows me to interpret my empirical findings. As explained above, without searching for new lenders, borrowers pay r_B on their bank loans. In order to understand why the advantageous selection comes from states with higher exemption levels, it is necessary to analyze the equilibrium in traditional credit markets.

From a bank's perspective, the expected loss in case of a borrower's default is higher when such borrower can keep a portion (or sometimes the entire) house of residence. Therefore, it is optimal to set a higher price $r_B(e)$. Severino and Brown (2017) note that risk-neutral banks' break-even condition for a loan of size k is:

$$(1+r_f)k = \tilde{\mathbb{E}}\big[\min\big\{(1+r_B(e))k, \max\{H-e,0\}\big\}\big] \tag{1.9}$$

With a constant left-hand side, as e increases banks must raise $r_B(e)$ to be remunerated for risk-bearing. They also provide empirical evidence for this prediction. They show that interest rates on unsecured credit tend to be higher in states with higher debt relief generosity. They explain these findings with a high demand for unsecured debt. Therefore, it is likely that LC borrowers face different outside options. In particular, LC borrowers living in states with high exemptions are likely to face higher $r_B(e)$.

The assumption that LC's interest rate r_{LC} does not depend on the state of residency seems strong, but the data back it. In the next section, I show that, while I cannot access LC's current credit scoring algorithm, I can exclude that the variation in the interest rate charged is affected by the state of residence. Moreover, looking at old versions of LC's website through the Internet Archive's Wayback Machine, it is clear that when the algorithm was published, none of the criteria was about the location of loan applicants.[9]

From the above logic, all else equal, the pool of individuals who are willing to borrow from traditional banks will be different across states. It will be worse in states with higher exemption because of the higher price. However, conditional on borrowing (as every LC borrower has also external debt), it follows that the safest borrower in high exemption states pays a higher premium over her symmetric information interest rate relative to an identical borrower in states with low exemptions. Therefore, it is intuitive to conclude that safe borrowers in high states

[9] https://web.archive.org/web/20121031160219/http://www.lendingclub.com/public/how-we-set-interest-rates.action

are more likely to look for better credit conditions elsewhere.

I can formalize this prediction by studying the sensitivity of the optimal effort exerted with respect to the interest paid on borrowers' current debt r_B. By applying the Implicit Function Theorem to the first order condition in Equation (1.8):

$$\frac{\partial t^*}{\partial r_B} = -\frac{\alpha'(t)(1-p_i)\left[u'(c_N^B)k - u'(c_N^{LC})(k - d(1-\delta))\right]}{\alpha''(t)(\mathbb{E}[U^{LC}] - \mathbb{E}[U^B]) - c''(t)} \tag{1.10}$$

If consumers' consumption is higher when they switch to LC, which is always true under the FOC in Equation (1.7), the numerator is positive. Moreover, from the FOC in Equation (1.8), it must be the case that

$$\mathbb{E}[U^{LC}] - \mathbb{E}[U^B] > 0$$

Therefore, for a concave $\alpha(t)$ and a convex $c(t)$, the above sensitivity is positive. Accordingly, all else equal, borrowers that face higher interest rates are more likely to exert more effort to find cheaper loans.

The positive selection arises from the fact that borrowers have to pay a search and a switching cost to save on their interest payments. Intuitively, for someone that is sure to default there is no reason to search for cheaper loans. Formally, taking the sensitivity of the optimal effort with respect to default probability yields the following:

$$\frac{\partial t^*}{\partial p_i} = -\frac{\alpha'(t)\left[u(c_N^B) - u(c_N^{LC})\right]}{\alpha''(t)(\mathbb{E}[U^{LC}] - \mathbb{E}[U^B]) - c''(t)} \tag{1.11}$$

As in the previous case, if consumption in the non-default state is higher when switching to LC, this derivative is negative. That is, keeping everything constant, a riskier borrower exerts a lower effort to find a cheaper lender. Thus, they are more likely to find themselves at the stage of choosing the loan amount, conditional on finding LC.

1.5.1 Testing Model Predictions

The data and the setting largely justify the assumptions I make in this model. First, I assume LC's interest rate r_{LC} does not depend on the state of residence. This pricing model is arguably suboptimal from LC's perspective. By charging higher prices, they could be able to make up for higher expected losses given default. In Figure A.8 I plot the R-squared of eight regressions of the same dependent variable (interest rate), on a richer and richer set of controls. Once I add states to previous controls (in *Model 7*), there is no gain in the explained variance of r_{LC}, thus hinting that borrowers do not pay different rates based on where they live.

It could also be that LC is simply pricing loans wrongly. In this case, borrowers in states with a high exemption are implicitly subsidized by those in more bankruptcy-hostile states. Yet, if this were true, I would also observe different post charge-off recovery, namely lower settlements in more generous states.

First, in Table A.13 I rule out mispricing by LC. In the first column, I report the result of a regression of the post charge-off recovery on a *High Exemption Dummy* and a rich set of controls. Similar borrowers do not settle for a different payment in the two groups of states, leading to the conclusion that the typical LC's borrower should not expect to be shielded against the platform's attempt to be repaid depending on her residency. The same conclusion follows from the second column. Here I use the percentage of the outstanding debt recovered, and it does not seem advantageous to live in states with higher exemption.

While I cannot directly observe the outside options that LC borrowers face (i.e., the interest rate on their outstanding loans), I can test whether they have different debt stocks before applying for a loan on LC. By assuming an inverse relation between loan demand and price, observing differential indebtedness would imply that interest rates faced are also different. Absent any difference in the rate charged by banks and in the exemption level, identical borrowers should not have different loan demand. If one allows the exemption to vary, keeping rate fixed, it should be optimal to borrow more in more generous states. Then, observing lower debt stock in *High Exemption* states could be evidence of the fact that banks mitigate excess borrowing by charging higher prices.

I test this hypothesis by running a regression of several debt classifications - the total outstanding debt (excluding mortgages), the installment loans balance, and the revolving balance - on a *High Exemption Dummy*, controlling for a large set of observable characteristics. I only use loans issued before March 2016 to avoid overlap with new menus offered, but results are robust when using an extended sample.

Results are reported in Table A.14. The dummy of interest is always negative and strongly significant, except the third column on revolving debt. In particular, the first column shows that households applying to LC living in generous states have $1,947 lower total debt stock than identical borrowers in other states. While I cannot perfectly disentangle the price effect from other confounding factors, such a differential take-up is evidence of a different interest rate faced.

Similarly, the second column shows results of the identical regression using the balance in installment loans (e.g., auto loans) as the dependent variable. As in the previous case, the coefficient on the exemption dummy is negative and strongly statistically significant. Borrowers in states with high exemption have $1,565 lower debt to be paid in installments, suggesting that also in this case households may face differential banks' rates. Only in the last column, using revolving debt balance as dependent variable, such dummy is not significant. All in all, given the above discussion about exemption level, bank rates, and loan demand, this evidence is suggestive of different prices faced by borrowers depending on the exemption they are subject to.

In the model, the nature of private information is the probability of realization of the bad state of the world. More realistically, borrowers' private information stems in their future ability to repay their debt. One measure of such ability is the FICO score. LC distributes a *Payments* dataset that contains the time series of all payments made, together with the FICO score at the time of each payment date. I construct a new variable:

$$\Delta FICO = FICO_{issue} - FICO_{last}$$

to study how borrowers' credit score evolves over the life of the loan. Then, I run the regression in Equation (1.2) using the above variable as the outcome. If it is the case that borrowers have private information about their future income shocks, these are likely to appear in worse FICO scores. Results are in Table A.15.

The first column of table A.15 shows the results using the entire sample. As the model predicts, selected borrowers have a FICO score decline at least 2.8 points higher than other borrowers. This point estimates are statistically significant at the 1% level. Interestingly, the average LC borrower sees her FICO score decreasing over the relationship with the platform. Yet, this evidence confirms that borrowers at the old limit deal with worse income shocks and thus are more likely to default.

The second and the third columns show the estimates of the identical regression, but this time dividing states above and below the median exemption. This evidence corroborates the fact that it is only in *High Exemption* states that LC members display private information about their future income shocks. The DID estimator in the second column, which only includes less generous states, does not show any relative difference in the FICO score of borrowers treated and in the control group. Instead, in the third column, when I only look at individuals in states above the median exemption, the DID estimator shows a relative decline of 3.3 scores for $35,000 loans, compared to other size bundles. Therefore, it is exactly in these states that private information leads borrowers to choose different loan size choices.

To provide further evidence on the private information underlying size choice, it is useful to look at the performance of $40,000 loans. If it is the case that borrowers choosing the new maximum are less likely to be exposed to negative income and creditworthiness shocks, then in turns, their ex-post default probability should be lower than borrowers choosing $35,000. It turns out that the default probability of $40,000 loans issued after March 2016 is roughly 11%, against 20.5% at $35,000. Yet, characteristics of borrowers choosing these two bundles could be different.

Thus, it is more convincing to test this hypothesis using the regression in Equation (1.2), this time including newly available loans in the sample. It is important to point out that I am not perfectly able to measure selection for new loans, as I do not have a proper counterfactual for these borrowers. Therefore, it is possible that the results are affected by individuals who applied for a loan only because $40,000 was available, and would not have signed up before March 2016. Although this is not likely to have happened, as they had the alternative to ask for a smaller loan, I cannot test this assumption. With this potential caveat in mind, I report evidence for this prediction in Table A.16

The estimates in the first column use the entire sample. The DID estimator shows a relative increase in default probability of selected borrowers by 2.31 percentage points compared to other loan sizes. It is statistically significant at the 1% level. To understand how the default probability of borrowers choosing new loans compares to the treated group, it is useful to compare this point estimate with its equivalent in the second column of Table A.6. In both cases, the regression compares identical borrowers choosing different loan amounts before and after the maximum amount increase. Table A.16 only adds loans larger than $35,000 in the sample. Thus, the only reason why the last point estimate is 0.42 percentage points larger than its parallel in Table A.6 is that the relative performance of selected loans is even worse when compared to larger ones.

The second and the third columns of Table A.16 display the results of the identical regression in states with low and high exemptions. I do this to verify that safer borrowers choose large loans especially in states where the interest rate differential is higher. If this is the case, I should not observe any relative difference with the results in Table A.10 in states with lower than the median exemption. It is only in more generous states that borrowers choose loan size based on their private information. Therefore, my model implies that large loans are less likely to default than smaller loans only in the second group of states.

Results confirm this thesis. The second column, which focuses on less generous states, does not show any differential performance of selected loans. This is the same result I show in the first column of Table A.10. On the contrary, the DID estimator in the third column is significant at the 1% level and suggests a relative increase in default probability of selected borrowers of 3.5 percentage points. To complement this analysis, this point estimate is 0.5 percentage points higher than its equivalent in Table A.10. This difference implies that once new loans are added to the sample the performance of borrowers choosing the old limit is even worse, indicating that large borrowers have a lower default probability.

1.6 Conclusion

The growing literature on fintech lending still lacks evidence on how borrowers choose marketplace lenders based on their private information. In this paper, I document advantageous selection in loan amount in the context of the online unsecured credit market. In particular, I show that safer borrowers choose large loans in bankruptcy-friendly states. In the theoretical model, I argue that this is because the main gain from switching to online markets to refinance debt is represented by interest rate savings. Under asymmetric information, traditional lenders set higher prices where exemptions are more generous, which lead to adverse selection. Therefore, when a low-cost competitor enters the market, the safest borrowers have the highest incentives to refinance their loans. Instead, risky borrowers are not willing to bear the costs and stay with their current lenders.

This work is the first to document empirically advantageous selection in unsecured loans. It is relevant as it shows how fintech lenders can beat conventional banks in attracting safe borrowers by offering simple contracts and sizeable price savings. It is important for policymakers to investigate which borrowers switch to less regulated online lenders, and why. In fact, as the market share of these players grows, so does the share of borrowers that are not subject to standard regulation. Understanding these incentives is crucial to limit (and possibly govern) their impact on financial stability.

2. Broadband and Bank Intermediation

Angelo D'Andrea, Marco Pelosi and Enrico Sette[1]

2.1 Introduction

The arrival of fast internet has been one of the most disruptive innovations in history and as such, it had a substantial impact on economic activity. The availability of a massive amount of information, together with the ability to communicate them quickly, transformed the size and the operations of many industries. As an information-intensive business, banking was particularly prone to this transformation.

When information flows are limited, banks face higher information asymmetries, communication costs, and more severe agency problems (Leland and Pyle, 1977). Innovation in information technology, from hardware such as computers and phones, to software such as credit scoring and client profitability programs, mitigates these frictions and can play a crucial role in shaping banking activity (Mishkin and Strahan, 1999). As proof of this, banks have long relied on cutting-edge technologies to deliver innovative products, streamline loan making processes and improve their back-office efficiency (Frame et al., 2018).

Despite its relevance, and the importance of the banking industry for the economy (Levine, 1997), the evidence on the effects of the arrival of fast internet on the activity of banks, in particular lending, is scant.[2] A key reason for this is the lack of high quality administrative data and of an identification strategy to deal with the endogeneity of the introduction of internet.[3]

[1] We benefited from helpful comments from Francesco Decarolis, Marco di Maggio, Nicola Gennaioli, Nicola Limodio, Daniel Paravisini, Nicolas Serrano-Velarde, Fabiano Schivardi, Emanuele Tarantino and seminar participants at Bocconi University and the Bank of Italy.

[2] D'Andrea and Limodio (2019) is an exception. They focus on the effects of high-speed internet in Africa, and show that fast internet favored new financial technologies in the interbank market, thus alleviating banks' liquidity risk and promoting lending to the private sector.

[3] By contrast, a large literature studies the effects on the banking industry of regulatory reforms (Bertrand et al., 2007), removals of barriers to entry (Cetorelli and Strahan, 2006), shocks of various

This paper studies the effect of access to broadband internet on bank credit to non-financial firms, and it sheds light on the mechanisms behind this effect. We focus on supply and show that fast internet affects banks' organizational design, with effects on productivity, market geography and local banking competition. We resort to a unique granular dataset from Italy, that includes detailed information on the dates of broadband internet arrival and on the geographical location of the necessary infrastructure. This information is matched with loan level data from the comprehensive Italian credit register and with other administrative details on the location of bank branches and on banks' assets, liabilities, and employees. We observe these data between 1998 and 2008, that are the years marking the expansion of broadband internet in Italy, and perform an econometric analysis based on instrumental variables (IV).

Measuring the impact of access to broadband on credit is challenging. Since high speed internet is not randomly assigned to municipalities, bank credit could be affected by hidden factors other than (but related to) broadband connection. To deal with this source of endogeneity, we use an instrumental variable strategy and leverage the position of the municipality in the pre-existing voice telecommunications (telephone) infrastructure, to instrument for broadband availability (Campante et al., 2018).

As fast internet services in Italy could only be offered in municipalities connected to high-order telecommunication exchanges via optic fiber, we use the distance between the municipality centroid and these exchange infrastructures (a proxy for the required investment to connect the municipality with the fiber) as a source of variation for the availability of high-speed internet (Ciapanna and Sabbatini, 2008). Because the pre-existing telecommunication network was not randomly distributed, our instrumental variable relies on the interaction between the above mentioned distance and a dummy variable for the period after broadband internet became available. Our identification assumption is that, whatever correlation existed between the distance and relevant municipality characteristics, this did not

nature (financial, real, natural disasters), institutional quality, and even the role of culture and ethnicity (Caprio et al., 2007; Grosjean, 2011; Calomiris and Carlson, 2016; Pascali, 2016; Fisman et al., 2017; D'Acunto et al., 2019).

change at the time of the introduction of broadband for reasons other than broadband itself. Results from our 2SLS estimates shed light on the causal effect of fast internet on bank credit.

The latter is an equilibrium outcome, that takes into account both demand and supply. The effect of internet on credit demand has been only indirectly documented in the literature, by focusing on firm's productivity (Akerman et al., 2015). On the other hand, and although there is consensus that information technologies have revolutionized the way lending is conducted by traditional banks (He et al., 2021), the evidence on the effects of fast internet on banks' internal activity and their organizational design remains relatively scant. For that reason, in this paper we are mostly interested in how broadband internet affects bank credit supply.

The granularity and the structure of our data help us isolating credit supply from other counfounding factors. Similar to Khwaja and Mian (2008), in the most demanding specification we exploit the panel structure of the data and the diffusion of multiple bank relationships in Italy (Gobbi and Sette, 2014) to compare the amounts of credit extended to firms that have relationships with banks in municipalities served differently by broadband. In this way, we are able to isolate credit supply from demand, and we can exclude that total credit variation is driven by firm-specific needs (which in turn may be affected by broadband availability). We use high dimensional fixed-effect regressions that control for time-varying firms' loan demand and for features specific to the firm-bank-municipality relationship. Finally, we add bank-year fixed effects to control for time-varying credit variation that originate from bank-specific policies.

Italy represents an ideal laboratory for our analysis. First, the long history of human settlement in the country allowed for the existence of several relatively small municipalities located at short distance from one another, often separated by geographical barriers (rivers, lakes, mountains). This creates large variation in the distribution of the infrastructure needed to bring broadband to different municipalities, which often have a very similar level of economic activity and development and are just a few miles away. Second, Italy in our sample period did not experience an especially fast growth in credit, nor it experienced a housing bubble, contrary for example to the US, UK, Ireland or Spain. Third, Italy is a developed economy, mostly bank dependent, with an economic structure similar to that of other major countries. Finally, Italy has very granular administrative micro-data that are crucial to implement our identification strategy.

The main findings from our empirical analysis can be summarized as follows.

We find a positive and statistically significant effect of broadband internet on the extensive and the intensive margin of the credit relationship. Going from zero to high broadband coverage is associated with an increase in the number of loans issued by banks of 12% (0.08 of a standard deviation, s.d.), and an increase in the amount of credit granted of 28% (0.13 of a s.d.).[4] Then, we find a negative and statistically significant effect of broadband on the price of bank credit. Moving from zero to high coverage of high-speed internet is associated with a decrease in the average interest rate of 30 b.p. (0.18 of a s.d.). This result is ex-ante not-obvious, and leads us to deeply investigate the effects of broadband on bank credit supply. When restricting the analysis on credit supply by controlling for firm-level demand for credit, we find that branches in municipalities reached by fast internet expand their amount of credit 19% more than other branches (0.14 of a s.d.).

To qualify the effect of broadband internet on bank credit, we focus on credit supply and study how banks' productivity, geographical scope and local competition are affected by fast internet. As far as the first is concerned, we show that the lending efficiency of banks, measured by banks' labor productivity and credit quality, increases as a consequence of broadband availability. Internet helps increasing credit extended per employee by 24%. Standard models suggest that a surge in loan supply may lead to a worse quality distribution (Berger and Udell, 2004; Foos et al., 2010). If anything, in our setting we find that credit quality improves, as the share of non-performing loans (NPLs) per bank decreases. These findings are in line with Petersen and Rajan (2002) and Berger (2003), who argue that richer hard information and more efficient back-office technology helps both ex-ante screening and ex-post monitoring.

[4] "High" broadband coverage means at least 75% of the population in the municipality connected to fast internet.

Moreover, the geographical reach of banks widens. Banks operating in municipalities reached by fast internet expand their markets beyond standard geographical borders, which typically coincide with provinces. In fact, we find that they are more likely to originate loans outside the province where they are located. In addition, the physical distance between banks' municipality and borrowers increases, too. These results are consistent with the view that improved screening and monitoring, together with a reduction in communication costs, allow banks to reduce the dis-economies of distance (Berger, 2003; Felici and Pagnini, 2008).

Finally, municipalities reached by fast internet experience a rise in banking competition. This is confirmed by the increase in the number of available bank brands in the municipality, together with the dynamics of standard proxies of competition. The concentration ratio of the top 5 and top 3 banks decreases, as well as the Herfindahl–Hirschman index (HHI) of deposits. Consistently, we find that increased competition pushes down loan prices (in line with Hauswald and Marquez (2003); Vives and Ye (2021)).

The arrival of fast internet has additional effects on banks and borrowers. To this extent, we find that banks tend to implement internal credit reallocation across municipalities, with new loans managed by larger, more distant branches. In the test using granular data, we show that broadband internet does not have any effect on credit supply from branches located in small municipalities.[5] Indeed, banks also tend to open new branches in places reached by fast internet, but not when these places are small. Access to high-speed internet creates digital highways that carry bank credit from connected peripheries to the center, i.e. from smaller municipalities connected to broadband towards bigger municipalities. Yet, this local credit desertification is not accompanied by slower economic growth. Broadband access boosts GDP per capita both in bigger and smaller municipalities, showing off the virtues of the credit flows that broadband contributes to create.

Our results are robust to several robustness checks, most notably, different measures of broadband coverage and the inclusion of several control variables at the

[5] We define a municipality as small if its population is below the in-sample median of 4,639 inhabitants.

municipality level, that aim to control for municipality time trends. We also run placebo IV specifications for the years before fast internet was available, and simulate internet availability as if we were in the post broadband period. Reassuringly, we find no impact of broadband internet on bank credit.

This paper builds and extends on different strands of the literature. It contributes to the broad literature on the effects of new telecommunication infrastructures on the economy (Roller and Waverman, 2001; Forman et al., 2009; Czernich et al., 2011; Kolko, 2012; Akerman et al., 2015; Pascali, 2017; DeStefano et al., 2018; Donaldson, 2018; Steinwender, 2018; Hjort and Poulsen, 2019). In this respect, it is one of the few that concentrate on the role of ICTs on banking. D'Andrea and Limodio (2019) exploit the staggered arrival of fiber-optic submarine cables in Africa and show that high-speed internet lifts banking. They highlight one possible mechanism behind this effect that is related with more efficient interbank markets. Lin et al. (2021) study China in the late 19th century and show that the telegraph significantly expanded banks' branch networks. This paper adds to the existing literature by focusing on a specific technology, broadband internet, and a specific instrument, enhanced bank credit to firms, and by showing the channels through which it operates.

The paper also contributes to the literature on information technology and banking, by showing the effects of broadband internet on bank lending and on banks' organizational design. Hauswald and Marquez (2003) provide a theoretical framework on the effects of new information technologies on loan prices and competition. They show that the advent of new information technologies generates ambiguous effects depending on whether these technologies are easily available to all competitors rather than being of exclusive use to some of them. Using similar arguments, Vives and Ye (2021) show that IT progress involves an increase in competition intensity when it weakens the influence of bank-borrower distance on monitoring costs. Petersen and Rajan (2002) and Berger (2003) provide intuitions and empirical evidence on the effects of new technologies on the distance between lenders and borrowers. New technologies allow financial intermediaries to substitute soft information with hard information, thus increasing the distance between borrowers and lenders. Related to this topic, Felici and Pagnini (2008) find that the geographical reach of entry decisions increases for those banks that resort more to information and communication technologies, and that the latter has important pro-competitive effects. On a similar vein, Degryse and Ongena (2005) and Keil and Ongena (2020) show the effects of technologies on bank organizational structure, with a particular emphasis on de-branching. In a recent paper, Ahnert et al. (2021) show that job creation by young firms in the US is stronger in counties that are more exposed to IT-intensive banks. The paper is also close to the literature on internet banking and bank performance (DeYoung, 2005; Ciciretti et al., 2009) and that of information technology and financial stability (Pierri and Timmer, 2020). Finally, the paper builds a bridge between the traditional literature on technology and banking and the fast growing literature on FinTech (De Roure et al., 2016; Buchak et al., 2018; Tang, 2019; Braggion et al., 2020; Di Maggio and Yao, 2020), which documents the economic effects of state of the art financial technologies.

To conclude, our paper contributes to the large literature on information in financial intermediation (Leland and Pyle, 1977; Campbell and Kracaw, 1980). Stiglitz and Weiss (1992) show that despite the richer strategy space available to lenders, market equilibria can be characterized by credit rationing if information asymmetries are relevant. New technologies such as credit scoring (Einav et al., 2013), fax machines, or internet can help reduce these information asymmetries and improve bank lending (Liberti et al., 2016; Liberti and Petersen, 2019).

The rest of the paper is organized as follows. Section 2 presents the institutional background in Italy. Section 3 describes our data. Section 4 presents the empirical specifications and the identification strategy. Section 5 shows the main results with robustness checks. Section 6 elucidates the mechanisms behind our findings. Section 7 discusses the effects of new digital highways on bank credit. Finally, Section 8 concludes.

2.2 Institutional Background

Italy represents an ideal laboratory for our analysis. First, the long history of human settlement in the country allowed for the existence of several relatively small municipalities located at short distance from one another, often separated by geographical barriers (rivers, lakes, mountains). This creates large variation in the distribution of the infrastructure needed to bring broadband to different municipalities, which often have a very similar level of economic activity and development and are just a few miles away. Second, Italy in our sample period did not experience an especially fast growth in credit, nor it experienced a housing bubble, contrary for example to the US, UK, Ireland or Spain. Third, Italy is a developed economy, mostly bank dependent, with an economic structure similar to that of other major countries. Finally, Italy has very granular administrative micro-data that are crucial to implement our identification strategy.

2.2.1 Broadband Internet

Broadband internet connection in Italy has been traditionally provided through asymmetric digital subscriber lines (ADSL). ADSL technology is a data communications technology that enables faster data transmission than a conventional voiceband modem, and it was introduced by the Italian telecommunications incumbent operator, Telecom Italia, in 1999. The development of the ADSL infrastructure was relatively slow in the first years. By the end of 2000, only 117 out of 8,100 Italian municipalities had access to the new technology. Instead, it sped up sensibly during the subsequent years. By the end of 2005, about half of all municipalities owned an ADSL line, accounting for approximately 86% of the population. Figure B.1 reports the time series of broadband adoption in terms of the number of municipalities with ADSL access, between 2000 and 2008. Given the low access and penetration rates until 2001, we consider this as the last "pre-broadband" year throughout our analysis.

ADSL technology relies on information transmission over conventional copper phone wires. Henceforth, ADSL access depends crucially on the user's position in the pre-existing voice telecommunications infrastructure. Technically, the voice telecommunications infrastructure consists of three levels: the Line Stage (LS), the Urban Group Stage (UGS), and the Transit Group Stage (TGS). The LS is the last structure where all the providers connect with their equipment, after which the famous last mile that reaches the end-users begins. In Italy, the 10,500 LSs are linked to one of the 628 UGS, which are connected to one of the 65 TGS. To complete the physical architecture of the network, some TGSs are tied to the three international gateways (Milan, Rome, and Palermo), which allow for international communications.

Two parameters are of specific importance for ADSL deployment and performance. The first is the distance between the end user's premises and the closest telecommunication exchange (the LS), known as the "local loop". If the length of the local loop is above a certain threshold, the ADSL connection cannot be implemented through traditional copper wires, but it needs fiber optic cables. The second is the distance between the LS and the closest higher-order telecommunication exchange (the UGS). Independently from the length of the local loop, for ADSL to be available, the connection between the LS and the UGS must be through fiber optic cables. In Italy, the length of the local loop has not constituted a limiting factor for the development of the broadband infrastructure. Since the local loop was a key element in the voice telecommunications network, its length was generally short and distribution capillary. Instead, the distance between the LS and the UGS, which was irrelevant for voice communication purposes, has become the primary determinant of the investment needed to provide ADSL to a given area and, consequently, of the timing of ADSL adoption (Ciapanna and Sabbatini, 2008). The latter is behind the choice of our instrumental variable.

To build our instrument, we exploit the fact that the 628 UGSs were inherited from the pre-existing voice telecommunication system, so their location was determined long before the advent of the internet.[6] As a consequence, the position of

[6] The network of physical infrastructures needed to provide voice telephony services to the Italian citizens was built in the post World War II period, between 1945 and 1960.

the telecommunication infrastructures was not influenced by the ADSL technology (Impiglia et al., 2004; AGCOM, 2011). Our IV builds on the assumption that ceteris paribus, the closer a municipality happened to be to a UGS when the ADSL became available, the more likely that municipality had access to high-speed internet earlier in the ADSL diffusion process.

2.2.2 The Italian Bank Credit Market

During the twenty years between 1980 and 2000, the Italian financial industry has changed substantially, modernizing its operations and performance. Following the implementation of the Second Banking Coordination Directive, Italy enacted a comprehensive banking law (”Testo Unico Bancario”) in 1993, which drastically reduced government ownership of banks. The share of assets in the hands of banks owned by central and local government or foundations accounted for 12%, from the 18% in 1998 and the 58% in 1990, (ABI, 2001). Under the joint effect of deregulation and technological changes, the system became much more ”market-oriented”, and substantial advances occurred in terms of the quantity, productivity and prices of banking services and the diversification, depth and efficiency of the markets (of Italy, 2003; Angelini and Cetorelli, 2003).

In the same period, the Italian banking system has undergone substantial restructuring. At the end of 2000, there were more than 800 banks, one-third of which were part of a banking group. The reorganization of the banking system took place mainly by ways of mergers and acquisitions, increasing the overall degree of concentration. The five largest groups in Italy held more than 50% of total banking assets, up from 35% in 1996. In this regard, Italy lined up with other European economies. However, the extension of individual banks’ branch networks and numerous competitors in the same markets heightened competition. A series of standard indicators confirms that the increasing concentration of the Italian credit system has come within a framework of intensifying competition.

The branch network at the beginning of the 2000s was also very dense. At the end of 2000, there was one branch for every 2,100 inhabitants (ABI, 2001), which means that about four-fifths of the population could choose between (at least) three banks in their town of residence. Similarly, the ATM network was very capillary, with 31,750 ATMs (more than one every 2,000 inhabitants) and POS terminals widespread, about 570,800 (more than one every 110 inhabitants).

The Italian financial system is mostly bank-dependent. In 2000, deposits and money market fund shares were equivalent to 87% of GDP. Loans accounted for about one-third of non-financial companies' outstanding liabilities, and this share was fast growing. Considering the average annual flows over the period 1998-2000, loans accounted for 55% of the total increase in firms' liabilities. The majority of these loans were granted by resident banks (of Italy, 2003).

Throughout the 2000s, Italian banks have supported the increasing demand for credit by non-financial firms through loosened supply conditions. Firm leverage has increased conspicuously between 2000 and 2007, moving from 34% to 39%. The available evidence (see Bugamelli et al. (2018) for a detailed review) suggests that the Italian banking system has sustained productivity growth before the great financial crisis by supporting firm-level innovation and exporting and by improving the allocation of capital across firms. These dynamics have been similar in other major European countries.

2.3 Data

Our final dataset combines information from several sources. It includes details on ADSL coverage and the infrastructural characteristics of the Italian municipalities between 2004 and 2008. It pairs these information with matched firm-bank data related to the period 1998-2008. We have information on the amount of credit granted by bank b to firm f, and the specific features of the credit relationship (loan type, presence of collateral, interest rate). Then, we also use data on the location and the opening and closing of bank branches during the period of analysis, together with data on bank employees and bank deposits. Finally, we gather information on the balance sheets of Italian non-financial incorporated firms.

Data on ADSL coverage are from the "Osservatorio Banda Larga", backed by the Italian Ministry of Telecommunications (Campante et al., 2018). The data include information on the percentage of households with access to ADSL-based services, for each municipality and year between 2005 and 2008, on an asymmetric six-point scale: 0%, 1%–50%, 51%–75%, 76%–85%, 86%–95%, and above 95%. No data are available for years prior to 2005. We view 2002 as the first year the ADSL technology became available (as discussed in the previous section). Hence, for years prior to 2002, we consider the percentage of households with access to ADSL equal to zero. We then use information from the Annual reports of the Italian Communications Authority (AGCOM) to retrieve data for 2004, whereas we treat 2002 and 2003 as missing.

Throughout the analysis, we use the asymmetric six-point-scale variable as our baseline measure of broadband internet access. However, in robustness checks we also experiment with alternative measures. First, we create dummy variables for *good access* (which takes value 1 if broadband access is above 50%) and *some access* (1 if broadband access is above 0%) to broadband. These measures facilitate the interpretation of the coefficients, as they do not rely on the asymmetric scale. Second, we define a proxy for good internet as the number of years since at least 50% of households in a municipality have had access to the ADSL. The latter has the advantage to provide a dynamic to the broadband effect but comes with the disadvantage of considering 2004 as the first year of ADSL adoption, introducing some noise in the first years of the sample.

Figure B.2 reports the distribution of access to broadband across Italian municipalities in 2004, the first year of data availability, and 2008, the last year considered in our sample, with darker colors indicating high or full access. The figure documents the rapid diffusion of high-speed internet throughout the country.

Data on ADSL coverage are complemented with those on the infrastructural features of the internet technology. In particular, we collect information on the number and geographical location of LSs and UGSs. Then, we compute the geodesic distance between the centroid of each municipality and the closest UGS and use this variable, interacted with a dummy post-2001, to instrument access to the broadband.[7]

Data on matched firm-bank relationships are from the Italian Credit Register (CR) held by the Bank of Italy. The CR contains information on the universe of loans and guarantees banks and financial companies issued to their customers above 30,000 euros (75,000 euros before 2008). For each credit relationship, we observe data on the bank and the firm involved, the total amount granted, the amount utilized, the composition of the credit (three loan types are distinguished: credit lines, term loans, loans backed by receivables), its status (performing or not) and the timing of the relationship. Moreover, we also observe the municipality of the branch that the borrower selects as the reference for the management of the credit relationship. This feature is essential as it allows to observe the bank's location with which the borrower interacts. It also allows us to match data on the loans issued to a firm by banks in different municipalities, with information on internet access in each municipality.

Data on interest rates are from *Taxia*, which is part of the CR. While a subset of the CR, it provides detailed information on interest rates covering more than 80% of total bank lending (Rodano et al., 2013). Such data include the rates charged on outstanding loans (distinguished into credit lines, term loans, and loans backed by receivables) and newly issued term loans.

Data on bank deposits and bank employees are from the Supervisory Reports that banks submit to the Bank of Italy, the banking supervisor of the country, during our sample period.

Data on bank branches are from the Bank of Italy "Lista succursali". For each bank branch, we observe its name, bank identifier, group to which it belongs (when relevant), location, and period of activity (initial and closing date).

We match data at the bank-level using the unique bank identifier.

Data on firm balance sheets are from the firm register collected by CERVED Group. These data provide balance sheets and income statements for the universe of incorporated firms in Italy from 1998-2008. Firms not covered are mainly small

[7] Data on the location of LSs and UGSs have been kindly provided by Francesco Sobbrio and are used in Campante et al. (2018).

firms (sole proprietorship or small household producers). Throughout our analysis, our sample includes all the firms covered by the CERVED database. We match these data with the credit data using the unique firm tax identifier.

Finally, we collect information on the local economic activity and the social background of different areas by using publicly available data from the Bank of Italy, the Italian national statistical institute (ISTAT), and the Ministry of Economy and Finance (MEF).

Table B.1 reports summary statistics associated with our final sample. Panel A refers to data at the municipality level and shows municipalities geographical distribution (*North*, *Center*, *South*), as well as statistics on access to broadband (*Internet*) and the ADSL underlying infrastructure (the number of LSs and their average distance from the municipality; the number of UGSs and their average distance from the municipality). Panel B refers to data at the bank-municipality level that we use as our baseline setting throughout the analysis. It shows the number of loans issued by a bank in a given municipality, the amount of credit granted, and the average interest rate charged. Finally, panel C refers to loan-level data and reports statistics on (granted) loan amounts.

2.4 Empirical Strategy

We test the effects of banks' broadband availability on several outcomes, at the bank-level, at the loan (bank-firm relationship) level, and the municipality level.

The existing literature provides some guidance on what could be the effect of broadband internet on bank credit. First, we expect credit outcomes, both the extensive (loans issued) and the intensive (amount of credit granted) margin, to be positively related to broadband (Petersen and Rajan, 2002; Berger, 2003). The effect of high-speed internet on interest rates, instead, is a priori ambiguous. Following the framework by Hauswald and Marquez (2003), the effect of new information technologies on credit price is negative (an increase in interest rates) when the informational advantage of the intermediary that gathers information leads to less competition in the credit market. On the contrary, the effect is positive (a decrease in interest rates) when access to information makes data much more widely and readily available to all competitors. In this case, an improvement in IT generates spillover effects that erode informational advantages and serve to level the playing field, with a consequent reduction of interest rates.

To estimate the intention-to-treat (ITT) effect of the increased access to broadband internet, we rely on the following specification:

$$Y_{(r)bmt} = \nu + \beta Broadband_{mt} + \gamma X_{(r)bmt} + \text{fixed effects} + \varepsilon_{(r)bmt} \tag{2.1}$$

where subscripts r, b, m, and t indicate, respectively, relationship, bank, municipality and year (r or b depending on whether the specification refers to credit relationship characteristics or aggregate bank characteristics); Y is the outcome variable; *Broadband* represents the percentage of households that have access to the ADSL in the municipality of the branch, based on the asymmetric six-point scale described in the previous section; X includes time-varying controls at the relationship or bank level; fixed effects are different sets of dummies depending on the outcome variable. The latter may include time (year), branch municipality, and bank-municipality of the branch fixed effects. Furthermore, in the regressions at the loan level, we saturate the model with the inclusion of bank-year, firm-year, and firm-bank-municipality of the branch fixed effects, in order to isolate the effect of broadband internet on credit supply.[8] We estimate equation (1) with standard errors clustered at the municipality level.

Our main outcome variables are the following. In the regressions aggregated at the bank level: i. the (log) number of loans issued by bank b, in municipality m, at time t, which measures the effect of broadband internet on the extensive margin of bank credit; ii. the total (log) amount of credit granted by bank b, in municipality m, at time t, which measures the intensive and the extensive margin combined; iii. the average interest rate charged by bank b in municipality m, which proxies for the price of credit.[9] In the regressions at the bank-firm relationship level, when we

[8] Notice that this is finer than firm-bank fixed effects, since it focuses on the relationship between the firm and the bank in a specific municipality.

[9] In the main analysis, we use a weighted average of bank's interest rates, where credit is not distinguished between different loan types (term loans, loans backed by receivables, credit lines)

turn on credit supply: i. the (log) amount of credit commitments by bank b to firm f, at time t, which evaluates the effect of broadband on the intensive margin of the credit relationship.

Next, we further explore the mechanisms through which broadband internet affects lending (section 6). We keep the same econometric specification at the bank level and focus on different outcome variables. We use indicators of bank productivity and credit quality to test the effect of broadband internet on the lending efficiency of the bank. We also study measures of distance and location of the borrowers to test for the geography of the loans once broadband internet is available. Then, we look at outcomes at the municipality level to test for local competition. In that regard, we concentrate on the number of local competitors and standard measures of market concentration. Finally, we use a mix of the above-mentioned specifications to provide evidence on the reallocation effects of broadband internet and the associated real effects.

A major concern with the estimates of a model in which we regress bank credit features on access to ADSL is that broadband availability is unlikely to be randomly allocated across municipalities, potentially generating a bias in the estimates of model parameters (Comin and Hobijn, 2004). We use an instrumental variable approach that exploits exogenous variation in broadband adoption across different geographical areas to deal with this concern.

We take advantage of the geographical distribution of ADSL physical infrastructures and leverage differences across Italian municipalities in the distance between a municipality and the closest UGS, where the latter represents a key determinant of the cost of supplying ADSL services. The underlying assumption is that the distance to the closest UGS affects the pattern of ADSL roll-out, with municipalities located farther away from UGSs getting access to broadband internet later on, ceteris paribus.

Even though the presence and the location of the UGSs precedes the development (and even the existence) of broadband in Italy, the spatial distribution of UGSs is itself non-random. To address this issue, we exploit the panel struc-

and weights depend on the amount of firms' utilized credit for each loan type.

ture of the data and add bank-municipality (or more granular, i.e., firm-bank-municipality) fixed effects to our specifications. Then, to account for potential time-varying confoundings, we interact the distance of a municipality to the closest UGS with a dummy for the post-2001 period (i.e., after the introduction of high-speed internet). The latter constitutes our instrument to ADSL coverage.

The main identifying assumption behind our IV is that whatever correlation existed between the distance to the closest UGS and relevant municipality characteristics, this did not change at the time of the introduction of the ADSL in the municipality. Indeed, we identify the effect of the change in the impact of distance on the outcomes of interest, under the assumption that any change in that impact occurs solely because of the availability of broadband internet (Campante et al., 2018).[10]

Hence, our econometric model relies on the following two-stage specification:

$$Broadband_{mt} = \rho + \delta DistanceUGS_m \times Post2001 + \omega X_{(r)bmt} + \text{fixed effects} + \xi_t + \epsilon_{(r)bmt} \tag{2.2}$$

$$Y_{(r)bmt} = \nu + \beta \widehat{Broadband}_{mt} + \gamma X_{(r)bmt} + \text{fixed effects} + \phi_t + \varepsilon_{(r)bmt} \tag{2.3}$$

where *Distance UGS* is the (time invariant) distance of the bank's municipality centroid to the closest high-order telecommunication exchange (UGS), and we interact this variable with a dummy *Post2001*, that takes value 1 for the years after 2001, and zero otherwise.[11] We estimate equation (3) via two-stages least squares (2SLS) regressions.

Table B.2, reports first stage estimates as presented in equation (2). Column 1 refers to our baseline measure of ADSL coverage. Columns 2 and 3 refer to the dummies *good access* and *some access*, which are equal to 1 if at least 50% or more than 0% of households have access to the internet, respectively. Finally, column 4 focuses on the *years since good internet* (the number of years since at least 50% of the population has ADSL access). As we can see from the table, coefficients are nega-

[10] In the appendix, table B.17, we provide a balance table that compares mean values of geographical and socioeconomic indicators for municipalities below and above the median of distance from the closest UGS.

[11] This approach is similar to that in Paravisini et al. (2015), Campante et al. (2018), Manacorda and Tesei (2020) and Guriev et al. (2021).

tive and statistically significant, in line with our underlying hypothesis. Moreover, F-statistics are generally high and well above the rule of thumb thresholds.

2.5 Results

This section shows the main results on the effects of broadband internet on bank credit in Italy. We first present motivating evidence using a difference-in-differences (DiD) setting. Next, we implement our preferred empirical strategy and show estimates from the instrumental variable (IV) analysis.

2.5.1 Motivating Evidence

To gain intuition, before implementing the two-stage specification as in equations (2) and (3), we run a standard DiD event study on the (log) number of loans and the (log) amount of credit granted by Italian banks.

We simulate a DiD setting and divide the sample into two groups. Treated banks are those in municipalities where at least 50% of the households were connected to the ADSL in 2006. Control banks are those in municipalities where ADSL was unavailable in 2006 or solely to a restricted share of the households. Then, we consider 2001 as the baseline year (in line with the main analysis) and show the heterogeneous effects of broadband internet at the extensive and intensive margin of bank credit.

Figure B.3 displays estimated event study coefficients, together with the corresponding 95% confidence interval. The top panel focuses on the number of loans. We consider the semi-dynamic regression proposed by Borusyak and Jaravel (2017), and drop the farthest (negative) relative year from the event, in addition to the baseline category.[12] First, we do not find evidence of pre-trends, meaning that the two groups of banks are on a parallel trend before the arrival of high-speed internet. Second, the treatment dynamics show the positive and statistically significant effect of broadband on bank credit. The effect of fast internet takes one year to

[12] Borusyak and Jaravel (2017) show that this is the minimum number of restrictions for point identification.

materialize and then monotonically increases, with a long-term effect amounting to roughly a 9% increase in the number of loans issued. The bottom panel focuses on the amount of credit extended. While standard errors are somewhat larger, the event study points to similar results. Treated and control banks are on a parallel trend before the treatment, and fast internet is associated with more credit.

Overall, these preliminary findings indicate a positive relationship between access to broadband internet and credit granted by banks.

2.5.2 The Effects of Broadband on Bank Credit

We now implement the baseline two-stage specification on our main dependent variables, following equations (2) and (3).

We first aggregate data at the bank-municipality-year level and focus on the extensive margin (the number of loans issued by banks to non-financial firms) and the intensive margin (the amount of credit issued) of bank credit. Results are shown in Table B.3. Columns 1 and 2 report the basic OLS estimates. The coefficients associated with broadband access are positive and statistically significant, indicating that the advent of high-speed internet is associated with more credit granted by banks. Interestingly, OLS coefficients are lower than the average treatment effect from the event study in section 5.1. As regards the extensive margin, the coefficient *broadband* is 0.007 (meaning that moving from zero to full broadband increases the number of loans granted by 3.9%), while the corresponding mean effect implied by the event study is around 6.7%.[13]

This difference may signal the short-term bias of the two-way fixed effects estimator when treatment effects take time to materialize.[14] In order to test this hypothesis, we follow De Chaisemartin and d'Haultfoeuille (2020) and compute the weights associated with the OLS estimates. Results are reported in Table B.4. In line with our hypothesis, more than 60% of the OLS weights are associated with

[13] Computed on the semi-dynamic model that also excludes 2002 and 2003, for which broadband data are missing in our dataset. Notice that estimates from a standard event study considering the first year of ADSL access provide a mean effect of 7%.

[14] Borusyak and Jaravel (2017) show that the OLS coefficient from two-way fixed effects staggered DiD is a weighted average of the OLS coefficients. When treatment effects are heterogeneous over time, those weights are higher for low values of the relative time from the event and may even become negative in the long run.

the first years after the treatment, and weights are strongly decreasing over time.[15]

Next, we instrument broadband access with the interaction between distance to the closest UGS and the post-2001 dummy and consider this as our preferred specification.[16] Columns 3 and 4 show the main results. We find a positive and statistically significant effect of broadband internet on bank credit. The effect is larger than in the basic OLS specification: moving from zero to high broadband coverage is associated with an increase in the number of loans issued of 12% (0.08 of a s.d.) and an increase in the amount of credit granted of 28% (0.13 of a s.d.).[17] This result is qualitatively in line with Liberti et al. (2016) that show an improvement in the allocation of credit when more and better information is available.

Then, we test the effects of having access to broadband internet on the price of bank credit. We aggregate data at the bank-municipality-year level and run a specification similar to equations (1) and (3). Results, shown in Table B.5, indicate that the average effect of high-speed internet on interest rates is negative and statistically significant. Passing from zero to high ADSL coverage is associated with a decrease in the average interest rate of 30 b.p. (0.18 of a s.d.).[18] The latter is in line with the theoretical argument by Hauswald and Marquez (2003) that, being broadband internet a general purpose information technology whose benefits are widely and easily available among bank competitors, the spillover effects from information dissemination dominate the negative effects from informational rents.[19]

[15] The downward bias of the OLS coefficient could also have different sources. First, coverage might be related to unobservable municipality characteristics associated with higher credit. Second, it could be related to the coarseness of the measure of broadband access, especially at the bottom, where going from 1% to 49% access would be entirely missed and yet have the strongest impact (if diminishing returns to broadband are present).

[16] Importantly, when we compute the weights for the reduced form regression of Y on our instrument, we find equal weights. Results are reported in Table B.4.

[17] "High" broadband coverage means a value of *internet* equal to 3 (at least 75% of the population connected). The latter is also the interquartile range of ADSL coverage for municipalities included in our sample during the period of broadband availability.

[18] This accounts for a reduction of 5% in the average rate, that is in line with Brown and Goolsbee (2002) which find that the growth of the internet has reduced term life prices by 8–15 percent.

[19] Schenone (2010) finds evidence of these spillover effects in initial public offerings (IPOs), using the Securities Data Company (SDC) dataset by Thomson Reuters.

2.5.3 The Supply Channel

The findings in the previous section represent equilibrium outcomes. Indeed, the effect of broadband internet on bank credit results from two forces acting simultaneously: credit demand (firms that ask for credit) and credit supply (banks that offer credit).

The effect of fast internet on credit demand has been indirectly investigated in the literature by studying firm productivity (Akerman et al., 2015; DeStefano et al., 2018). Although there is consensus on the fact that banks use cutting-edge technologies to deliver innovative products, streamline loan-making processes, and improve back-office efficiency (He et al., 2021), evidence on the direct effect of broadband on banks' productivity is relatively scant.[20] Moreover, the fact that lending increases while interest rates go down is difficult to reconcile with a pure demand-side story.

In this paper, we further contribute to the existing literature by isolating the component of the total effect of broadband on bank credit due to credit supply.

We perform this exercise in Table B.6, where we exploit the granularity of our dataset to further characterize the effect of broadband on the intensive margin of the credit relationship. We use loan-level data and leverage variation within firm-bank (Ali Choudhary and Limodio, 2021).[21] Columns 1 and 3 report OLS and 2SLS estimates related to this specification. Passing from zero to high broadband coverage is associated with an increase in the amount of credit granted by bank b to firm f of 41% (0.28 of a s.d).

Next, following Khwaja and Mian (2008), we saturate the model with the inclusion of firm by year fixed effects that capture the component of the total effect that is related to demand. Importantly, we also add bank-time fixed effects to attenuate the concern that the (exogenous) arrival of broadband internet simultaneously

[20] In the appendix, figures B.5 and B.6, we report descriptive statistics on the % growth of the use of web technologies within bank branches, during the period 2001-2007. These statistics, obtained from the Economic Analysis of the Italian Banking Association (ABI), suggest a sensible increase in the use of web technologies in the back office activities of banks during the examined period.

[21] These specifications at the firm-bank-municipality-year level always include controls for the loan (one-year lagged) share of revolving loans of the firm, and the loan share of extended credit of the issuing bank

affects both the supply and the demand for credit. Finally, this specification also includes firm-bank-municipality of the branch fixed effects to capture time-invariant characteristics of the credit relationship and possible confounders related to the specialization of different bank branches.[22]

Estimates are reported in columns 2 and 4 and refer to the effect of broadband internet on credit supply. As we can see from the table, almost half of the total effect is a pure supply-side effect.[23] Going from zero to high broadband coverage is associated with an increase in bank credit supply of 19% (0.14 of a s.d.).

2.5.4 Robustness

We subject our main results to several robustness checks.

First, we verify that our instrument passes the Angrist and Imbens (1995) instrument's monotonicity test. By instrumenting our endogenous variable (D_{mt}, Z=[0,1]) with the distance to the closest UGS interacted with the dummy post-2001 (Z_{mt}), we are implicitly assuming that the effect of distance on access to broadband is monotone, that is, either $D_{mt}^{high} \geq D_{mt}^{low}$ or $D_{mt}^{high} \leq D_{mt}^{low}$, $\forall mt$. The assumption is not verifiable, but has testable implications on the CDFs of internet for municipalities close or far from the UGS, that is, they should never cross. In fact, if $D_{mt}^{high} \geq D_{mt}^{low}$ with probability 1, then $Pr(D^{high} \geq j) \geq Pr(D^{low} \geq j), \forall j \in supp\ D$ (Decarolis and Rovigatti). Figure B.7 plots the CDFs of internet for banks close to the UGS (blue solid line) and far from the UGS (red solid line).[24] Since the two CDFs never cross, the instrument passes the test.

Second, our basic identification assumption would be violated if there are underlying trends affecting the outcomes of interest that correlate with our instrument. To control for these confounding factors, we augment our specifications in equations (2) and (3) with several economic and socio-demographic municipal characteristics available from the 2001 Census, interacted with a second-order

[22] Notice that this specification is more demanding than that used in Paravisini et al. (2015) as we observe the municipality of the specific branch that manages the loan.

[23] These results always have to be interpreted with the caveat that the variation used in these specifications is that of firms that borrow from multiple banks.

[24] Notice that, for this exercise, we use a dummy variable below/above the median of UGS distance to proxy for our instrument (that is a continuous variable).

polynomial-time trend. In this way, we control flexibly for the possibility of differential time trends. The baseline group of controls includes the natural logarithm of the total population, elderly population, the number of private firms operating in the municipality, the number of employees, the distance from the provincial capital.[25] Table B.18 reports the estimates related to the extensive and the intensive margin of bank credit. All the coefficients keep the same sign as in the baseline specification, and they remain statistically significant at standard levels. Furthermore, the magnitudes are not sensibly affected by the inclusion of these control variables.

Next, we test the robustness of the main results by using different measures of ADSL coverage. Table B.19 focuses on the extensive margin of the credit relationship. In column 1, we report our baseline IV estimate for reference. In column 2, we use *Years Since Good Internet* and find a similarly positive effect. One more year of good internet is associated with an increase of 6.3% in the number of loans. Considering an average of four years of good internet, the total effect of broadband on credit is an increase of 25%. In this case, OLS and IV estimates get much closer to one another, showing the importance of considering the treatment effect dynamic in our analysis. In columns 3 and 4, we consider two alternative dummy variables: *good access* and *some access*. The coefficients are close in magnitude to our baseline *Broadband* measure, although they slightly vary between each other.

In Table B.20, we address the possibility that our results are picking up some underlying trend in credit that happens to be correlated with the diffusion of broadband. We run placebo IV specifications for the years from 1998 to 2003, assuming that the level of ADSL access in 2006 was already present in 2001 (and the following years). Reassuringly, we see no impact of this fictitious introduction of broadband internet, supporting the view that pre-existing trends do not drive our findings.

[25] *Population* is the only variable for which a yearly time series is available.

2.6 Mechanisms

Results so far show that the arrival of broadband internet leads to an expansion of bank credit. This materializes on the extensive margin (number of loans granted), on the intensive margin, and on the price of credit issued by banks (interest rates go down). Moreover, part of the total effect is due to factors independent of credit demand, as we have seen that loan amounts are affected by access to broadband, which also controls for firm-specific time-varying unobservables. In what follows, we further explore the channels through which all these effects take place.

2.6.1 Lending Efficiency

It is often argued that IT advances play a substantial role in boosting productivity (Draca et al., 2007). It is thus essential to test whether bank lending efficiency increases as a consequence of broadband internet availability.

We measure bank lending efficiency using two different indicators. First, loans per bank employee, which is a measure of the bank's labor productivity. Second, the share of non-performing loans (NPLs), which proxies banks' credit quality.

Petersen and Rajan (2002) suggest that new technologies allow banks to collect more hard information about borrowers, enabling them to change the nature of lending from an emphasis on strict ex-ante screening and costly ex-post monitoring, to fine-tuned screening and frequent ex-post monitoring. Similarly, Berger (2003) documents the increase in profit productivity due to improvements in "back-office" technologies, as well as consumer benefits from new "front-office" technologies. Since high-speed internet enhances screening and monitoring, we expect the effect of broadband on productivity to be positive and significant. On the other hand, the effect of broadband on credit quality is a priori ambiguous. Since marginal borrowers are generally worse than incumbent customers, credit quality could worsen as a consequence of credit expansion. However, hard information on borrowers becomes richer and timely once new technologies are available. Improved screening and monitoring activities can thus offset the negative effect of credit growth.

Table B.7 shows the results. Columns 1 and 3 show that bank's access to broadband internet has a positive effect on its labor productivity, measured as the amount of extended credit per bank employee. Going from zero to high broadband coverage is associated with an increase in credit per employee of 24%. On the other hand, columns 2 and 4 show that high-speed internet is associated with a slight decrease in the share of NPLs, meaning that credit quality, on average, improves with the expansion of credit.[26]

Taken together, these two findings support the thesis that banks' overall lending efficiency sensibly increases after the introduction of broadband internet.[27]

2.6.2 Banks' Geographical Reach

Lending is traditionally a "local" business, and the distance between lenders and borrowers is a crucial factor shaping credit relationships, especially those that involve SMEs (Degryse and Ongena, 2005). However, Petersen and Rajan (2002) suggest that technology helps break the "tyranny of distance". By improving screening and monitoring activities of banks, new technologies allow for increasing capital intensity of lending and thus lending to more distant borrowers. Along the same lines, Berger (2003) shows that technological progress facilitates the geographic expansion of banking organizations by reducing distance-related dis-economies. New services created by technological progress with higher value added, traditional banking services delivered more efficiently, bank monitoring and the control of risk exposures at longer distances and lower costs, and reduced managerial diseconomies of distance all contribute to ease the way banks find and finance new clients.

On the other hand, Wilhelm (2001) argues that advances in communication technology and increased capacity for information do not imply greater exchange

[26] The latter is in line with Pierri and Timmer (2020), which study the implications of IT in banking for financial stability. The authors find that pre-crisis IT adoption that was higher by one standard deviation led to 10% fewer NPLs during the 2007–2008 financial crisis.

[27] Casolaro and Gobbi (2007) find that banks adopting IT capital intensive techniques are more efficient and interpret the latter as evidence of a catching-up effect consistent with the usual pattern of diffusion of new technologies.

of information inside the bank. This is due to the limited incentives for loan officers to transfer information on which they hold monopoly power. Similarly, advances in communication technology may not lead to more exchange of information between firms and banks (Bhattacharya and Chiesa, 1995) and between different banks (Padilla and Pagano, 1997). In this regard, technological developments may have no effects on the distance between lenders and borrowers.[28]

To address these critical issues, we look at the effects of broadband internet on the geography of the credit relationship by focusing on new loans originated by Italian banks during the period of our analysis.[29] We define a dummy variable for the loan being originated outside the province of the branch (*Diff. Province*) to measure the effect of broadband internet on the geographical borders (the market) within which the bank operates.[30] Then, we create a direct measure of the distance between lenders and borrowers by computing the geodesic distance between the centroid of the municipality of the branch that manages the loan and the exact location of the firm.

Estimates from model (3), aggregated at the bank-municipality-year, are reported in Table B.8. Columns 1 and 3 refer to the share of new loans originated out of the province of the branch. Columns 2 and 4 refer to the inverse hyperbolic sine of the geodesic distance between the municipality of the branch and the firm. The table shows that access to high-speed internet increases the probability that the bank extends credit outside its province. At the same time, broadband internet is associated with firm-bank relationships exhibiting longer distances.[31] Findings are in line with the literature and document the shrinking effects of new information technologies on the distance between lenders and borrowers (Petersen and Rajan, 2002; Berger, 2003; Felici and Pagnini, 2008).

[28] See also Degryse and Ongena (2005) for empirical evidence on the static nature of the relationship between firms and banks in Belgium, between 1973 and 1997.

[29] To substantiate our hypothesis that broadband reduces communication costs and increases "proximity", in Table B.21 of the appendix, we preliminary show that broadband access is positively associated with the share of loans granted to firms connected themselves to broadband.

[30] In Italy, before the advent of fast internet, provinces defined the borders of bank credit markets (Crawford et al., 2018).

[31] In Table B.22 of the appendix we propose the same exercise at the firm level, where we can control for firm fixed effects (that capture time-invariant firm characteristics). As we can see, results are qualitatively in line with those just presented.

These results also suggest that broadband internet can trigger a re-definition of local credit markets, with all the consequences for agents involved in the credit relationship and for regulatory and supervisory authorities.

2.6.3 Competition

The expansion of credit following the arrival of broadband internet may be driven by tougher local competition. Hauswald and Marquez (2003) find that when the information gap between banks becomes smaller, because of ICT diffusion, there is a softening of the winner's curse that leads to an increase in the intensity of bank competition. Similarly, Vives and Ye (2021) find that when IT progress involves a weakening in the influence of bank-borrower distance on monitoring costs, then banks' competition intensity increases. Finally, Felici and Pagnini (2008) show that new communication and information technologies have significant pro-competitive effects in local banking markets. By increasing the geographical reach of bank entry decisions, these new technologies augment local credit market contestability.

The effects of broadband internet, and more efficient information technologies in general, on banking competition, is an interesting question in itself. We explore this question in what follows, focusing on two measures of competition: the number of (physical) bank competitors in the municipality; and measures of concentration of the local credit market.

Vesala (2000) shows that loan mark-ups were decreasing in Finland, in lock-step with the rapid development of the internet. On the other hand, Gropp et al. (2009) find only a small increase in contestability in the European loan markets despite the impressive technological advances experienced in many countries.

Our results on the effects of broadband on local competition are shown in Table B.9 and Table B.10. In Table B.9, the main dependent variable is the (log) number of bank competitors in a municipality. Estimates from Table B.9 show a significant increase in local competition.[32] Indeed, the number of banks competing in the market increases when the municipality is reached by fast internet.

In Table B.10, the main dependent variables are standard indicators of concen-

[32] Competition within the municipality.

tration of the local (municipality) credit market: the concentration ratio of the top 5 and 3 banks, and the Herfindahl–Hirschman Index (HHI), computed using data on bank deposits.[33] As we can see from the table, all the coefficients are negative and statistically significant at the 99% level.

Overall, high-speed internet is associated with an increase in competition in the local credit markets. Results are in line with Frame et al. (2018) on the effects of new technologies on banking.

2.7 Extensions: Digital Highways, Credit, and the Real Effects of Broadband

In this section, we present additional results associated with the arrival of broadband internet on the activity of banks, with a focus on credit allocation and the spatial distribution of the effects.

2.7.1 Credit Reallocation

An extension that helps characterize our main results relates to the effects of access to broadband internet on credit allocation. The expansion of credit documented in the previous sections may coincide with an enlargement of the set of borrowers served by the banking industry. Otherwise, it can result from an increased amount of credit granted to borrowers that already benefited from banking services. The latter is particularly likely to occur in the case of credit flows from branches of the same bank, where transaction and operating costs are relatively lower (Cetorelli and Goldberg, 2012a,b).

We analyze these possibilities in Table B.11. Columns 1 and 3 test whether new loans originated by banks reached by fast internet are issued to firms already having a credit relationship. Columns 2 and 4 focus on new loans originated towards firms that already have a credit relationship with the same bank (in a different municipality). Results show that banks connected to high-speed internet have a higher

[33] As robustness, we replicate the same estimates by computing standard indicators of competition using data on extended credit. Results are similar to those in Table B.10.

probability of granting a loan to firms out of the credit market before. Moreover, they also have a higher probability of serving firms with a relationship with the same bank.

These findings provide evidence on the complementary effects of broadband on bank credit allocation. First, access to broadband internet determines an expansion in the set of borrowers served. Second, it allows banks to implement internal capital reallocation.

2.7.2 The Spatial Distribution of the Effects: Does Broadband Internet Create Banking Deserts?

The previous sections show that fast internet leads banks to expand their geographical reach (outside of their province) and reallocate part of their credit within their internal organization. These dynamics may lead to a movement away from small municipalities, typically in the countryside, towards larger municipalities.

Physical infrastructures as highways and railroads, for example, are known to have affected the spatial allocation of economic activities. The direction of their effects has been heterogeneous between rural and urban areas, the periphery and the center, places crossed by the facility and adjacent areas (Rephann and Isserman, 1994; Chandra and Thompson, 2000; Baum-Snow, 2007; Michaels, 2008; Atack et al., 2010; Banerjee et al., 2012; Duranton and Turner, 2012; Donaldson, 2018). Digital infrastructures, sharing some of the underlying features, may have similar effects.

To explore this issue, in this section, we focus on the heterogeneity that distinguishes between small and bigger municipalities and show the existence of potential winners and losers from the process of broadband diffusion (Akerman et al., 2015). This heterogeneity is of particular interest because it assesses whether technological progress can determine credit desertification and, eventually, local economic stagnation. Furthermore, it is relevant in the political economy literature as the internet has been identified as a potential driver of polarization of the political spectrum and behind the recent rise of populist parties.

Table B.12 replicates the analysis in Table B.6 for small municipalities. We exploit the granularity of our dataset and test the effect of broadband internet on the intensive margin of the credit relationship. We use loan-level data and leverage variation within firm-bank. Columns 1 and 3 report OLS and 2SLS estimates related to this specification. Then, we follow Khwaja and Mian (2008) and concentrate on the supply of loans by saturating the model with the inclusion of firm-year fixed effects, which capture demand-side confounders. The coefficients associated with this specification are reported in columns 2 and 4 and isolate the effect of broadband internet on credit supply. As we can see from the table, there seems to be no effect of fast internet on the amount of credit granted to non-financial firms in small municipalities.[34]

In the appendix, tables B.24 to B.28, we also replicate the analysis on the mechanisms behind our results and show that for banks in small municipalities: lending efficiency goes down (in particular, productivity goes down), competition increases, and banks do not react in terms of loans geographical expansion and credit reallocation.[35]

The test in Table B.12 highlights how broadband does not have any positive effect on credit supply for branches located in small municipalities. In this section, we argue that a reshaping can explain part of this null effect in the markets served by banks in different municipalities.

As we have argued in the introduction, broadband internet is a multi-dimensional information technology that reduces information asymmetries, communication costs, and agency problems in the banking industry. To work properly and activate a cheaper and timely communication channel between lenders and borrowers, both agents need to be connected. One possible explanation behind the null effect of broadband on credit in small municipalities is that firms in such places get access to larger credit markets. Firms in small municipalities reached by fast internet may become the "easiest target" for banks in bigger municipalities connected to broad-

[34] In Table B.23 of the appendix, we report the results based on the specification on interest rates. Our results show that access to broadband internet is associated with an increase in the price of credit (contrary to the specification that refers to the entire sample). The increase in the average rate in small municipalities connected to high-speed internet can be explained, in part, by the nature of the credit relationships that remain anchored in those municipalities. These credit relationships usually rely on soft information and are less sensitive to credit price (Ioannidou and Ongena, 2010).

[35] Even if some of the IV coefficients are not statistically significant at standard levels.

band. As a result, they can exploit this situation to borrow a larger amount of money from outside the municipality and rip better credit prices.

We test this hypothesis in Table B.13. Columns 1 and 2 report OLS and IV estimates where the dependent variable is a dummy equal to one when the loan is from a bank in a bigger municipality to a firm in a small municipality, out of the province of the bank, and connected to fast internet.

The results from Table B.13 provide supporting evidence that improved lender-borrower communication increases the probability of credit relationship formation. Firms in small municipalities reached by fast internet face new (and broader) investment opportunities and reallocate part of their borrowing out of the municipality towards banks that operate in larger cities (which offer higher amounts of credit, at a lower cost). Consequently, both borrowers in small municipalities and banks in bigger municipalities benefit from the arrival of broadband internet. The former, as they face higher credit supply at a lower price. The latter can reach borrowers that were previously out of their market and can expand their customer base.

This phenomenon of credit centralization has the potential to create local banking deserts. In the same direction, it is interesting to analyze whether banks react to the new technological advancements with a change in their geographical location. We investigate this aspect by looking at the evolution of bank branches.

Italy was in a phase of sensible expansion of the number of branches since the beginning of the 90s. Figure B.4 provides evidence in this direction. Furthermore, although we are used to associate informatization with de-branching, following the idea that automatic lending diminishes the value of geographical proximity and so the relevance of local branch presence (Kroszner and Strahan, 1999; Berger and DeYoung, 2001; Petersen and Rajan, 2002), the substitution of a brick-and-mortar model with a click-and-mortar model of banking is a very recent phenomenon.[36]

[36] The latter is probably due to the limited diffusion of smartphones and software allowing for sufficient cyber-security, until recent times. Different studies have attempted to measure the impact of technology on branching empirically. Degryse and Ongena (2004) argue that new technologies may have only a limited impact on branch presence because of the importance of bank branch proximity for customers. Keil and Ongena (2020) show that broadband and mobile internet access explain well the recent de-branching of banks at the country level, but not that at the US county or bank branch level.

Therefore, during the period of our analysis, the choice that Italian banks faced was not whether to open or close branches but rather whether and where to open new branches.

In Table B.14 we test the (heterogeneous) effects of having broadband internet on the number of branches in a municipality. Columns 1 and 3 refer to the whole sample, and columns 2 and 4 focus on the sample of small municipalities. Results from the table show that banks tend to increase their branch presence in places reached by broadband, where they can exploit the potential of fast internet. However, this does not happen in small municipalities, in which the effect of broadband is null. Indeed, we consider the latter as another sign of credit centralization from the perspective of the bank.

To conclude our analysis, we check whether these developments, credit flows towards bigger municipalities that lead to local credit desertification, are accompanied by economic underdevelopment in small areas.

Table B.15 and Table B.16 replicate the analysis in equations (2) and (3), where the dependent variables are: the natural logarithm of population, the natural logarithm of income, and the natural logarithm of income per capita in the municipality.[37] Table B.15 accounts for the entire sample. Table B.16 focuses on the restricted sample of small municipalities. As we can see from the tables, the effect of broadband internet on the real economy is generally positive and more pronounced in small municipalities. Even if not conclusive, these findings suggest that new credit flows allowed by high-speed internet are the expression of broader investment opportunities and alleviated financial frictions, rather than bank desertification followed by local economic underdevelopment.

2.8 Conclusion

In this research, we provide empirical evidence on the effects of broadband internet on bank credit to non-financial firms. To address this point, we combine data on

[37] Data on income are from the publicly available dataset of the Italian Ministry of Economy and Finance (MEF). The time series starts in 2000, meaning that we have no information for the period 1998-1999.

access to the ADSL technology in Italy with firm-bank matched data from the Bank of Italy. We follow 901 banks in 5271 municipalities, during the period 1998-2008, and show the effects of broadband at the extensive and at the intensive margin of the credit relationship and on credit price.

Our quasi-experimental design relies on the staggered adoption of the ADSL technology across Italian municipalities and an instrumental variable strategy that exploits the municipality's position in the pre-existing voice telecommunications infrastructure.

To explore our research question, we implement two-stages least squares analysis and focus on the effects of broadband on credit by isolating the effects on credit supply, on interest rates, and the underlying mechanisms that elucidate our main results.

Our findings highlight that high-speed internet fosters bank credit towards non-financial firms. The total amount of credit increases with broadband availability, while the average interest rate goes down. Many channels contribute to this aggregate effect. The internal efficiency of banks goes up as a consequence of broadband access. Banks reached by fast internet expand their markets and increase the distance towards their borrowers. At the same time, local competition increases, as reflected by the growth in the number of physical branches and competitors and by standard indicators of competition. Finally, banks connected to broadband internet tend to reach new borrowers and implement internal credit reallocation.

The effect of broadband, however, is heterogeneous. Access to high-speed internet creates digital highways that carry bank credit from the periphery to the center (i.e., from small municipalities to bigger municipalities). Nevertheless, credit desertification in small municipalities does not lead to economic underdevelopment, showing off the virtues of the credit flows generated by internet technologies.

Overall, our results are consistent with high-speed internet promoting bank credit and creating new credit opportunities for non-financial firms.

To conclude, our paper directly speaks to policymakers as we document the multifaceted effects of investments in new hi-tech infrastructures. The latter can serve as a guide for the introduction of future technologies, as the ultra-high-speed internet and the 5G mobile technology. Moreover, the issue of the effects of technological innovation on the operativity and the structure of the banking system is of utmost importance for central banks, both as monetary policy authorities (since it involves banks' lending activity) and as micro and macroprudential supervisors (since it involves banks' risk profiling).

3. Living on the Edge: the Salience of Property Taxes in the UK Housing Market

FRANCESCO NICOLAI, MARCO PELOSI AND SIMONA RISTESKA[1]

3.1 Introduction

A standard tenet of economic theory is that the statutory incidence of taxes is irrelevant to their economic incidence.[2] It should also be the case that whether a tax is paid at the moment of transaction or later is irrelevant to its incidence, as long as we consider the time value of money and the riskiness of the cash flows. By looking at the UK residential property market, this paper shows that this is not the case and that deferred taxes have a markedly lower incidence compared to taxes paid at the time of decision-making.

Together with France, the United Kingdom is one of the few countries receiving a sizeable fraction of revenues from property taxes, amounting to about 4.3% of GDP or more than £84 billion in 2016 (European Commission, 2018). The two main taxes levied on domestic properties are the Stamp Duty Land Tax and the council tax. The former is a tax levied on the transaction value of land and any buildings and structures thereon. The fact that its statutory incidence falls on the buyer, who is required to pay the tax liability to the HM Revenue and Customs within very few weeks from the completion of the transaction, and the fact that the tax represents a lump sum ranging between 1% and 7% of the property value are features that make the stamp duty tax particularly salient at the moment of purchase. The latter,

[1] We are grateful to Vicente Cuñat, Daniel Ferreira, Dirk Jenter, Christian Julliard, Daniel Paravisini, Andrea Tamoni, Michela Verardo and the participants at the LSE PhD seminar for the useful comments on the paper. We thank Vittorio Raoul Tavolaro for invaluable research assistance. The paper contains HM Land Registry data © Crown copyright and database right 2019. The data is licensed under the Open Government Licence v3.0. 1. We thank the University of Glasgow - Urban Big Data Centre for providing Zoopla property data. Zoopla Limited, © 2019. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data, 2019 [data collection]. University of Glasgow - Urban Big Data Centre.

[2] Kotlikoff and Summers (1987) provide a detailed review of classical theory on tax incidence.

which will be the focus of the present paper, is a tax levied by the local government on a yearly basis.

The council tax is levied on the resident, as opposed to the house owner, and is based on the property value in 1991. While the council tax is extremely salient when it needs to be paid, we will show that this is not the case when properties are purchased even though, in present value terms, it is similar to or even larger than the stamp duty tax.

By using the geographical discontinuity at the border of different local authorities in the London area, we can estimate the incidence of the council tax on property prices and contrast it with the incidence of the stamp duty tax estimated, among others, by Best and Kleven (2018). We rely on two empirical strategies. First, we divide the area of London into a grid of squares of equal size. We then limit our sample to squares that contain similar properties sold in the same year in two different councils, and assign a fixed effect to each square of the grid, so that we can estimate the council tax incidence using within-square regressions. For a sufficiently small square, our identification assumption is that the only difference between the two properties is the associated council tax.[3]

The second identification approach uses a matching estimator. In this case, we restrict the sample to dwellings whose distance to the nearest border is lower than 500 meters. For each house, to find the closest match on the opposite side of a border, we rely on the Euclidean distance between the two houses' features and on a linear model (that uses the same features) to estimate the model-implied prices. In both cases, we choose as a match the property with the lowest Euclidean distance or with the smallest difference between the linear model-implied prices. Finally, similar to the previous approach, we run within-pair regressions to compute the incidence of the council tax.

The London area is particularly suitable for the estimation because of the sharp nature of the council borders and the large dispersion in council tax rates across Boroughs. For instance, Figure C.1 depicts a road at the border of the Borough of Westminster and the Borough of Kensington and Chelsea. As can be seen from the

[3] The smallest square has a 200 meters side length.

picture, the houses on both sides of the street are otherwise identical except for the fact that they pay quite a different council tax amounts: the ones on the left pay £2,279 per year in council tax while those on the right pay £1,421 per year.

Suppose we discount the future payments as a perpetuity at a rate of 4%, similar to the mortgage rates observed in the sample. In that case, the difference between the two present values amounts to £21,450 (about $28,000). The tax differentials become even more significant once we consider the fact that many London Boroughs share services, such as waste management, and that many other amenities, such as access to parks, schooling, and religious facilities, are not strictly limited to residents of a given Borough. In Section 3.4 we show that the price of similar properties on opposite sides of a border does not adjust for differentials in council tax amounts.

By employing the above estimators, we establish that the council tax incidence is never statistically negative. We then proceed in Section 3.5 to set up a model where down payment-constrained households purchase a house and pay two sets of taxes: a lump sum stamp duty tax levied at the moment of purchase and a periodic council tax. We move on to perform a Bayesian analysis in Section 3.5.1 where we provide a posterior range for council tax incidence using priors that are economically motivated by our rational model.

In all these estimates, the incidence of council tax on property prices is too low relative to existing estimates of the incidence of other property taxes, even after accounting for the time value of money and the fact that discount rates might be larger because of borrowing constraints. These findings can be rationalised in a model where agents neglect taxes levied in the future.

We show in Section 3.5.2 that then a trade-off between the two types of taxes arises: the stamp duty tax is distortionary because agents are liquidity constrained; on the other hand, the council tax leads agents to over-consume the housing good and, therefore, distorts their consumption choices by reducing available income. As a result, we demonstrate that the Government can optimally tune the two taxes to minimize distortions for a given level of revenue.

The present paper adds to the burgeoning literature on behavioral public finance and the salience of taxes (or the lack thereof). Chetty et al. (2009) is the first paper to empirically estimate how salience differences can alter economic agents' behavior. They intervene in a grocery store to modify the salience of sales taxes and show that the incidence on buyers is largely reduced when taxes are made entirely salient. In a second experiment, they compare the effect of excises taxes, which are included in posted prices, and sales taxes, which are not explicitly included, on alcohol demand and again show that tax salience plays an important role in consumer behavior.

The setting in the present paper is quite similar to the second experiment in Chetty et al. (2009), given that the stamp duty tax is paid upfront while the council tax is deferred and thus less salient. For policy reasons, however, the question of property taxes is of greater importance because of the large amounts of money involved and the fact that it is very difficult for agents to learn since buying a new property is typically a once-in-a-lifetime event.

Following Chetty et al. (2009), other papers have also explored the question of tax salience. For instance, Feldman and Ruffle (2015) and Feldman et al. (2018) have replicated the findings in laboratory experiments, while Finkelstein (2009) similarly shows that the introduction of electronic toll payments raises toll expenditures. Taubinsky and Rees-Jones (2018) further explore the topic by showing a significant variation in how agents react to tax salience and investigate policy implications.

The present paper is also similar to Allcott (2011) who demonstrates a similar bias in the automobile market. Namely, car buyers fail to correctly price in the future energy cost at the time of purchase. As in Allcott (2011), our conclusions also rely on the choice of an appropriate discount factor. We show in Section 3.5.1 that the bias persists even after allowing for large discount rates. In a similar vein, using Norwegian data, Agarwal and Karapetyan (2016) explore the effect of non-salient debt features on households' purchasing decisions and show that they do not fully factor in the added cost. The authors show that the mispricing was eliminated once these features became fully salient.

Finally, the paper extends the literature on property taxes; among others, we use the results of Besley et al. (2014) and, in particular, Best and Kleven (2018) to compare our estimates of the council tax incidence with their stamp duty incidence estimates to highlight the lack of salience of the former.

The rest of the paper is organized as follows: Section 3.2 describes the data and the institutional setting; Section 3.3 gives evidence of the geographical distribution of council taxes and points out that this can significantly bias our estimates if not appropriately taken care of, before proceeding with the details of our identification strategies; Section 3.4 presents the empirical estimates of the council tax incidence; Section 3.5 develops a stylized model to help interpret the findings and show that the estimated incidence is too low to be consistent with fully-salient taxes, and explores some policy implications; and finally, Section 3.6 summarises and concludes the paper.

3.2 Data

To estimate the incidence of council tax, we need access to data on property characteristics and house prices, as well as council taxes paid. Data on house transaction prices are readily available from the HM Land Registry website. This dataset contains information about all residential properties transacted in England and Wales from 1995 that have been sold for full market value.[4] The dataset comprises of the price paid, the transaction date, and, most importantly, the house address, allowing us to pinpoint the exact location of every property. Additionally, the data provide us with information on the property type, which can be one of five possible categories (a detached, semi-detached, or terraced house, a flat/maisonette, and other), the age of the property (classified into old or new to distinguish between newly built properties and already established buildings) and the duration of tenure, i.e., whether the property is under freehold or leasehold.[5]

Since we would ideally like to compare properties as similar to each other as possible, we need more information on property characteristics. For this purpose,

[4] Data excluded from the dataset include commercial transactions, property transactions that have not been lodged in with HM Land Registry, and transactions made below market value. For more details on the property sales not included in the dataset, the reader can visit the HM Land Registry website: https://www.gov.uk/guidance/about-the-price-paid-data.

[5] Note that leases of seven years or less are not recorded in the dataset.

we use two additional datasets: the Zoopla Property data and Domestic Energy Performance Certificates.

Zoopla has collected the Zoopla Property data, one of the UK's leading providers of property data for consumers and property professionals, giving free access to information on 27,000,000 property records and up to 1,000,000 property listings and 15 years of sold prices data.[6]

The dataset covers the period between 1st January 2010 and 31st March 2019 for Great Britain (England, Wales, Scotland) properties. The dataset contains details on characteristics such as property location, property type, whether the property has been categorized as residential or commercial, number of bedrooms, number of floors, number of bathrooms, number of receptions, and whether the property is listed for sale or rent.[7] In addition, we also have access to the asking price for both rents and sales. However, we use the more accurate transaction price from the HM Land Registry dataset.

The second source of house characteristics comes from the Ministry of Housing, Communities, and Local Government. One can access the Energy Performance Certificates (EPC) for domestic and non-domestic buildings on their website. For domestic properties, before 2008, certificates could be lodged on a voluntary basis. From 2008 onwards, however, it has become mandatory for accredited energy assessors to lodge the energy certificates. Consequently, the data coverage drastically improves around that time, as does our ability to match these data with the price paid data. More specifically, the matching rate jumps from about 50 percent to over 90 percent around 2008. The dataset contains information on the location, property type, total floor area, number of storeys, number of rooms, floor level,

[6] The access to the dataset has been kindly provided by the University of Glasgow - Urban Big Data Centre. Access to the dataset for research purposes can be obtained directly through the Urban Big Data Centre. Zoopla has collected the data. Zoopla Limited, © 2019. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data, 2019 [data collection]. University of Glasgow - Urban Big Data Centre.

[7] Property types include: barn conversion, block of flats, bungalow, business park, chalet, château, cottage, country house, detached bungalow, detached house, end terrace house, equestrian property, farm, farmhouse, finca, flat, hotel/guest house, houseboat, industrial, land, leisure/hospitality, light industrial, link-detached house, lodge, longère, maisonette, mews house, mobile/park home, office, parking/garage, pub/bar, restaurant/cafe, retail premises, riad, semi-detached bungalow, semi-detached house, studio, terraced bungalow, terraced house, townhouse, unknown, villa and warehouse. We keep only properties categorized as residential and listed for sale.

and height, along with many indicators of energy efficiency and quality of glazed surfaces. The final piece of data needed to conduct our analysis is related to council tax data; in the following section, we will describe in more detail the functioning of this property tax and the relevant data.

3.2.1 Council Tax

The taxation of properties in the United Kingdom is peculiar compared to other OECD countries, representing a rather large source of both central Government and local authorities revenues. The three main taxes levied on properties are the council tax, business rates, and stamp duty taxes.

Council taxes are levied on each occupier rather than the owner of domestic properties. The tax is one of the few levies in Great Britain being both set and collected by local authorities (Boroughs in the case of London), and it represents one of their primary sources of revenue (around one-third of total revenue), the other sources being commercial property taxes (business rates) and transfers from the central Government. The tax is based on a classification in eight bands (A-H) based on the property's value as established by the Valuation Office in 1991; newly built properties are assigned to a band after having their current value converted into the value of a comparable property in 1991.

Each London Borough is responsible for setting the annual tax amount to be paid by a property in band D every year; the amount to be paid by other bands is automatically set as a ratio to the amount for band D.[8] Bands C and D represent the largest fraction of dwellings (about 50 percent of the total), but there is variation across Boroughs with central properties being skewed towards higher valued bands compared to properties in outer Boroughs. Figure C.2 shows the time series of the council tax payable per band per Borough. Each panel in the figure depicts the amount payable by different bands showing that, by construction, the tax moves in locksteps across bands. More interestingly, it should be noted that there is a wide dispersion in amounts payable across Boroughs, even though the

[8] The ratios are constant across Boroughs and are as follows: band A 6/9, band B 7/9, band C 8/9, band D 1, band E 10/9, band F 13/9, band G 15/9, band H 2.

ranking across different local authorities remains almost constant with the only exception being the Borough of Hammersmith and Fulham where taxes have been slashed starting from the late 2000s. After a marked increase in council tax rates in the early 2000s, the freeze mandated by the central Government after the 2008 financial crisis is visible in the time series; since 2011, taxes can be raised only by a centrally set amount unless a local referendum allows the authority to do so.

We will show in Section 3.3.1 that the geographical distribution of council tax rates is not random and could severely bias any estimate of incidence, given that central (and pricier) Boroughs tend to set lower council tax rates. This is mainly because central Boroughs tend to have larger fraction of properties in higher bands; for instance, the Borough of Kensington and Chelsea raises more than fifty percent of its revenues from bands G and H, while Barking and Dagenham raise less than five percent from such bands.

We obtain information on council tax band assignment from the website of the Valuation Office Agency, which provides data on the complete address and the council tax band for each property in Great Britain. The average amount to be paid in each London Borough by each band in the period 1999-2018 is obtained from the London Datastore managed by the Greater London Authority.

In the following section, we provide some descriptive statistics of the data we have mentioned so far.

3.2.2 Descriptive Statistics

Figure C.3 shows the distribution of transaction prices for domestic properties in London, truncated to exclude extremely high property prices, which are, however, included in the analysis. The data consist of 889,925 observations between 1999 and 2018, for which property characteristics and council tax information is available. We confirm that the distribution is highly skewed, with the average and median property values being £366, 528 and £250, 000, respectively.

It is immediately obvious that there is a large degree of bunching in prices, as noted, for instance, in Best and Kleven (2018). The bunching mainly happens just before stamp duty notches, which allows Best and Kleven (2018) to estimate the local incidence of this tax. Figure C.4, for instance, shows the large extent of bunching at the threshold of £250,000 (upper panel) and £500,000 (lower panel) where the stamp duty tax jumps from 1% to 3% and from 3% to 4%, respectively.

Best and Kleven (2018) estimate a rather large incidence of stamp duty tax on property prices and argue in favor of evidence of rather strict borrowing constraints; we will use their estimates to inform our analysis of the incidence of the council tax, allowing us to disentangle how much of the incidence is due to borrowing constraints (or the lack thereof) and how much is attributable to pure time discount.

Figure C.5 shows the distribution of house prices per band. The vertical red lines depict the median price within each band. As one should expect, higher bands tend to have houses with higher average prices, although there is a large dispersion within bands. This is because prices have increased significantly over the past twenty years, especially for more central and higher-banded properties. This makes it essential that we compare only transactions occurring in close periods. Moreover, one should notice that the number of properties belonging to bands C and D dominates the rest, as previously mentioned.

In Figures C.6, C.7, C.8, C.9 and C.10 we show that there is a wide dispersion of transaction prices based on house characteristics such as property type, number of rooms, property age and duration. There is a disproportionate amount of flats in our sample, which we see as an advantage in our estimation, as flats are much more likely to be similar to each other relative to other property types. Detached houses are the most expensive, with a median price of £525,000, followed by semi-detached houses (£319,950) and terraced houses (£270,000), and finally, flats are the cheapest category (£195,000). Naturally, the house price is increasing in the number of rooms, with the median value of each additional room being about £40,000 in the full sample. Newly-built properties represent a minority in our sample and trade at a small discount relative to established buildings. This is due to the geographic distribution of the housing stock in London, where older properties tend to be in the more sought-after central areas. However, there is some heterogeneity when we look at the year of construction: properties built before 1949 sold at a median of £287,000 close to those built after 2003 (£275,000), while properties built in the period 1950-1982 and 1983-2002 were sold at lower prices (£215,000 and £200,000, respectively). This pattern can be explained by differences in type and location across groups. Finally, it can be noted that properties under freehold ownership have a higher median price (£305,000) compared to leasehold properties (£195,000).

After having described the data, we next proceed to the discussion of our empirical strategy in the next section.

3.3 Empirical Strategy

3.3.1 Evidence of Selection

The main issue that arises when estimating the incidence of council taxes is that the cross-sectional distribution of council tax amounts across Boroughs strongly correlates with other characteristics that affect house prices.

To see this, Figure C.11 shows a map of the distribution of Band D council tax amounts payable for each London Borough along with the respective distribution of house prices. Panel C.11a shows the distribution of council taxes in 2000, where taxes increase moving from yellow to red; Panel C.11b the distribution of house prices in the same year, where prices increase moving from light blue to brown. Panel C.11c shows the distribution of council taxes in 2018, while panel C.11d shows the distribution of house prices in the same year.

It is visually striking that councils with lower taxes tend to have higher house prices. For instance, the City of Westminster had the lowest Band D council tax in 2000 (£375.17) and the second highest average house price (£357,925), after the Borough of Kensington and Chelsea (£726,908) which had the fourth lowest council tax (£623.38). In 2018 the same holds true, with the City of Westminster having the lowest council tax (£710.50) and the second highest average price (£1,612,231), after Kensington and Chelsea (£3,040,547), which had the fifth lowest council tax

(£1,139.41).

In general, it is clear from the map that Boroughs that lie further from the center tend to have higher council taxes and lower prices, while the more central Boroughs tend to exhibit the opposite pattern. To confirm the intuition obtained from Figure C.11, we can run a naïve regression of house prices on house characteristics and council tax payable without controlling for the geographical location of the property, i.e.:

$$p_{idbt} = \beta\tau_{dbt} + \delta_{bt} + \zeta' x_{idbt} + \varepsilon_{idbt} \tag{3.1}$$

where p_{idbt} is the price of house i in Borough d, band b at time t; τ_{dbt} is the council tax amount for a house in Borough d, band b at time t; δ_{bt} are year-band fixed effects; and x_{idbt} are controls which include the property size measured in squared meters, number of rooms, property type, age, duration and month which controls for seasonality in the housing market (Ngai and Tenreyro, 2014).

Table C.1 reports the results of regression (3.1); the first column provides the baseline result where month and year-band fixed effects are included to remove the mechanical correlation between increasing property prices and taxes over time and the fact that moving from band A to band H goes hand in hand with higher house prices. If we took this evidence at face value, we would conclude that the incidence of council tax is extremely large and statistically significant, with a point estimate of -231.2.

To give intuition, using a discount rate of $r = 4\%$ (similar to the risk-free rate observed in sample) this would roughly imply that an extra £1 in present value of taxes would lead to a drop in prices of $r \times \beta = 4\% \times 231.2 = £9.25$. It is obvious that this figure is only the artifact of the negative correlation between the value of properties and the average tax within councils, as observed in Figure C.11. Extremely negative coefficients are obtained in columns (2), (3), and (4), where we control for the property size, the number of rooms, property type, whether the property is newly-built, and whether it is a leasehold. The smallest of these coefficients in absolute value, i.e., -228.7 in column (3), would imply an incidence of $r \times \beta = 4\% \times 228.7 = £9.15$ which is still unreasonably high. Table C.2 shows similar estimates when we include all the variables available as controls.

To further corroborate the negative correlation between property prices and council taxes due geographical selection, we provide the results of the following two-step estimation. First, we regress house prices on characteristics to obtain hedonic residuals:

$$p_{idbt} = \zeta' x_{idbt} + \varepsilon_{idbt} \tag{3.2}$$

For each Borough, band, year, we compute the median residual price ε^{med}_{dbt} and proceed to regress it on council tax amount payable including year-band fixed effects:

$$\varepsilon^{med}_{dbt} = \beta\tau_{dbt} + \delta_{bt} + \eta_{dbt} \tag{3.3}$$

The results are reported in Table C.3. The vector of predictors x_{idbt} in first-stage hedonic regression includes month fixed effects in column (1); month, property size, number of rooms in column (2); month, property size, number of rooms and property type in column (3); and month, property size, number of rooms, property type and indicators for whether the property is newly-built and a leashold in column (4). Similarly, Table C.4 reports results when the dependent variable in the second stage is the average hedonic residual $\bar{\varepsilon}_{dbt}$ per Borough, band, year, i.e.:

$$\bar{\varepsilon}_{dbt} = \beta\tau_{dbt} + \delta_{bt} + \eta_{dbt} \tag{3.4}$$

Both tables confirm the previous finding that Boroughs with higher house values tend to impose lower council tax bills: the coefficients are negative and statistically significant, ranging from -183.6 to -368.4.

The results provided so far imply that special care needs to be taken before using the geographical variation in council taxes to estimate their incidence on house prices. For this reason, in our identification strategy, we will compare only houses that lie extremely close, i.e., no more than 500 meters and mainly closer than 200 meters, to the border between two adjacent Boroughs to disentangle the actual incidence of the tax from the geographical distribution of taxes across Boroughs.

Throughout the rest of the paper, the reader should remember that the geographical distribution of council taxes entails that any estimated incidence will be, at most, an upper bound for the *true* incidence. This is because if buyers value certain characteristics upon purchasing a house, these should be capitalized in the house price, which, in this case, acts almost like a sufficient statistic for their value; the results of Figure C.11 and Tables C.1-C.4 signal that houses with more highly valued characteristics (and higher prices) tend to be located in Boroughs with lower taxes, thus inflating any estimate of tax incidence.

A second and more subtle reason why we can only estimate an upper bound for the incidence has to do with our identification strategy. By comparing similar dwellings on opposite sides of a border, we implicitly assume that the buyer always has an outside option during the price bargaining process. As a result, the buyer would be much more elastic than an otherwise identical buyer involved in purchasing a house located in the heart of a Borough where there is no outside option in terms of council tax.

We will show in Section 3.5 that the seller will bear the whole incidence of the tax at the border, while that will not necessarily be the case at an interior point. In general, even in the absence of perfect substitutes across council borders, it is reasonable to conjecture that the incidence will still be much larger at the border compared to the council center, where the agent would have to move long distance to pay a different council tax rate.

In the next section, we will describe the identification strategies that will allow us to estimate the incidence of council taxes as precisely as possible given the present setting, bearing in mind that any attempt is likely to result in an overestimation of the *true* incidence.

3.3.2 Identification Strategies

We use two different identification strategies to measure an upper bound of the incidence of council tax on property prices: regressions grids and a matching algorithm.

Regression Grids

The first strategy compares houses in close proximity by dividing the area of London in a grid and assigning a fixed effect to each square in the grid. By doing so, we are de-facto comparing two houses that are otherwise identical but lie on opposite sides of a given border between two Boroughs.

Figure C.12 graphically depicts our first approach. Panel C.12a shows a grid of squares with equal sizes superposed on a map of London. Panel C.12b shows a more detailed picture of the Boroughs in the center.[9]

We then select the squares with two houses that are sold in the same year, in the same council tax band and lie on opposite sides of the border; Panel C.12b displays in blue examples of such squares. It can be noticed that we discard observations for which the border is located on the Thames River bank. To avoid relying on an arbitrary division, we have used three different grids. Namely, one grid divides the area in 50×50 squares, another divides it in 100×100 squares, and, finally, the last grid is a 150×150 one. These squares have side lengths of approximately 800 meters, 400 meters, and 250 meters, respectively. While the maximal possible distance between houses can be inferred as $\sqrt{2} \times$ square side length, we have decided to remove observations that were more than 500 meters far from the border.

Figure C.13 shows the distribution of distances to the border for our different specifications. As mentioned, no house lies more than 500 meters from the border, and most observations are about 200 meters away from the closest border. As we refine our grids by subdividing into a larger number of squares, we can see that we lose observations in the 200 meters-500 meters range; this will reduce our power significantly, but it will ensure that we compare houses that are indeed in very close proximity.

Our strategy consists of running within square regressions whereby we compare houses that are sold in the same year and in the same council tax band, specifically:

$$p_{ibgdt} = \beta\tau_{bdt} + \delta_{bgt} + \zeta' x_{ibgdt} + \varepsilon_{ibgdt} \tag{3.5}$$

[9] The three main Boroughs depicted in the picture are, starting from left, Hammersmith and Fulham, Kensington and Chelsea, and the City of Westminster.

where p_{ibgdt} is the price of house i, in council tax band b, grid square g, Borough d, and year t; τ_{bdt} is the council tax amount for band b, Borough d in year t; and x_{ibgdt} are house-specific controls. The presence of the band-grid square-year fixed effects δ_{bgt} guarantees that the regression compares houses that are in the same square, same council tax band, and are sold in the same year, implying that our identification assumption is that they systematically differ only due to the amount of council tax paid, after partialling out the effect of house characteristics that we add to increase our precision.

It should be noticed that, as mentioned above, *better* Boroughs, i.e., Boroughs with higher average prices, tend to have lower council taxes, implying that - if we leave some hidden characteristic out of our regression - the estimate of β is most likely going to overstate the *true* incidence. To give an example, while highly unlikely given the sharp nature of the borders, one could argue that there is a name tag value of living in certain Boroughs over others. For instance, a house in Westminster commands a premium over a similar house on the other side of the border in Brent. The fact that Westminster has a lower tax compared to Brent implies that we will overestimate the incidence of the tax because of the name tag value of living in Westminster. In general, to reverse this bias and claim that the *true* incidence might be higher than the one we estimate, the reader should think of some hidden characteristic that systematically causes people to prefer living in a Borough with worse amenities compared to a Borough with better ones.

The following section will present our second identification strategy, which relies on a matching estimator rather than grid squares fixed effects.

Matching Estimator

Our second identification approach consists of the pairwise matching of houses on opposite sides of a given border. To find the closest match, we need to define a distance: in what follows, we rely on a Euclidean distance and a distance based on a linear model. Under the first one, we restrict the possible matches to be: no more than 500 meters away from each other, sold in the same year, in the same council tax band, and to both be either old or newly-built and freehold or leasehold properties. For each property we then choose the closest match as the one minimising the Euclidean distance $d(i, j) = \sqrt{\sum_{k=1}^{K}(x_{ik} - x_{jk})^2}$, where i is the original property, j indexes the possible matches on the other side of the border, x_{ik} are house i characteristics and x_{jk} are house j characteristics. We then run within-pair regressions:

$$p_{ibdt} = \beta\tau_{bdt} + \delta_{ij} + \zeta' x_{ibdt} + \varepsilon_{ibdt} \tag{3.6}$$

where δ_{ij} are ij-pair dummies and x_{ibdt} are house i-specific features. The second choice of distance is based on a linear pricing model:

$$p_{it} = \alpha + \beta' x_{it} + \varepsilon_{it} \tag{3.7}$$

where x_{it} similarly contains house-specific characteristics as above. We then compute the model-predicted price $\hat{p}_{it} = \hat{\alpha} + \hat{\beta}' x_{it}$. As before, we restrict the pairing to houses sold in the same year, band, old/new, and leasehold/freehold categories and no further than 500 meters from each other. For each property i we pick the closest match j as the one that minimises the following distance: $d(i, j) = |\hat{p}_{it} - \hat{p}_{jt}|$. To estimate the incidence, we run within pair-regressions as in equation (3.6) where the δ_{ij} dummies are determined according to the new matching algorithm. As in Section 3.3.2 the identification will be valid as long as the only systematic difference within pairs is the amount of council tax. As previously explained, any other omitted variable will likely lead us to estimate an upper bound for the incidence, given the geographical distribution of council taxes.

3.4 Results

3.4.1 Grid Estimator

Table C.5 presents the results of the grid regressions described in Section 3.3.2 where we have used a 50×50 grid and included band-grid square ID-year fixed effects to compare the effect of council taxes on properties in the same band, sold in the same year, located in the same grid square but on opposite sides of a border as in Equation (3.5). The controls we include are as follows: column (1) uses month fixed effects to control for housing market seasonality; column (2) adds the number of rooms fixed effects and controls for property size; column (3) also adds property type fixed effects, and; column (4) includes an indicator for newly-built and leasehold properties. These will be our default specifications throughout the rest of the paper.

In all columns, the coefficient on council taxes is statistically indistinguishable from zero and always with the wrong sign. The lack of significance cannot be attributed to a lack of statistical power in the regressions, given that other control variables are always strongly statistically significant. For instance, the effect of one additional squared meter ranges between £$4,537$ and £$4,627$, newly-built properties command a premium of about £$33,400$, and freehold properties sell for £$76,000$ more relative to leaseholds.

The same conclusion can be drawn from Table C.6 where we expand the regressions to include all available house price predictors, showing that even relatively minor characteristics such as the number of lighting outlets or the presence of fireplaces in the property have a significant effect on prices.

Table C.7 displays the grid regression results for grids of different sizes: column (1) uses a grid that divides the London area into 50×50 squares, column (2) 100×100, and column (3) 150×150. This might help to alleviate concerns that grids made of large squares might be comparing houses rather distant from each other. The specification is otherwise the same as the one in column (4) of Table C.5. The coefficient on council tax remains statistically insignificant, and the point estimate varies from positive to negative across columns: this is precisely what we should expect when a regressor has no effect on the outcome variable and simply reacts to the noise in the sample. The fact that the R-squared is very high (between 77% and 83%) and that all other coefficients are precisely estimated confirms our previous finding that the incidence of the council tax is indistinguishable from zero.

In Table C.8 we augment the regressions by adding all additional house characteristics: the coefficient on council tax ranges from -11.8 to 75.4 and is never statistically lower than zero.

To make sure that the confounding effect of the stamp duty notches does not play a role in our estimation results, Table C.9 presents the results of the grid regressions when we remove the two main stamp duty notches at £250, 000 and £500, 000. Column (1) excludes only the first notch, column (2) the second, and column (3) removes both. The results are virtually unchanged, with the incidence still being statistically insignificant, small in magnitude, and always displaying the wrong sign. As previously mentioned, the large R-squared and the fact that the remaining coefficients are precisely estimated guarantees that this is not due to a lack of power.

Finally, Tables C.10 and C.11 provide estimates of council tax incidence using a similar two-step approach as in Tables C.3 and C.4, i.e., by first obtaining residual hedonic prices as follows:

$$p_{ibdgt} = \zeta' x_{ibdgt} + \varepsilon_{ibdgt} \tag{3.8}$$

and subsequently regressing the median or average hedonic residuals for each Borough, band, grid square and year on council tax amounts:

$$\varepsilon^{med}_{bdgt} = \beta \tau_{bdt} + \delta_{bgt} + \eta_{bdgt} \tag{3.9}$$

$$\bar{\varepsilon}_{bdgt} = \beta \tau_{bdt} + \delta_{bgt} + \eta_{bdgt} \tag{3.10}$$

where δ_{bgt} are band-grid square-year fixed effects included to ensure that we are comparing values of houses in the same council tax band, sold in the same year, and located in the same square of the grid. As usual, we restrict the analysis to grid squares with at least two houses located on different sides of a border and present the four standard specifications. The results confirm the previous finding: both the median and average hedonic residuals are not decreasing in the council tax amount paid, suggesting that the incidence of this tax on house prices is not different from zero.

In the following section, we will supplement the evidence by presenting results using our second identification strategy.

3.4.2 Matching Estimator

Tables C.12, C.13 and C.14 show the results of our second estimation approach where we explicitly match similar dwellings on opposite sides of a border as described in Section 3.3.2. As previously mentioned, all the results are obtained using housing pairs on opposite sides of a border no more than 500 meters apart, sold in the same year, in the same council tax band, and which are both either old or newly-built and leasehold or freehold properties.

Table C.12 displays the results where closest pairs have been determined by minimising the Euclidean distance $d(i,j) = \sqrt{\sum_{k=1}^{K}(x_{ik} - x_{jk})^2}$, where the vectors x_i and x_j consist of property size and number of rooms in columns (1) and (2), and also energy cost in columns (3) and (4). All the variables have been standardized to be comparable. This procedure leads to 57,612 and 57,323 observations of property pairs with 71,578 and 71,656 unique transactions in columns (1)-(2) and (3)-(4), respectively.[10] After obtaining the pairs, we run the regression specified in Equation (3.6). The presence of δ_{ij} pair fixed effects amounts to regressing the difference in prices of matched houses on the difference in council tax paid, controlling for other property characteristics along which the matched properties may differ.

Consistent with the results obtained with the grid estimator, none of the coefficients on council tax is statistically significantly negative. As pointed out before, this result is not attributable to a lack of statistical power. For instance, the coefficient on size is highly statistically significant and has the same order of magnitude as the ones obtained with the earlier estimator.[11]

Table C.13 confirms these findings under linear matching algorithm where pairs have been chosen by minimising the distance $d(i,j) = |\hat{p}_{it} - \hat{p}_{jt}|$, where the predicted prices $\hat{p}_{it}$ and $\hat{p}_{jt}$ have been obtained from a linear model as in Equation (3.7). As before, columns (1) and (2) match properties based on size and number of rooms, while columns (3) and (4) add energy cost.

[10] Notice that any given transaction can be the closest match for more than one property. We cluster standard errors at the transaction ID level to take care of this redundancy.

[11] Notice that, compared to the default specifications used in Tables C.1, C.5, C.7 and C.9, the indicators for newly-built and leasehold properties have been dropped given that properties have been constrained to be identical along these dimensions.

Finally, Table C.14 presents the last set of results for the linear model where we have allowed each property to be paired with more than one similar property on the other side of the border, as long as the absolute difference in predicted prices is less than 30% of the largest predicted price, namely: $|\hat{p}_{it} - \hat{p}_{jt}| < 0.3 \times \max\{\hat{p}_{it}, \hat{p}_{jt}\}$. While the point estimates range between -5.24 and -8.19, none of the coefficients is statistically different from zero as in all previous specifications. We will shed more light on the interpretation of these and the previous results in Section 3.5.1.

The empirical findings above demonstrate that council tax differences never significantly explain house price differences. Moreover, while the absence of evidence, namely the fact that agents seem insensitive to taxes that are postponed to the future, does not directly imply evidence of absence, many point estimates are positive and hence with the wrong sign. Bearing these estimates in mind, in the next section we develop a simple model that will allow us to propose a plausible explanation for the above results. We then calibrate the model using a Bayesian approach informed by all of the above estimates and briefly discuss policy implications.

3.5 Model

In what follows, we present a simple multi-period model of housing-consumption choice to calibrate the above results. We begin with the optimisation problem of an agent who chooses at time $t = 0$ an infinite stream of consumption $\{c_t\}_{t=0}^{\infty}$ and a composite housing good h:

$$\max_{\{\{c_t,d_t\}_{t=0}^{\infty},h,\mathbb{1}_{\{A\}},\mathbb{1}_{\{B\}}\}} \quad U(\{c_t\}_{t=0}^{\infty}, h) = c_0 + \sum_{t=1}^{\infty} \beta^t u(c_t) + \sum_{t=0}^{\infty} \beta^t \log(h) \tag{3.11}$$

$$s.t. \quad c_0 + h(p_{A0}\mathbb{1}_{\{A\}} + p_{B0}\mathbb{1}_{\{B\}} + \tau_S) \leq w_0 + d_0 \tag{3.12}$$

$$c_t + h(\tau_{At}\mathbb{1}_{\{A\}} + \tau_{Bt}\mathbb{1}_{\{B\}}) + d_{t-1}(1+r) \leq w_t + d_t \quad t = 1, 2, 3, ... \tag{3.13}$$

$$d_t \leq \alpha h(p_{At}\mathbb{1}_{\{A\}} + p_{Bt}\mathbb{1}_{\{B\}}) \quad t = 0, 1, 2... \tag{3.14}$$

For simplicity, we choose the agent's utility to be time-separable and separable in consumption and housing. The utility function is quasi-linear in c_0 to eliminate income effects, as is standard practice in the public finance literature. For tractability and to separate the impact of stamp duty and council tax, the agent purchases the housing good only once at $t = 0$. There are two Boroughs, A and B, with exogenously chosen and potentially different council tax rates. We assume that there is an equal supply of housing in both Boroughs.[12]

Equation (3.12) is the first-period budget constraint: the agent spends his initial endowment w_0 on consumption c_0 and the after-tax cost of his housing demand h. When he buys a house, the agent pays the pre-tax price $p_{i0}, i = A, B$, and, in addition, he also needs to pay the stamp duty tax τ_S hereby assumed to be proportional to the quality-adjusted level of housing demand. If his total demand exceeds his initial endowment, the agent can borrow additional funds d_0 for one period at the risk-free rate. The budget constraints for all subsequent periods are identical and given by Equation (3.13): from time $t = 1$ onwards, the agent will need to spend his endowment w_t on his optimal consumption choice c_t and pay the council tax that corresponds to the Borough where he chose to locate $\tau_{it}, i = A, B$. He will also need to repay his short-term debt from the previous period, including interest $d_{t-1}(1 + r)$, and will be allowed to borrow again at the same terms to balance his budget constraint. Finally, the last constraint in Equation (3.14) is the financing constraint: the agent cannot borrow more than a fraction α of the pre-tax cost of his housing demand. This can potentially generate a very large incidence for the stamp duty tax since the lump sum nature of this tax will tighten the leverage constraint.

[12] This assumption is crucial, and de-facto eliminates the potential for a differential elasticity of supply with respect to council taxes at the border. We consider this assumption quite reasonable, given that the vast majority of the housing stock in London had been constructed well before the introduction of this tax in the early 90s, as shown in Figures C.8 and C.9.

The Lagrangian for the above problem can be written as:

$$\begin{aligned}\mathcal{L} =& U(\{c_t\}_{t=0}^{\infty}, h) - \lambda_0(c_0 + h(p_{B0} + \tau_S) - w_0 - d_0) - \\ & \sum_{t=1}^{\infty} \lambda_t(c_t + h\tau_{Bt} + d_{t-1}(1+r) - w_t - d_t) - \sum_{t=0}^{\infty} \mu_t(d_t - \\ & \alpha h p_{Bt}) - h\mathbb{1}_{\{A\}} \left[\lambda_0(p_{A0} - p_{B0}) + \sum_{t=1}^{\infty} \lambda_t(\tau_{At} - \tau_{Bt}) - \alpha \sum_{t=0}^{\infty} \mu_t(p_{At} - p_{Bt})\right]\end{aligned} \tag{3.15}$$

where we have used the fact that $\mathbb{1}_{\{B\}} = 1 - \mathbb{1}_{\{A\}}$. Notice that the Lagrangian is monotone in the choice of Borough $\mathbb{1}_{\{A\}}$, therefore, the choice of where to locate can be separated from the consumption and housing-quality choices.

The agent chooses to live in Borough A if:

$$p_{A0} - p_{B0} \leq -\sum_{t=1}^{\infty} \frac{\lambda_t}{\lambda_0}(\tau_{At} - \tau_{Bt}) + \alpha \sum_{t=0}^{\infty} \frac{\mu_t}{\lambda_0}(p_{At} - p_{Bt}) \tag{3.16}$$

i.e., if the price differential between the same-quality house in Boroughs A and B more than compensates for the present value of the difference in future council tax payments and the collateral value of the house.

In equilibrium, markets clear if Equation (3.16) holds with equality which, from now onwards, we assume to be the case. Assuming that the agent is indifferent between living in Boroughs A and B, we proceed by suppressing the Borough subscripts and denote the house price as p_t and the council tax as τ_t.

The first-order conditions for an interior solution are:

$$1 = \lambda_0 \tag{3.17}$$

$$\beta^t u'(c_t) = \lambda_t \quad \forall t = 1, 2, 3, ... \tag{3.18}$$

$$-\lambda_t + \lambda_{t+1}(1+r) + \mu_t = 0 \quad \forall t = 0, 1, 2, ... \tag{3.19}$$

$$\frac{h^{-1}}{(1-\beta)} = \lambda_0 (p_0 - \alpha \frac{\mu_0}{\lambda_0} p_0 + \tau_S) + \sum_{t=0}^{\infty} \lambda_{t+1} \tau_{t+1} - \sum_{t=0}^{\infty} \lambda_{t+2} \frac{\mu_{t+1}}{\lambda_{t+2}} \alpha p_{t+1} \tag{3.20}$$

Combining the first-order conditions for consumption and for the optimal debt choice, we obtain the following Euler equation:

$$\frac{\lambda_{t+1}}{\lambda_t} = \beta \frac{u'(c_{t+1})}{u'(c_t)} = \frac{1}{1 + r + \frac{\mu_t}{\lambda_{t+1}}} \tag{3.21}$$

The above Euler equation implies that the agent's discount factor is equal to the inverse of the risk-free rate and a liquidity premium $\frac{\mu_t}{\lambda_{t+1}}$, arising from the fact that the house has some collateral value. In order to simplify the exposition, we assume that in equilibrium the liquidity premium is constant and equal to $\frac{\mu_t}{\lambda_{t+1}} = k$, that house prices grow at a constant rate g, i.e., $p_{it} = p_{i0}(1+g)^t$, and council tax amounts grow at a constant rate $\tilde{g}$, i.e., $\tau_{it} = \tau_{i1}(1+\tilde{g})^{t-1}$.

Re-arranging Equations (3.16), (3.20) and (3.21), we obtain the final no-arbitrage condition and housing demand:

$$(p_{A0} - p_{B0}) \left(1 - \frac{\alpha k}{r + k - g}\right) = -(\tau_{A1} - \tau_{B1}) \frac{1}{r + k - \tilde{g}} \tag{3.22}$$

$$\frac{h^{-1}}{(1-\beta)} = p_0 \left(1 - \frac{\alpha k}{r + k - g}\right) + \tau_S + \frac{\tau_1}{r + k - \tilde{g}} \tag{3.23}$$

The first equation is the equilibrium condition of how house prices should behave across Boroughs: the house price differential, after having taken into account the collateral value $\frac{\alpha k}{r+k-g}$, needs to match (the negative of) the present value of the council tax differential. The second equation states that the agent's marginal utility of housing is equal to the house price inclusive of (the present value of) all taxes and collateral value.

It is important to note that the no-arbitrage condition (3.22) in general gives a different incidence compared to the one obtained from the housing demand (3.23). This is because the former holds only at the border between two Boroughs where the outside option, i.e., the option to buy an otherwise identical house on the other side of the border, implies that the supply bears the whole burden of the tax. In particular, from Equation (3.22) we obtain an incidence of:

$$\frac{dp_0}{d\tau_1} = -\frac{1}{r+k-\tilde{g}} \times \frac{r+k-g}{r+(1-\alpha)k-g} \tag{3.24}$$

On the other hand, for both houses on the border as well as houses in the middle of a given Borough we can define the optimal demand from Equation (3.23) as $D(p_0, \tau_1, \tau_S) = h^*(p_0, \tau_1, \tau_S)$. Equating with the optimal supply, $S(p_0) = D(p_0, \tau_1, \tau_S)$, and after total differentiation we obtain the standard formula for the incidence:

$$\frac{dp_0}{d\tau_1} = -\frac{\frac{\partial D}{\partial \tau_1}}{\frac{\partial D}{\partial p_0} - \frac{\partial S}{\partial p_0}} = -\frac{1}{r+k-\tilde{g}} \times \frac{1}{\frac{r+(1-\alpha)k-g}{r+k-g} + \tilde{\eta}_S} \tag{3.25}$$

where $\tilde{\eta}_S = \frac{\partial S}{\partial p_0}\frac{p_0}{S}\frac{p_0\left(1-\frac{\alpha k}{r+k-g}\right)+\tau_S+\frac{\tau_1}{r+k-\tilde{g}}}{p_0} = \eta_S \frac{p_0\left(1-\frac{\alpha k}{r+k-g}\right)+\tau_S+\frac{\tau_1}{r+k-\tilde{g}}}{p_0}$ is a slightly modified version of the supply elasticity η_S that takes into account the price inclusive of taxes and collateral value. In general, we have that:

$$\frac{1}{\frac{r+(1-\alpha)k-g}{r+k-g} + \tilde{\eta}_S} \leq \frac{r+k-g}{r+(1-\alpha)k-g} \tag{3.26}$$

implying that the incidence at the border between Boroughs is an upper bound for the *true* council tax incidence as long as the modified elasticity of supply is non-negative, i.e., $\tilde{\eta}_S \geq 0$. Notice that the modified elasticity of supply $\tilde{\eta}_S$ is positive as long as the true elasticity of supply η_S is positive.

3.5.1 Calibration

The model in the previous section allows us to better interpret the empirical results of Section 3.4. By using Equations (3.22), (3.23) and (3.24) we get:

$$\frac{dp_0}{d\tau_1} = \frac{dp_0}{d\tau_S} \times \frac{1}{r+k-\tilde{g}} \tag{3.27}$$

i.e., the incidence of the council tax can be interpreted as the present value of the sum of the incidence of the stamp duty tax discounted at the liquidity-adjusted cost of capital $r + k$ with growth rate $\tilde{g}$.[13]

In what follows we are going to use the results in Tables C.5 - C.14 and provide some further direction on how to interpret them. We treat each estimate as a separate model m.

Conditional on the model being true and given a common prior distribution $p(\beta_\tau|m) = p(\beta_\tau)$ about the true incidence of council tax and the likelihood function of the data $p(y|\beta_\tau, m)$ we can use Bayes' rule to express the posterior distribution for the incidence under each model m as:

$$p(\beta_\tau|y, m) = \frac{p(y|\beta_\tau, m) \times p(\beta_\tau)}{\int p(y|\beta_\tau, m) \times p(\beta_\tau) d\beta_\tau} \tag{3.28}$$

We then proceed to obtain the model-averaged posterior distribution as:

$$p(\beta_\tau|y) = \sum_m p(\beta_\tau|y, m) p(m|y) \tag{3.29}$$

The computational burden of Equation (3.29) is significant, therefore, we proceed with the simplifying assumptions described in Appendix C.4. We always start from a normally-distributed prior $\beta_\tau \sim \mathcal{N}(b_\tau, \sigma_\tau^2)$ and likelihood function which leads to a normal posterior. As detailed in Appendix C.4 the mean of the prior is chosen by calibrating the parameters g, $\tilde{g}$, r and α based on historical data and matching the stamp duty incidence to results in Best and Kleven (2018). For robustness we also vary the precision of the prior and provide results for five different specifications: $p(\beta_\tau) = \mathcal{N}(-150, 50^2), \mathcal{N}(-100, 50^2), \mathcal{N}(-50, 50^2), \mathcal{N}(-150, 75^2), \mathcal{N}(-50, 25^2)$.

Figure C.15 plots the model-averaged density of the posterior distribution for the council tax incidence. Panel (a) displays the posterior density for a constant standard deviation of the prior of 50, while (b) for a standard deviation equal to half the prior mean. It can be noted that the shape of the posterior is similar across specifications and that it displays a significant shift of mass toward zero. Table C.15 provides the quantiles, the mode, and the mean of the posterior distribution of the incidence.

The median posterior incidence ranges between -22.87 and -2.17, well below

[13] This assumes that $\tilde{\eta}_S = 0$, i.e., that the housing supply is fixed in the short term.

the median implied by the model calibration, which has informed the prior. The last column reports the ratio between the two, giving the implied attenuation bias displayed by agents. Given the model parameters, the price reaction to council taxes is between 4% and 37% of what the price reaction to the stamp duty tax would imply from agents who fully perceive the tax.

The results above become striking once coupled with the extent to which house buyers react to stamp duty taxes. When buyers are liquidity-constrained, their effective discount rates become large, and, therefore, one might be tempted to attribute the previous evidence solely to extreme discounting of future cash flows. If we are willing to take this view, we would have to assume discount rates ranging between 23.4% and 231.9% to fit the posterior estimates of the council tax incidence.

Moreover, it should be noted that every estimate of the council tax incidence is conditioned on an estimate of the stamp duty incidence, i.e., the discount rate is not a free parameter in the calibration. To put it differently, changing the discount rate to match a reasonable incidence for the council tax would lead to an incidence of the stamp duty tax that is inconsistent with current estimates in the literature. The fact that the incidence of the stamp duty is large but not extreme implies that the liquidity premium cannot be the only source of the low council tax incidence.

Third, in our estimation, we have used relatively concentrated priors around the model-informed incidence; had we allowed the likelihood to dominate by assigning diffuse priors, we would have obtained much lower estimates compared to the conservative ones we have provided so far.

One way to explain these findings is by hypothesizing that, when buying their properties, agents discount tax payments that happen in the future disproportionately compared to those that occur concurrently with the purchase. It is difficult to argue that this might be due to uncertainty associated with council tax payments, given that differences in council tax amounts across Boroughs are very smooth and predictable, as shown in Figure C.2.

This leaves us with another plausible alternative explanation: agents fail to fully internalize the (difference in) council tax payments across Boroughs upon purchasing a property, either because this is much less salient compared to the stamp duty tax or because they fail to appreciate the magnitude of its present value.[14]

Notice also that the results so far suggest that there is somebody who is not taking the council tax differentials into account in a fully-rational way, but this does not need to be the house buyer: our previous analysis goes through even if the buyer is fully aware of the tax and hopes to shift its incidence onto the subsequent buyer, or the renter in the case of buy-to-let property transactions.[15]

Motivated by these findings, we are going to explore some policy implications in the following section.

3.5.2 Implications for Tax Policy

Given the results in the previous section, it seems reasonable to argue that agents fail to fully perceive deferred taxes. As a result, we propose a modified version of the model above that allows for non-fully salient taxes.

We extend our analysis to properties that are potentially far from the border and, therefore, allow the elasticity of supply η_S to be non-zero. Recall that the incidence estimates coming from the border in Section 3.4 are an upper bound for the incidence in the middle of Boroughs. For simplicity, let us assume we are in an equilibrium where the leverage constraint (3.14) is binding, i.e., $d_t = \alpha h p_t$. If we multiply each of the constraints (3.12) and (3.13) by $\frac{1}{(1+r+k)^t}$ and add them together, we obtain the following consolidated budget constraint:

$$c_0 + \frac{c_1}{(1+r+k)} + \frac{c_2}{(1+r+k)^2} + ... + \tilde{p}h = w_0 + \frac{w_1}{(1+r+k)} + ... = I \tag{3.30}$$

where $\tilde{p} = p_0\left(1 - \frac{\alpha k}{r+k-g}\right) + \tau_S + \frac{\tau_1}{r+k-\tilde{g}}$ is the tax-inclusive house price. For simplicity of exposition, define $p = p_0\left(1 - \frac{\alpha k}{r+k-g}\right)$ and $\tau = \frac{\tau_1}{r+k-\tilde{g}}$, so that we can

[14] It is also possible that the tax is fully salient to agents, but due to mental accounting, they fail to integrate its present value into the house price they are willing to pay. Other explanations could be related to search costs and cognitive costs. For a property in band D worth, say, £300,000, the stamp duty tax in 2018 would amount to £9,000. If the buyer could choose whether to buy the property in the Borough of Camden or the Borough of Westminster, the difference in council tax would amount to about £778 in 2018, which, in present value using a discount rate of 4%, would be equal to £19,450, more than twice the value of the stamp duty tax.

[15] Note that we have largely interpreted the results as evidence of overpricing. Another possibility is that the properties on the low council tax side of borders are relatively underpriced, and it is, therefore, sellers who fail to incorporate the tax discount into their ask price.

rewrite $\tilde{p} = p + \tau_S + \tau$. Following Chetty et al. (2009), Farhi and Gabaix (2020) and Goldin (2015), we assume that the agent misperceives taxes with attenuation factor γ, i.e., he solves the following maximisation problem:

$$\max_{\{\{c_t\}_{t=0}^{\infty},h\}} U(\{c_t\}_{t=0}^{\infty}, h) = c_0 + \log(h) + \sum_{t=1}^{\infty} \beta^t \left(u(c_t) + \log(h)\right) \tag{3.31}$$

s.t.

$$c_0 + \frac{c_1}{(1+r+k)} + \frac{c_2}{(1+r+k)^2} + ... + \tilde{p}_\gamma h = w_0 + \frac{w_1}{(1+r+k)} + ... = I \tag{3.32}$$

where the perceived house price is:

$$\tilde{p}_\gamma = p + \tau_S + \gamma\tau, \quad \gamma \in [0,1] \tag{3.33}$$

Recall from the previous section that the attenuation factor for the council tax implied by the data ranges between 0.04 and 0.37. Notice that while the agent perceives the above budget constraint, he has to satisfy the actual budget constraint (3.30) given by the rational model. As pointed out in Reck (2016), it is crucial to decide what choice variable will bear the burden of adjustment. Given our assumption about the quasi-linear utility function in first-period consumption c_0, it is natural to let c_0 be the shock absorber. This choice amounts to assuming the following train of events: 1) the agent misperceives the council tax he will have to pay going forward and, as a result, buys "too much" quality-adjusted housing; 2) following this, he realizes that the actual amount of taxes he will have to pay is beyond his budget; 3) consequently, the agent adjusts his consumption in the first period keeping everything else constant.

Denoting the observed demands as $\hat{c_0}, \hat{c_t}, \hat{h}$, and the optimal demands absent any behavioural frictions as c_0^*, c_t^*, h^*, we have the following first-order conditions:

$$\hat{c_t} = [u']^{-1}\left(\frac{1}{(\beta(1+r+k))^t}\right) = c_t^* \tag{3.34}$$

$$\hat{h} = \left[(1-\beta)\tilde{p}_\gamma\right]^{-1} \neq \left[(1-\beta)\tilde{p}\right]^{-1} = h^* \tag{3.35}$$

$$\hat{c_0} = I - \sum_{t=1}^{\infty} \frac{\hat{c}_t}{(1+r+k)^t} - \hat{h}\tilde{p} \neq c_0^* \tag{3.36}$$

As previously mentioned, the optimality condition for future consumption remains as before. However, Equation (3.35) shows that the agent will demand "too much" housing because the perceived price $\tilde{p}_\gamma$ is lower than the true price $\tilde{p}$, as long as $\gamma < 1$. As a result, because of quasi-linearity in the utility function, $\hat{c_0}$ will adjust to absorb the reduced available income.

The previous discussion highlights that misperception of the house price will affect both consumption and housing demand, albeit in opposite directions. This implies that a benevolent social planner needs to carefully balance the two distortions when setting the optimal tax policy.

To see this more formally, let us adopt the approach of Goldin (2015) and assume that the Government will choose the optimal (property) tax combination to raise a fixed amount of revenue and maximize the buyer's utility.[16] For convenience, define the present value of council tax revenue from the Government's point of view, discounted at the risk-free rate, as $\tilde{\tau} = \frac{\tau_1}{r-\tilde{g}}$.

The total revenue raised from a given buyer is:

$$R = (\tau_S + \tilde{\tau})h = \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)h \tag{3.37}$$

The second equality of the above equation shows that the Government discounts the revenue raised through council taxes at a lower rate than agents due to the presence of borrowing constraints. The Government can twick the two taxes to maintain revenue-neutrality. In particular, a revenue-neutral tax change will be such that:

$$\left[h + \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)\frac{\partial h}{\partial \tau_S}\right]\Delta\tau_S = -\left[\frac{r+k-\tilde{g}}{r-\tilde{g}}h + \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)\frac{\partial h}{\partial \tau}\right]\Delta\tau \tag{3.38}$$

This implies that the change in stamp duty per unit change in council tax needed

[16] In what follows, we will abstract from analyzing the effect on the utility of the seller.

to maintain revenue-neutrality will be:

$$\frac{\Delta\tau_S}{\Delta\tau} = -\frac{\frac{r+k-\tilde{g}}{r-\tilde{g}}h + \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)\frac{\partial h}{\partial\tau}}{h + \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)\frac{\partial h}{\partial\tau_S}} = -\frac{\frac{r+k-\tilde{g}}{r-\tilde{g}}h + \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)\theta_\tau\frac{\partial h}{\partial p}}{h + \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)\theta_{\tau_S}\frac{\partial h}{\partial p}} \quad (3.39)$$

where $\theta_{\tau_S} = \frac{\frac{\partial h}{\partial\tau_S}}{\frac{\partial h}{\partial p}}$ and $\theta_\tau = \frac{\frac{\partial h}{\partial\tau}}{\frac{\partial h}{\partial p}}$ tell us how responsive the demand is with respect to taxes relative to pre-tax prices. From Equations (3.33) and (3.35) we infer that $\theta_{\tau_S} = 1$ and $\theta_\tau = \gamma$ in our model. The indirect utility function for an inattentive agent will be:

$$V(p,\tau_S,\tau) = I - \sum_{t=1}^{\infty}\frac{\hat{c}_t}{(1+r+k)^t} - \hat{h}(p+\tau_S+\tau) + \sum_{t=1}^{\infty}\beta^t u(\hat{c}_t) + \frac{\log(\hat{h})}{(1-\beta)} \quad (3.40)$$

where $\hat{c}_t = [u']^{-1}\left(\frac{1}{(\beta(1+r+k))^t}\right)$ and $\hat{h} = \hat{h}(p,\tau_S,\tau) = [(1-\beta)(p+\tau_S+\gamma\tau)]^{-1}$ from the agent's first-order conditions. Differentiate the indirect utility function above to obtain:

$$\frac{dV}{d\tau} = -\hat{h}\left(\frac{dp}{d\tau} + \frac{\partial\tau_S}{\partial\tau} + 1\right) + \left[\frac{\partial U}{\partial h} - (p+\tau_S+\tau)\right]\left[\frac{dp}{d\tau} + \theta_{\tau_S}\frac{\partial\tau_S}{\partial\tau} + \theta_\tau\right]\frac{\partial\hat{h}}{\partial p} \quad (3.41)$$

where $\frac{dp}{d\tau} = \frac{\partial p}{\partial\tau} + \frac{\partial p}{\partial\tau_S}\frac{\partial\tau_S}{\partial\tau}$ is the total incidence of the council tax after having taken into account the shift in stamp duty to guarantee revenue neutrality.

As in Goldin (2015), the change in welfare can be decomposed into four components: the first part, i.e., $-\hat{h}\left(\frac{dp}{d\tau} + \frac{\partial\tau_S}{\partial\tau} + 1\right)$ measures the direct welfare effect of a tax shift due to the alleviation of the borrowing constraint; the second part, i.e., $\left[\frac{\partial U}{\partial h} - (p+\tau_S+\tau)\right]$ is the behavioural wedge and it represents the difference between perceived and actual prices; the third component, i.e., $\left[\frac{dp}{d\tau} + \theta_{\tau_S}\frac{\partial\tau_S}{\partial\tau} + \theta_\tau\right]$ is equal to the change in prices as perceived by the agent; and the fourth component, i.e., $\frac{\partial\hat{h}}{\partial p}$ is the impact of a change in prices on demand for housing.

With no bias, i.e., when $\gamma = 1$ the perceived price is equal to the actual price and the envelope theorem ensures that the second component above is equal to zero. As a consequence, the optimal tax policy will depend on the sign of the first term.[17] If this is positive, it is optimal for the government to set $\tau_S = 0$, if negative,

[17] Notice that $\frac{\partial\tau_S}{\partial\tau} < -1$ because $r+k-\tilde{g} > r-\tilde{g}$, $\theta_\tau < \theta_{\tau_S}$ and $\frac{\partial h}{\partial p} < 0$. The above assumes

$\tau_S = R$. It is easy to show that when $\gamma = 1$ this term is unambiguously positive as long as $\eta_S > 0$. The Government will then choose a zero stamp duty tax in order to alleviate the agent's liquidity constraint.

In the presence of biases, however, there is a trade-off between the two inefficiencies: 1) the liquidity constraint and differences in salience make increasing the stamp duty tax less efficient than raising the council tax; 2) on the other hand, raising the council tax causes a shift in demand away from c_0 which in our example is the shock absorber. In the extreme case when there are no liquidity constraints, it is optimal to impose no council tax. Otherwise, the problem of the social planner will be to choose the optimal combination of stamp duty and council tax to jointly solve the following two equations:

$$\hat{h}\left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right) = R \tag{3.42}$$

$$\frac{dV}{d\tau} = 0 \tag{3.43}$$

Figure C.16 reports the optimal mix of taxes computed for a house worth £430, 000, which is the median value of properties in band D in 2017. The property pays a stamp duty of £11, 500, and we assume that it pays a yearly council tax of £1, 419.73, the in-sample median amount in the corresponding band and year.

The figures confirm the above intuition. The upper panel shows how the optimal combination varies as a function of the discount rate $r + k$, while the bottom panel varies the attenuation parameter γ. From Figure C.16a we can see that when the liquidity premium is zero, the optimal policy is to levy only the stamp duty tax. For a small liquidity premium, there is an optimal mix that includes positive amounts of both taxes. However, the borrowing constraints become dominant fairly quickly and make it optimal to set a zero stamp duty tax.

Figure C.16b, on the other hand, focuses on the effect of salience. Even when the council tax is entirely non-salient, i.e., $\gamma = 0$, it is still optimal to raise a little

that $\frac{r+k-\tilde{g}}{r-\tilde{g}}h + \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)\frac{\partial h}{\partial \tau} > 0$ and $h + \left(\tau_S + \tau\frac{r+k-\tilde{g}}{r-\tilde{g}}\right)\frac{\partial h}{\partial \tau_S} > 0$, i.e., the Government is on the upward sloping part of the Laffer curve. The term $\frac{\partial p}{\partial \tau} + \frac{\partial p}{\partial \tau_S}\frac{\partial \tau_S}{\partial \tau}$ will usually be positive since agents react less to a decrease in council tax relative to a revenue-neutral increase in the stamp duty.

over 20% of revenue through it. As the tax becomes more salient, its distortionary effect on c_0 decreases. Therefore, its proportion should increase up to the point where it becomes the only form of taxation for γ greater than 0.25. It should be noted, however, that this assumes that tax policy changes do not affect any of the parameters. In practice, changing the tax mix can change the inattention parameter γ.

3.6 Conclusions

This paper has studied the incidence of property taxes in the UK housing market. Using a geographical discontinuity approach, exploiting the considerable difference in council tax rates across London Boroughs, we have shown that agents significantly underreact to council taxes.

Our empirical estimates of council tax incidence on house prices are never significantly negative, and this lack of significance cannot be attributed to a lack of statistical power. This is in sharp contrast to the large stamp duty incidence estimated by Best and Kleven (2018) and suggests that agents do not pay sufficient attention to taxes deferred to the future or possibly point to evidence of very large search frictions or other cognitive costs.

In Section 3.5.2, we have touched upon the policy implications of our findings. However, one should be aware of issues arising when manipulating tax rates given that there is no guarantee that policy changes are not followed by changes in tax salience and, therefore, behavior. The analysis in this paper relies on data from the residential property market. However, it can also be extended to other domains of tax policy.

One general takeaway from the present work is that transaction taxes, such as the stamp duty tax, have a large incidence on transaction prices. In contrast, deferred taxes, such as the council tax, have a lower effect on prices but potentially a higher impact on consumption choices. This implies that the optimal mix of taxes may be some combination of the two. The analysis can be extended, for instance, to financial securities where the fact that a transaction tax might be very distortionary does not imply that it is optimal to raise revenues only through capital gains or dividend taxes.[18]

The findings in the paper keep open the question of the nature of the channels through which inattentive households correct their mistakes and adjust their consumption policies once neglected taxes materialize. Access to disaggregated expenditure data could help shed light on this matter: this can be done by analyzing differences in consumption responses at the border between Boroughs, which we should expect to arise whenever agents fail to optimally account for tax differences and are forced to adjust their expenditures ex-post to meet their budget constraints.

[18] While the capital gains tax is a transaction tax, the fact that it is borne by the seller of the asset suggests that agents could still underreact to it as it is a deferred tax and, therefore, less salient compared to a tax charged at the moment of purchase like the stamp duty tax.

Bibliography

ABI. Italy. bank's initiatives in b2b digital marketplace. 2001.

R. Adams, T. Dore, C. Greene, T. Mach, and J. Premo. U.s. consumers' awareness and use of marketplace lending. *Working Paper*, 2017.

R. M. Adams. Do marketplace lending platforms offer lower rates to consumers? *FED Notes*, 2018.

S. Agarwal and A. Karapetyan. Salience and mispricing: Homebuyers' housing decisions. *Norges Bank Working Paper 21*, 2016.

AGCOM. Annual report. 2011.

T. Ahnert, S. Doerr, N. Pierri, and Y. Timmer. Does it help? information technology in banking and entrepreneurship. Technical report, International Monetary Fund, 2021.

A. Akerman, I. Gaarder, and M. Mogstad. The skill complementarity of broadband internet. *The Quarterly Journal of Economics*, 130(4):1781–1824, 2015.

M. Ali Choudhary and N. Limodio. Liquidity risk and long-term finance: Evidence from a natural experiment. *The Review of Economic Studies*, 2021.

H. Allcott. Consumers' Perceptions and Misperceptions of Energy Costs. *American Economic Review*, 101(3):98–104, May 2011.

P. Angelini and N. Cetorelli. The effects of regulatory reform on competition in the banking industry. *journal of Money, credit and banking*, pages 663–684, 2003.

J. D. Angrist and G. W. Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90(430):431–442, 1995.

J. Atack, F. Bateman, M. Haines, and R. A. Margo. Did railroads induce or follow economic growth?: Urbanization and population growth in the american midwest, 1850–1860. *Social Science History*, 34(2):171–197, 2010.

A. Auclert, W. S. Dobbie, and P. Goldsmith-Pinkham. Macroeconomic effects of debt relief: Consumer bankruptcy protections in the great recession. *Working paper*, 2019.

L. M. Ausubel. The failure of competition in the credit card market. *The American Economic Review*, pages 50–81, 1991.

A. Banerjee, E. Duflo, and N. Qian. On the road: Access to transportation infrastructure and economic growth in china. Technical report, National Bureau of Economic Research, 2012.

N. Baum-Snow. Did highways cause suburbanization? *The quarterly journal of economics*, 122(2):775–805, 2007.

A. N. Berger. The economic effects of technological progress: evidence from the banking industry. *Journal of Money, credit and Banking*, pages 141–176, 2003.

A. N. Berger and R. DeYoung. The effects of geographic expansion on bank efficiency. *Journal of financial services research*, 19(2):163–184, 2001.

A. N. Berger and G. F. Udell. The institutional memory hypothesis and the procyclicality of bank lending behavior. *Journal of financial intermediation*, 13(4):458–495, 2004.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1): 249–275, 2004.

M. Bertrand, A. Schoar, and D. Thesmar. Banking deregulation and industry structure: Evidence from the french banking reforms of 1985. *The Journal of Finance*, 62(2):597–628, 2007.

T. Besley, N. Meads, and P. Surico. The Incidence of Transaction Taxes: Evidence from a Stamp Duty Holiday. *Journal of Public Economics*, 119(C):61–70, 2014.

M. Best and H. J. Kleven. Housing market responses to transaction taxes: Evidence from notches and stimulus in the u.k. *Review of Economic Studies*, 85(1):157–193, 2018.

S. Bhattacharya and G. Chiesa. Proprietary information, financial intermediation, and research incentives. *Journal of financial Intermediation*, 4(4):328–357, 1995.

K. Borusyak and X. Jaravel. Revisiting event study designs. *Available at SSRN 2826228*, 2017.

F. Braggion, A. Manconi, N. Pavanini, and H. Zhu. The value of. *European Corporate Governance Institute–Finance Working Paper*, 719, 2020.

J. R. Brown and A. Goolsbee. Does the internet make markets more competitive? evidence from the life insurance industry. *Journal of political economy*, 110(3):481–507, 2002.

G. Buchak, G. Matvos, T. Piskorski, and A. Seru. Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453–483, 2018.

M. Bugamelli, F. Lotti, M. Amici, E. Ciapanna, F. Colonna, F. D'Amuri, S. Giacomelli, A. Linarello, F. Manaresi, G. Palumbo, et al. Productivity growth in italy: a tale of a slow-motion change. *Bank of Italy Occasional Paper*, (422), 2018.

C. W. Calomiris and M. Carlson. Corporate governance and risk management at unprotected banks: National banks in the 1890s. *Journal of Financial Economics*, 119(3):512–532, 2016.

F. Campante, R. Durante, and F. Sobbrio. Politics 2.0: The multifaceted effect of broadband internet on political participation. *Journal of the European Economic Association*, 16(4):1094–1136, 2018.

T. S. Campbell and W. A. Kracaw. Information production, market signalling, and the theory of financial intermediation. *The Journal of Finance*, 35(4):863–882, 1980.

G. Caprio, L. Laeven, and R. Levine. Governance and bank valuation. *Journal of Financial Intermediation*, 16(4):584–617, 2007.

L. Casolaro and G. Gobbi. Information technology and productivity changes in the banking industry. *Economic Notes*, 36(1):43–76, 2007.

J. Cespedes. Heterogeneous sensitivities to interest rate changes: Evidence from consumer loans. *Working paper*, 2019.

N. Cetorelli and L. S. Goldberg. Liquidity management of us global banks: Internal capital markets in the great recession. *Journal of International Economics*, 88(2): 299–311, 2012a.

N. Cetorelli and L. S. Goldberg. Banking globalization and monetary transmission. *The Journal of Finance*, 67(5):1811–1843, 2012b.

N. Cetorelli and P. E. Strahan. Finance as a barrier to entry: Bank competition and industry structure in local us markets. *The Journal of Finance*, 61(1):437–461, 2006.

A. Chandra and E. Thompson. Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional science and urban economics*, 30(4):457–490, 2000.

R. Chetty, A. Looney, and K. Kroft. Salience and Taxation: Theory and Evidence. *American Economic Review*, 99(4):1145–1177, September 2009.

E. Ciapanna and D. Sabbatini. Broadband in italy: an overview. *Bank of Italy Occasional Paper*, (34), 2008.

R. Ciciretti, I. Hasan, and C. Zazzara. Do internet activities add value? evidence from the traditional banks. *Journal of financial services research*, 35(1):81–98, 2009.

D. Comin and B. Hobijn. Cross-country technology adoption: making the theories face the facts. *Journal of monetary Economics*, 51(1):39–83, 2004.

G. S. Crawford, N. Pavanini, and F. Schivardi. Asymmetric information and imperfect competition in lending markets. *American Economic Review*, 108(7):1659–1701, 2018.

N. Czernich, O. Falck, T. Kretschmer, and L. Woessmann. Broadband infrastructure and economic growth. *The Economic Journal*, 121(552):505–532, 2011.

A. D'Andrea and N. Limodio. High-speed internet, financial technology and banking in africa. *BAFFI CAREFIN Centre Research Paper*, (2019-124), 2019.

C. De Chaisemartin and X. d'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96, 2020.

D. De Meza and D. C. Webb. Too much investment: a problem of asymmetric information. *The Quarterly Journal of Economics*, 102(2):281–292, 1987.

D. De Meza and D. C. Webb. Advantageous selection in insurance markets. *RAND Journal of Economics*, pages 249–262, 2001.

D. de Meza, F. Reito, et al. Too little lending: A problem of symmetric information. Technical report, Working Paper, 2019.

C. De Roure, L. Pelizzon, and P. Tasca. How does p2p lending fit into the consumer credit market? 2016.

F. Decarolis and G. Rovigatti. From mad men to maths men: Concentration and buyer power in online advertising. *American Economic Review*.

H. Degryse and S. Ongena. The impact of technology and regulation on the geographical scope of banking. *Oxford review of economic policy*, 20(4):571–590, 2004.

H. Degryse and S. Ongena. Distance, lending relationships, and competition. *The Journal of Finance*, 60(1):231–266, 2005.

T. DeStefano, R. Kneller, and J. Timmis. Broadband infrastructure, ict use and firm performance: Evidence for uk firms. *Journal of Economic Behavior & Organization*, 155:110–139, 2018.

R. DeYoung. The performance of internet-based business models: Evidence from the banking industry. *The Journal of Business*, 78(3):893–948, 2005.

M. Di Maggio and V. Yao. Fintech Borrowers: Lax Screening or Cream-Skimming? *The Review of Financial Studies*, 2020.

W. Dobbie and P. M. Skiba. Information asymmetries in consumer credit markets: Evidence from payday lending. *American Economic Journal: Applied Economics*, 5 (4):256–82, 2013.

W. Dobbie and J. Song. Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *American Economic Review*, 105(3):1272–1311, 2015.

D. Donaldson. Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934, 2018.

M. Draca, R. Sadun, and J. Van Reenen. Productivity and icts: A review of the evidence. *The Oxford handbook of information and communication technologies*, 2007.

G. Duranton and M. A. Turner. Urban growth and transportation. *Review of Economic Studies*, 79(4):1407–1440, 2012.

F. D'Acunto, M. Prokopczuk, and M. Weber. Historical antisemitism, ethnic specialization, and financial development. *The Review of Economic Studies*, 86(3): 1170–1206, 2019.

W. Edelberg. Risk-based pricing of interest rates in household loan markets. *Available at SSRN 484522*, 2003.

W. Edelberg. Testing for adverse selection and moral hazard in consumer loan markets. *Available at SSRN 515903*, 2004.

L. Einav and A. Finkelstein. Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, 25(1):115–38, 2011.

L. Einav, M. Jenkins, and J. Levin. Contract pricing in consumer credit markets. *Econometrica*, 80(4):1387–1432, 2012.

L. Einav, M. Jenkins, and J. Levin. The impact of credit scoring on consumer lending. *The RAND Journal of Economics*, 44(2):249–274, 2013.

European Commission. Taxation Trends in the European Union. *DG Taxation and Customs Union*, 2018.

E. Farhi and X. Gabaix. Optimal taxation with behavioral agents. *American Economic Review*, 110(1):298–336, 2020.

N. Feldman, J. Goldin, and T. Homonoff. Raising the Stakes: Experimental Evidence on the Endogeneity of Taxpayer Mistakes. *National Tax Journal*, 71(2): 201–230, June 2018.

N. E. Feldman and B. J. Ruffle. The Impact of Including, Adding, and Subtracting a Tax on Demand. *American Economic Journal: Economic Policy*, 7(1):95–118, February 2015.

R. Felici and M. Pagnini. Distance, bank heterogeneity and entry in local banking markets. *The Journal of Industrial Economics*, 56(3):500–534, 2008.

A. Finkelstein. E-ztax: Tax Salience and Tax Rates. *The Quarterly Journal of Economics*, 124(3):969–1010, 2009.

R. Fisman, D. Paravisini, and V. Vig. Cultural proximity and loan outcomes. *American Economic Review*, 107(2):457–92, 2017.

D. Foos, L. Norden, and M. Weber. Loan growth and riskiness of banks. *Journal of Banking & Finance*, 34(12):2929–2940, 2010.

C. Forman, A. Goldfarb, and S. Greenstein. The internet and local wages: Convergence or divergence? Technical report, National Bureau of Economic Research, 2009.

W. S. Frame, L. D. Wall, and L. J. White. Technological change and financial innovation in banking: Some implications for fintech. 2018.

G. Gobbi and E. Sette. Do firms benefit from concentrating their borrowing? evidence from the great recession. *Review of Finance*, 18(2):527–560, 2014.

J. Goldin. Optimal Tax Salience. *Journal of Public Economics*, 131(C):115–123, 2015.

P. Goodman. The emergence of homestead exemption in the united states: Accommodation and resistance to the market revolution, 1840-1880. *The Journal of American History*, 80(2):470–498, 1993.

R. E. Gropp, S. Corvoisier, et al. Contestability, technology and banking. Technical report, ZEW-Leibniz Centre for European Economic Research, 2009.

P. Grosjean. The institutional legacy of the ottoman empire: Islamic rule and financial development in south eastern europe. *Journal of Comparative Economics*, 39 (1):1–16, 2011.

S. Guriev, N. Melnikov, and E. Zhuravskaya. 3g internet and confidence in government. *The Quarterly Journal of Economics*, 2021.

R. Hauswald and R. Marquez. Information technology and financial services competition. *The Review of Financial Studies*, 16(3):921–948, 2003.

Z. He, S. Jiang, D. Xu, and X. Yin. Investing in lending technology: It spending in banking. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2021-116), 2021.

A. Hertzberg, A. Liberman, and D. Paravisini. Screening on loan terms: Evidence from maturity choice in consumer credit. *The Review of Financial Studies*, 31(9): 3532–3567, 2018.

J. Hjort and J. Poulsen. The arrival of fast internet and employment in africa. *American Economic Review*, 109(3):1032–79, 2019.

R. M. Hynes, A. Malani, and E. A. Posner. The political economy of property exemption laws. *The Journal of Law and Economics*, 47(1):19–43, 2004.

G. W. Imbens and J. M. Wooldridge. Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1):5–86, 2009.

P. Impiglia, F. Laurenti, L. Magnone, and R. Persico. La rete in rame di telecom italia: caratteristiche e potenzialità per lo sviluppo delle tecnologie xds. *Notiziario Tecnico Telecom Italia*, 13(1):74–89, 2004.

S. Indarte. The impact of debt relief generosity and liquid wealth on household bankruptcy. *Working paper*, 2019.

V. Ioannidou and S. Ongena. "time for a change": loan conditions and bank behavior when firms switch banks. *The Journal of Finance*, 65(5):1847–1877, 2010.

D. M. Jaffee and T. Russell. Imperfect information, uncertainty, and credit rationing. *The Quarterly Journal of Economics*, 90(4):651–666, 1976.

V. Jambulapati and J. Stavins. Credit card act of 2009: What did banks do? *Journal of Banking & Finance*, 46:21–30, 2014.

D. Karlan and J. Zinman. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, 77(6):1993–2008, 2009.

J. Keil and S. Ongena. It's the end of bank branching as we know it (and we feel fine). *September*, 29:20–83, 2020.

A. I. Khwaja and A. Mian. Tracing the impact of bank liquidity shocks: Evidence from an emerging market. *American Economic Review*, 98(4):1413–42, 2008.

J. Kolko. Broadband and local growth. *Journal of Urban Economics*, 71(1):100–113, 2012.

L. Kotlikoff and L. Summers. Tax incidence. In A. J. Auerbach and M. Feldstein, editors, *Handbook of Public Economics*, volume 2, chapter 16, pages 1043–1092. Elsevier, 1 edition, 1987.

R. S. Kroszner and P. E. Strahan. What drives deregulation? economics and politics of the relaxation of bank branching restrictions. *The Quarterly Journal of Economics*, 114(4):1437–1467, 1999.

H. E. Leland and D. H. Pyle. Informational asymmetries, financial structure, and financial intermediation. *The journal of Finance*, 32(2):371–387, 1977.

R. Levine. Financial development and economic growth: views and agenda. *Journal of economic literature*, 35(2):688–726, 1997.

J. M. Liberti and M. A. Petersen. Information: Hard and soft. *Review of Corporate Finance Studies*, 8(1):1–41, 2019.

J. M. Liberti, A. Seru, and V. Vig. Information, credit, and organization. *Credit, and Organization (June 2016)*, 2016.

C. Lin, C. Ma, Y. Sun, and Y. Xu. The telegraph and modern banking development, 1881–1936. *Journal of Financial Economics*, 141(2):730–749, 2021.

M. Manacorda and A. Tesei. Liberation technology: Mobile phones and political mobilization in africa. *Econometrica*, 88(2):533–567, 2020.

G. Michaels. The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4):683–701, 2008.

F. S. Mishkin and P. E. Strahan. What will technology do to financial structure? *NBER working paper*, (w6892), 1999.

L. R. Ngai and S. Tenreyro. Hot and cold seasons in the housing market. *American Economic Review*, 104(12):3991–4026, 2014.

B. of Italy. The italian financial system: trends and perspectives. 2003.

A. J. Padilla and M. Pagano. Endogenous communication among lenders and entrepreneurial incentives. *The review of financial studies*, 10(1):205–236, 1997.

D. Paravisini, V. Rappoport, P. Schnabl, and D. Wolfenzon. Dissecting the effect of credit supply on trade: Evidence from matched credit-export data. *The Review of Economic Studies*, 82(1):333–359, 2015.

L. Pascali. Banks and development: Jewish communities in the italian renaissance and current economic performance. *Review of Economics and Statistics*, 98(1):140–158, 2016.

L. Pascali. The wind of change: Maritime technology, trade, and economic development. *American Economic Review*, 107(9):2821–54, 2017.

M. A. Petersen and R. G. Rajan. Does distance still matter? the information revolution in small business lending. *The journal of Finance*, 57(6):2533–2570, 2002.

M. N. Pierri and M. Y. Timmer. *Tech in fin before fintech: Blessing or curse for financial stability?* International Monetary Fund, 2020.

D. Reck. Taxes and Mistakes: What's in a Sufficient Statistic? *Working Paper*, April 2016.

T. Rephann and A. Isserman. New highways as economic development tools: An evaluation using quasi-experimental matching methods. *Regional Science and Urban Economics*, 24(6):723–751, 1994.

G. Rodano, N. Serrano-Velarde, and E. Tarantino. Bank financing in times of crisis: Evidence from rdd time series estimates. 2013.

L.-H. Roller and L. Waverman. Telecommunications infrastructure and economic development: A simultaneous approach. *American economic review*, 91(4):909–923, 2001.

C. Schenone. Lending relationships and information rents: Do banks exploit their information advantages? *The Review of Financial Studies*, 23(3):1149–1199, 2010.

F. Severino and M. Brown. Personal bankruptcy protection and household debt. *Working Paper*, 2017.

D. A. Skeel. *Debt's Dominion*. Princeton University Press, 2001.

C. Steinwender. Real effects of information frictions: When the states and the kingdom became united. *American Economic Review*, 108(3):657–96, 2018.

J. E. Stiglitz and A. Weiss. Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3):393–410, 1981.

J. E. Stiglitz and A. Weiss. Asymmetric information in credit markets and its implications for macro-economics. *Oxford Economic Papers*, 44(4):694–724, 1992.

J. Stroebel. Asymmetric information about collateral values. *The Journal of Finance*, 71(3):1071–1112, 2016.

H. Tang. Peer-to-peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies*, 32(5):1900–1938, 2019.

D. Taubinsky and A. Rees-Jones. Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment. *Review of Economic Studies*, 85(4):2462–2496, 2018.

B. Vallee and Y. Zeng. Marketplace lending: a new banking paradigm? *The Review of Financial Studies*, 32(5):1939–1982, 2019.

J. Vesala. *Technological transformation and retail banking competition: Implications and measurement*. 2000.

X. Vives and Z. Ye. Information technology and bank competition. *Available at SSRN 3863988*, 2021.

W. J. Wilhelm, Jr. The internet and financial market structure. *Oxford Review of Economic Policy*, 17(2):235–247, 2001.

J. Zinman. Household debt: Facts, puzzles, theories, and policies. *Annual Review of Economics*, 7(1):251–276, 2015.

A. Appendix to Advantageous Selection in Fintech Loans

A.1 Tables

Table A.1: Homestead exemptions by State

State	Homestead Exemption	State	Homestead Exemption
Alabama	15,000	Montana	250,000
Alaska	72,900	Nebraska	60,000
Arizona	150,000	Nevada	550,000
Arkansas	Unlimited	New Hampshire	100,000
California	75,000	New Jersey	0
Colorado	75,000	New Mexico	60,000
Connecticut	75,000	New York	165,550
Delaware	125,000	North Carolina	35,000
District of Columbia	Unlimited	North Dakota	100,000
Florida	Unlimited	Ohio	136,925
Georgia	21,500	Oklahoma	Unlimited
Hawaii	20,000	Oregon	40,000
Idaho	100,000	Pennsylvania	0
Illinois	15,000	Rhode Island	500,000
Indiana	19,300	South Carolina	58,255
Iowa	Unlimited	South Dakota	Unlimited
Kansas	Unlimited	Tennessee	5,000
Kentucky	5,000	Texas	Unlimited
Louisiana	35,000	Utah	20,000
Maine	47,500	Vermont	125,000
Maryland	22,975	Virginia	5,000
Massachusetts	500,000	Washington	125,000
Michigan	30,000	West Virginia	25,000
Minnesota	390,000	Wisconsin	75,000
Mississippi	75,000	Wyoming	20,000
Missouri	15,000		

This table shows the homestead exemption level in all the United States.

Table A.2: Summary Statistics

	Mean	Median	SD
Panel A: Loan Characteristics			
Loan amount	26,143.09	25,000	5,208.29
Interest rate (%)	12.96	12.59	4.65
Installment	739.76	704.10	199.36
Long Maturity	0.50		
Default	0.183		
Panel B: Borrowers Characteristics			
FICO	699.71	695	31.22
Debt-to-Income (%)	19.45	8.63	18.94
Annual Income ($)	109,602.40	92,000	90,750.80
Revolving balance	29,412.97	21,981	35,520.93
Revolving utilization (%)	56.43	57.7	23.72
Delinquencies (2yrs)	0.21		
Home owner	0.10		
Mortgage	0.61		
Rent	0.29		
N		63,300	

This table shows summary statistics of the main sample of Lending Club borrowers in pre-expansion months, which includes all loans whose listing date is between October 2015 and February 2016, for an amount between $20,000 and $35,000.

Table A.3: Distribution of Loan Purposes

Purpose	Frequency	Percent
Car	422	0.35
Credit Card	29,220	24.40
Debt Consolidation	74,131	61.91
Home Improvement	7,979	6.66
House	491	0.41
Major Purchase	1,963	1.64
Medical	463	0.39
Moving	211	0.18
Other	3,352	2.80
Renewable Energy	36	0.03
Small Business	1,354	1.13
Vacation	125	0.10
N	119,747	100

This table shows the distribution of self-reported loan purposes of the main sample of Lending Club borrowers, which includes all loans whose listing date is between October 2015 and July 2016, for an amount between $20,000 and $35,000.

Table A.4: Summary Statistics of USD 35,000 loans

	35,000 Pre	35,000 Post
Panel A: Loan Characteristics		
Installment	1,011.41	1,078.82
Long Maturity	0.51	0.40
Default	0.19	0.21
Panel B: Borrowers Characteristics		
FICO	700.1	698.21
Debt-to-Income (%)	18.82	18.84
Annual Income ($)	148,471	147,173.5
Revolving balance ($)	43,428	40,244.24
Revolving utilization (%)	59.58	58.56
Delinquencies (2yrs)	0.22	0.24
Home owner	0.11	0.12
Mortgage	0.65	0.64
Rent	0.24	0.23
N	9,791	6,712

This table shows summary statistics of Lending Club borrowers choosing $35,000 loans in pre-expansion months, which includes loans whose listing date is between October 2015 and February 2016.

Table A.5: Event study of takeup

	(1)	(2)	(3)
	35k Loan	35k Loan	35k Loan
POST	-0.0369***	-0.0554***	-0.0941***
	(0.00179)	(0.00333)	(0.00539)
Interest Rate	1.860***	2.230***	2.208***
	(0.0412)	(0.0542)	(0.0567)
Default Bin FE	Yes	Yes	Yes
Maturity FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Employment FE	Yes	Yes	Yes
Home Ownership FE	Yes	Yes	Yes
Fico FE	Yes	Yes	Yes
Income Bin FE	Yes	Yes	Yes
Delinquencies FE	Yes	Yes	Yes
DTI Bin FE	Yes	Yes	Yes
Cluster	State	State	State
Sample	20-35	25-35	30-35
Mean	.15	.29	.53
N	117329	60691	33637

This table shows the result of regression in Equation (1.1) and measures the extent of the selection into new loans after March 2016. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan amount equals $35,000, and zero otherwise. The independent variable POST is a dummy that equals one if the loan is issued after March 2016, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio fixed effects. Loans in all columns are issued between October 2015 and July 2016. The first column contains loans between $20,000 and $35,000. The second column contains loans between $25,000 and $35,000. The last column contains loans between $30,000 and $35,000. Standard errors are robust to heteroskedasticity. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table A.6: Regression of default on selected loans

	(1) Default	(2) Default	(3) % Repaid	(4) % Repaid
DID	0.0233***	0.0189***	-0.0139***	-0.0122**
	(0.00532)	(0.00563)	(0.00472)	(0.00486)
Amount Bin FE	Yes	Yes	Yes	Yes
Default Bin-Month FE	No	Yes	No	Yes
Maturity-Month FE	Yes	Yes	Yes	Yes
State-Month FE	Yes	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes	Yes
Cluster	State	State	State	State
Sample	20-35	20-35	20-35	20-35
Mean	.18	.18	.78	.78
R-squared	0.0592	0.0791	0.222	0.233
N	117135	117135	117135	117135

This table shows the result of regression in Equation (1.2) and measures the relative change in the default rate (columns 1 and 2) and in the share of the initial loan repaid (columns 3 and 4) of $35,000 loans after March 2016, relative to smaller loans. The unit of observation is a loan. In columns 1 and 2, the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. In columns 3 and 4, the dependent variable is the percentage of the initial loan eventually repaid. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is $35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regressions in columns 1 and 3 contain $5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regressions in columns 2 and 4 add default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between $20,000 and $35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table A.7: Regression of default on 30,000 loans

	(1)	(2)
	Default	Default
DID 30k	0.00466	0.00332
	(0.00432)	(0.00464)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-34.975	20-34.975
Mean	.18	.18
N	Yes	Yes

This table shows the result of a robustness test. Using the same regression in Equation (1.2) I test whether there is a relative change in default rate of $30,000 loans after March 2016, relative to other loans (but excluding $35,000 loans). The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID 30k is a dummy that equals one if the loan is issued after March 2016 and its amount is $30,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains $5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between $20,000 and $34,975. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table A.8: Summary Statistics of loans and borrowers in low and high exemption states before March 2016

	Low Exemption	High Exemption
Panel A: Loan Characteristics		
Loan amount	26,201.02	26,105
Interest rate (%)	13.03	12.91
Installment	734.28	743.32
Long Maturity	0.53	0.48
Default	0.18	0.18
Panel B: Borrowers Characteristics		
FICO	699.90	699.58
Debt-to-Income (%)	19.99	19.11
Annual Income ($)	106,635.8	111,527.7
Outstanding Debt (no mortgage) ($)	74073.17	71270.66
Delinquencies (2yrs)	0.34	0.33
Home owner	0.12	0.11
Mortgage	0.66	0.56
Rent	0.22	0.33
N	24,924	38,406

This table shows summary statistics of Lending Club borrowers in pre-expansion months, which includes all loans whose listing date is between October 2015 and February 2016, for an amount between $20,000 and $35,000. The first column shows statistics for borrowers living in states with exemption lower or equal than median ($75,000). The second column shows statistics for borrowers living in states with higher than median exemption.

Table A.9: Take up of 35,000 loans on exemption level

	(1) 35k Loan	(2) 35k Loan
DID HSE	0.000891	0.000545
	(0.00373)	(0.00374)
Interest Rate	2.038***	1.918***
	(0.0457)	(0.0406)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	State	State
Cluster	20-35	20-35
Sample	.14	.14
Mean	117135	117135

This table shows the result of regression in Equation (1.3) and measures whether the extent of the selection into new loans after March 2016 is different in states with higher exemption, relative to states with lower exemptions. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan amount equals $35,000, and zero otherwise. The independent variable DID HSE is a dummy that equals one if the loan is issued after March 2016 in a state with higher than median exemption, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains $5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between $20,000 and $35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table A.10: Regression of default on selected loans in different HSE samples

	(1)	(2)	(3)	(4)
	Default	Default	% Repaid	% Repaid
DID	-0.000528	0.0299***	-0.00183	-0.0178***
	(0.0103)	(0.00514)	(0.00986)	(0.00421)
Amount Bin FE	Yes	Yes	Yes	Yes
Default Bin-Month FE	Yes	Yes	Yes	Yes
Maturity-Month FE	Yes	Yes	Yes	Yes
State-Month FE	Yes	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes	Yes
Cluster	State	State	State	State
Sample	20-35	20-35	20-35	20-35
Exemption	Low	High	Low	High
Mean	.18	.18	.77	.79
R-squared	0.0896	0.0856	0.251	0.231
N	45968	70988	45968	70988

This table shows the result of regression in Equation (1.2) and measures the relative change in the default rate (columns 1 and 2) and in the share of the initial loan repaid (columns 3 and 4) of $35,000 loans after March 2016, relative to smaller loans, in samples with different exemptions levels. The unit of observation is a loan. In columns 1 and 2, the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. In columns 3 and 4, the dependent variable is the percentage of the initial loan eventually repaid. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is $35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain $5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The sample in the first column contains loans issued in states with exemption lower or equal than median ($75,000). The sample in the second columns contains loans issued in states with higher than median exemption. Loans in all columns are issued between October 2015 and July 2016 for amounts between $20,000 and $35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table A.11: Regression of default on selected loans in different HSE samples

	(1)	(2)	(3)	(4)
	Default	Default	Default	Default
DID	0.00517	-0.0133	0.0308***	0.0289**
	(0.0122)	(0.0219)	(0.00722)	(0.0103)
Amount Bin FE	Yes	Yes	Yes	Yes
Default Bin-Month FE	Yes	Yes	Yes	Yes
Maturity-Month FE	Yes	Yes	Yes	Yes
State-Month FE	Yes	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes	Yes
Cluster	State	State	State	State
Sample	20-35	20-35	20-35	20-35
HSE Quartile	First	Second	Third	Fourth
Mean	.18	.18	.18	.19
N	26769	19033	42653	28185

This table shows the result of regression in Equation (1.2) and measures the relative change in default rate of $35,000 loans after March 2016, relative to smaller loans, in samples with different exemptions levels. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is $35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain $5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The sample in the first column contains loans issued in states with exemption lower than the first quartile ($21,500). The sample in the second column contains loans issued in states whose exemption is between the first and the second quartile ($75,000). The sample in the third column contains loans issued in states whose exemption is between the second and the third quartile ($250,000). Finally, the sample in the last column contains loans issued in states whose exemption is above or equal than the third quartile. Loans in all columns are issued between October 2015 and July 2016 for amounts between $20,000 and $35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table A.12: Regression of Default on Selected Loans for Other Purposes

	(1) Default	(2) Default	(3) Default
DID	-0.00240	-0.0622	0.0268
	(0.0242)	(0.0422)	(0.0305)
Amount Bin FE	Yes	Yes	Yes
Default Bin-Month FE	Yes	Yes	Yes
Maturity-Month FE	Yes	Yes	Yes
State-Month FE	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes
Cluster	State	State	State
Sample	No Ref	No Ref	No Ref
Exemption		Low	High
Mean	.2	.2	.2
R-squared	0.154	0.255	0.178
N	15737	5699	9847

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table A.13: Regression of Recovery on High Exemption Dummy

	(1)	(2)
	Recovery	% Recovered
High HSE Dummy	58.87	0.00181
	(50.87)	(0.00260)
Default Bin-Month FE	Yes	Yes
Amount Bin FE	Yes	Yes
Maturity-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
SE	Robust	Robust
Sample	Pre March 2016	Pre March 2016
Mean	2202.67	.12
R-squared	0.0926	0.0559
N	11503	11503

This table tests whether post charge-off recoveries are different for states below and above the median exemption. In the first column the dependent variable is the post charge off recovery (in dollars). The dependent variable in the second column is the ratio between the amount recovered and the debt outstanding at the time of default. In both columns High HSE Dummy is equal to one if the loan is issued in a state with higher or equal than median exemption ($75,000). I compute 20 default probability bins to form Default Bin FE. All regressions contain $5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. Loans in all columns are issued in the pre-expansion period, between October 2015 and February 2016 and for amounts between $20,000 and $35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table A.14: Regression of outstanding existing debt on High Exemption Dummy

	(1)	(2)	(3)
	Balance (No Mort)	Install Loans	Revolving
High HSE Dummy	-1946.8***	-1565.8***	-229.0
	(372.3)	(498.6)	(237.9)
Amount Bin FE	Yes	Yes	Yes
Default Bin-Month FE	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes
SE	Robust	Robust	Robust
Sample	Pre March 2016	Pre March 2016	Pre March 2016
Mean	72148.71	43426.33	29303.12
N	63207	30305	63207

This table tests whether borrowers in states with higher exemptions have different levels of indebtedness outside LC. In the first column the dependent variable is total debt (in dollars, excluding mortgages). The dependent variable in the second column is debt in installment loans (in dollars). The dependent variable in the third column is revolving debt balance (in dollars). In all columns High HSE Dummy is equal to one if the loan is issued in a state with higher or equal than median exemption ($75,000). I compute 20 default probability bins to form Default Bin FE. All regressions contain $5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. Loans in all columns are issued in the pre-expansion period, between October 2015 and February 2016 and for amounts between $20,000 and $35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table A.15: Regression of FICO(end)-FICO(start) on selected loans

	(1)	(2)	(3)
	Δ(FICO)	Δ(FICO)	Δ(FICO)
DID	-2.784**	-1.516	-3.332**
	(1.119)	(1.582)	(1.400)
Amount Bin FE	Yes	Yes	Yes
Default Bin-Month FE	Yes	Yes	Yes
Maturity-Month FE	Yes	Yes	Yes
State-Month FE	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes
Cluster	State	State	State
Sample	20-35	20-35	20-35
Mean	-14.2	-14.47	-13.99
R-squared	0.0766	0.0868	0.0849
N	114213	44827	69211

This table shows the result of regression in Equation (1.2) and measures the relative change in the 4-digit range FICO score by $35,000 loans after March 2016, relative to smaller loans. The unit of observation is a loan. In both columns the dependent variable is difference between the last and the first recorded 4-digit FICO score band. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is $35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains $5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, 10% debt-to-income ratio, and default bin (all interacted with month except loan amount bin) fixed effects. The sample in the second column only includes loans issued in states with exemption lower or equal than median ($75,000). The sample in the third column contains loans issued in states with higher than median exemption. Loans in all columns are issued between October 2015 and July 2016 for amounts between $20,000 and $35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

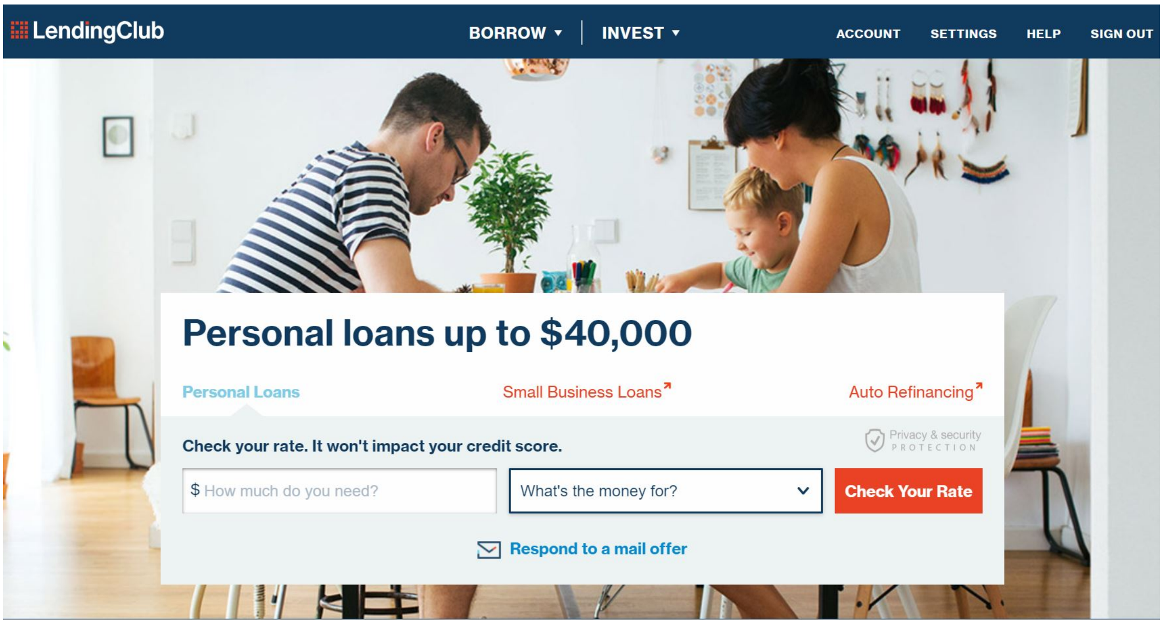
Table A.16: Regression of default on selected loans (up to 40,000)

	(1)	(2)	(3)
	Default	Default	Default
DID	0.0231***	0.00188	0.0349***
	(0.00537)	(0.00941)	(0.00515)
Amount Bin FE	Yes	Yes	Yes
Default Bin-Month FE	Yes	Yes	Yes
Maturity-Month FE	Yes	Yes	Yes
State-Month FE	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes
Cluster	State	State	State
Sample	20-40	20-40	20-40
Mean	.18	-14.47	-13.98
R-squared	0.0792	0.0895	0.0854
N	119506	46935	72395

This table shows the result of regression in Equation (1.2) and measures the relative change in default rate of $35,000 loans after March 2016, relative to other loans (including newly available $40,000). The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is $35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains $5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, 10% debt-to-income ratio, and default bin (all interacted with month except loan amount bin) fixed effects. The sample in the second column only includes loans issued in states with exemption lower or equal than median ($75,000). The sample in the third column contains loans issued in states with higher than median exemption. Loans in all columns are issued between October 2015 and July 2016 for amounts between $20,000 and $40,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

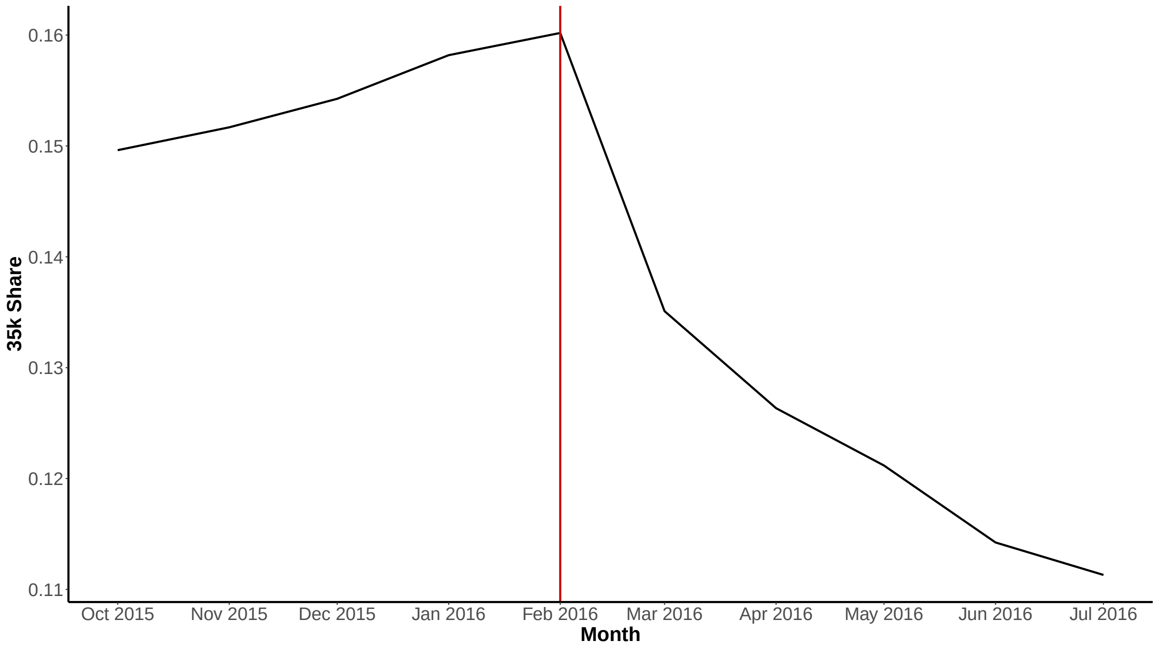
A.2 Figures

Figure A.1: Lending Club Homepage



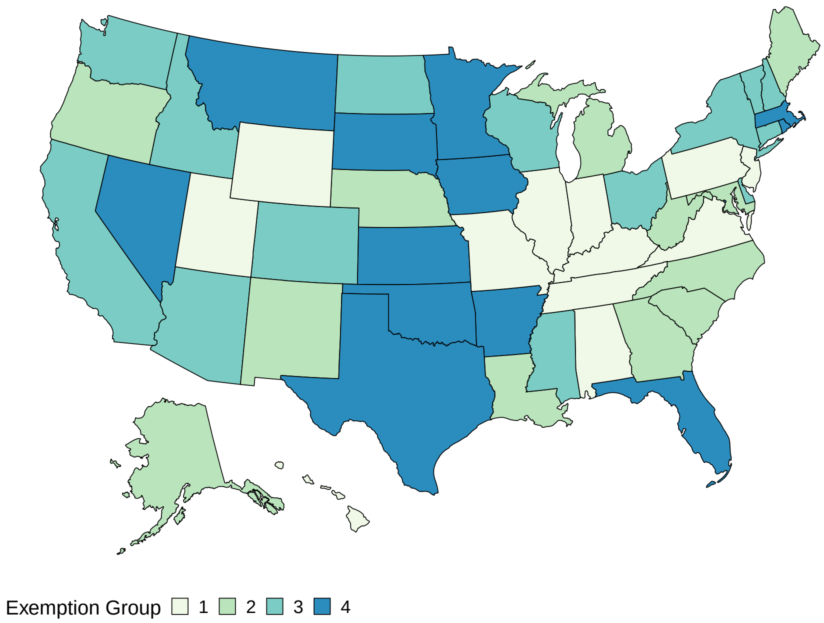
This figure is a screenshot of the Lending Club's website. On the homepage the platform only asks for the amount needed and the purpose. Following steps require more information, such as complete name, address, and Social Security Number.

Figure A.2: Share of 35k loans



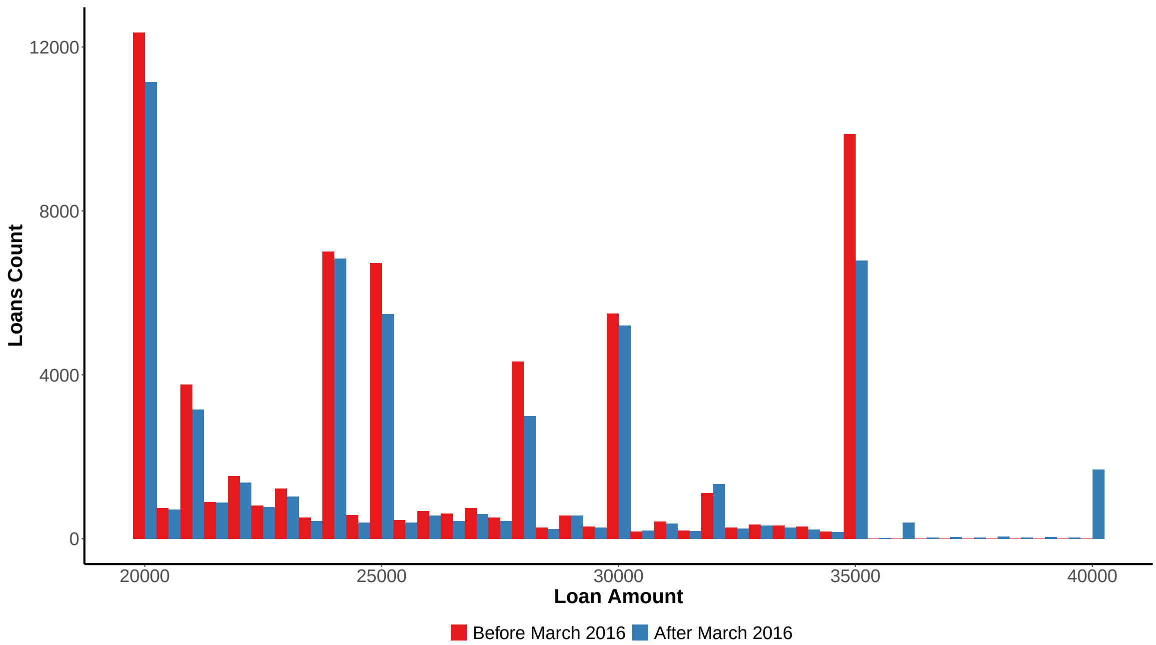
This figure presents the time-series of the share of $35,000 loans issued by Lending Club between October 2015 and July 2016.

Figure A.3: Homestead Exemption Map



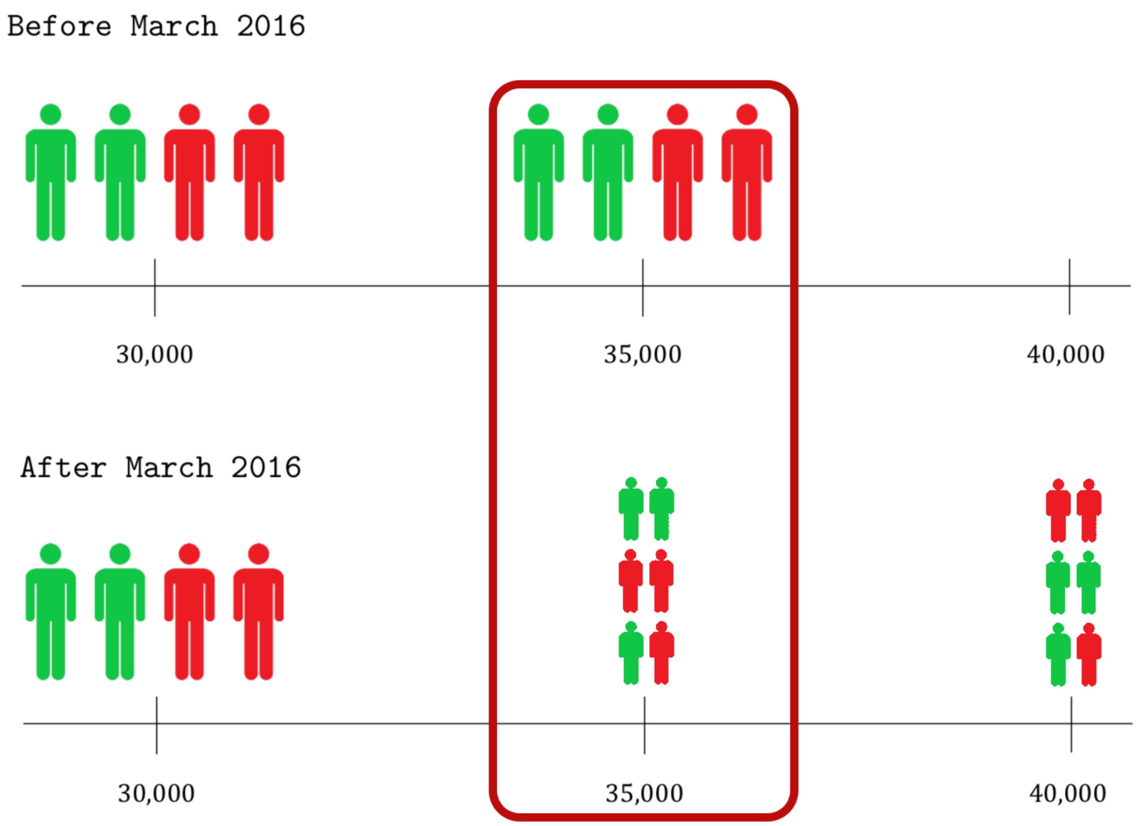
This figure represents the different level of Homestead Exemption in the United States. States in the first group are those with lower exemption levels (such as New Jersey and Pennsylvania), while states in the last group offer the highest level of debt relief (e.g. Texas and Florida).

Figure A.4: Loan amount distributions



This histogram represents the number of loans issued by Lending Club in $500 bands. The sample includes loans issued between October 2015 and July 2016.

Figure A.5: Research Design



This figure depicts the empirical design. In my difference-in-differences specification I compare the change in the default probability of $35,000 loans before and after March 2016 (the squared group) with the same change for smaller loans (to the left). Without loss of generality, I fill safe borrowers in green and risky borrowers in red.

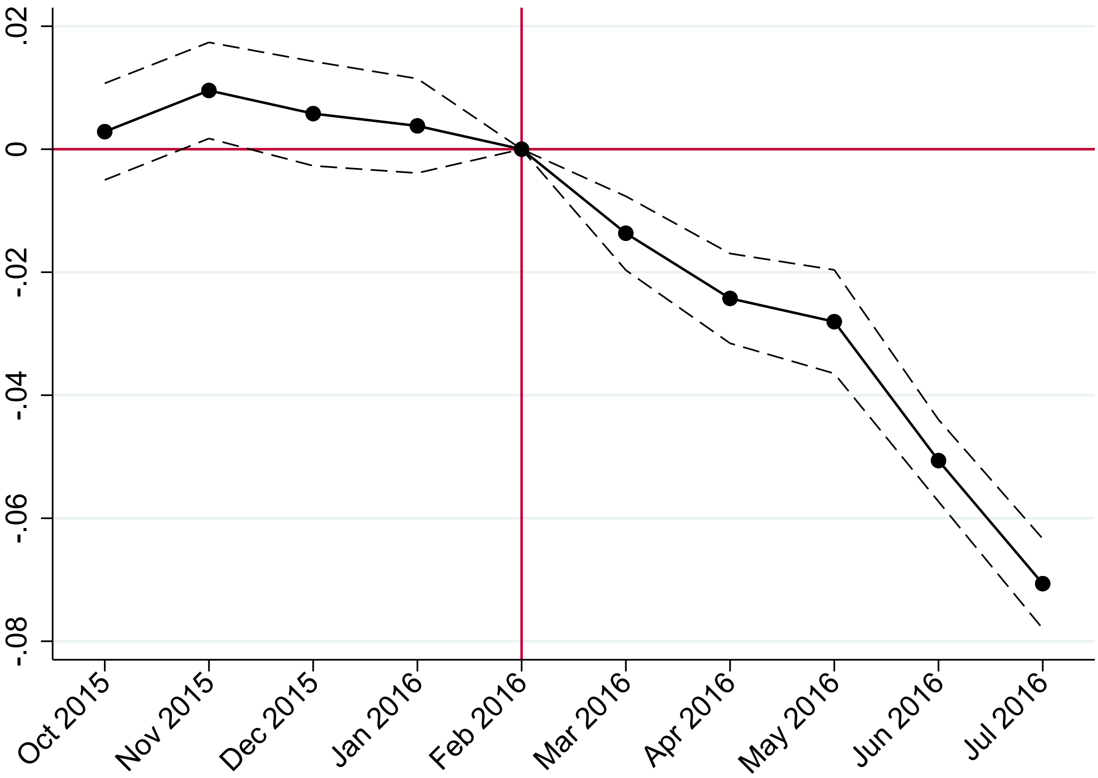


Figure A.6: Pre-trend of $35,000 Origination

This figure presents point estimates and 95% confidence intervals of the regression in Equation 1.1, where the dummy POST has been replaced by monthly dummies (excluding February 2016). The unit of observation is a loan. The dependent variable is a dummy that equals one if the loan amount equals $35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression contains default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio fixed effects. Loans in the sample are issued between October 2015 and July 2016 for amounts between $20,000 and $35,000. Standard errors are robust to heteroskedasticity.

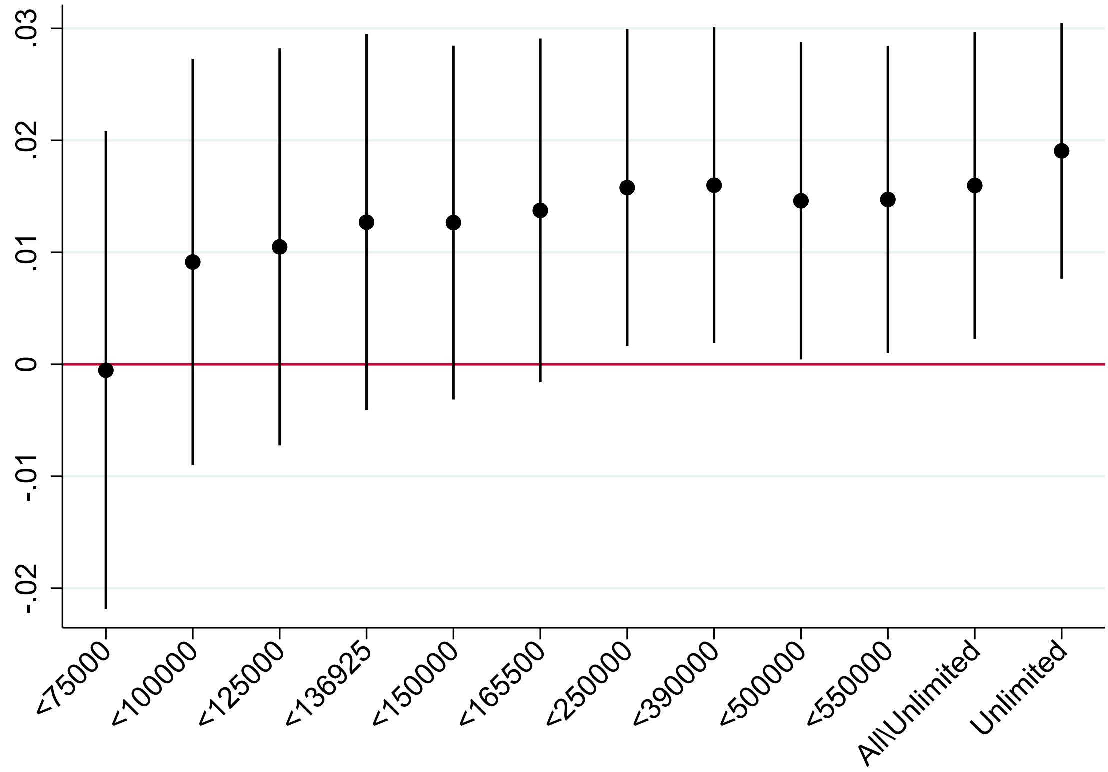
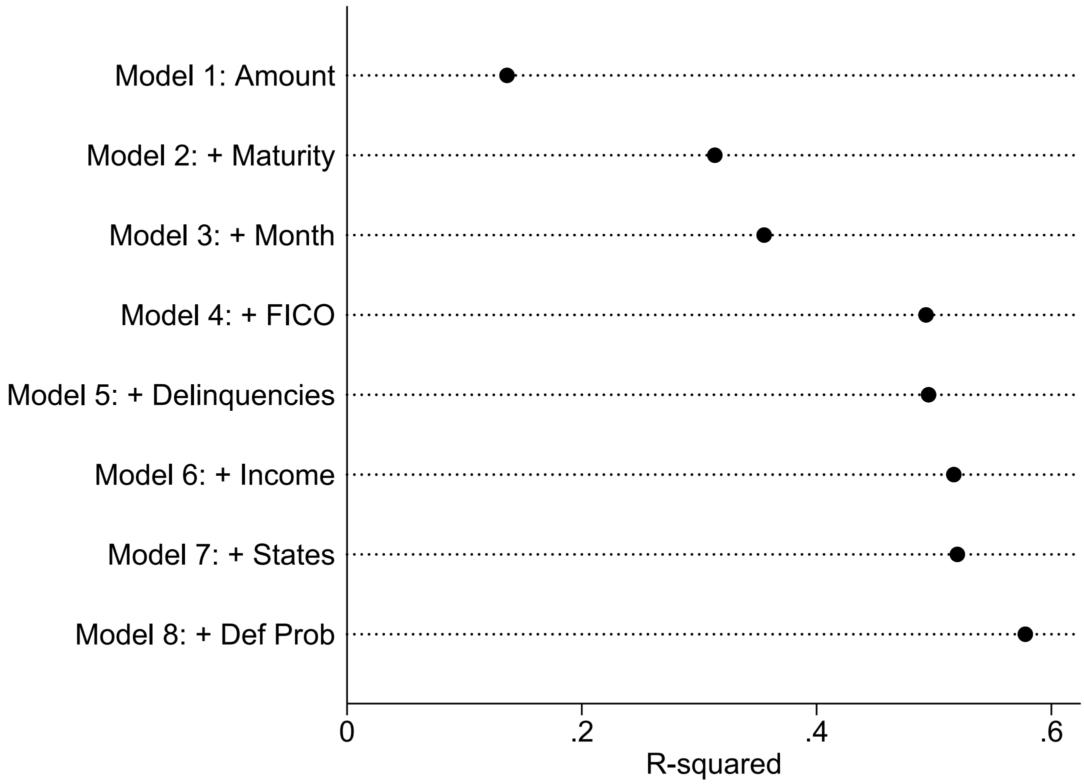


Figure A.7: DID point estimate in different samples

This figure presents point estimates and 95% confidence intervals of the main regression in Equation 1.2, in sample with different Homestead Exemptions (on the x-axis). The unit of observation is a loan. In all regressions the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is $35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain $5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, $10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. All regressions include loans issued between October 2015 and July 2016 for amounts between $20,000 and $35,000. Standard errors are clustered at state level.

Figure A.8: R-squared of LC Interest Rate controlling for different observables



This figure depicts the R-squared of eight regressions of the same dependent variable (interest rate), on a richer and richer set of controls. Model 1 contains only loan size fixed effects. Model 2 adds maturity fixed effects. Model 3 adds month FE. Model 4 adds 4-digit FICO score range fixed effects. Model 5 adds delinquencies in the last two years fixed effects. Model 6 adds $10,000 income bin fixed effects. Model 7 adds state fixed effects. Finally, Model 8 adds default probability bin fixed effects.

B. Broadband and Bank Intermediation

B.1 Tables

Table B.1: Summary Statistics

	Mean	sd	p50	N
	Panel A: Municipality			
Municipalities				5,271
Years				11
North	0.61	0.49	1.00	51,290
Center	0.15	0.36	0.00	51,290
South	0.24	0.43	0.00	51,290
Internet	2.04	2.35	0.00	42,058
Number SLs	1.79	4.04	1.00	51,290
Distance SL	0.40	1.23	0.00	51,290
Number UGSs	0.13	1.10	0.00	51,290
Distance UGS	12.49	8.87	11.07	51,290
Distance prov. capital	21.96	12.93	20.00	50,859
	Panel B: Bank-municipality			
Number of loans	28.23	147.37	8	153,120
Extended credit	29,086.22	282,980.90	3,584.40	153,120
Average interest rate	6.10	1.70	5.98	86,382
	Panel C: Loan			
Extended credit	1,028.48	8,159.02	299.32	4,330,369

Notes: This table reports summary statistics for our final dataset. Panel A refers to data at the municipality level. We provide information on the municipality geographical distribution, as well as on access to broadband and the ADSL underlying infrastructure. Panel B refers to data at the bank-municipality level, that we use quite intensively throughout our analysis. We provide information on the number of loans issued by a bank in a given municipality, the amount of credit granted (in thousands of euros), and the average interest rate charged. Finally, panel C refers to loan level data and reports the credit amount.

Table B.2: First stage regressions

	0-5 (Internet)	Dummy (Good access)	Dummy (Some access)	Years since good internet
distance UGS	-0.053***	-0.010***	-0.009***	-0.035***
$\times$ post 2001	(0.007)	(0.001)	(0.001)	(0.005)
Mun FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Method	OLS	OLS	OLS	OLS
F-statistic	52.6	46.5	46.6	47.2
Mean	2.041	0.437	0.451	1.206
R-squared	0.760	0.750	0.763	0.818
N	41932	41932	41932	41932

Notes: This table reports estimates from OLS as presented in equation (2). The dataset is at the municipality-year level. The dependent variables are: *Internet*, the percentage of households with access to ADSL-based services, in municipality m and year t, on an asymmetric six-point scale; *Good access*, a dummy variable that takes value 1 if broadband access is above 50%, and zero otherwise; *Some access*, a dummy variable that takes value 1 if broadband access is above 0%, and zero otherwise; and *Years since good internet*, a variable that counts the number of years since the percentage of households with access to the ADSL was above 50%. The main predictor is our instrument: the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the F-statistic from the regression; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.3: Regressions of Internet on Banks' Number of Loans and Extended Credit

	(1)	(2)	(3)	(4)
	Ln	Ln	Ln	Ln
	(N. loans)	(Ext. credit)	(N. loans)	(Ext. credit)
Internet	0.007***	0.009***	0.039**	0.081***
	(0.002)	(0.004)	(0.016)	(0.024)
Bank-Mun FE	X	X	X	X
Year FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			134.9	133.6
Mean	28.79	30094.456	28.79	30094.456
R-squared	0.901	0.860		
N	124243	123762	124243	123762

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level. The dependent variables are: *Ln(N. loans)*, the natural logarithm of the number of loans issued by bank b in year t; and *Ln(Ext. Credit)*, the natural logarithm of the amount of loans granted by bank b in year t. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.4: Weights OLS estimates on the extensive margin of the credit relationship

Internet		Instrument	
Year	weight	Year	weight
2004	0.34	2004	0.20
2005	0.27	2005	0.20
2006	0.19	2006	0.20
2007	0.11	2007	0.20
2008	0.08	2008	0.20

Notes: This table reports the weights associated with OLS estimates from equation (1). Estimates are reported for years from 2004 to 2008 only (post broadband). The left panel accounts for the regression of the number of loans (extensive margin of the credit relationship) on *Internet*. The right panel reports weights for the reduced form regression of the number of loans on the instrument, *Distance from UGS* $\times$ *post2001*. The weights associated with the coefficients of *Internet* are decreasing over time. Those associated with the coefficients of the instrument are, instead, constant.

Table B.5: Regressions of Internet on Average Interest Rates

	(1)	(2)
	Average Rate	Average Rate
Internet	-0.018***	-0.107**
	(0.007)	(0.045)
Controls	X	X
Bank-Mun FE	X	X
Year FE	X	X
Method	OLS	IV
F-statistic		318.4
Mean	6.81	6.81
R-squared	0.678	
N	112834	112834

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level. The dependent variable is *Average Rate*, the (weighted) average interest rate on loans issued by bank *b* in year *t*. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Column 1 refers to the basic OLS estimate. Column 2 and refers to the 2SLS estimate, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.6: Regressions of Internet on Firms' Extended Credit

	(1) Ln (Ext. Credit)	(2) Ln (Ext. Credit)	(3) Ln (Ext. Credit)	(4) Ln (Ext. Credit)
Internet	0.007***	0.004***	0.114***	0.058***
	(0.002)	(0.001)	(0.017)	(0.009)
Controls	X	X	X	X
Bank-Year FE	X	X	X	X
Bank-Mun FE	X		X	
Firm-Year FE		X		X
Firm-Branch FE		X		X
Method	OLS	OLS	IV	IV
F-statistic			259.8	335.9
Mean	1057.814	1180.751	1057.814	1180.751
R-squared	0.153	0.948		
N	2115962	1643157	2115962	1643157

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the firm-bank-municipality-year level. The dependent variable is *Ln(Ext. Credit)*, the natural logarithm of the amount of loans granted by bank b to firm f, in year t. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Controls* refer to the loan (one-year lagged) share of revolving loans of the firm, and the loan share of extended credit of the issuing bank. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and bank-year level, in columns 1 and 3. The model is saturated with firm-bank-municipality fixed effects and firm-year fixed effects in columns 2 and 4. The latter aims at isolating the supply effect. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.7: Regressions of Internet on Banks' Lending efficiency (productivity and quality)

	(1)	(2)	(3)	(4)
	Ln	Asinh	Ln	Asinh
	(Ext./Empl.)	(NPLs/N. loans)	(Ext./Empl.)	(NPLs/N. loans)
Internet	0.010**	-0.000	0.072***	-0.002**
	(0.004)	(0.000)	(0.023)	(0.001)
Bank-Mun FE	X	X	X	X
Year FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			123.4	134.9
Mean	1487.241	0.020	1487.241	0.020
R-squared	0.759	0.597		
N	116743	124243	116743	124243

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level. The dependent variables are: *Ln(Ext./Empl.)*, the natural logarithm of the amount of credit issued by bank employee; and *Asinh(NPLs/N. loans)*, the inverse hyperbolic sine of the share of non performing loans on total loans. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.8: Regressions of Internet on Loan Geography

	(1) Share ($\neq$ Prov.)	(2) Asinh (Avg. Distance)	(3) Share ($\neq$ Prov.)	(4) Asinh (Avg. Distance)
Internet	0.001	0.003	0.021***	0.062*
	(0.001)	(0.006)	(0.006)	(0.032)
Bank-Mun FE	X	X	X	X
Year FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			94.0	133.5
Mean			0.16	18.17
R-squared	0.292	0.324		
N	81851	79425	81851	79425

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level, and focuses on new loans. The dependent variables are: *Share(Diff. Province)*, the share of the loans originated outside the province of the bank; and *Asinh(Avg. Distance)*, the inverse hyperbolic sine of the average geodesic distance between the centroid of the municipality of the bank and the location of the firm. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.9: Regressions of Internet on Bank Competitors in the municipality

	(1) Ln (Competitors)	(2) Ln (Competitors)
Internet	0.010***	0.071***
	(0.001)	(0.009)
Mun FE	X	X
Year FE	X	X
Method	OLS	IV
F-statistic		52.3
Mean	3.28	3.28
R-squared	0.962	
N	41858	41858

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the municipality-year level. The dependent variables is *Ln(competitors)*, the natural logarithm of the number of bank (physical) competitors in municipality m, in year t. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Column 1 refers to basic OLS estimates. Column 2 refers to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.10: Regressions of Internet on Competition (Deposits)

	(1) HHI	(2) Top 3 Share	(3) Top 5 Share	(4) HHI	(5) Top 3 Share	(6) Top 5 Share
Internet	-0.004***	-0.005***	-0.003***	-0.025***	-0.026***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.003)	(0.002)	(0.002)
Mun FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Method	OLS	OLS	OLS	IV	IV	IV
F-statistic				88.4	88.4	88.4
Mean	0.68	0.96	0.99	0.68	0.96	0.99
R-squared	0.924	0.651	0.326			
N	49566	49566	49566	49566	49566	49566

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the municipality-year level. The dependent variables are: *HHI*, the Herfindahl–Hirschman Index of bank deposits in municipality *m* and year *t*; *Top 3 share*, the share of deposits owned by top 3 banks in the municipality; and *Top 5 share*, the share of deposits owned by top 5 banks in the municipality. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 to 3 refer to basic OLS estimates. Columns 4 to 6 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.11: Regressions of Internet on Existing Relationships

	(1) Dummy (Multiple)	(2) Dummy (Multiple Bank)	(3) Dummy (Multiple)	(4) Dummy (Multiple Bank)
Internet	0.001	0.001	-0.031***	0.034**
	(0.001)	(0.002)	(0.009)	(0.014)
Bank-Year FE	X	X	X	X
Bank-Mun FE	X	X	X	X
Firm FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			191.5	191.5
Mean	0.91	0.18	0.91	0.18
R-squared	0.650	0.525		
N	633732	633732	633732	633732

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the firm-bank-municipality-year level, and focuses on new loans. The dependent variables are: *Dummy(Multiple)*, a dummy variable for the loan issued to a firm already having a credit relationship; and *Dummy(Multiple Bank)*, a dummy variable for the loan issued to a firm already having a credit relationship with the same bank (in a different place). The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality, firm and bank-year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.12: Regressions of Internet on Firms' Ext. Credit - Small Municipalities

	(1) Ln (Ext. Credit)	(2) Ln (Ext. Credit)	(3) Ln (Ext. Credit)	(4) Ln (Ext. Credit)
Internet	0.001	-0.001	-0.031	-0.005
	(0.003)	(0.004)	(0.029)	(0.046)
Controls	X	X	X	X
Bank-Year FE	X	X	X	X
Bank-Mun FE	X		X	
Firm-Year FE		X		X
Firm-Branch FE		X		X
Method	OLS	OLS	IV	IV
F-statistic			22.4	18.9
Mean	665.857	746.809	665.857	746.809
R-squared	0.243	0.967		
N	130647	47709	130647	47709

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the firm-bank-municipality-year level, and includes information on small municipalities (below the median of population) only. The dependent variable is *Ln(Ext. Credit)*, the natural logarithm of the amount of loans granted by bank b to firm f, in year t. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Controls* refer to the loan (one-year lagged) share of revolving loans of the firm, and the loan share of extended credit of the issuing bank. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and bank-year level, in columns 1 and 3. The model is saturated with firm-bank-municipality fixed effects and firm-year fixed effects in columns 2 and 4. The latter aims at isolating the supply effect. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.13: Regressions of Internet on the Easiest Target

	Dummy (Firm in small muni,out prov, with internet)	Dummy (Firm in small muni,out prov, with internet)
Internet	0.007***	0.054***
	(0.002)	(0.018)
Bank-Year FE	X	X
Bank-Mun FE	X	X
Firm FE	X	X
Method	OLS	IV
F-statistic		86.2
Mean	0.04	0.04
R-squared	0.598	
N	550197	550197

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the firm-bank-municipality-year level, and focuses on new loans. The dependent variable is a dummy that identifies loans to firms in small municipalities connected to fast internet, out of the province of the bank, granted by banks in municipalities with at least 10,000 inhabitants. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Column 1 refers to basic OLS estimates. Column 2 refers to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality, firm and bank-year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.14: Internet on Bank Branches in the municipality

	(1)	(2)	(3)	(4)
	Ln (Branches)	Ln (Branches) Small	Ln (Branches)	Ln (Branches) Small
Internet	0.004***	0.000	0.054***	0.003
	(0.001)	(0.000)	(0.006)	(0.002)
Bank-Mun FE	X	X	X	X
Year FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			111.9	166.2
Mean	1.88	1.04	1.88	1.04
R-squared	0.950	0.894		
N	137691	45837	137691	45837

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level. The dependent variables is *Ln(branches)*, the natural logarithm of the number of branches of bank b in municipality m, in year t. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. Columns 1 and 3 refer to the all sample. Columns 2 and 4 refer to the sample of small municipalities. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality (or municipality) and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.15: Regressions of Internet on Income and Population

	(1) ln (Income)	(2) ln (Pop.)	(3) ln (Income p.c.)	(4) ln (Income)	(5) ln (Pop.)	(6) ln (Income p.c.)
Internet	0.003***	0.002***	0.002***	0.040***	0.028***	0.017***
	(0.000)	(0.000)	(0.000)	(0.003)	(0.002)	(0.002)
Mun FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Method	OLS	OLS	OLS	IV	IV	IV
F-statistic				52.6	53.7	53.7
Mean	123.914	11318	10.062	123.914	11318	10.062
R-squared	0.998	0.999	0.986			
N	33268	41932	33268	33268	41932	33268

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the municipality-year level. The dependent variables are: *Ln(Income)*, the natural logarithm of income; *Ln(Pop.)*, the natural logarithm of population; and *Ln(Income p.c.)*, the natural logarithm of income per capita. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 to 3 refer to basic OLS estimates. Columns 4 to 6 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

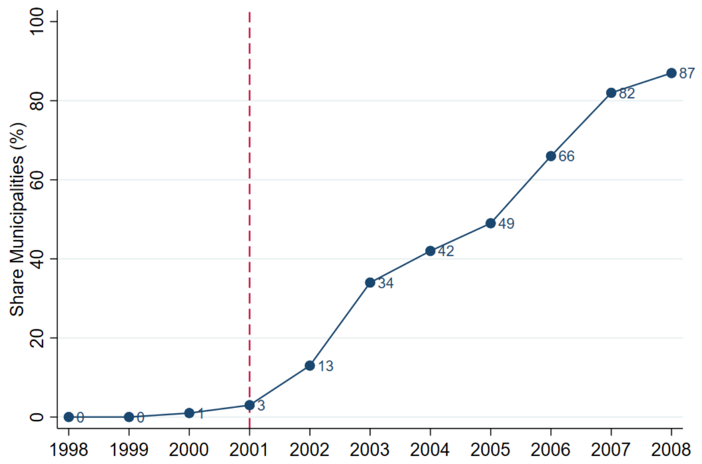
Table B.16: Regressions of Internet on Income and Population - Small municipalities

	(1) ln (Income)	(2) ln (Pop.)	(3) ln (Income p.c.)	(4) ln (Income)	(5) ln (Pop.)	(6) ln (Income p.c.)
Internet	0.004***	0.002***	0.002***	0.064***	0.053***	0.021***
	(0.000)	(0.000)	(0.000)	(0.007)	(0.005)	(0.004)
Mun FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Method	OLS	OLS	OLS	IV	IV	IV
F-statistic				122.8	126.4	122.8
Mean	25.125	2517	9.988	25.125	2517	9.988
R-squared	0.992	0.995	0.979			
N	16630	20913	16630	16630	20913	16630

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the municipality-year level, and includes information on small municipalities (below the median of population) only. The dependent variables are: *Ln(Income)*, the natural logarithm of income; *Ln(Pop.)*, the natural logarithm of population; and *Ln(Income p.c.)*, the natural logarithm of income per capita. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 to 3 refer to basic OLS estimates. Columns 4 to 6 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

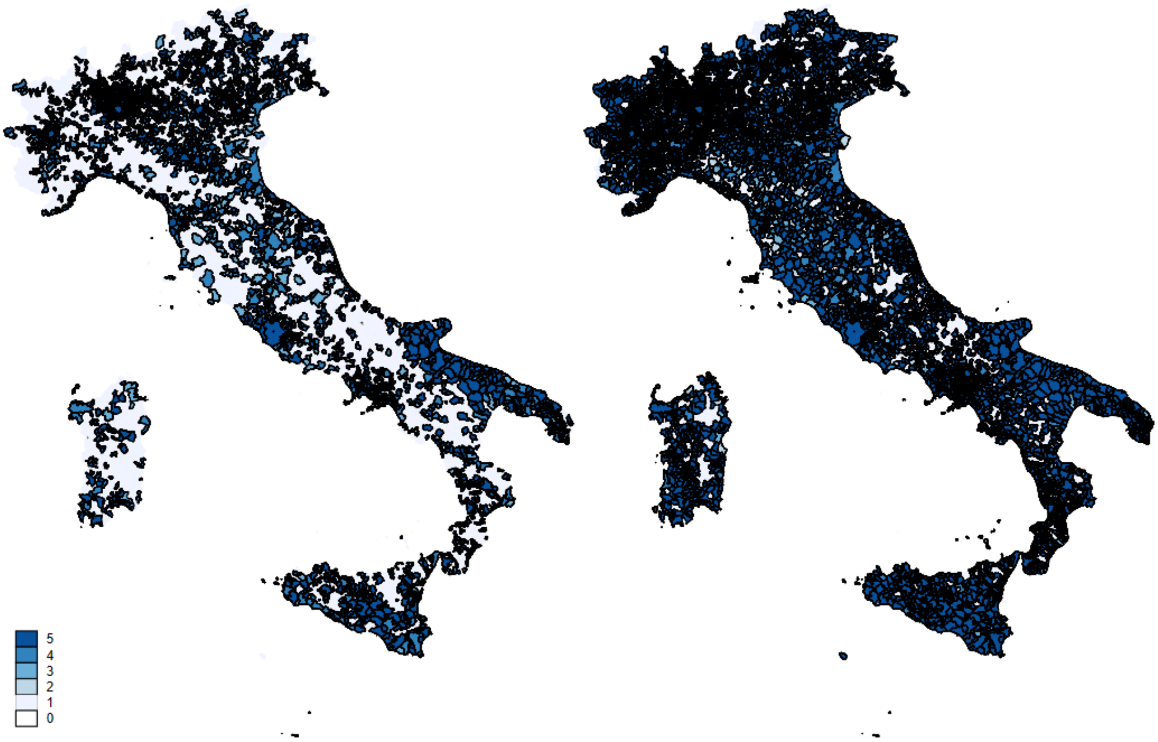
B.2 Figures

Figure B.1: Broadband internet in Italy



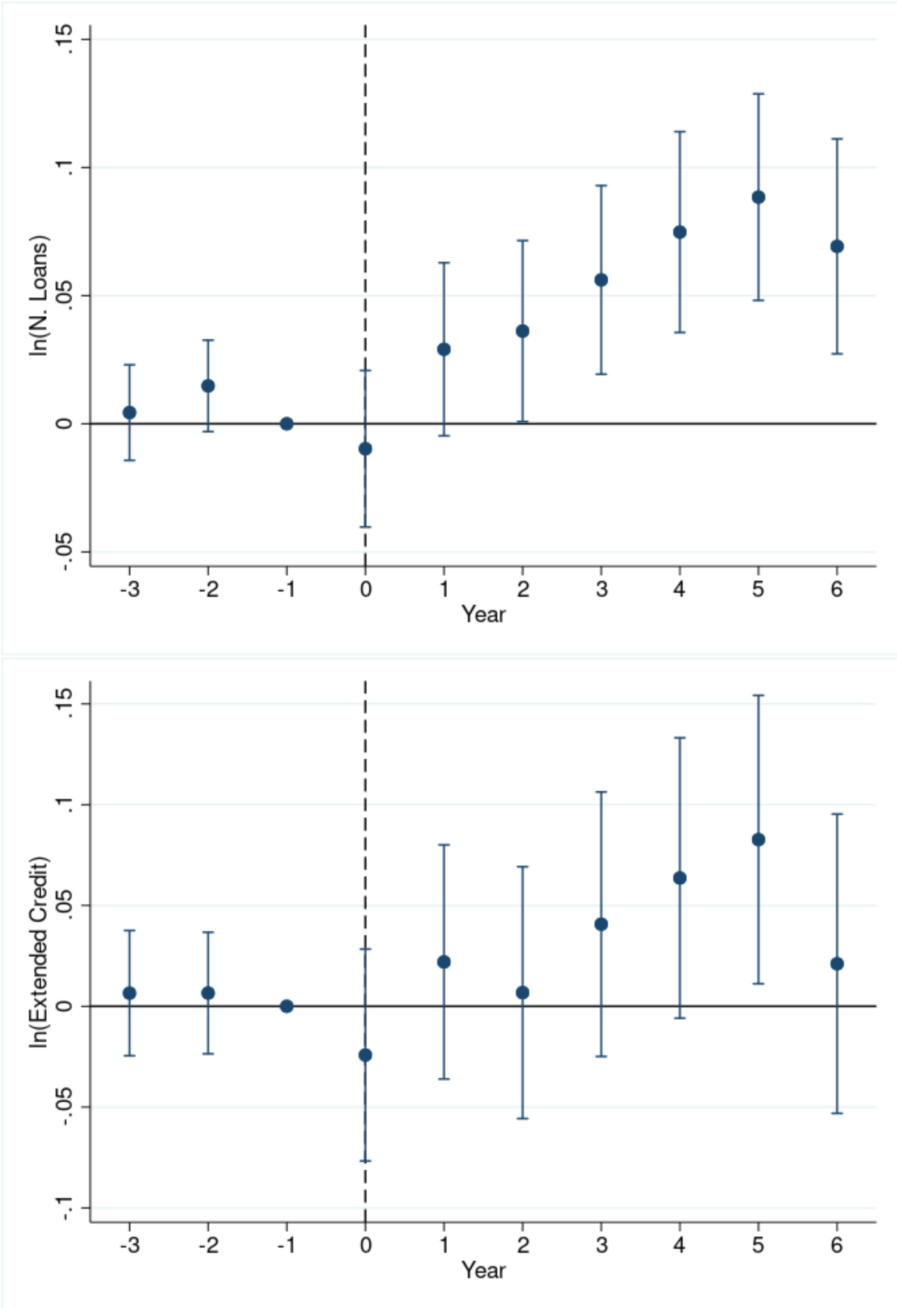
Notes: Broadband diffusion in Italy between 2000 and 2008. On the y-axis, we report the share of municipality with access to the ADSL technology. On the x-axis, the years. The dashed vertical line indicates the separation between the pre-broadband and the post-broadband period, that we make coincide with 2001.

Figure B.2: Geographical distribution of Broadband internet



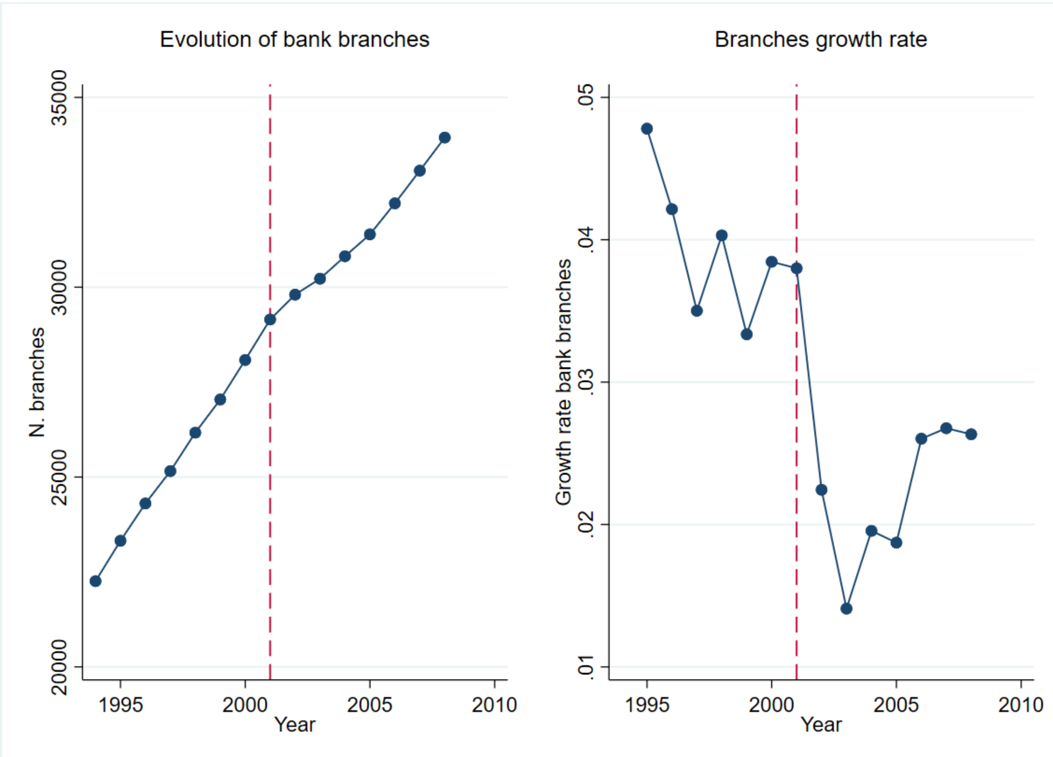
Notes: Geographical distribution of ADSL access in Italian municipalities. The left panel refers to 2004, the first year for which data on ADSL are available. The right panel refers to 2008, the last year in our sample. The measure of broadband internet is the baseline six-point asymmetric scale of ADSL coverage. Lighter colors indicate no or low access. Darker colors indicate high or full access.

Figure B.3: DiD Event study: number of loans and credit amount



Notes: DiD setting. The treatment group is made by banks in municipalities with access to ADSL in 2006 (early adopters). The control group is made by banks in municipalities with no access to ADSL in 2006 (late adopters). Year 0 corresponds to 2002, the first year in which broadband internet is available. In the top panel, on the y-axis is $ln(N.\ loans)$, the natural logarithm of the number of loans issued by each bank. In the bottom panel, on the y-axis is $ln(Ext.\ credit)$, the natural logarithm of the total amount of credit granted by each bank. Both panels follow the indications of Borusyak and Jaravel (2017) and drop 1998, in addition to the baseline year (2001), from the computations.

Figure B.4: Bank Branches in Italy, 1995-2010



Notes: This figure plots the evolution of bank branches in Italy during our sample period. On the left is reported the time series of the total number of branches. On the right is the growth rate of branches by year.

B.3 Additional Tables

Table B.17: Balance Table

	Close	Far	Norm. diff.	N.
Surface	35.31	55.06	(-.24)	4400
Altitude	205.08	336.16	**(-.37)**	4400
North	.69	.55	(.21)	4528
SL per capita	.24	.43	**(-.39)**	4528
Dist. province capital	2.67	3.23	**(-.61)**	4492
Pop. growth	.07	.04	**(.3)**	4361
Adults growth	.05	.02	(.21)	4528
Graduate growth	.88	.8	(.15)	4528
Foreigners growth	2.53	2.83	(-.09)	4524
Buildings growth	.13	.1	(.13)	4528
Firms growth	.1	.04	**(.31)**	4528
Employees growth	.05	.05	(.02)	4528
Income p.c. growth	.19	.17	(.19)	4361

Notes: balance table. This table compares several geographical and socioeconomic indicators, for municipalities that are at different distances from the necessary infrastructure for broadband. We distinguish between municipalities close (below the median of distance) and far (above the median of distance) from the closest UGS. Column 1 reports the average value of each variable for municipalities close to the UGS. Column 2 reports the average value of each variable for municipalities far from the UGS. Column 3 reports the normalized difference as in Imbens and Wooldridge (2009). Values above 0.25 can be considered problematic. Finally, column 4 reports the total number of observations (municipalities).

Table B.18: Regressions of Internet on Banks' Number of Loans and Extended Credit with Controls

	(1)	(2)	(3)	(4)
	Ln	Ln	Ln	Ln
	(N. loans)	(Ext. credit)	(N. loans)	(Ext. credit)
Internet	0.006**	0.007	0.029*	0.064***
	(0.002)	(0.004)	(0.016)	(0.023)
Controls	X	X	X	X
Bank-Mun FE	X	X	X	X
Year FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			142.5	141.2
Mean	28.9	30240.062	28.9	30240.062
R-squared	0.902	0.861		
N	123350	122869	123350	122869

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level. The dependent variables are: *Ln(N. loans)*, the natural logarithm of the number of loans issued by bank b in year t; and *Ln(Ext. Credit)*, the natural logarithm of the amount of loans granted by bank b in year t. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Controls* refers to municipality-level variables: the natural logarithm of total population; elderly population, number of private firms, number of employees, distance from the provincial capital, interacted with a second-order polynomial-time trend. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.19: Regressions of Internet on Banks' Number of Loans

	Ln (N. loans)	Ln (N. loans)	Ln (N. loans)	Ln (N. loans)
Internet	0.039**			
	(0.016)			
Years Since Good Internet		0.063**		
		(0.026)		
Good access			0.226**	
			(0.095)	
Some access				0.254**
				(0.107)
Bank-Mun FE	X	X	X	X
Year FE	X	X	X	X
Method	IV	IV	IV	IV
F-statistic	134.9	121.7	117.9	118.3
Mean	28.79	28.79	28.79	28.79
R-squared				
N	124243	124243	124243	124243

Notes: This table reports estimates from 2SLS as presented in equation (3). The dataset is at the bank-municipality-year level. The dependent variable is *Ln(N. loans)*, the natural logarithm of the number of loans issued by bank b in year t. The main predictors are: *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale; *Years since good internet*, a variable that counts the number of years since the percentage of households with access to the ADSL was above 50%; *Good access*, a dummy variable that takes value 1 if broadband access is above 50%, and zero otherwise; and *Some access*, a dummy variable that takes value 1 if broadband access is above 0%, and zero otherwise. All our predictors are instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.20: Regressions of Internet Placebo on Banks' Number of Lines and Firm's Extended Credit

	(1) Ln (N. loans)	(2) Ln (Ext. Credit)	(3) Ln (N. loans)	(4) Ln (Ext. Credit)
Internet placebo	0.000	-0.005	0.008	0.001
	(0.003)	(0.005)	(0.009)	(0.015)
Bank-Mun FE	X	X	X	X
Year FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			100.4	99.6
Mean	26.12	24395.638	26.12	24395.638
R-squared	0.932	0.906		
N	72277	71905	72277	71905

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level. The dependent variables are: *Ln(N. loans)*, the natural logarithm of the number of loans issued by bank *b* in year *t*; and *Ln(Ext. Credit)*, the natural logarithm of the amount of loans granted by bank *b* in year *t*. The main predictor is *Internet placebo*, a fake measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. The sample includes years from 1998 to 2003, where we assign ADSL data of 2006 to years from 2001 to 2003. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.21: Regressions of Internet on the Share of Connected Firms out of the municipality

	(1)	(2)
	Share	Share
	(Connected firms)	(Connected firms)
Internet	0.021***	0.116***
	(0.002)	(0.010)
Bank-Mun FE	X	X
Year FE	X	X
Method	OLS	IV
F-statistic		132.2
Mean	0.31	0.31
R-squared	0.515	
N	92654	92654

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level, and focuses on new loans. The dependent variables is the *Share(Connected firms)*, the share of the loans originated with firms connected to broadband. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Column 1 refers to basic OLS estimates. Column 2 refers to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.22: Regressions of Internet on Loan Geography

	(1) Dummy (Diff. Province)	(2) Asinh (Distance)	(3) Dummy (Diff. Province)	(4) Asinh (Distance)
Internet	0.001	0.003	0.033***	0.074**
	(0.002)	(0.007)	(0.011)	(0.037)
Bank-Year FE	X	X	X	X
Bank-Mun FE	X	X	X	X
Firm FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			191.5	215.4
Mean	0.42	77.35	0.42	77.35
R-squared	0.672	0.756		
N	633732	567594	633732	567594

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the firm-bank-municipality-year level, and focuses on new loans. The dependent variables are: *Dummy(Diff. Province)*, a dummy variable for the loan being originated outside the province of the bank; and *Asinh(Distance)*, the inverse hyperbolic sine of the geodesic distance between the centroid of the municipality of the bank and the location of the firm. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality, firm and bank-year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.23: Regressions of Internet on Average Interest Rates - Small Municipalities

	(1)	(2)
	Average Rate	Average Rate
Internet	0.009	0.309**
	(0.012)	(0.139)
Bank-Mun FE	X	X
Year FE	X	X
Method	OLS	IV
F-statistic		43.0
Mean	6.25	6.25
R-squared	0.546	
N	16637	16637

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level, and includes information on small municipalities (below the median of population) only. The dependent variable is *Average Rate*, the (weighted) average interest rate on loans issued by bank b in year t. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Column 1 refers to the basic OLS estimate. Column 2 and refers to the 2SLS estimate, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.24: Regressions of Internet on Banks' Internal efficiency (productivity and quality) - Small Municipalities

	(1) Ln (Ext./Empl.)	(2) Asinh (NPLs/N. loans)	(3) Ln (Ext./Empl.)	(4) Asinh (NPLs/N. loans)
Internet	-0.014**	0.000	-0.165***	-0.002
	(0.006)	(0.000)	(0.055)	(0.003)
Bank-Mun FE	X	X	X	X
Year FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			81.6	90.6
Mean	750.624	0.01	750.624	0.01
R-squared	0.720	0.337		
N	29465	31045	29465	31045

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the bank-municipality-year level, and includes information on small municipalities (below the median of population) only. The dependent variables are: *Ln(Ext./Empl.)*, the natural logarithm of the amount of credit issued by bank employee; and *Asinh(NPLs/N. loans)*, the inverse hyperbolic sine of the share of non performing loans on total loans. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.25: Regressions of Internet on Bank Competitors in the municipality - Small Municipalities

	(1) Ln (Competitors)	(2) Ln (Competitors)
Internet	0.002**	-0.007
	(0.001)	(0.010)
Mun FE	X	X
Year FE	X	X
Method	OLS	IV
F statistic		125.4
Mean	2.03	2.03
R-squared	0.908	
N	20839	20839

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the municipality-year level, and includes information on small municipalities (below the median of population) only. The dependent variable is *Ln(competitors)*, the natural logarithm of the number of bank (physical) competitors in municipality m, in year t. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Column 1 refers to basic OLS estimates. Column 2 refers to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.26: Regressions of Internet on Competition (Deposits) - Small Municipalities

	(1) HHI	(2) Top 3 Share	(3) Top 5 Share	(4) HHI	(5) Top 3 Share	(6) Top 5 Share
Internet	-0.003***	-0.002***	-0.002***	-0.032***	-0.023***	-0.023***
	(0.001)	(0.000)	(0.000)	(0.005)	(0.003)	(0.003)
Mun FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Method	OLS	OLS	OLS	IV	IV	IV
F-statistic				215.0	215.0	215.0
Mean	0.84	0.99	0.99	0.84	0.99	0.99
R-squared	0.824	0.192	0.180			
N	28161	28161	28161	28161	28161	28161

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the municipality-year level, and includes information on small municipalities (below the median of population) only. The dependent variables are: *HHI*, the Herfindahl–Hirschman Index of bank deposits in municipality m and year t; *Top 3 share*, the share of deposits owned by top 3 banks in the municipality; and *Top 5 share*, the share of deposits owned by top 5 banks in the municipality. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 to 3 refer to basic OLS estimates. Columns 4 to 6 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the municipality and year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table B.27: Regressions of Internet on Loan Geography - Small Municipalities

	(1)	(2)	(3)	(4)
	Dummy	Asinh	Dummy	Asinh
	(Diff. Province)	(Distance)	(Diff. Province)	(Distance)
Internet	0.004**	-0.008	0.019	0.039
	(0.002)	(0.009)	(0.021)	(0.068)
Bank-Year FE	X	X	X	X
Bank-Mun FE	X	X	X	X
Firm FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			4.7	4.0
Mean	0.15	14.61	0.15	14.61
R-squared	0.948	0.955		
N	9572	8120	9572	8120

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the firm-bank-municipality-year level. It focuses on new loans, and includes information on small municipalities (below the median of population) only. The dependent variables are: *Dummy(Diff. Province)*, a dummy variable for the loan being originated outside the province of the bank; and *Asinh(Distance)*, the inverse hyperbolic sine of the geodesic distance between the centroid of the municipality of the bank and the location of the firm. The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality, firm and bank-year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

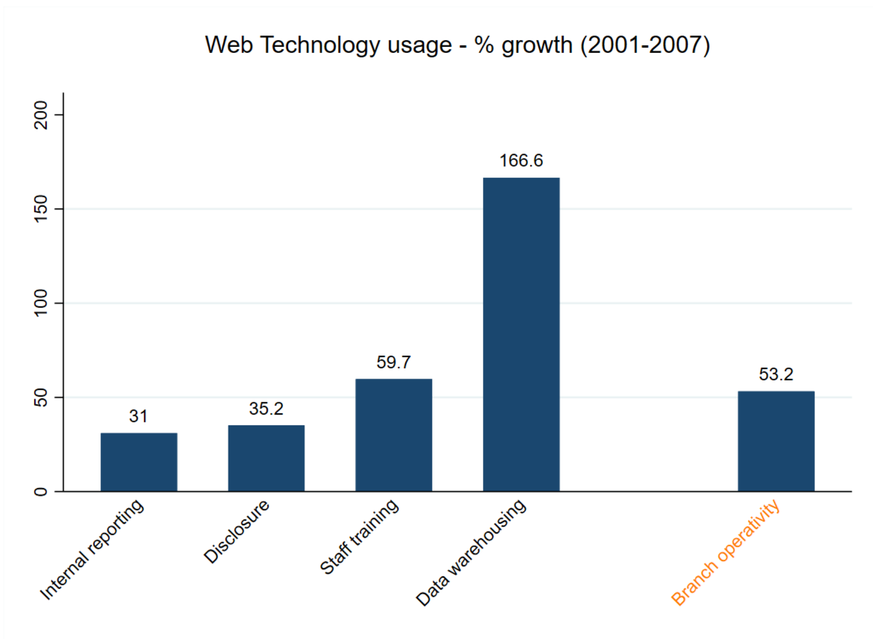
Table B.28: Regressions of Internet on Existing Relationships - Small Municipalities

	(1) Dummy (Multiple)	(2) Dummy (Multiple Bank)	(3) Dummy (Multiple)	(4) Dummy (Multiple Bank)
Internet	0.011*	0.002	-0.014	-0.059
	(0.006)	(0.005)	(0.057)	(0.047)
Bank-Year FE	X	X	X	X
Bank-Mun FE	X	X	X	X
Firm FE	X	X	X	X
Method	OLS	OLS	IV	IV
F-statistic			4.7	4.7
Mean	0.86	0.1	0.86	0.1
R-squared	0.836	0.793		
N	9572	9572	9572	9572

Notes: This table reports estimates from OLS and 2SLS as presented in equations (1) and (3). The dataset is at the firm-bank-municipality-year level. It focuses on new loans, and includes information on small municipalities (below the median of population) only. The dependent variables are: *Dummy(Multiple)*, a dummy variable for the loan issued to a firm already having a credit relationship; and *Dummy(Multiple Bank)*, a dummy variable for the loan issued to a firm already having a credit relationship with the same bank (in a different place). The main predictor is *Internet*, a measure of ADSL coverage in the municipality, based on a six-point asymmetric scale. Columns 1 and 2 refer to basic OLS estimates. Columns 3 and 4 refer to 2SLS estimates, where the variable *Internet* is instrumented by the interaction between *Distance from UGS* and a dummy variable *post2001*. *Method* reports the used estimator; *F-statistic* reports the Sanderson-Windmeijer multivariate F-statistic, when the 2SLS methodology is adopted; *Mean* refers to the mean of the dependent variable; *R-squared* is the adjusted R^2; and *N.* refers to the number of observations. Fixed effects are at the bank-municipality, firm and bank-year level. Standard errors, in parentheses, are clustered at the municipality level. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

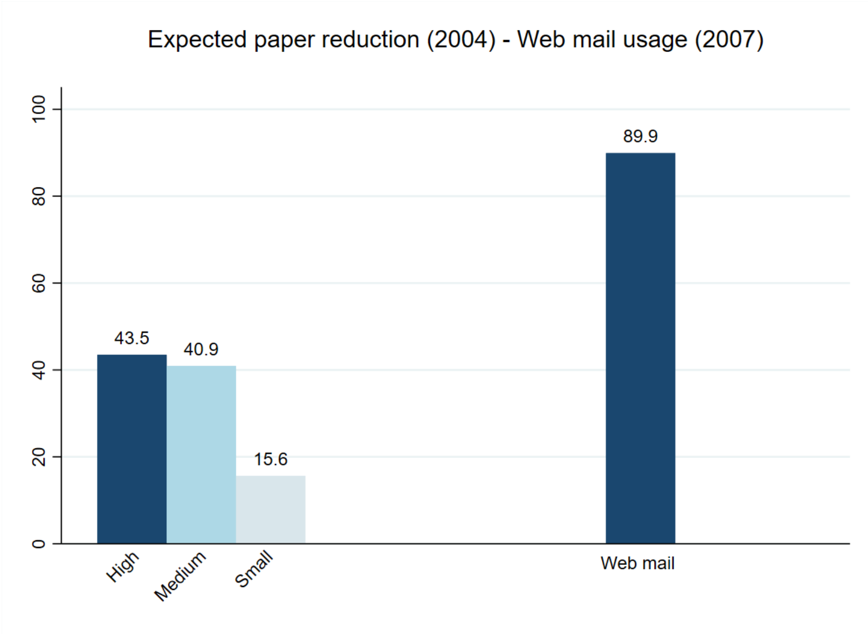
B.4 Additional Figures

Figure B.5: Usage of web technologies



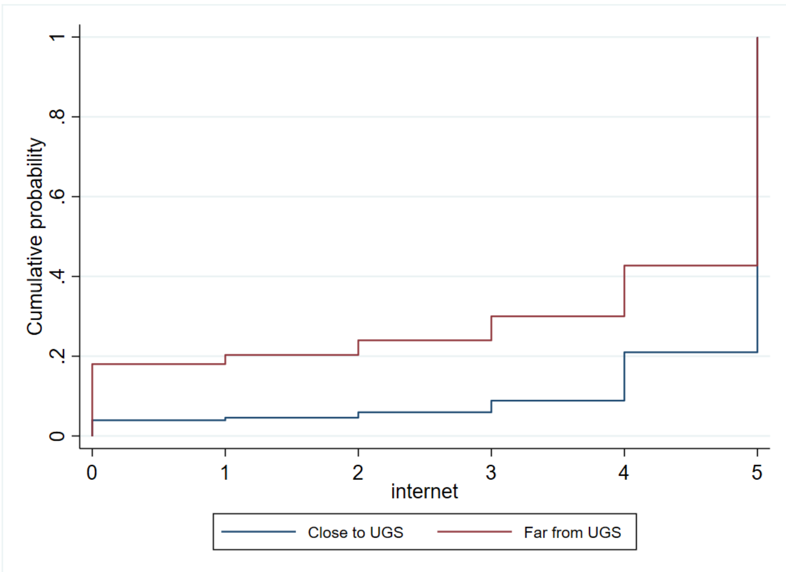
Notes: This figure reports the % growth of the use of web technologies, during the period 2001-2007, for different functions within Italian banks. The source is the Economic Analysis, 2002-2008, from the Italian Banking Association (ABI). Notice that for *Branch operativity*, we do not have the % growth but simply the % usage of web technology in 2007.

Figure B.6: Web e-mails



Notes: This figure reports expectations about paper reduction, in 2004, within bank branches. Much of the banks included in the survey expected a high or medium reduction of paper. At the same time and for the same unit of analysis, the figure reports the % usage of web e-mails in 2007. The source is the Economic Analysis, 2002-2008, from the Italian Banking Association (ABI).

Figure B.7: Instrument Monotonicity Test



Notes: Angrist and Imbens (1995) instrument's monotonicity test. The assumption of monotonicity of the LATE is not verifiable, but has testable implications on the CDFs of internet for municipalities close or far from the UGS. That is, they should never cross. Here, we plot the CDFs of internet for banks close to the UGS (blue solid line) and far from the UGS (red solid line). In order to separate the two groups, we use a dummy variable below/above the median of UGS distance, that proxy well for our continuous instrument. Values of the CDFs refer to the post broadband period. Since the two CDFs never cross, the instrument passes the test.

C. Appendix to Living on the Edge: the Salience of Property Taxes in the UK Housing Market

C.1 Variable Definitions

Variable Name	Description
Price	Transaction price for the property as recorded by HM Land Registry
Council Tax	Amount of council tax payable per year
Band	Council tax band. One of: A, B, C, D, E, F, G, H
Year	Calendar year of the transaction
Month	Calendar month of the transaction
Size	Total floor area measured in squared meters
No. Rooms	Number of habitable rooms in the property as defined in the EPC
Property Type	One of: detached, semi-detached or terraced house and flat
Newly-built	Equals 1 if the property is newly-built
Leasehold	Equals 1 if the property is under a leasehold agreement
Energy Cost	Sum of the annual heating, hot water and lighting costs for the property
	One of very low, low, medium, high and very high expenditures
	Baseline = very low
CO_2 Emissions	CO_2 emissions in tonnes/year
	One of very low, low, medium, high and very high
	Baseline = very low
No. Lighting Outlets	Number of fixed lighting outlets in the property, standardised

Continued on next page

Continued from previous page

Energy Rating	A-G energy rating fixed effects with A being the most efficient
Glazed Type	Indicates the type of glazing
	Various categories of single, double or triple glazing according to
	the British Fenestration Rating Council or manufacturer declaration
No. Storeys > 3	Equals 1 if the building has more than 3 storeys
Glazed Area	Estimate of total glazed area of the property
	One of: Normal, Less than Normal, More than Normal
	Baseline = Normal
Fireplaces	Equals 1 if the property has open fireplaces
No. Extensions	Number of extensions added to the property
	One of: 0, 1, 2, 3, 4
Floor Height	Average storey height in metres
	One of: less than 2.3, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3 or more
Built in	Age band when the building was constructed
	One of: before 1949, 1950-1982, 1983-2002, after 2003
Grid ID	An indicator for the grid square in which the property is located
Pair ID	An indicator for the pair of matched properties

C.2 Tables

Table C.1: Evidence of Selection

	(1)	(2)	(3)	(4)
Council Tax	-231.2***	-263.3***	-228.7***	-229.2***
	(71.8)	(86.4)	(78.0)	(78.3)
Size		2,233.7***	2,271.7***	2,270.8***
		(724.4)	(731.2)	(730.9)
Newly-built				14,054.3**
				(5,619.8)
Leasehold				-8,681.7
				(10,801.3)
Fixed-effects				
Band × Year	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
No. Rooms	No	Yes	Yes	Yes
Property Type	No	No	Yes	Yes
Obs.	889,925	889,925	889,925	889,925
R^2	0.530	0.573	0.578	0.578
Within R^2	0.022	0.064	0.058	0.058

The table shows the estimates of a simple regression of house prices on council tax amounts, namely: $p_{ibdt} = \beta\tau_{bdt} + \delta_{bt} + \zeta' x_{ibdt} + \varepsilon_{ibdt}$ where p_{ibdt} is the price of house i in band b, Borough d at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{bt} are band-year fixed effects; and x_{ibdt} are controls. All columns include band-year and month fixed effects. All other variables are defined in Section C.1. Standard errors double-clustered at the Borough and year level are reported in parentheses.***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.2: Evidence of Selection - Additional Controls

	(1)	(2)	(3)	(4)
Council Tax	-255.9***	-225.0**	-259.6***	-220.1**
	(85.6)	(80.0)	(88.2)	(78.2)
Size	2,747.2***	2,266.9***	2,534.4***	2,310.4***
	(911.3)	(734.5)	(780.3)	(784.4)

Continued on next page

Table C.2 – *Continued from previous page*

	(1)	(2)	(3)	(4)
Energy Cost Low	-26,896.1*			-15,049.6**
	(13,515.0)			(7,107.5)
Energy Cost Medium	-47,312.9**			-24,385.3**
	(22,380.7)			(11,482.6)
Energy Cost High	-69,359.2**			-32,869.7**
	(30,818.3)			(15,291.8)
Energy Cost Very High	-94,269.8*			-39,075.4
	(45,563.6)			(22,987.4)
CO_2 Emisions Low	-17,677.3**			-14,199.3***
	(7,006.9)			(4,857.3)
CO_2 Emissions Medium	-26,558.8**			-23,257.6**
	(11,971.2)			(8,475.0)
CO_2 Emissions High	-36,052.5*			-31,559.2**
	(18,323.2)			(12,521.6)
CO_2 Emissions Very High	-32,523.2			-26,343.9
	(28,385.1)			(17,461.3)
No. Lighting Outlets	20,870.4***			19,659.8***
	(5,833.9)			(5,317.1)
No. Storeys > 3		-3,140.4		632.9
		(5,841.7)		(6,385.2)
Glazed Area Less than Normal		6,923.6		851.3
		(11,930.2)		(10,981.8)
Glazed Area More than Normal		16,669.1***		13,729.2***
		(3,337.1)		(3,490.2)

Continued on next page

Table C.2 – *Continued from previous page*

	(1)	(2)	(3)	(4)
Fireplaces		42,454.0***		33,624.0***
		(9,985.6)		(9,114.6)
Newly-built			23,567.3***	29,368.6***
			(5,295.0)	(4,958.9)
Leasehold			24,601.4*	-13,104.3
			(13,060.7)	(11,681.6)
Built in 1950-1982			-43,868.7***	-29,435.6***
			(8,169.9)	(5,870.5)
Built in 1983-2002			-22,756.2**	-30,012.4***
			(9,533.2)	(8,919.4)
Built after 2003			-21,575.1	-31,925.1**
			(13,706.9)	(15,196.7)
Fixed-effects				
Band × Year	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
Energy Rating	Yes	No	No	Yes
Glazed Type	Yes	No	No	Yes
No. Rooms	No	Yes	No	Yes
Property Type	No	Yes	No	Yes
No. Extensions	No	Yes	No	Yes
Floor Height	No	Yes	No	Yes
Obs.	889,925	889,925	889,925	889,925
R^2	0.566	0.580	0.564	0.583

Continued on next page

Table C.2 – *Continued from previous page*

	(1)	(2)	(3)	(4)
Within R^2	0.095	0.059	0.092	0.063

The table shows the estimates of a simple regression of house prices on council tax amounts, namely: $p_{ibdt} = \beta\tau_{bdt} + \delta_{bt} + \zeta' x_{ibdt} + \varepsilon_{ibdt}$ where p_{ibdt} is the price of house i in band b, Borough d at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{bt} are band-year fixed effects; and x_{ibdt} are controls. All columns include band-year and month fixed effects and control for the property size. All other variables are defined in Section C.1. Standard errors double-clustered at the Borough and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.3: Evidence of Selection - Median Price per Borough, Band, Year

	(1)	(2)	(3)	(4)
Council Tax	-183.6***	-334.0***	-324.3***	-325.1***
	(56.4)	(84.7)	(83.5)	(83.1)
Fixed-effects				
Band × Year	Yes	Yes	Yes	Yes
First-stage controls				
Month	Yes	Yes	Yes	Yes
Size	No	Yes	Yes	Yes
No. Rooms	No	Yes	Yes	Yes
Property Type	No	No	Yes	Yes
Newly-built	No	No	No	Yes
Leasehold	No	No	No	Yes
Obs.	5,014	5,014	5,014	5,014
R^2	0.804	0.501	0.503	0.500
Within R^2	0.055	0.122	0.117	0.118

The table shows the estimates of the following regression: $\varepsilon^{med}_{bdt} = \beta\tau_{bdt} + \delta_{bt} + \eta_{bdt}$, where ε^{med}_{bdt} is the median residual price of all houses in band b, Borough d at time t obtained from a hedonic regression of prices on house characteristics; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; and δ_{bt} are band-year fixed effects. The explanatory variables used to computed the hedonic residuals are reported in the panel First-stage controls. All variables are defined in Section C.1. Standard errors double-clustered at the Borough and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.4: Evidence of Selection - Average Price per Borough, Band, Year

	(1)	(2)	(3)	(4)
Council Tax	-195.6***	-368.4***	-358.6***	-358.9***
	(64.9)	(93.9)	(92.8)	(92.5)
Fixed-effects				
Band × Year	Yes	Yes	Yes	Yes
First-stage controls				
Month	Yes	Yes	Yes	Yes
Size	No	Yes	Yes	Yes
No. Rooms	No	Yes	Yes	Yes
Property Type	No	No	Yes	Yes
Newly-built	No	No	No	Yes
Leasehold	No	No	No	Yes
Obs.	5,014	5,014	5,014	5,014
R^2	0.797	0.512	0.513	0.511
Within R^2	0.053	0.123	0.118	0.118

The table shows the estimates of the following regression: $\bar{\varepsilon}_{bdt} = \beta\tau_{bdt} + \delta_{bt} + \eta_{bdt}$, where $\bar{\varepsilon}_{bdt}$ is the average residual price of all houses in band b, Borough d at time t obtained from a hedonic regression of prices on house characteristics; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; and δ_{bt} are band-year fixed effects. The explanatory variables used to computed the hedonic residuals are reported in the panel First-stage controls. All variables are defined in Section C.1. Standard errors double-clustered at the Borough and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.5: Grid Regressions

	(1)	(2)	(3)	(4)
Council Tax	50.3	12.6	13.4	14.3
	(50.9)	(48.0)	(45.3)	(44.7)
Size		4,626.9***	4,547.6***	4,537.0***
		(1,380.6)	(1,368.4)	(1,366.9)
Newly-built				33,398.5***
				(9,937.9)
Leasehold				-75,924.3**
				(27,874.0)
Band $\times$ Grid ID $\times$ Year	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
No. Rooms	No	Yes	Yes	Yes
Property Type	No	No	Yes	Yes
Obs.	71,734	71,734	71,734	71,734
R^2	0.696	0.771	0.773	0.773
Within R^2	0.000	0.103	0.010	0.101

The table shows the estimates of a regression of house prices on council tax amounts, namely: $p_{ibdgt} = \beta\tau_{bdt} + \delta_{bgt} + \zeta' x_{ibdgt} + \varepsilon_{ibdgt}$, where p_{ibdgt} is the price of house i, in band b, Borough d, grid square g at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{bgt} are band-grid ID-year fixed effects; and x_{ibdgt} are controls. All columns include band-grid ID-year and month fixed effects. The squares are constructed from a 50 $\times$ 50 grid of London. All other variables are defined in Section C.1. Standard errors double-clustered at the grid-ID and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.6: Grid Regressions - Additional Controls

	(1)	(2)	(3)	(4)
Council Tax	7.98	15.0	9.19	17.5
	(42.3)	(45.4)	(40.1)	(43.4)
Size	5,855.9***	4,522.7***	5,318.7***	4,787.6***
	(1,585.1)	(1,366.3)	(1,353.6)	(1,548.8)
Energy Cost Low	-66,317.8**			-35,546.7**
	(23,418.5)			(14,809.1)
Energy Cost Medium	-108,665.0**			-57,508.8**

Continued on next page

Table C.6 – *Continued from previous page*

	(1)	(2)	(3)	(4)
	(38,438.8)			(23,978.0)
Energy Cost High	-147,580.7**			-77,036.8**
	(52,657.2)			(33,410.5)
Energy Cost Very High	-195,178.8**			-106,670.2*
	(79,418.3)			(52,374.7)
CO_2 Emissions Low	-32,054.7***			-28,497.1***
	(10,727.2)			(9,893.5)
CO_2 Emissions Medium	-48,961.2**			-47,738.6***
	(17,139.3)			(16,123.8)
CO_2 Emissions High	-75,329.8**			-71,914.7***
	(27,138.1)			(24,295.9)
CO_2 Emissions Very High	-69,844.7			-66,123.0*
	(41,209.7)			(33,075.1)
No. Lighting Outlets	21,965.9**			19,370.8**
	(8,176.9)			(7,921.1)
No. Storeys > 3		-20,775.9***		-22,481.8***
		(6,704.8)		(7,545.2)
Glazed Area Less than Normal		-25,624.5		-18,393.5
		(18,311.0)		(17,084.0)
Glazed Area More than Normal		13,298.3		12,980.2
		(8,438.6)		(8,365.9)
Fireplaces		34,202.3***		32,533.3***
		(6,533.8)		(6,832.4)
Newly-built			23,232.4**	23,142.1*

Continued on next page

Table C.6 – *Continued from previous page*

	(1)	(2)	(3)	(4)
			(9,272.9)	(11,929.6)
Leasehold			-64,925.4***	-81,662.8***
			(17,738.1)	(28,274.4)
Built in 1950-1982			-33,647.4***	-31,823.5***
			(8,513.9)	(9,833.6)
Built in 1983-2002			43,030.8**	5,862.8
			(19,392.1)	(10,375.6)
Built after 2003			37,169.2**	-1,989.2
			(16,898.5)	(15,825.0)
Fixed-effects				
Band × Grid ID × Year	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
Energy Rating	Yes	No	No	Yes
Glazed Type	Yes	No	No	Yes
No. Rooms	No	Yes	No	Yes
Property Type	No	Yes	No	Yes
No. Extensions	No	Yes	No	Yes
Floor Height	No	Yes	No	Yes
Obs.	71,734	71,734	71,734	71,734
R^2	0.762	0.774	0.759	0.777
Within R^2	0.216	0.010	0.209	0.110

Continued on next page

Table C.6 – *Continued from previous page*

	(1)	(2)	(3)	(4)

The table shows the estimates of a regression of house prices on council tax amounts, namely: $p_{ibdgt} = \beta\tau_{bdt} + \delta_{bgt} + \zeta' x_{ibdgt} + \varepsilon_{ibdgt}$, where p_{ibdgt} is the price of house i, in band b, Borough d, grid square g at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{bgt} are band-grid ID-year fixed effects; and x_{ibdgt} are controls. All columns include band-grid ID-year and month fixed effects and control for the property size. The squares are constructed from a 50×50 grid of London. All other variables are defined in Section C.1. Standard errors double-clustered at the grid-ID and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.7: Grid Regressions for Different Grids

	(1)	(2)	(3)
Council Tax	14.3	-16.2	28.2
	(44.7)	(58.4)	(32.4)
Size	4,537.0***	6,988.4***	7,737.4***
	(1,366.9)	(1,794.8)	(2,319.6)
Newly-built	33,398.5***	22,929.9	-28,536.5
	(9,937.9)	(20,993.7)	(24,742.0)
Leasehold	-75,924.3**	-82,738.9*	-151,551.3**
	(27,874.0)	(44,068.4)	(69,763.6)
Fixed-effects			
Band × Grid ID × Year	Yes	Yes	Yes
Month	Yes	Yes	Yes
No. Rooms	Yes	Yes	Yes
Property Type	Yes	Yes	Yes
Obs.	71,734	21,446	6,954
R^2	0.773	0.792	0.827
Within R^2	0.101	0.139	0.154
Grid	50×50	100×100	150×150

The table shows the estimates of a regression of house prices on council tax amounts, namely: $p_{ibdgt} = \beta\tau_{bdt} + \delta_{bgt} + \zeta' x_{ibdgt} + \varepsilon_{ibdgt}$, where p_{ibdgt} is the price of house i, in band b, Borough d, grid square g at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{bgt} are band-grid ID-year fixed effects; and x_{ibdgt} are controls. The grids divide London into 50×50, 100×100 and 150×150 squares in columns (1), (2) and (3), respectively. All columns include band-grid ID-year, month, number of rooms, property type, newly-built and leasehold fixed effects, as well as a control for the property size. All variables are defined in Section C.1. Standard errors double-clustered at the grid-ID and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.8: Grid Regressions for Different Grids - Additional Controls

	(1)	(2)	(3)
Council Tax	17.5	-11.8	75.4**
	(43.4)	(62.0)	(33.1)
Size	4,787.6***	7,579.1***	7,516.7***
	(1,548.8)	(1,968.4)	(2,014.1)
Newly-built	23,142.1*	23,012.5	-17,873.0
	(11,929.6)	(23,337.2)	(11,886.9)
Leasehold	-81,662.8***	-100,958.8*	-180,856.2*
	(28,274.4)	(48,348.3)	(92,255.8)
Built in 1950-1982	-31,823.5***	-36,770.1**	-46,677.3*
	(9,833.6)	(13,202.8)	(22,562.5)
Built in 1983-2002	5,862.8	26,842.2	-20,068.4
	(10,375.6)	(22,698.3)	(26,632.8)
Built after 2003	-1,989.2	-30,612.1	-68,073.3
	(15,825.0)	(29,315.1)	(66,304.7)
No. Storeys > 3	-22,481.8***	-19,920.1**	-10,496.4
	(7,545.2)	(9,293.2)	(10,564.8)
Glazed Area Less than Normal	-18,393.5	-37,901.5	41,021.3
	(17,084.0)	(25,339.4)	(69,940.7)
Glazed Area More than Normal	12,980.2	4,587.5	-102,420.2
	(8,365.9)	(19,443.3)	(63,902.4)
Fireplaces	32,533.3***	41,107.7***	49,004.1***
	(6,832.4)	(12,251.9)	(16,591.1)

Continued on next page

Table C.8 – *Continued from previous page*

	(1)	(2)	(3)
Energy Cost Low	-35,546.7**	-55,601.8**	-62,161.9***
	(14,809.1)	(20,340.8)	(16,726.8)
Energy Cost Medium	-57,508.8**	-93,685.8***	-85,741.3***
	(23,978.0)	(31,367.3)	(23,759.2)
Energy Cost High	-77,036.8**	-141,100.4***	-161,362.9***
	(33,410.5)	(46,540.7)	(41,048.0)
Energy Cost Very High	-106,670.2*	-170,909.0**	-189,343.0**
	(52,374.7)	(66,032.5)	(72,613.7)
CO_2 Emissions Low	-28,497.1***	-46,607.7***	-43,592.6***
	(9,893.5)	(14,538.3)	(13,312.9)
CO_2 Emissions Medium	-47,738.6***	-73,467.2***	-96,311.1**
	(16,123.8)	(23,582.7)	(43,159.3)
CO_2 Emissions High	-71,914.7***	-104,329.4***	-141,727.0**
	(24,295.9)	(32,544.5)	(58,712.5)
CO_2 Emissions Very High	-66,123.0*	-133,060.2***	-150,608.5*
	(33,075.1)	(44,218.2)	(77,511.2)
No. Lighting Outlets	19,370.8**	22,306.9	53,060.6*
	(7,921.1)	(16,072.2)	(30,324.9)
Fixed-effects			
Band × Grid ID × Year	Yes	Yes	Yes
Month	Yes	Yes	Yes
No. Rooms	Yes	Yes	Yes

Continued on next page

Table C.8 – *Continued from previous page*

	(1)	(2)	(3)
Property Type	Yes	Yes	Yes
No. Extensions	Yes	Yes	Yes
Floor Height	Yes	Yes	Yes
Energy Rating	Yes	Yes	Yes
Glazed Type	Yes	Yes	Yes
Obs.	71,734	21,446	6,954
R^2	0.777	0.798	0.846
Within R^2	0.110	0.150	0.165
Grid	50×50	100×100	150×150

The table shows the estimates of a regression of house prices on council tax amounts, namely: $p_{ibdgt} = \beta\tau_{bdt} + \delta_{bgt} + \zeta' x_{ibdgt} + \varepsilon_{ibdgt}$, where p_{ibdgt} is the price of house i, in band b, Borough d, grid square g at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{bgt} are band-grid ID-year fixed effects; and x_{ibdgt} are controls. The grids divide London into 50×50, 100×100 and 150×150 squares in columns (1), (2) and (3), respectively. All columns include band-grid ID-year fixed effects. All control variables are identical across columns and are as defined in Section C.1. Standard errors double-clustered at the grid-ID and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.9: Grid Regressions - Without Stamp Duty Notches

	(1)	(2)	(3)
Council Tax	16.1	16.7	18.8
	(46.9)	(45.3)	(47.7)
Size	4,715.8***	4,586.7***	4,765.9***
	(1,446.6)	(1,403.6)	(1,487.4)
Newly-built	37,062.1***	30,619.3***	33,964.1***
	(9,998.4)	(9,694.0)	(9,549.4)
Leasehold	-80,083.0**	-75,897.2**	-80,141.1**
	(30,142.0)	(28,682.7)	(31,002.5)
Fixed-effects			
Band $\times$ Grid ID $\times$ Year	Yes	Yes	Yes
Month	Yes	Yes	Yes
No. Rooms	Yes	Yes	Yes
Property Type	Yes	Yes	Yes
Obs.	65,328	70,012	63,606
R^2	0.775	0.776	0.779
Within R^2	0.105	0.102	0.106
p $\notin$	[240k-270k]	[490k-520k]	[240k-270k] & [490k-520k]

The table shows the estimates of a regression of house prices on council tax amounts, namely: $p_{ibdgt} = \beta\tau_{bdt} + \delta_{bgt} + \zeta' x_{ibdgt} + \varepsilon_{ibdgt}$, where p_{ibdgt} is the price of house i, in band b, Borough d, grid square g at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{bgt} are band-grid ID-year fixed effects; and x_{ibdgt} are controls. All columns include band-grid ID-year, month, number of rooms, property type, newly-built and leasehold fixed effects, as well as a control for property size. The squares are constructed from a 50 $\times$ 50 grid of London. Column (1) excludes properties sold at a price between £240,000 and £270,000; column (2) properties sold for between £490,000 and £520,000; and column (3) excludes both properties sold in the £240,000 - £270,000 and £490,000 - £520,000 price range. All variables are defined in Section C.1. Standard errors double-clustered at the grid-ID and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.10: Grid Regressions - Median Price per Borough, Band, Grid, Year

	(1)	(2)	(3)	(4)
Council Tax	92.1*	15.4	19.9	19.1
	(50.8)	(35.4)	(36.3)	(36.6)
Fixed-effects				
Band × Grid ID × Year	Yes	Yes	Yes	Yes
Fit statistics				
Obs.	19,377	19,377	19,377	19,377
R^2	0.866	0.833	0.825	0.823
Within R^2	0.006	0.000	0.000	0.000
First-stage controls				
Month	Yes	Yes	Yes	Yes
Size	No	Yes	Yes	Yes
No. Rooms	No	Yes	Yes	Yes
Property Type	No	No	Yes	Yes
Newly-built	No	No	No	Yes
Leasehold	No	No	No	Yes

The table shows the estimates of the following regression: $\varepsilon^{med}_{bdgt} = \beta\tau_{bdt} + \delta_{bgt} + \eta_{bdgt}$, where ε^{med}_{bdgt} is the median residual price of all houses in band b, Borough d, grid square g at time t obtained from a hedonic regression of prices on house characteristics; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; and δ_{bgt} are band-grid ID-year fixed effects. The squares are constructed from a 50×50 grid of London. The explanatory variables used to computed the hedonic residuals are reported in the panel First-stage controls. All variables are defined in Section C.1. Standard errors double-clustered at the grid ID and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.11: Grid Regressions - Average Price per Borough, Band, Grid, Year

	(1)	(2)	(3)	(4)
Council Tax	104.5**	23.6	28.2	26.6
	(48.0)	(32.7)	(33.5)	(33.9)
Fixed-effects				
Band × Grid ID × Year	Yes	Yes	Yes	Yes
Obs.	19,377	19,377	19,377	19,377
R^2	0.875	0.835	0.827	0.825
Within R^2	0.007	0.001	0.001	0.001
First-stage controls				
Month	Yes	Yes	Yes	Yes
Size	No	Yes	Yes	Yes
No. Rooms	No	Yes	Yes	Yes
Property Type	No	No	Yes	Yes
Newly-built	No	No	No	Yes
Leasehold	No	No	No	Yes

The table shows the estimates of the following regression: $\bar{\varepsilon}_{bdgt} = \beta\tau_{bdt} + \delta_{bgt} + \eta_{bdgt}$, where $\bar{\varepsilon}_{bdgt}$ is the average residual price of all houses in band b, Borough d, grid square g at time t obtained from a hedonic regression of prices on house characteristics; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; and δ_{bgt} are band-grid ID-year fixed effects. The squares are constructed from a 50 × 50 grid of London. The explanatory variables used to computed the hedonic residuals are reported in the panel First-stage controls. All variables are defined in Section C.1. Standard errors double-clustered at the grid ID and year level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.12: Matching Regressions - Euclidean Distance

	(1)	(2)	(3)	(4)
Council Tax	53.8**	12.9	50.7**	9.00
	(23.4)	(18.3)	(23.8)	(18.8)
Size		3,770.6***		3,750.2***
		(763.8)		(734.2)
Fixed-effects				
Pair ID	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
No. Rooms	No	Yes	No	Yes
Property Type	No	Yes	No	Yes
Obs.	115,224	115,224	114,646	114,646
Unique Transaction IDs	71,578	71,578	71,656	71,656
R^2	0.799	0.836	0.796	0.834
Within R^2	0.001	0.042	0.001	0.042
Distance	Euclidean 1	Euclidean 1	Euclidean 2	Euclidean 2

The table shows the estimates of the following regression: $p_{ibdt} = \beta\tau_{bdt} + \delta_{ij} + \zeta' x_{ibdt} + \varepsilon_{ibdt}$, where p_{ibdt} is the price of house i in band b, Borough d at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{ij} are pair fixed effects; and x_{ibdt} are controls. Housing pairs from opposite sides of a given border are constrained to be no more than 500 metres away, sold in the same year, in the same council tax band and to both be either old or newly-built and freehold or leasehold properties. The closest match for each property is chosen as the one minimising the Euclidean distance $d(i,j) = \sqrt{\sum_{k=1}^{K}(x_{ik} - x_{jk})^2}$. The vectors x_i and x_j in columns (1) and (2) include size and number of rooms, while columns (3) and (4) add the energy cost. All variables are defined in Section C.1. Standard errors clustered at the transaction ID level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.13: Matching Regressions - Linear Distance

	(1)	(2)	(3)	(4)
Council Tax	56.8**	15.3	55.7**	14.6
	(23.4)	(18.1)	(23.7)	(18.7)
Size		3,879.2***		3,809.8***
		(778.8)		(762.1)
Fixed-effects				
Pair ID	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
No. Rooms	No	Yes	No	Yes
Property Type	No	Yes	No	Yes
Obs.	114,904	114,904	113,854	113,854
Unique Transaction IDs	71,588	71,588	71,649	71,649
R^2	0.799	0.837	0.798	0.835
Within R^2	0.001	0.045	0.001	0.043
Distance	Linear 1	Linear 1	Linear 2	Linear 2

The table shows the estimates of the following regression: $p_{ibdt} = \beta\tau_{bdt} + \delta_{ij} + \zeta' x_{ibdt} + \varepsilon_{ibdt}$, where p_{ibdt} is the price of house i in band b, Borough d at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{ij} are pair fixed effects; and x_{ibdt} are controls. Housing pairs from opposite sides of a given border are constrained to be no more than 500 metres away, sold in the same year, in the same council tax band and to both be either old or newly-built and freehold or leasehold properties. The closest match for each property is chosen as the one minimising the following distance: $d(i,j) = |\hat{p}_{it} - \hat{p}_{jt}|$, where $\hat{p}_{it}$ and $\hat{p}_{jt}$ are model-predicted prices for two matched property transactions i and j based on a linear model: $p_{it} = \alpha + \beta' x_{it} + \varepsilon_{it}$. The vectors x_{it} and x_{jt} in columns (1) and (2) include size and number of rooms, while columns (3) and (4) add the energy cost. All variables are defined in Section C.1. Standard errors clustered at the transaction ID level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.14: Matching Regressions - Linear Distance Less than 30% of Predicted Prices

	(1)	(2)	(3)	(4)
Council Tax	-8.19	-5.24	-7.65	-8.14
	(10.1)	(9.68)	(11.0)	(10.3)
Size		3,980.1***		3,982.4***
		(295.8)		(349.3)
Fixed-effects				
Pair ID	Yes	Yes	Yes	Yes
Month	Yes	Yes	Yes	Yes
No. Rooms	No	Yes	No	Yes
Property Type	No	Yes	No	Yes
Obs.	175,639	175,639	167,704	167,704
Unique Transaction IDs	59,722	59,722	58,917	58,917
R^2	0.871	0.875	0.855	0.859
Within R^2	0.000	0.017	0.000	0.018
Distance	Linear 1	Linear 1	Linear 2	Linear 2

The table shows the estimates of the following regression: $p_{ibdt} = \beta\tau_{bdt} + \delta_{ij} + \zeta' x_{ibdt} + \varepsilon_{ibdt}$, where p_{ibdt} is the price of house i in band b, Borough d at time t; τ_{bdt} is the council tax amount for a house in band b, Borough d at time t; δ_{ij} are pair fixed effects; and x_{ibdt} are controls. Housing pairs from opposite sides of a given border are constrained to be no more than 500 metres away, sold in the same year, in the same council tax band and to both be either old or newly-built and freehold or leasehold properties. Each house i is matched to all possible candidates j that satisfy the following constraint: $d(i,j) = |\hat{p}_{it} - \hat{p}_{jt}| < 0.3 \times \max\{\hat{p}_{it}, \hat{p}_{jt}\}$, where $\hat{p}_{it}$ and $\hat{p}_{jt}$ are model-predicted prices for two matched property transactions i and j based on a linear model: $p_{it} = \alpha + \beta' x_{it} + \varepsilon_{it}$. The vectors x_{it} and x_{jt} in columns (1) and (2) include size and number of rooms, while columns (3) and (4) add the energy cost. All variables are defined in Section C.1. Standard errors clustered at the transaction ID level are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table C.15: Model-averaged Posterior Distributions for the Council Tax Incidence

Prior	1%	5%	10%	25%	50%	75%	90%	95%	99%	mode	mean	γ
$\mathcal{N}(-150, 50^2)$	-143.50	-110.66	-93.21	-61.85	-22.87	-1.98	18.67	31.04	51.90	-12.08	-31.85	0.15
$\mathcal{N}(-100, 50^2)$	-116.75	-85.88	-69.23	-39.43	-12.79	7.60	29.51	41.86	62.81	-9.86	-16.81	0.13
$\mathcal{N}(-50, 50^2)$	-90.71	-61.45	-45.54	-20.99	-2.17	20.33	42.09	54.20	75.25	-6.78	-1.76	0.04
$\mathcal{N}(-150, 75^2)$	-126.67	-87.78	-67.60	-32.92	-7.49	16.49	41.31	54.86	78.03	-8.24	-10.46	0.05
$\mathcal{N}(-50, 25^2)$	-82.09	-64.43	-54.54	-36.79	-18.64	-4.40	9.15	17.64	32.93	-13.86	-20.87	0.37

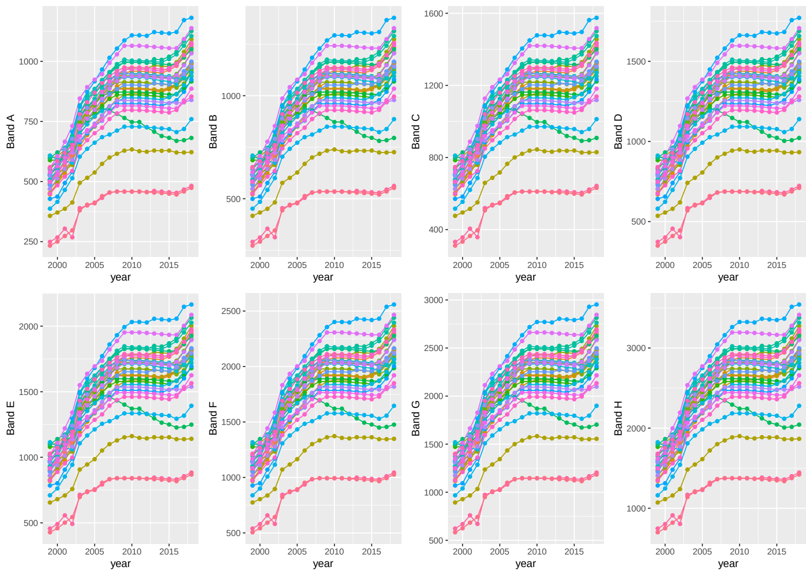
The table displays 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, 99% quantiles, the modal and mean values of the average posterior distribution for the council tax incidence obtained by using the estimates from Tables C.5-C.9 and C.12-C.14. The last column reports the attenuation factor γ computed as the ratio of the posterior and prior median. Each row refers to a different choice of prior.

C.3 Figures

Figure C.1: A Typical Border

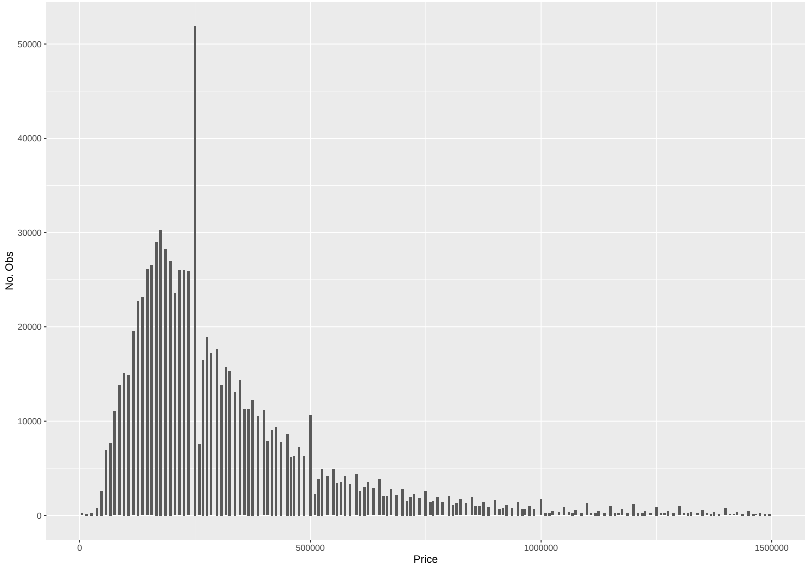
The figure shows an example of a border between two Boroughs in London. Houses on the left side of the West Eaton Place road belong to the Borough of Kensington and Chelsea and have an annual council tax bill of £2, 279, while houses on the right side belong to the Borough of Westminster and have an annual council tax bill of only £1, 421.

Figure C.2: Time Series of Council Taxes



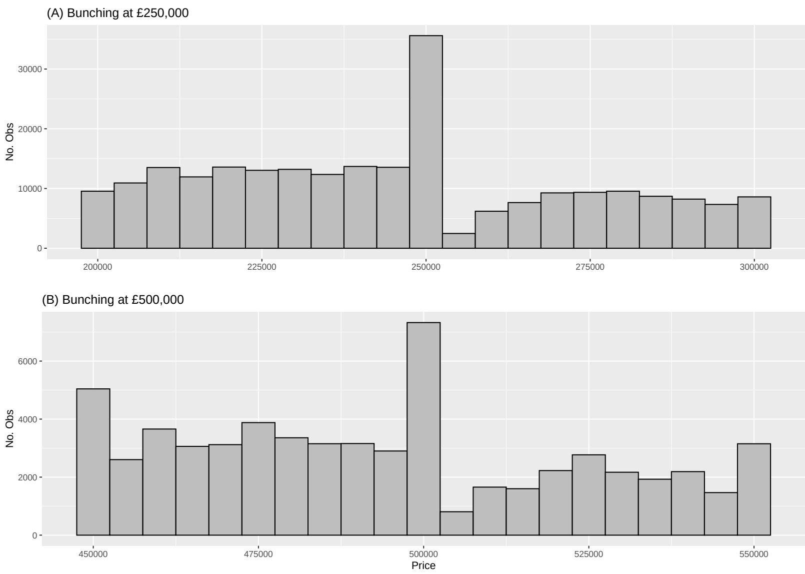
The figure reports the time series of council tax amounts payable across Boroughs. Each panel refers to a different band, while the lines in each panel represent different Boroughs.

Figure C.3: Histogram of Property Prices in London



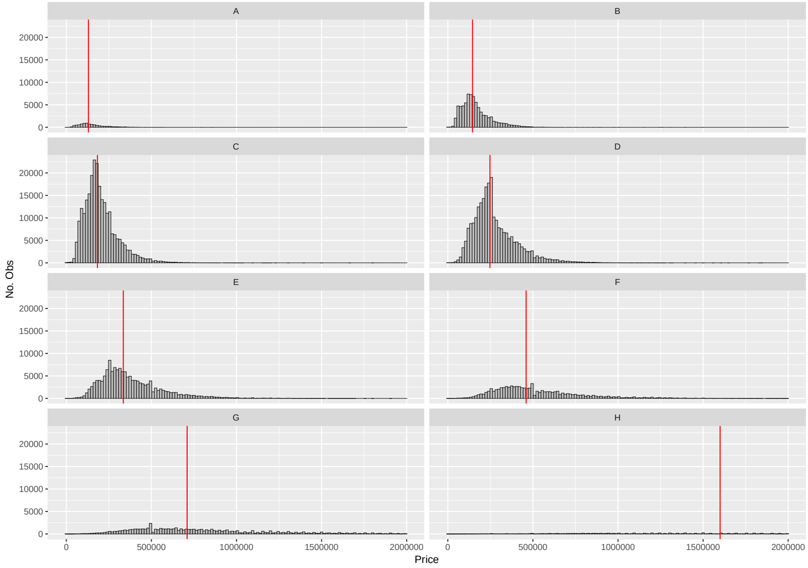
The figure presents a histogram of the distribution of house transaction prices in London. The distribution is truncated at £1,500,000.

Figure C.4: Bunching at Stamp Duty notches



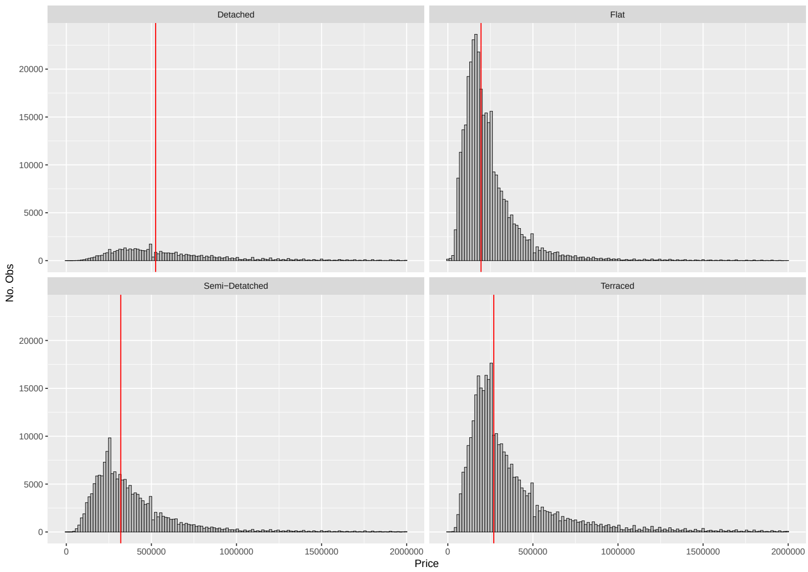
The figure presents a histogram of the distribution of house transaction prices in London around stamp duty notches. Panel (A) refers to the notch at £250,000, while panel (B) at £500,000.

Figure C.5: Histogram of Prices by Band



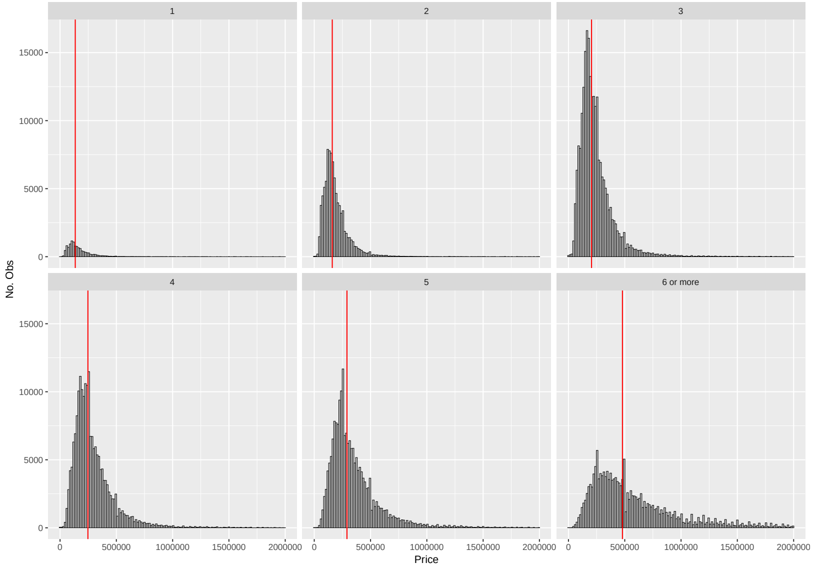
The figure presents a histogram of the distribution of house transaction prices in London per band. Each panel refers to properties belonging to different bands. The distribution is truncated at £2,000,000. The red vertical lines represent the median values computed using the full sample.

Figure C.6: Histogram of Prices by Property Type



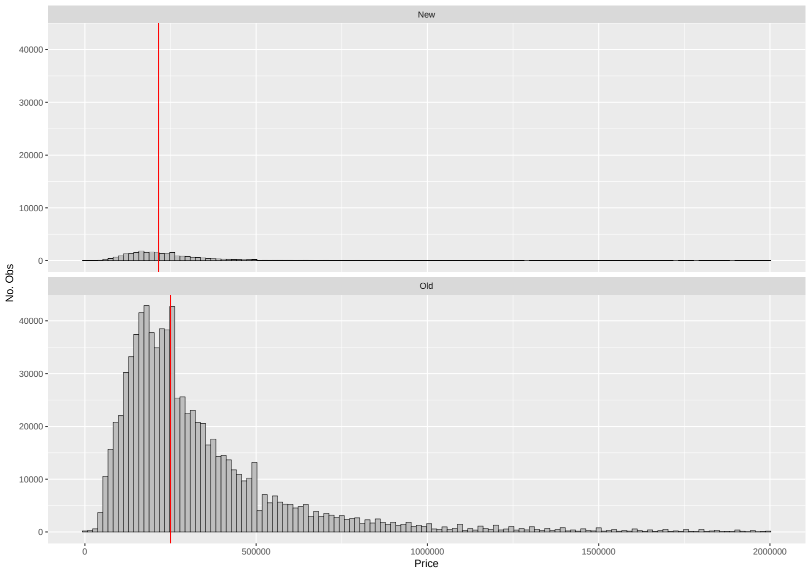
The figure presents a histogram of the distribution of house transaction prices in London by property type. The distribution is truncated at £2,000,000. The red vertical lines represent the median values computed using the full sample.

Figure C.7: Histogram of Prices by Number of Rooms



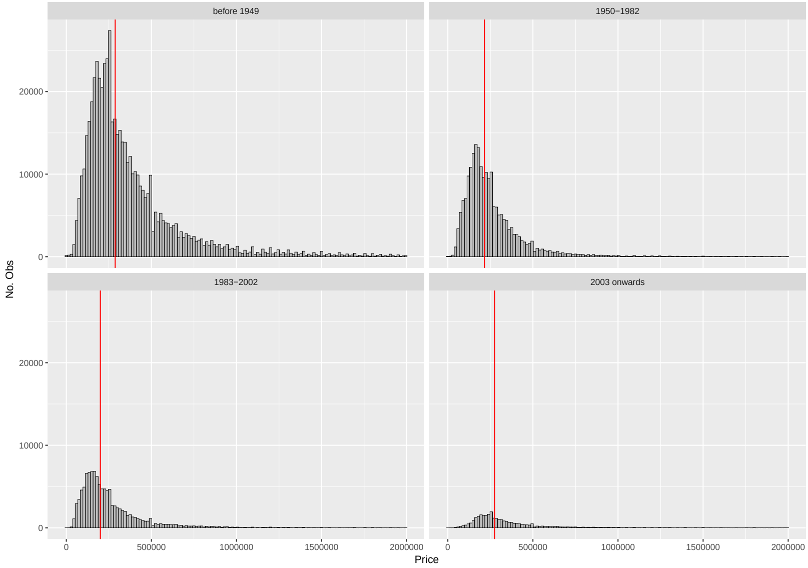
The figure presents a histogram of the distribution of house transaction prices in London by number of rooms. The distribution is truncated at £2, 000, 000. The red vertical lines represent the median values computed using the full sample.

Figure C.8: Histogram of Prices by Age



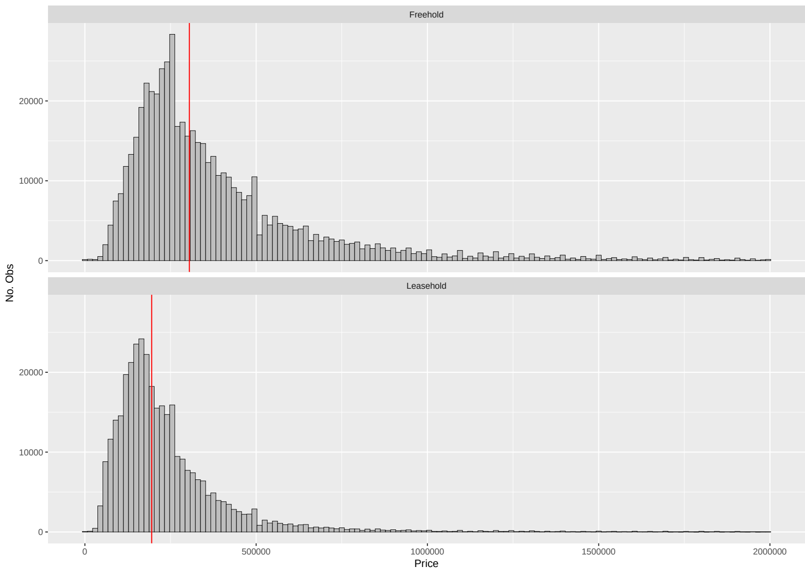
The figure presents a histogram of the distribution of house transaction prices in London by age. The top panel reports the histogram of prices for newly-built properties, while the bottom for established residential buildings. The distribution is truncated at £2, 000, 000. The red vertical lines represent the median values computed using the full sample.

Figure C.9: Histogram of Prices by Year of Construction



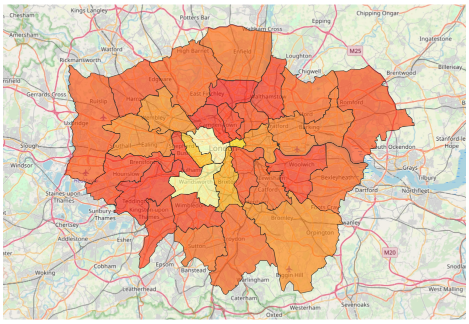
The figure presents a histogram of the distribution of house transaction prices in London by year of construction. The distribution is truncated at £2, 000, 000. The red vertical lines represent the median values computed using the full sample.

Figure C.10: Histogram of Prices by Duration

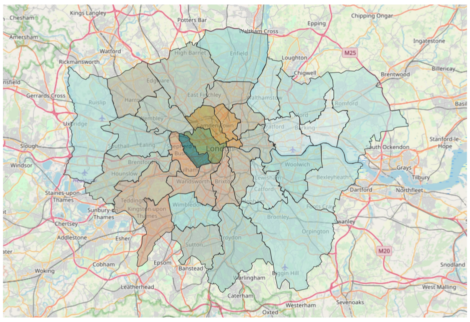


The figure presents a histogram of the distribution of house transaction prices in London by tenure duration. The top panel reports the histogram of prices for freehold properties, while the bottom for leasehold properties. The distribution is truncated at £2, 000, 000. The red vertical lines represent the median values computed using the full sample.

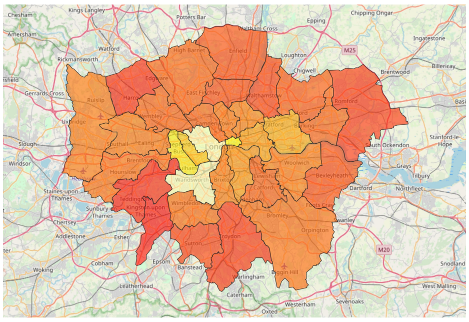
Figure C.11: Council Taxes and House Prices



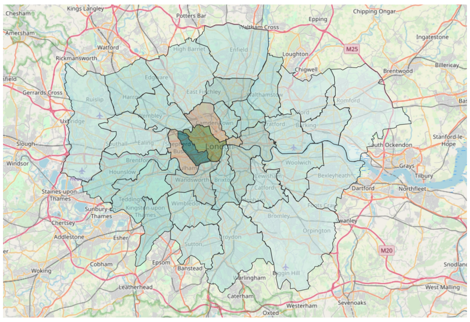
(a) : Council Taxes in 2000



(b) : House Prices in 2000



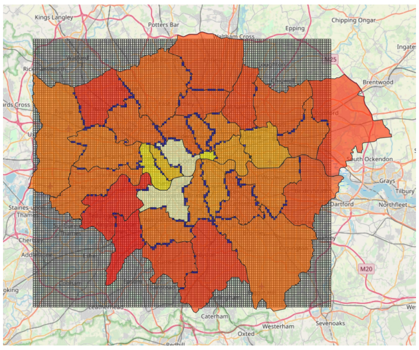
(c) : Council Taxes in 2018



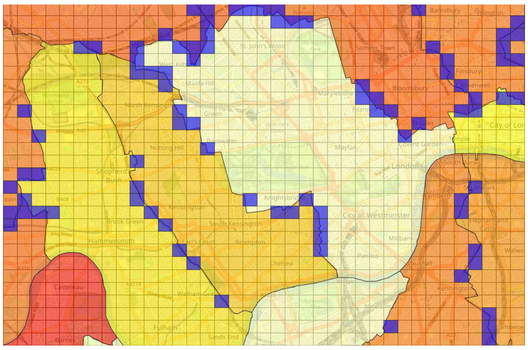
(d) : House Prices in 2018

The maps show the distribution of council tax payable for properties in band D for each London Borough, along with the respective distribution of house prices in 2000 and 2018.

Figure C.12: Grids



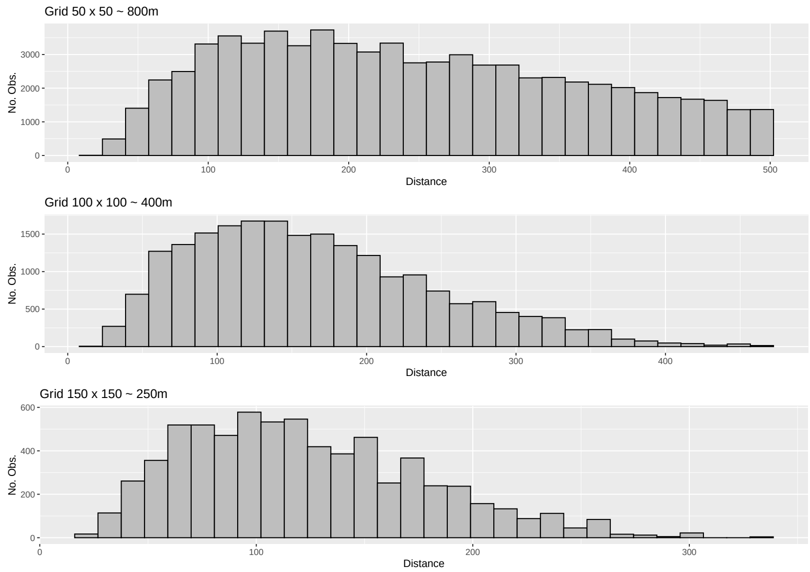
(a) Grid



(b) Enlargement of the Centre

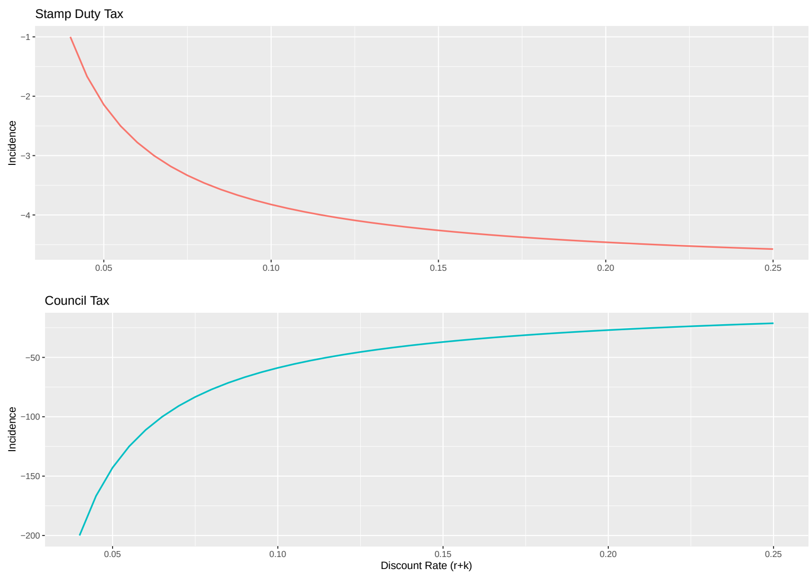
The maps depict our first identification strategy of dividing London in a grid of equally sized squares. Panel C.12a shows a grid of 150×150 squares superposed on the map of the city; Panel C.12b shows an enlargement of the central Boroughs. The blue squares denote areas which contain at least two similar properties located on opposite sides of a border.

Figure C.13: Distribution of Distances for the Grid Regressions



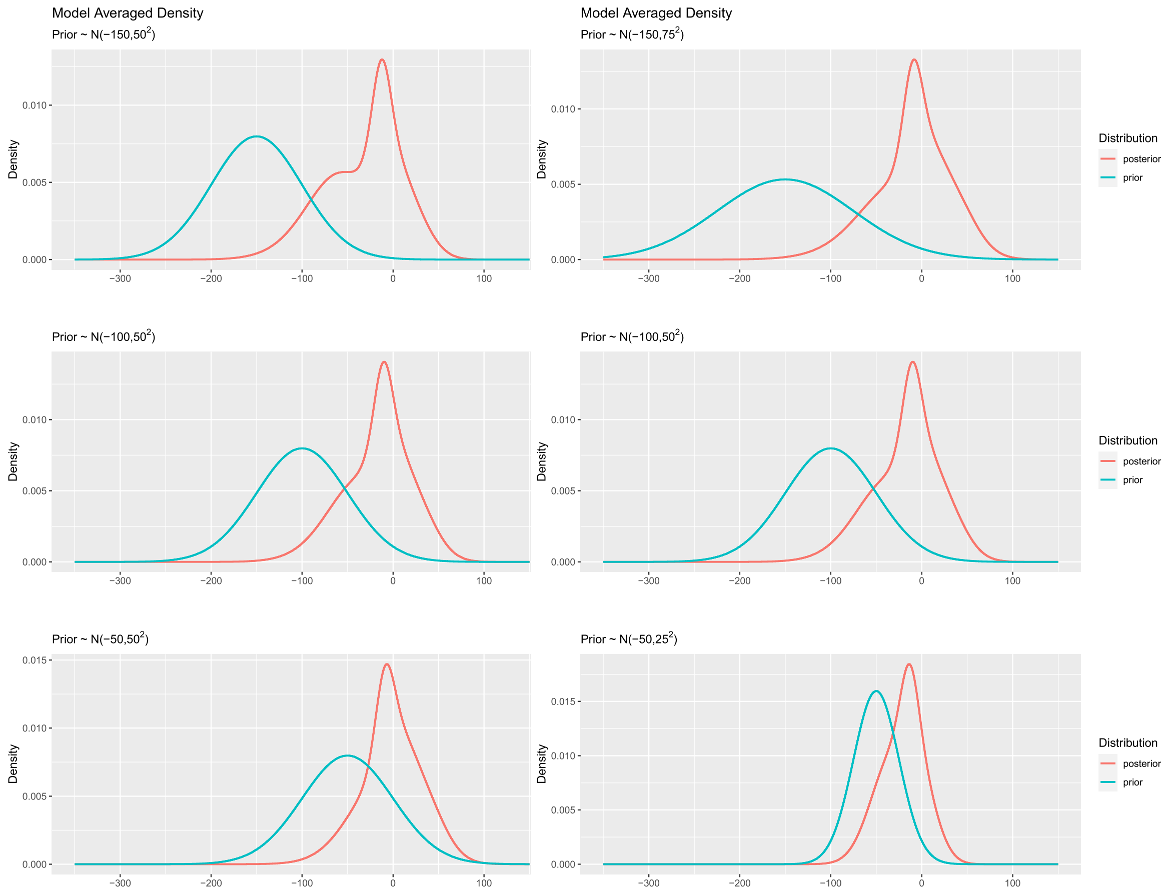
The figure depicts histograms for the distribution of distances between houses on opposite sides of a border that are used in our grid regressions. We report the distributions for three different grids, namely grids where we have divided London in 50 × 50 squares, 100 × 100 and, finally, 150 × 150. For each histogram we report the approximate size of the square sides in meters.

Figure C.14: Model-implied Incidence



The figure plots the relationship between tax incidence on house prices and discount rates, where the discount rate is defined as $r + k$ as in Section 3.5. The upper panel shows the incidence of the stamp duty, while the bottom panel the incidence of the council tax.

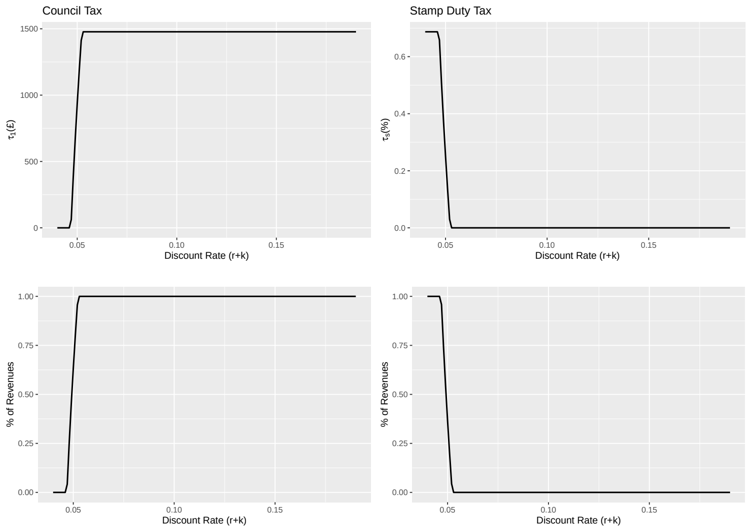
Figure C.15: Model-averaged Estimate of the Posterior Council Tax Incidence



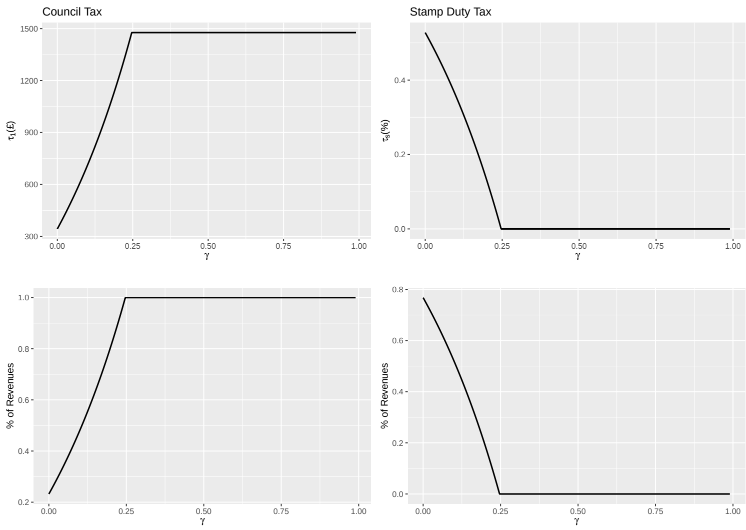
(a) : Constant Variance (b) : Proportional Variance

The figure plots the density of the council tax incidence obtained by taking the model-average of the posteriors as described in Sections 3.5.1 and C.4. The priors are normally distributed $\mathcal{N}(b_0, \sigma_0^2)$ in all figures. In panel (a) the priors have constant standard deviation $\sigma_0 = 50$ and varying means of $b_0 = -150, -100, -50$, respectively. In panel (b) the standard deviation of the priors is proportional to the mean, i.e., $\sigma_0 = |b_0|/2$.

Figure C.16: Optimal Tax Policy



(a) : Optimal Taxes as a Function of $r+k$



(b) : Optimal Taxes as a Function of γ

The figure plots the optimal mix of stamp duty and council tax the Government should choose to maximise the utility of buyers and maintain revenue-neutrality. Panel (a) displays the variables as a function of the discount rate $r+k$, while panel (b) as a function of the attenuation parameter γ. The top plots of each panel show the optimal amount of council tax in £ and stamp duty tax as percentage of house price, respectively. The bottom plots provide the relative percentages of revenue raised through council and stamp duty tax, respectively. In the upper panel we calibrate the parameters as follows: $\alpha = 0.8$, $g = \tilde{g} = 3.5\%$, $\eta_S = 0.5$, $\beta = 0.99$, $\gamma = 0.15$; in the bottom panel: $\alpha = 0.8$, $g = \tilde{g} = 3.5\%$, $\eta_S = 0.5$, $\beta = 0.99$, $r+k = 5\%$.

C.4 Computation of the Model-averaged Posterior Incidence of Council Tax

In Section 3.4 we have estimated models of the type:

$$y = X^m \beta^m + \varepsilon^m \tag{C.1}$$

where $\varepsilon^m | m \sim \mathcal{N}(0, \Omega^m)$, with Ω^m being the population covariance matrix of the errors under model m. We partition the parameters as $\beta^m = (\beta_0, \beta^m_{-0})$, where $\beta^m_{-0} = (\beta^m_1, \beta^m_2, ...)$ and β_0 is the parameter of interest. We then make the (strong) simplifying assumption that Ω^m is known and assume that the prior distribution of the parameters is: $\beta^m | m \sim \mathcal{N}(b^m, \Sigma^m)$. We also assume that the marginal prior distribution of the parameter of interest is common across models, i.e., $p(\beta_0|m) = p(\beta_0) = \mathcal{N}(b_0, \sigma_0^2)$. It follows that the posterior is: $\beta^m | y, m \sim \mathcal{N}(((\Sigma^m)^{-1} + X^{m\prime}(\Omega^m)^{-1}X^m)^{-1}(X^{m\prime}(\Omega^m)^{-1}y + (\Sigma^m)^{-1}b^m), ((\Sigma^m)^{-1} + X^{m\prime}(\Omega^m)^{-1}X^m)^{-1})$. We then proceed by making the following approximations:

$$((\Sigma^m)^{-1} + X^{m\prime}(\Omega^m)^{-1}X^m)^{-1}_{[1,1]} \approx (\sigma_0^{-2} + \widehat{Var}(\hat{\beta}^m)^{-1}_{[1,1]})^{-1} \tag{C.2}$$

$$\begin{aligned} (((\Sigma^m)^{-1} + X^{m\prime}(\Omega^m)^{-1}X^m)^{-1}(X^{m\prime}(\Omega^m)^{-1}y + (\Sigma^m)^{-1}b^m))_{[1]} \approx \\ (\sigma_0^{-2} + \widehat{Var}(\hat{\beta}^m)^{-1}_{[1,1]})^{-1}(\widehat{Var}(\hat{\beta}^m)^{-1}_{[1,1]}\hat{\beta}^m_0 + \sigma_0^{-2}b_0) \end{aligned} \tag{C.3}$$

where $A_{[i,j]}$ and $a_{[i]}$ indicate the i-th, j-th element of matrix A and the i-th element of vector a, respectively. This leads, therefore, to the following approximate posterior distribution for the parameter of interest:

$$\begin{aligned} & p(\beta_0|y, m) = \\ & \mathcal{N}\left((\sigma_0^{-2} + \widehat{Var}(\hat{\beta}^m)^{-1}_{[1,1]})^{-1}(\widehat{Var}(\hat{\beta}^m)^{-1}_{[1,1]}\hat{\beta}^m_0 + \sigma_0^{-2}b_0), (\sigma_0^{-2} + \widehat{Var}(\hat{\beta}^m)^{-1}_{[1,1]})^{-1}\right) \end{aligned} \tag{C.4}$$

After having obtained the posterior distribution for β_0 for each model we average using a flat prior across models to obtain the final density $p(\beta_0|y) = \frac{1}{M}\sum_{m=1}^{M} p(\beta_0|y, m)$.

Returning to the choice of prior distribution for the parameter of interest, we are guided by the model-implied incidence from Section 3.5. We calibrate the following parameters: $g = 0.035$, $\tilde{g} = 0.035$, $r = 0.04$ and $\alpha = 0.8$[1]. Given these values we pick three different means for the prior distribution to match the range of incidence of the stamp duty tax obtained in Best and Kleven (2018), namely, $b_0 = -150, -100, -50$, which roughly correspond to stamp duty incidences of: $\frac{dp}{d\tau_S} = -2, -3, -4$. We choose the standard deviations of the prior to be equal to $\sigma_0 = 50$ or $\sigma_0 = \frac{|b_0|}{2}$ to obtain five prior distributions.

[1]The parameters r and $\tilde{g}$ are consistent with the in-sample average mortgage rate and growth rate of council taxes in the UK, respectively; α is consistent with a downpayment of 20% which is common in the UK. We use a conservative expected growth rate of house prices of 3.5% compared to the in-sample average of 7.3%.