Information and financialization:

Credit markets as a new source of inequality

Torben Iversen

Harvard University

Department of Government

iversen@fas.harvard.edu

Philipp Rehm

Ohio State University

Department of Political Science

Rehm.16@osu.edu

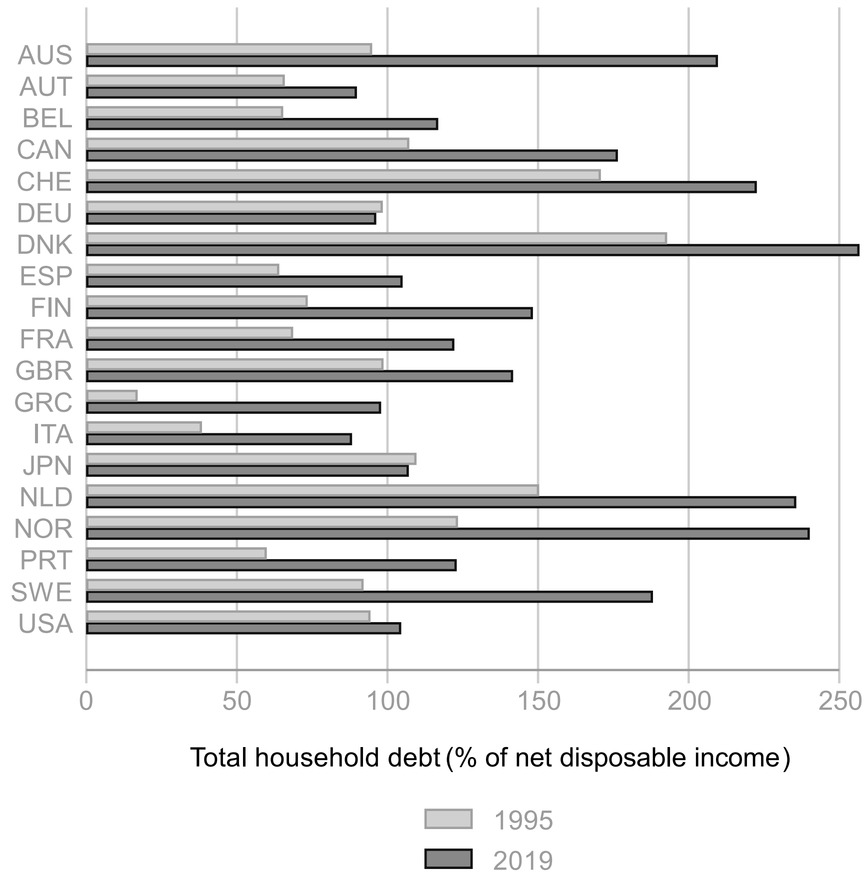
September 2021

1 Introduction

At the turn of the last century banking was personal. Banks made lending decisions based on personal knowledge of borrowers; a fact that made credit often haphazard and not infrequently biased towards friends and family, and against minorities. The small-town banker and horse trader David Harum, the main character in Edward Noyes Westcott's 1898 novel by the same name, described his approach to lending in the 1932 movie adaptation of the novel (played by Will Rogers): "I go a long way on a man's character. And then I go a longer way on his collateral. And if he's got character and collateral both, I let him have about half what he asked for … anybody can get along on half of what they think they can."

The use of information has come a long way since then, but the objective is the same: separate good risks from bad and lend to the former on the best possible terms (for the bank). The massive improvement in data, a large expansion of risk-sharing financial instruments, securitization, and a huge increase in demand have resulted in loans and credit to the household sector expanding exponentially (Figure 1). In less than 25 years, from 1995 to 2019, private debt in advanced democracies increased from an average of 90 percent to about 150 percent of disposable income (with some notable exceptions), extending a trend that started in the 1980s in the US and the UK (the rise in consumer lending mostly occurred earlier in these countries). A growing portion of personal income now goes to servicing debt, and this has a sizable effect on discretionary income. With an average interest rate of 10 percent, it would amount to 15 percent of disposable income, but obviously with huge variation across countries, time, and individuals.

Figure 1: Household debt as a percentage of disposable income



Note: Second data point refers to 2018 in JPN, NOR, USA, and 2020 in CAN.
Source: OECD National Accounts Statistics: National Accounts at a Glance (https://doi.org/10.1787/f03b6469-en, last accessed June 3, 2021 [https://perma.cc/HE5R-NR7X]).

Moreover, access to credit has become an important determinant of individual welfare in a new economy where credit is used to smooth income across increasingly nonlinear life-cycles, with frequent changes in jobs, time off for retraining and additional schooling, and moves back and forth between work and family (Iversen & Soskice, 2019, Chapter 4; Wiedemann, 2021). This also makes the terms of access more important for welfare, and such access has become more unequal with more data.

Owning a home has also become a marker of middle-class success in many countries, and access to mortgage finance is therefore increasingly seen as an important tool for aspirational voters. For all these reasons, access to credit and the cost of such access are emerging as important determinants of individual welfare. This paper explores the consequences of financialization of the household sector for economic inequality.

Specifically, we argue that credit has become a significant—and largely overlooked—driver of inequality. This is because terms of access to credit vary with individual risks of default, which is tied to socio-economic status. Risk assessments in turn depend on individual data on the likelihood of experiencing catastrophic life events—significant loss of income due to unemployment, illness, or involuntary job switches—and ability to financially weather such events. Such data have been greatly facilitated by the information revolution. "Big Data" combine information disclosed by borrowers with a trove of data on residence, demographic indicators, credit history, income, employment history, and so on, which are often kept in central credit registries. As Mikella and Adebayo (2017, p. 151) note, the credit-scoring industry takes an "all data is credit data" approach, "combining conventional credit information with thousands of data points mined from consumers' offline and online activities."

As the data available to lenders improve, they can make more differentiated risk-of-default assessments, which means that interest rates increasingly reflect the underlying risk distribution. As interest payments come out of disposable income, and insofar as disposable income is negatively correlated with default risk, the distribution of discretionary income becomes more unequal. Those deemed too risky will not qualify for large loans (such as home mortgages) in the first place, and while “cheap money” is plentiful, access has been rationed in the wake of the financial crisis.

While nearly all research on inequality focuses on market or disposable income, and increasingly on wealth inequality, what ultimately matters to the welfare of most people is their ability to consume the goods and services that define a middle-class life-style (Fligstein & Goldstein, 2015). Financialization combined with the data revolution have enabled many to do so, but it has also increased discretionary income inequality (even when disposable income inequality is held constant), and those who are excluded from credit markets cannot enjoy the benefits of either income-smoothing or home ownership.[1]

The combined effect of information and financialization, however, is conditioned by national-level financial and social institutions. Income losses are cushioned by the social protection system, and financial regulations can absorb some of the default risk by subsidizing debt repayments or providing lender-of-last-resort guarantees. A notable example of the latter is when governments step in to buy and securitize mortgage debt, thereby absorbing risks that would otherwise fall on lenders. They are in turn able to offer loans to more people, and on more equal

[1] Our argument is focused on the way information interacts with consumer financialization to increase inequality in the terms of credit. But we recognize that financialization has other important distributive consequences that we do not consider here, such as rising rents for “financiers,” winners from a greater emphasis on “shareholder value,” and a deepening indebtedness of the poor (see Froud et al., 2010; and Godechot, 2020 for an overview).

terms. Quasi-public financial institutions such as Fannie Mae and Freddie Mac in the US are major cases in point.

The welfare state also matters. Specifically, when people become unemployed some of their lost income is replaced by unemployment benefits, and the higher the replacement rate the more likely the unemployed are to service their debt. Lenders know this and compete down rates to reflect the lower risk of default. The welfare state, by directly reducing the effect of adverse life events on disposable income thus has a similar effect on discretionary income as state loan guarantees or interest subsidies do by improving the terms of borrowing at the lower part of the distribution.

The underlying political coalitions that sustain these different institutions have distinct historical origins, creating sometimes surprising cross-national patterns that do not conform to standard typologies of welfare states or varieties of capitalism. Institutions and policies in Denmark and the US produce surprisingly similar results, and while the differences in social protection institutions are familiar, it is Denmark and not the US that has the most liberal, market-based mortgage system. These are not outliers. The organization of the mortgage lending market varies a great deal across advanced democracies, and it does not correlate with any widely used typologies of political economies (Blackwell & Kohl, 2018, 2019; Fernandez & Aalbers, 2016; Schwartz & Seabrooke, 2008; Wood, 2019).

Our paper contributes to a burgeoning literature on the politics of financialization. Ahlquist and Ansell (2017) argue that borrowing is used to compensate for high inequality, and that credit has been expanded more in inegalitarian countries as a result. In a similar vein, Wiedemann (2021) argues that increasing access to credit has been used to insure against new life-cycle risks, which in turn undercuts support for the traditional welfare state. Ansell (2014) shows that house ownership can serve as a form of long-term insurance that reduces demand for redistribution. Consistent with this idea, Hariri et al. (2020) find that when people are cut off from credit, short-term liquidity constraints drive up demand for social transfers.

Our paper builds on this work but shifts the focus away from the relationship between credit markets and social insurance towards the direct effects of financialization and information on inequality, and these effects are conditioned by existing social institutions. Specifically, we make four contributions to the existing literature: (i) we develop a model that shows how financialization shapes discretionary income inequality; (ii) we show that this effect is magnified by better lender information about borrowers; (iii) we show how social protection and financial regulation can mitigate some of the inequalizing effects of financialization and data; and (iv) we present (quasi-experimental) evidence for large effects of data, financial regulation, and social protection on both access to credit and the terms of such access.

2 The logic

We base our presentation on a simple formal model that is developed in more detail in Appendix A. Much lending is for purposes of "income smoothing," where current consumption (say, of a car or daycare services) or investment (say, in further education) are enabled by anticipation of higher income in the future (Hall, 1988). This is known as the permanent income hypothesis and presents no problems in terms of repayment of loans. People who stay on their anticipated income path would never default. Unanticipated drops in income because of long-term layoffs or illness, however, can lead to default, and this is what lenders worry about. That is true as well when borrowing to buy a home, which for most people is a long-term investment that greatly enhances their welfare while also potentially generating wealth (Ansell, 2014). As in the case of other loans, discretionary income is reduced by the interest on the mortgage. We consider a world in which people's welfare is enhanced by access to credit, but where the terms of credit vary because lenders assess a premium on higher-risk individuals. Our aim is to understand the distributive effects of how risk is assessed.

2.1 Discretionary income and welfare

Utility to the individual is equal to discretionary income (D_i), which is disposable income minus spending on necessities and interest payments, plus the utility of the consumption that borrowing in credit markets $u(L_i)$ enables:

(1) $$U_i = D_i + u(L_i).$$

We focus on discretionary income, but since access to loans is determined by the same factors that shape the terms of borrowing, we generalize the logic to access.[2] We assume that spending on necessities is constant in order to identify the effect of borrowing.[3] Discretionary income for individual *i*, D_i, over the term of a loan is then equal to:

[2] Since access to borrowing is an important source of wealth accumulation—especially real estate—access is related to wealth inequality, but we do not attempt to model the complex relationship between credit and wealth.

[3] When we later consider borrowing to buy private housing (mortgages), the constant spending on necessities assumption requires everyone to be an owner. In the first of two empirical applications we restrict the analysis to owners accordingly. In the second application, we assume that restrictions on access to ownership reduces welfare

(2) $$D_i = Y_i - L_i \cdot r_i,$$

where Y_i is disposable income excluding necessities, and r_i is the interest rate.

Assuming that people borrow at an optimal rate, as long as the elasticity of demand for credit is higher than -1, higher interest rates will lead to lower discretionary income. The standard assumption is that the elasticity of demand for credit is close to zero.[4] Since utility is rising in both credit and discretionary income, and since a rise in interest rates reduces discretionary income as well as borrowing, it also reduces utility.

Given this demand function, the loan amount and the total cost of borrowing are determined by the interest rate, and discretionary income will be a function of the risk of defaulting, p_i. We show in Appendix A that:

(3) $$D_i = Y_i \cdot \left(1 - \alpha \cdot \frac{\bar{r} + 2p_i}{1 + \bar{r} + p_i}\right),$$

where $\bar{r}$ is the competitive rate in a market with no default risk, and α is a weight that determines the demand for credit. We see that $dD_i/dp_i < 0$, so that discretionary spending is decreasing in the default risk. If the probability of default is declining in income—which is strongly supported by the data (more on this below, and in Appendix C)—then *the greater the*

for some by forcing them into the rental market. The direction of this effect is not controversial, but the full welfare implications would require a general equilibrium model of the interaction of rental and ownership markets.

[4] DeFusco and Paciorek (2017) estimate the elasticity of demand for mortgages to be around -.02, but it may be notably higher (in nominal terms) for credit card debt, where the estimate by Gross and Souleles (2002) is around -.85.

dispersion of the distribution of risk, the greater the dispersion of the distribution of discretionary income. Indeed, discretionary income is always more dispersed than disposable income. This is our first result, and it shows a heretofore overlooked effect, via interest rates, of increasing inequality of risk, even though the latter is well-documented (Hacker, 2008; Hacker et al., 2013; Häusermann et al., 2015; Palier, 2010; Rueda, 2007).

Lenders are not always able to lend at the optimal rate because they are constrained by usury laws or other lending regulations, or because it is too difficult to determine actual default risks above a certain level. In consequence, the lender may adopt a cutoff rule to limit exposure to bad loans. In the presence of such a rule, an increase in the dispersion of observed risk will also lead to more people being denied credit. At the same time, the dispersion of the distribution among those who can obtain loans will increase (under standard assumptions about the shape of the risk distribution).[5] Consistent with this logic, in our data we observe an effect of risk distributions on both access to loans and interest rate dispersion.

It directly follows from our first result that countries with more unequal risk of default distributions have more unequal discretionary income distributions, *after* controlling for disposable income inequality. This is not captured by the effect of risk on (expected) future income (as in standard insurance models); it is a direct effect of credit markets on the distribution of current consumption.

[5] Imagine a normal distribution with a fixed lower cutoff point. A means-preserving increase in dispersion implies that a greater proportion of the distribution is below the cutoff, at the same time as the dispersion of the distribution above the threshold increases.

2.2 The effect of information

We have assumed that borrowers and lenders are all fully informed about the risk of default. In practice, default risks are hard to observe for the lender, and usually not possible to signal credibly by the borrower. This creates a classic adverse selection problem. If the lender has no information about risk type, it will have to set an average interest rate that is proportional in equilibrium to the amount of defaulted loans among all borrowers (which *can* be observed). This average rate is denoted $\bar{\bar{r}}$ (distinct from $\bar{r}$, which is the competitive rate charged *if* all loans were repaid with interest).[6]

This common interest rate means that high-risk types will face lower interest rates than low-risk types compared to the case of full information. The consequence is a shift in lending towards high-risk types so that the total amount of defaulted debt increases, and the average interest rate rises. This is an efficiency cost, but, at the same time, it *reduces* inequality in discretionary spending because those with higher income and lower risks now pay more for credit while those with lower income and higher risks pay less.

The inequalizing effect of information can be established more generally if we assume that lenders learn about individual risks by observing credit history. Such history is constructed by

[6] Note that since the individual loan amount depends on income, if p_i is (negatively) related to income the average loan amount among those who end up in the bad state is not the same as among those who stay in the good state. Hence $\sum p_i \cdot L_i \neq \bar{p} \cdot \bar{L}$.

collecting information about the speed of debt accumulation, timeliness of repayments, past instances of default, etc., and we can conceive of such information as signals in a Bayesian updating game where "observed" risk is a weighted function of a prior and a signal. If $p_l^o = [p_{\min}^o, p_{\max}^o]$ is observed risk of individual *i* by lender *l*, we can write:

(5) $$p_l^o = \iota \cdot p_i^s + (1-\iota) \cdot \overline{p} \,,$$

where p_i^s is a noisy signal drawn from a distribution that is centered on the individual's true risk, p_i, and $\overline{p}$ is the mean among all borrowers, which is the prior. The parameter ι is a measure of the "precision" of the signal, which equals the information about *i* available to the lender. With no information ($\iota = 0$) the lender only observes the population mean, $p_i^o = \overline{p}$, and the range is therefore zero. At the other extreme, with complete information, $p_i^o = p_i$, the range equals the difference between the individual with the lowest and the individual with the highest underlying risk.

If we use the range as a measure of dispersion, we therefore have that:

(5) $$[p_{\min}^o, p_{\max}^o] < [p_{\min}^i, p_{\max}^i],$$

and the difference in the range is falling in information:

(6) $$(p_{\max}^i - p_{\min}^i) - (p_{\max}^o - p_{\min}^o) = f(\bar{\iota})\,.$$

Alternatively, we could treat the difference in the *variance* of underlying and observed risks as a function of information. Keeping in mind that discretionary income is a function of default risk, the implication is that *more information increases the inequality of discretionary income* (i.e., increases the range or the variance in income). Closely related, since a more dispersed distribution increases the share who are above the lending threshold (i.e., considered too risky), more information also leads to more inequality in access. This is the *second implication* of the model.[7]

2.3 The role of the welfare state

So far, we have assumed that any "catastrophic" loss of income leads to default, but people have an incentive to try hard to avoid defaults, which will cut them off from future borrowing (or significantly raise the cost of such borrowing). In the case of defaulting on a mortgage, people will lose their home. We do not explicitly model the individual decision to default but instead assume that if private funds available in the bad state of the world, k_i, are at or below some threshold, T_i, the borrower will default; otherwise not:

$$\text{(7)} \qquad \text{If } \left\{ \begin{array}{l} k_i \leq T_i \text{ then default} \\ k_i > T_i \text{ then do not default} \end{array} \right\}.$$

We can think of k_i as income from savings, selling assets or bringing forward long-term pension accumulations, etc. It is natural to think that k_i must be high enough to cover basic needs as well as essential fixed expenses (such as medicine) before debt servicing is possible. But there are clearly also subjective aspects to what individuals consider acceptable sacrifices, and the lender

[7] Because more information raises the borrowing costs for those at high risk it should generate a partially offsetting reduction in the magnitude of borrowing. But as argued by Lazarus (2020), low-end borrowers often do not possess the financial literary to reduce their debt exposure to appropriate levels, which adds to their default risks. This further magnifies interest rate inequities. We thank an anonymous reviewer for pointing this out to us.

cannot observe these directly. Some people will try to repay their loans at great sacrifice; others will be more willing to let go.

In Appendix A we derive the interest rate for the cases where (i) the lender cannot observe either risk of income loss, p_i, or individual thresholds, T_i, and (ii) the lender knows p_i but not individual thresholds. In the former case there will be a common interest rate for all (see Equation A11 in Appendix A), but in the latter case it will vary according to:

$$r_i = \frac{\overline{r} + 2 \cdot p_i \cdot p_{(k_i<T)}}{1 - p_i \cdot p_{(k_i<T)}}. \tag{8}$$

Intuitively, the interest rate is rising in individual risks and the probability of default. Since the latter depends on personal assets, k_i, such assets are a source of discretionary income inequality, even in the good state, as long as they are rising in income.

Social protection mediates this relationship, however, by adding a transfer, b_i, to personal funds in the bad state, which has the exact same effect as raising k_i (and reducing $p_{(k_i<T)}$). Even if b_i is a lump-sum benefit paid to everyone by a flat-rate tax (as in a Meltzer-Richard model), we show in Appendix B that the distribution of interest rates, and hence the distribution of discretionary income, becomes less dispersed as b_i rises.[8] This holds for a flat rate benefit; *ipso facto* it also holds for benefits that are targeted to those with low income (“means-tested”).

[8] The intuition is that a flat-rate benefit shifts the distribution of income in the bad state to the right, while the distribution of default thresholds stays constant. If the default threshold distribution is normal, this means that the

The conclusion is that the welfare state dampens the inequalizing effects of financialization and information, and that this effect is *in addition* to the direct effect of the welfare state on disposable income inequality. This is our *third result*. The existing literature only considers the direct effect of social spending on disposable income through redistribution; not the indirect effect through interest rates.

2.4 The role of financial regulation

Social protection systems were not created to reduce default rates or to equalize discretionary spending through a lower dispersion of interest rates. They were created to alleviate poverty or to mitigate the risk of income loss, and it is only with financialization that the indirect effect of the welfare state on income has become important. For this reason, we treat social spending as an exogenous variable that is not caused by the credit regime.

Financial regulation, on the other hand, is specifically designed to shape the terms of lending, as well as the risks that lenders and borrowers assume. Regulations are complex, but what concerns us here is the extent to which they facilitate the transfer of default risk to the state. A common form is credit guarantee schemes (CGSs), where a state agency steps in to provide collateral and some repayment guarantees (which can be less than 100 percent). State-guaranteed educational loans or government-backed loans to small businesses are examples. If these guarantees are

---

bottom portion of the income distribution, say the bottom decile, moves into the "thicker" portion of the default threshold distribution with more people now able and willing to service their debt.

credible, it reduces the risk of lending, and since risks are concentrated at the bottom of the income distribution it has the same pro-poor / pro high-risk effect as government transfers.

To illustrate the logic, we use the regulation of the American mortgage market as an example; it is perhaps the most important case of transferring default risks to the state. At the center of the system are two government-sponsored enterprises (GSEs)—Fannie Mae and Freddie Mac (FM/FM)—which are required by law to purchase all mortgages that meet certain minimum requirements issued by commercial banks, savings and loan associations (S&Ls), and other originators, and to securitize them by issuing bonds in the secondary bond market. Before recent reforms, the quasi-public role of FM/FM had two effects on private lenders. First, they became less concerned about default risks because these were largely absorbed by FM/FM. Lenders were given considerable discretion and minimum requirements were often finessed by the banks since they knew loans were rarely returned. Second, less concerned about risk, lenders stopped acquiring detailed and costly information about individual borrowers and effectively treated all would-be homeowners equally (over and above the vague minimum requirements set by FM/FM). Once approved, 'conforming loans' were offered at essentially the same terms to nearly everyone.[9]

[9] We recognize that there is a large and important literature on racial discrimination in lending. A comprehensive review of the evidence is provided in Goering and Wienk (2018). Data-driven algorithms are clearly not exempt from problems of bias, which is why the use of zip codes has been banned. But it seems clear that GSEs had the effect of broadening access to lending. Discrimination of any kind, including discrimination based on actual risk of default (which is legal), was probably reduced as long as GSEs bought up mortgages in a competitive market of originators that could largely ignore the risk of "put-backs".

This equalizing effect masks significant subsidization of high-risk (usually lower income) borrowers.[10] The 1990 amendment of the Fannie Mae and Freddie Mac charter made it an explicit goal to "facilitate the financing of affordable housing for low- and moderate-income families," a provision used aggressively under the Clinton administration to expand loans to low-income families (Acharya et al., 2011). It was thought, or at least hoped, that FM/FM's strong market position and the large margins they had been able to sustain between borrowing costs in the securities market and mortgage interest rates were enough to cushion them from the risks of bad debt. Although private corporations since 1968, it was also widely believed that FM/FM loans were implicitly guaranteed by the government, which enabled the GSEs to borrow very cheaply. Apparently confirming this logic, China and other countries with saving surpluses poured large sums of money into the FM/FM-issued bonds, pushing average interest rates down (Eichengreen, 2008).

This cozy consensus was shattered with the crash of the sub-prime mortgage market, after which the stock prices of FM/FM collapsed. FM/FM were subsequently placed into conservatorship in September 2008. Before and after the government takeover, a series of reforms were implemented to reduce the risk-exposure of FM/FM and shift more of it to banks and other mortgage originators, as well as to a third government entity, Ginnie Mae, which securitizes

[10] Lax bankruptcy rules in the US may also be seen as part of an "accommodating" credit regime that subsidizes high risks. Yet, because it helps borrowers rather than lenders (and gives the latter a reason to restrict lending), the effect on inequality in credit is ex ante ambiguous.

mortgages directly guaranteed by the Federal Housing Administration (FHA). Below, we use these measures as a natural experiment whereby lenders are strongly incentivized to acquire more information and use it to screen out or raise interest rates on risky borrowers. The financial crisis is thus a window into both the effect of government regulation and the effect of information (one is a cause of the other).

3 Empirical tests

Our theoretical model makes three empirical predictions:

- (H1) More information increases the spread of interest rates (and hence the inequality of discretionary income).
- (H2) The government acting as a backstop in loan markets reduces the spread of interest rates (and hence the inequality of discretionary income).
- (H3) More generous public income support facilitates access to loans and decreases the spread of interest rates.

For the model and all three hypotheses the underlying assumption is that risks are correlated with income. We do not think that this is a controversial assumption, but we show in Appendix C that it is strongly supported by the data.

3.1 A note on the relationship between information and regulatory incentives

It is difficult to test H1 and H2 separately because while information is increasing over time as a result of the data revolution, discontinuous exogenous shifts in information typically occur only as a result of regulatory changes that incentivize lenders to seek more information (or not). Conversely, changes in public subsidies for lending changes the risks that lenders face but at the same time also their incentives to acquire information. In this section, we briefly show that changes in incentives, under certain assumptions, can be treated as equivalent to changes in information. We use this equivalence to infer the effect of information from sharp regulatory changes.

Like Westcott's small-town banker, lenders crave information because it allows them to separate good risks from bad, and thus to (i) cut out potential borrowers who are likely to default, and (ii) differentiate interest rates among borrowerrs to reflect individual risks. Yet, the benefits of information have to be weighted against the cost of acquiring information. Furthermore, when the state assumes some of the default risk, the incentive to acquire information falls. We can capture this logic using a very simply lender utility function:

$$U_L = \iota(\delta) - c(\iota, A) \ ,$$

where the benefit of information (ι measures information as before) is a negative function of δ, which we can think of as the probability that the regulator will assume responsibility for defaulted loans. The cost of information is a rising function of the level of information, moderated by an "information technology" factor, *A*. Big Data, faster processors, and better algorithms, make the rise in cost "flatter." A simple concave representation of this utility function is:

$$U_L = \iota \cdot (1-\delta) - c(A) \cdot (\iota^2),$$

which implies a maximum investment in information of:

$$\iota^* = \frac{1-\delta}{2 \cdot c(A)}.$$

The expression shows that changes in the regulatory framework that affect the cost to the lender of defaults, δ, have the same effect on information as changes in the cost of information, c, due to new technology. The latter is mostly driven by secular changes in ICT technology that reduce the costs of compiling and analyzing data. The former is driven by regulatory changes that can be abrupt. We know that the cost of information is declining—as implied by Moore's Law—but the gradual nature of this decline makes it hard to identify its effect on interest rates. Sudden changes in the regulatory framework, by contrast, can be used to gauge the causal effect of information, even if it can only capture this effect indirectly through changes in the incentives to acquire information. In the next section we provide a simultaneous test of H1 and H2 using this logic. We are also able to confirm that the gradual drop in the cost of information is correlated with a gradual increase in the dispersion of interest rates, although it is of course not possible to establish causality using this evidence.

3.2 Regulation, information, and inequality in mortgage interest rates

To test H1 and H2 we use a dataset that contains all single-family loans that Freddie Mac purchased or guaranteed from the first quarter of 1999 to the last quarter of 2019 – 36,269,139 mortgages. As described above, Freddie Mac is one of the two main government-sponsored enterprises (GSEs)—along with Ginnie Mae, a government agency—that purchase "conforming mortgages" from lenders and sell them in the secondary bond market.

The main reason GSEs return mortgages is delinquency or default—even several years after closing—but it is at the discretion of the GSE. Mortgages closed in 2007 and 2008 saw a dramatic rise in put-back rates, which far exceeded the rise in defaults (Goodman et al., 2014, p.

60); the aggregate amount of repurchase requests increased tenfold. In terms of the notation used in the previous section, an increase in put-backs is equivalent to a decrease in δ: the probability that the regulator will assume the default risk. Moreover, the GSEs tightened the underwriting guidelines for conforming mortgages that lenders had to adhere to,[11] and increased their quality controls in various ways.

These changes were rolled out starting in early 2008, and the beginning of that year therefore serves as a break after which lenders had strong incentives to use more information to accurately assess mortgage applications. From the perspective of our theoretical framework the subprime mortgage crisis is a discontinuity, at which the effort lenders expend and the amount of information they use to assess mortgage quality sharply increased. Again, the trigger for lenders to acquire more information was regulatory reforms, put-backs in particular, that raised the costs of not accurately identifying default risks. For this reason, we expect the spread of interest rates to increase at the discontinuity.[12]

This increased scrutiny and intensified information collection clearly shows up in the data as a sharp rise in the number of days to close a loan. In an interesting account of the role of

[11] For example: "In light of [deteriorating] market conditions, we are reinforcing our appraisal standards and underwriting expectations related to maximum financing in declining markets" (Freddie Mac Bulletin 11/15/07, p. 3).

[12] The effect is reduced, however, by the extent to which Ginnie Mae (a pure government entity) increased its share of mortgage-backed securities, since this reduced the exposure of Fannie and Freddie to high-risk, low-income lending.

technological innovation in mortgage underwriting, Foote, Loewenstein and Paul (2019) show that mortgage processing times dramatically dropped between 1995 and 1998—from close to 50 (1994) to under 30 (1998) days—and continued to trend downward until 2005—to about 17 days.[13] They attribute this decline in processing times to technology-augmented innovation; very consistent with the cost of information gradually dropping. More interestingly for our purposes is the sharp increase in processing times in 2008 and 2009, from about 18 (2007) to about 26 (2008) to almost 40 days (2009). It is worth citing their explanation in some detail:

> "After the US housing boom ended, refinance timelines increase sharply as various lender and governmental policies changed. One of the most significant policy changes involved the repurchase policies of the GSEs. Fannie Mae and Freddie Mac occasionally require mortgage originators to repurchase loans that do not meet the agencies' underwriting guidelines. After housing prices fell, both Fannie and Freddie increased their repurchase requests to originators that had incorrectly underwritten loans. This prompted originators to follow GSE policies more carefully, which likely lengthened origination timelines" (Foote et al., 2019, p. 14).

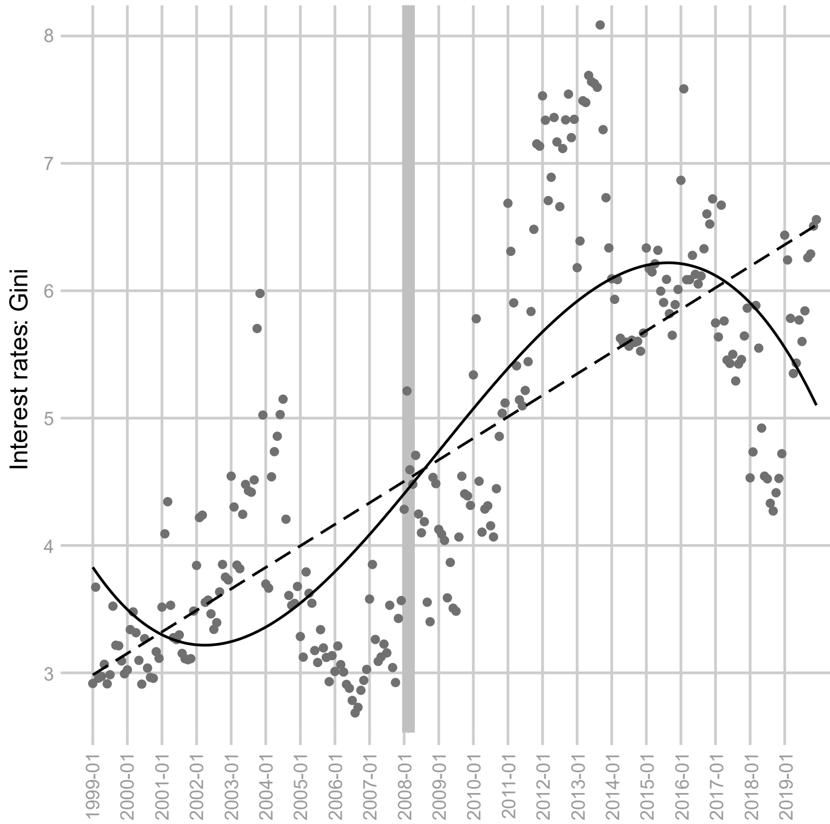
We consider these findings and conclusions by Foote, Loewenstein and Paul (2019) as evidence in favor of our assertion that the beginning of 2008 marks a discontinuity at which lenders were

[13] The rely on the Home Mortgage Disclosure Act (HMDA) micro-data (using confidential variables) and focus on processing times for refinance loans. They report "average processing time [in days] by year after stripping out any variation explained by the size of the lender, the borrower's race and gender, whether the borrower has a coapplicant, and the concurrent monthly application volume. The processing times are calculated as of the year of application and include both closed loans and denials" (Foote et al., 2019, p. 37 (note for Figure 7)).

very strongly incentivized to seek more and better information on mortgage applications. As argued, we expect an increase in the spread of interest rates at this discontinuity.

To assess the propositions that the interest rate spread increases over time in general, and at the discontinuity in particular, we start by calculating the Gini coefficient of interest rates for each year-month between 1999 and 2019, using Freddie Mac's "Single Family Loan-Level Dataset." On average, each cell (year-month) contains about 60,000 mortgages (the median cell size is 55,714, the minimum and maximum are 11,910 and 207,049, respectively). Figure 2 plots these Gini coefficients over time. While the figure shows an upward trend, there only seems to be a short-lived increase in the spread of interest rates at the discontinuity (January 2008).

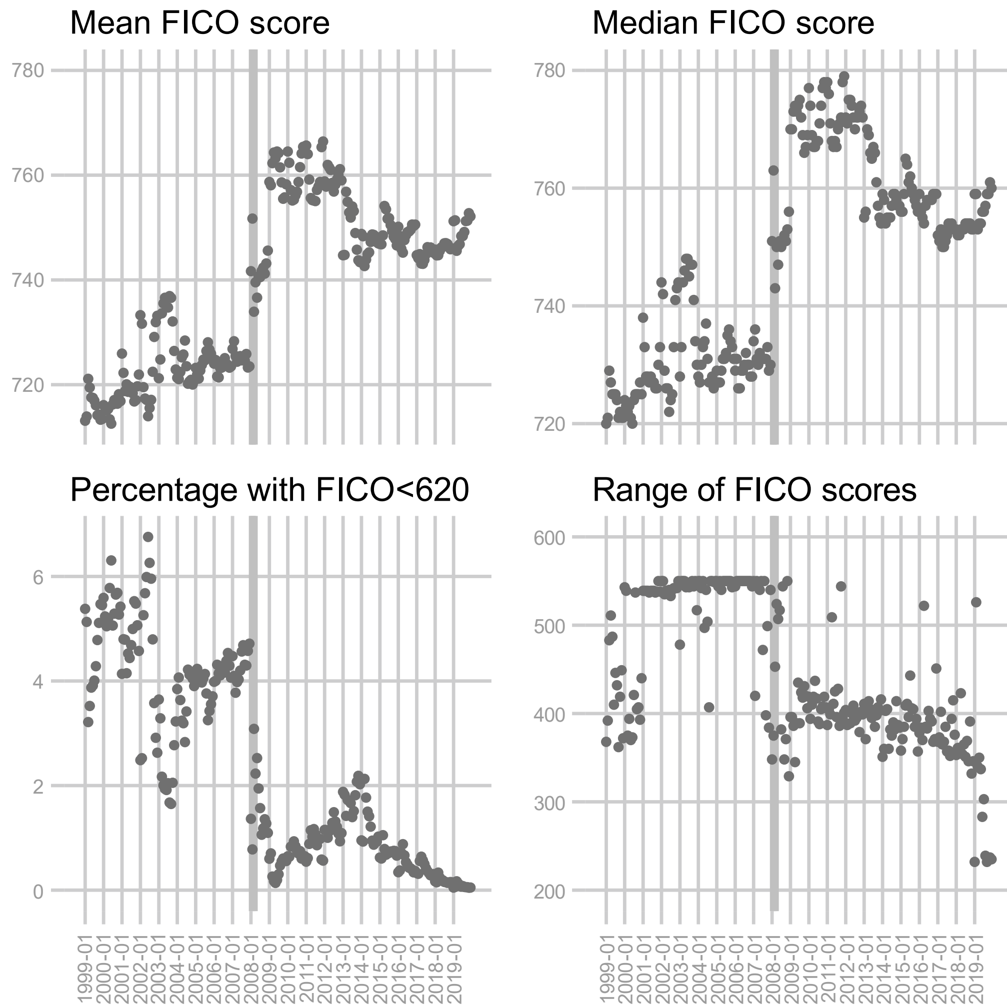
Figure 2: Interest rate spread over time (year-month level)



Note: Shown are the Gini coefficients of interest rates for each year-month between 1999 and 2019. The dashed line is a linear fit line while the solid line is a cubic fit line. The shaded area indicates the first quarter of 2008.

However, balance tests reveal that the samples to the left and right of the discontinuity are very different. Most importantly, the composition changes in terms of the distribution of FICO scores (and FICO scores are highly correlated with interest rates), right at the discontinuity. This can be seen in Figure 3.

Figure 3: Average FICO scores in sample, over time



Note: Shown are mean and median FICO scores, as well as the percentage of FICO-scores below 620 and the range of FICO scores within the full sample of the dataset. The shaded area indicates the first quarter of 2008.

This rise in FICO scores is itself very consistent with the claim that early 2008 was a discontinuity at which lenders engaged in more careful screening since it implies that an increased number of potential borrowers with low FICO scores were denied loans.[14] But in addition to such censoring, lenders began to differentiate more between borrowers with good FICO scores in the terms they were offered. The obvious interpretation is that lenders acquired additional information among borrowers with similar FICO scores. We focus our analysis on the change in the spread *within* FICO tranches to circumvent the potential problem of a changing composition of borrowers. Note that this gives us a conservative estimate of the effect of the discontinuity because (i) we do not capture the rise in rejected mortgage applications (which would otherwise increase dispersion), and (ii) we do not capture the rise in the interest-rate spread *across* FICO groups.

Specifically, to balance the samples before and after the discontinuity—to compare apples with apples—we restrict the sample to mortgages that fulfill the following criteria, and we also shift the analysis from the year-month-level to the year-month-FICO-2d level:

- Credit (FICO) scores in the range of 620 to 819. We drop cases with scores below 620 because this is the minimum score required by Freddie Mac to qualify for a conforming mortgage, at least under normal conditions in most years (this drops 2.19% of the sample).
- We drop cases with FICO scores in the 820-850 range (the very top-end of the FICO-score distribution) because only few mortgages are in this category, and they are unevenly distributed over time – leading to unreliable and infrequent estimates of the spread of interest rates for FICO scores above 819 (this drops 0.03% of the original sample).
- 30-year mortgages (applies to 67.1% of the original sample).

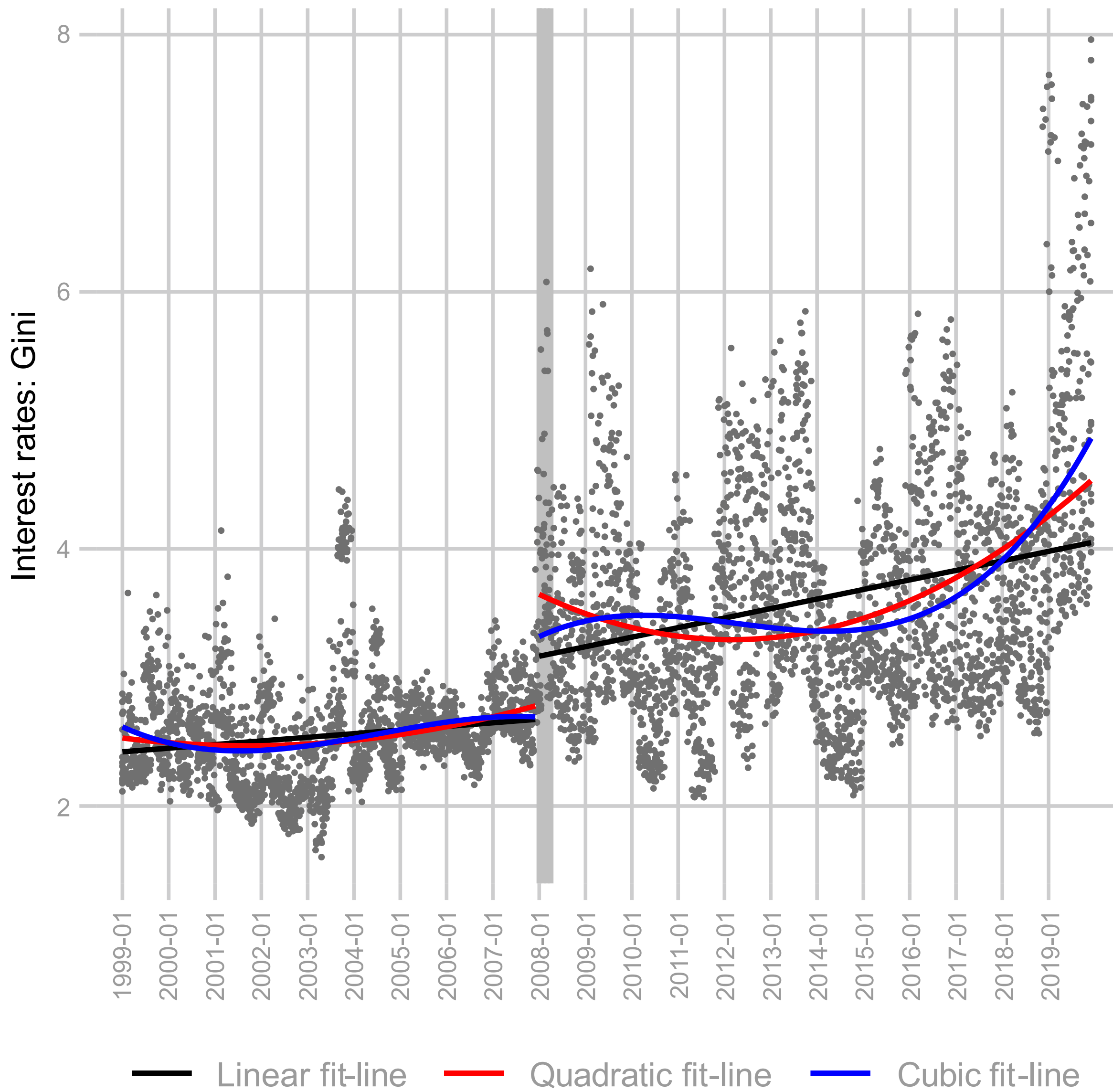
[14] Our dataset does not include denied mortgage applications, but the patterns in Figure 3 clearly suggest that lenders screened out applicants with low FICO scores.

- Fixed-rate mortgages (applies to 100% of the original sample).
- No mortgage insurance (applies to 82.9% of the original sample).
- Loan-to-value ratio of a maximum of 80 percent (i.e., minimum of 20 percent down payment) (applies to 80% of the original sample).
- Single-family units that are owner occupied (applies to 90.6% of the original sample).
- US states only (applies to 99.8% of the original sample).

This leaves us with a sample of 15,299,343 mortgages. Balance tests show that the composition of mortgages before and after January 2008 is very similar even across bins, and assuredly so within bins. We use this dataset to calculate measures of interest rate dispersion—such as the Gini coefficient, the Coefficient of Variation, and others—at the year-month-FICO-2d level.[15] Figure 4 plots the Gini coefficient of interest rates within FICO-2d-bins over time. Therefore, a dot in the figure represents the Gini coefficient of a year-month-FICO-2d-bin.

[15] By FICO-2d level, we refer to the first two digits of FICO scores, which range from 620 to 819 in our sample. For example, FICO-2d score 62 refers to FICO scores 620-629.

Figure 4: Interest rate spread over time (year-month-FICO-2d level)



Note: Shown are the Gini coefficients of interest rates for each year-month between 1999 and 2019, within FICO-2d levels (mildly jittered), along with local polynomial fit-lines of orders 1 to 3. The shaded area indicates the first quarter of 2008.
Source: Freddie Mac's "Single Family Loan-Level Dataset" (http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.page, last accessed June 3, 2021 [https://perma.cc/KU32-9TXH]), Q1-1999 to Q4-2019.

At least three aspects are noteworthy about the patterns in Figure 4. First, the spread of interest rates—as measured by the Gini coefficient within 10-point FICO-bands—clearly increases over time, as hypothesized. The spread of interest rates roughly doubled within the 20-year period under consideration.

Second, there is a clear increase of the spread of interest rates beginning in January 2008 (the discontinuity). The figure shows local polynomial fit lines of orders 1 through 3 (fitted over the entire support of the pre- and post-treatment time periods, respectively). All indicate a visual break in the series. To test whether there is, indeed, a break in the spread of interest rates, we rely on regression discontinuity (in time) analysis. The outcome variable is the interest rate spread (measured by Gini coefficients within FICO-2d bins at the month-year level). The score / running variable is month-years, with January 2008 as the discontinuity. We employ a (data-driven) mean square error (MSE) optimal bandwidth selection procedure—imposing the same bandwidth on each side of the cutoff—for local polynomial estimation of and inference on treatment effects and report robust bias-corrected confidence intervals, with standard errors clustered at the FICO-2d level. Observations are weighted via a triangular kernel function (i.e., observations closer to the cut-off are weighted more heavily). Our dataset has repeated observations in the running variable (20 FICO scores per year-month), for which the estimator controls. Table 1 reports the RD estimates based on local polynomials of orders 1 through 4.

Table 1: Regression discontinuity estimates

	(1)	(2)	(3)	(4)
Order of Local polynomial	1	2	3	4
RD estimate	1.003***	1.300***	1.032***	1.063***
	(0.0976)	(0.113)	(0.105)	(0.112)
Robust 95% CI	[.901 ; 1.283]	[1.14 ; 1.609]	[.75 ; 1.176]	[.782 ; 1.259]
BW type	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Triangular	Triangular
Order Loc. Poly. (p)	1	2	3	4
Order bias (q)	2	3	4	5
N	5025	5025	5025	5025
N (l)	2147	2147	2147	2147
N (r)	2878	2878	2878	2878
Eff. N (l)	340	300	320	500
Eff. N (r)	360	320	340	520
BW est. (l)	17.15	15.78	16.52	25.35
BW est. (r)	17.15	15.78	16.52	25.35
BW bias (l)	28.13	25.24	26.03	38.35
BW bias (r)	28.13	25.24	26.03	38.35

Note: Standard errors in parentheses (clustered at the FICO-2d level).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Estimates adjusted for mass points in the running variable.
The outcome variable is the Gini coefficient of interest rates at the FICO-2d level.
The running variable is time (month-year) with the cutoff in January 2008.
These estimates are based on the user-written Stata commands—rdrobust—(version 8.2.0 from 2021-05-18) (Calonico et al., 2017).
Source: Freddie Mac's "Single Family Loan-Level Dataset" (http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.page, last accessed June 3, 2021 [https://perma.cc/KU32-9TXH]), Q1-1999 to Q3-2020.

Table 1 shows that the RD estimate ranges from about 1 to about 1.3, depending on the order of the local polynomial. These estimates are statistically significant at p<0.001. The literature usually recommends (Gelman & Imbens, 2019) and chooses (Pei et al., 2021) lower-order over higher-order polynomials. Therefore, we prefer the first model (which uses local linear regressions). The RD estimate of about 1 implies roughly a 40% increase at the threshold in the interest rate spread (from about 2.5 before the threshold on average).

In appendix D, we perform a wide variety of additional tests and show that the finding of a statistically significant (and substantively meaningful) increase in the interest rate spread at the discontinuity is very robust. In particular, we perform the following additional/robustness checks:

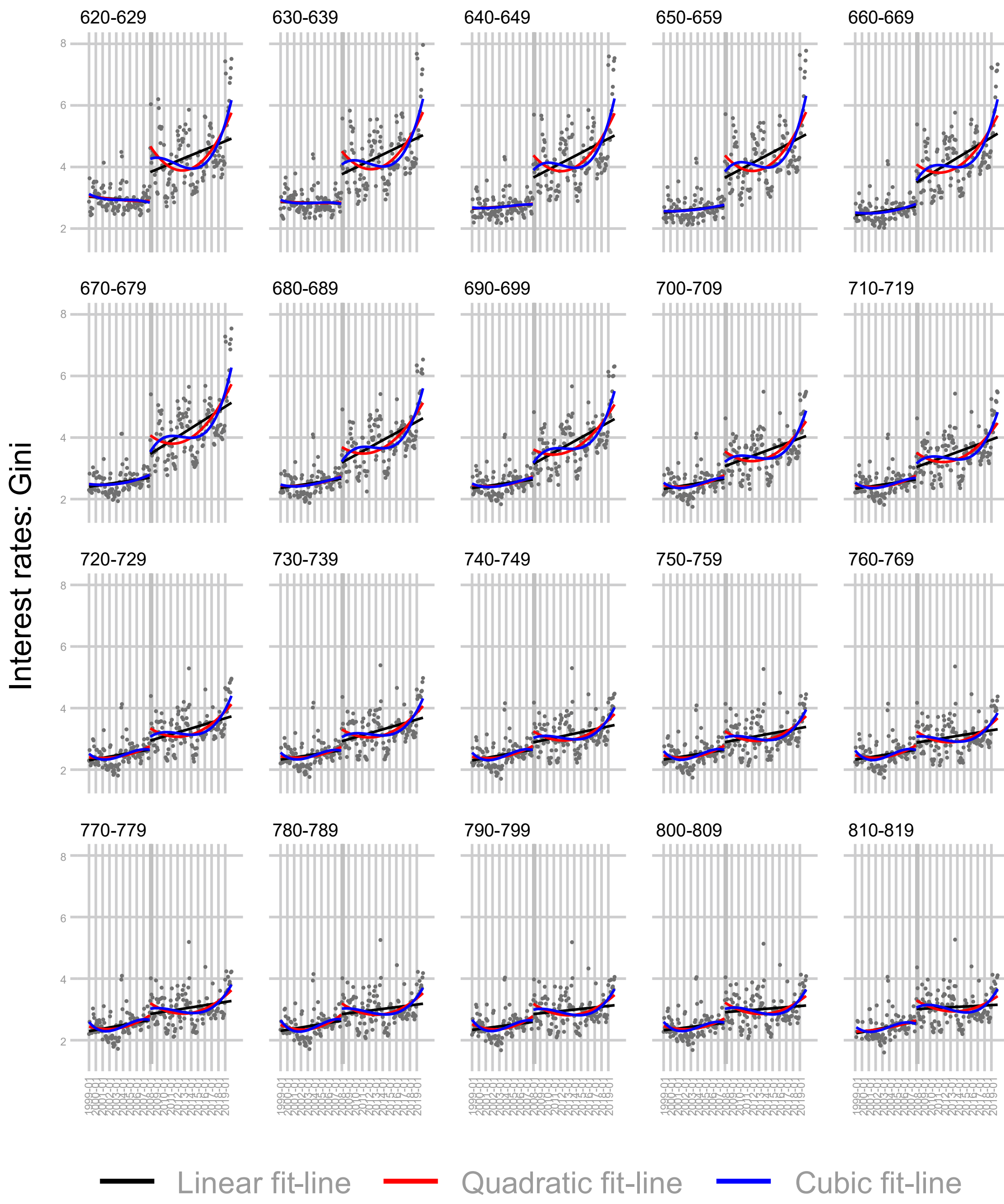
- Different bandwidth selection procedures
- Different kernel functions
- Covariate adjusted estimates (controlling for month dummies; lagged dependent variable; average interest rate) (Hausman & Rapson, 2018)
- Different specification of the running variable (months, quarters, trimesters, half-years, years), using the year-month-FICO-2d-level data
- Different specification of the running variable (months, quarters, trimesters, half-years, years), with the outcome variable recalculated at the respective unit level (months, quarters, trimesters, half-years, years)
- Sensitivity to observations near the cutoff (donut hole approach)
- Placebo outcomes
- Placebo cutoffs
- Masspoints adjustments

The third noteworthy aspect in Figure 4 is the increasing range of the interest rate spread, with a clear jump at the discontinuity. For example, in 2007, the Gini coefficient of interest rates ranged from 2.3 to 3.5 while it ranged from 2.4 to 6 in 2008. This might suggest that the increase in the interest rate spread at the discontinuity was higher for some of the 20 FICO-2d groups. The obvious hypothesis is that lenders focused their increased screening efforts on applicants with lower FICO scores because it is well-documented that the spread of risks, measured by default rates, is greater for lower FICO tranches (VantageScore, 2020).[16] If so, lower FICO scores essentially received a higher dosage of the treatment (scrutiny from lenders). To explore this supposition, Figure 5 reproduces Figure 4, but with separate panels for each of the 10-point FICO bands. Therefore, Figure 4 is a pooled version of Figure 5 and a dot within the panels of Figure 4 indicates a year-month. Within each panel, the mortgages are very similar. Most importantly, their FICO-scores are within 10 points of each other (by construction) and the samples before and after January 2008 are balanced well – the figure therefore offers something close to an apples-to-apples comparison.

Visual inspection of Figure 5 suggest that the increase in the spread of interest rates at the threshold was particularly pronounced at lower FICO-scores – roughly in the 620-679 range. In contrast, at higher FICO levels, the increases seem to be more minor.

[16] Taking FICO-scores as a proxy for default risk, the risk distribution is right-skewed.

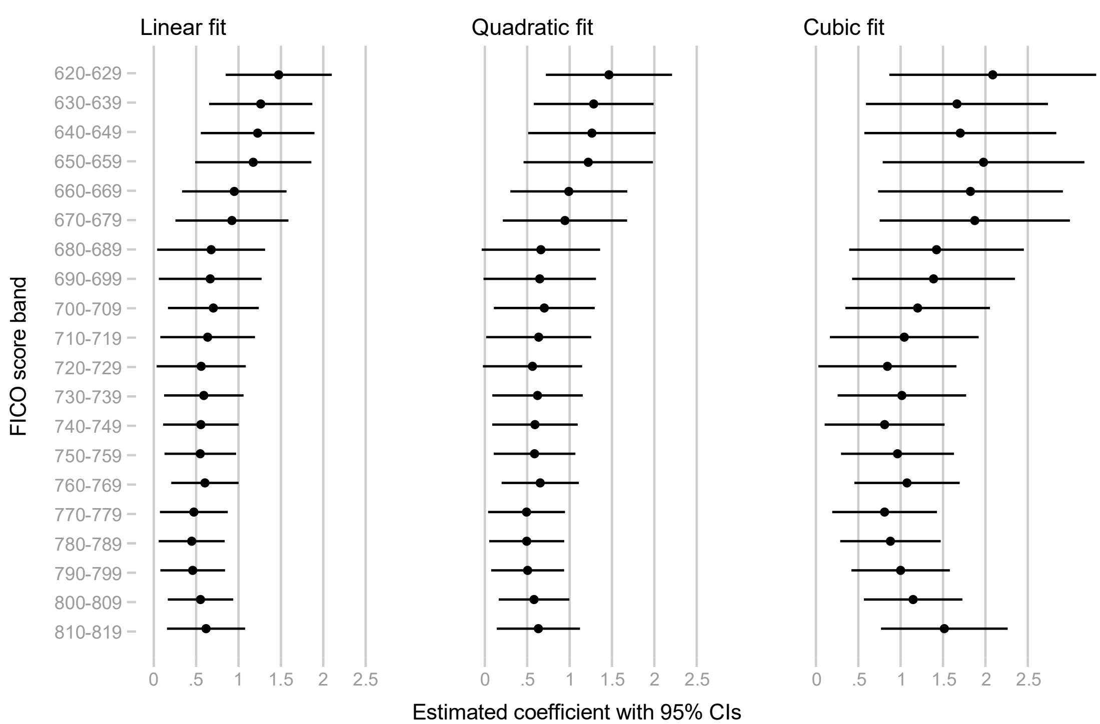
Figure 5: Interest rate spread over time (year-month-FICO2d level), at FICO2d



Note: Shown are the Gini coefficients of interest rates for each year-month between 1999 and 2019, by FICO-2d level, along with local polynomial fit-lines of orders 1 to 3. The shaded area indicates the first quarter of 2008. The figure contains the same datapoints as Figure 4, but arranges it by FICO-bands.

Statistical discontinuity tests confirm this pattern. In particular, Figure 6 summarizes the regression discontinuity estimates for each of the 20 FICO-bands as coefficient plots. The three columns display estimates based on linear, quadratic, and cubic local fit lines (the estimates are equivalent to the models 1, 2, and 3 in Table 1—they employ the same bandwidth selection procedure, kernel function, and so on—but they are derived from each FICO-2d bin separately). The figure shows that, while all estimates are positive and almost all estimates are statistically significant, they tend to be larger at the lower end of the FICO score distribution. This test therefore adds further evidence that is consistent with the hypotheses. It also ameliorates one of the weaknesses of regression discontinuity in time designs by adding cross-sectional evidence (Hausman & Rapson, 2018).

Figure 6: RD estimate at FICO-2d-levels, different polynomials



We interpret the general upward trend in interest rate inequality in general (Figure 4 and Figure 5), the increase in early 2008 (Figure 4), and the sharp increase in early 2008 among lower FICO scores (Figure 5 and Figure 6) as evidence consistent with our framework and hypotheses H1 and H2 above. Increasing information, whether gradually rising over time or induced by abrupt regulatory change, does indeed seem to increase interest rate dispersion, as predicted.

Regression discontinuity-in-time designs—such as our approach above—face challenges (Hausman & Rapson, 2018). For example, January 2008 was during a tumultuous time and there might be other candidate explanations for the increase in interest rate inequality. But starting in early January, the evidence clearly suggests the importance of information: the time to close sharply increased; the average FICO scores of loans supported by Freddie Mac 's lower FICO score loans became much less common in Freddie Mac's portfolio; and lenders began to make more fine-grained distinctions between borrowers. We can infer that they relied on information that went well beyond FICO scores since the spread in rates increased notably even within narrow (2-digit) FICO tranches. Lenders did this, we argue, because regulators provided them with powerful new incentives to separate good from bad risks.

But there are good reasons to believe that the financial industry has been continuously improving its information both before and after 2008. It is a frontrunner in adopting new ICT technologies for that purpose (Foote et al., 2019), and even with the increased role of Ginnie Mae (which reduces industry exposure to bad risks), there is a clear upward trend in the spread. This is consistent with the price of information falling over time, which is theoretically predicted to have the same effect on lender behavior as a rise in the cost of defaults (section 3.1).

3.3 The welfare state and home ownership

Our model's third prediction is that more (less) generous public income support expands (contracts) access to lending and decreases (increases) the spread of interest rates. The unemployed and those at high risk of unemployment are a greater risk to lenders unless a generous unemployment benefit system enables people to keep servicing their debt. Because unemployment risks are higher for lower-skilled, lower-paid workers, low replacement rates will disproportionately raise borrowing costs and rejection rates at the bottom of the income distribution. Those at the higher end will instead benefit from lenders screening potential borrowers more carefully.

For an initial exploration of this hypothesis, we exploit the profound changes in the German unemployment benefit system resulting from the Hartz-IV reforms in 2005 (Arent & Nagl, 2013). Because the reform affected the ability of unemployed to service debt in the event of unemployment, we can compare changes in homeownership rates across groups (un)affected by the reform, from before to after the reform; a difference-in-difference approach.

Unlike the US system, government entities play no direct role in the lending market in Germany,[17] and banks offer mortgages, which are typically fixed-rate, on a competitive basis. The system has strong build-in prudential safeguards, including low loan-to-value ratios and limited equity release options, so any changes in the assessed creditworthiness of borrowers

[17] An exception is the state-owned promotional bank "Kreditanstalt für Wiederaufbau" (KfW) that has various programs to support home ownership, but it is not allowed to compete with commercial banks. There are also subsidies incentivizing homeownership through the state-run aid for pension schemes (Wohn-Riester).

show up immediately in lending decisions, and because of civil usury law, lenders tend to cut off risky prospects rather than charge high interest rates.[18]

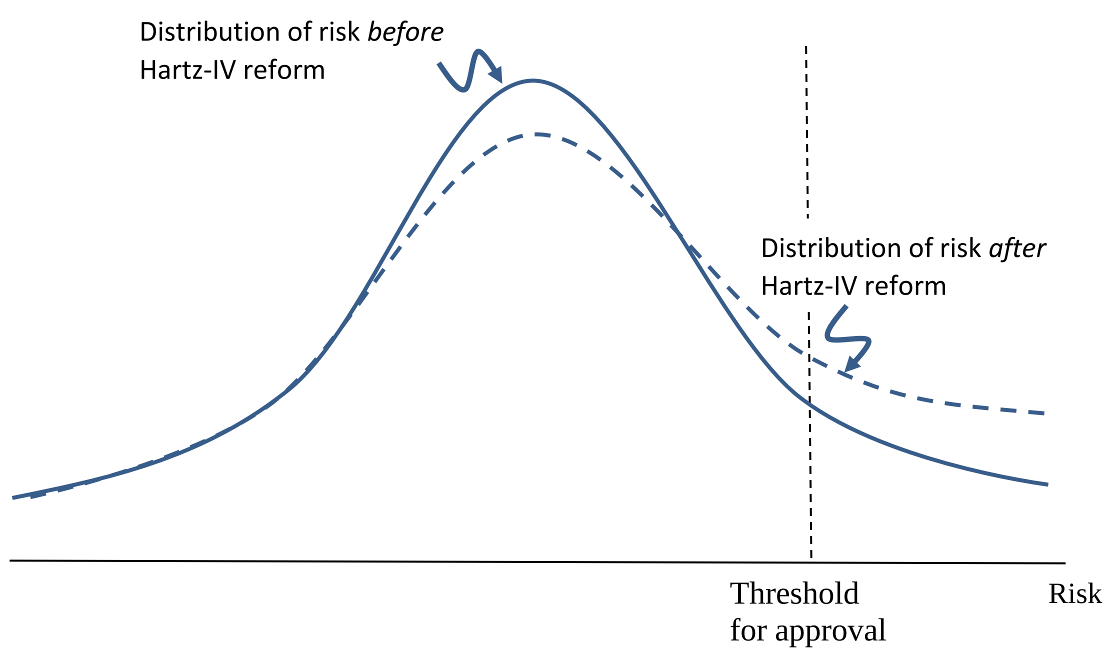
The most consequential changes of the Hartz-IV reform occur after a year of unemployment benefit recipiency. Before the reform, the unemployed could qualify for ‘unemployment assistance,’ a benefit proportional to previous wages—67% in the first year and ca. 57% in the second—that could be collected indefinitely, subject to annual renewal. To qualify, assets had to be below a certain threshold (~520 Euro times age), though certain assets were protected (“Schonvermögen”), including owner-occupied housing of reasonable size. For banks, default was not a major concern if the income-to-loan ratio was high enough because unemployment assistance was proportional to previous wages at a fairly generous rate and paid indefinitely.

After the reform, the reasonably generous, proportional, perpetual unemployment assistance in the second year was replaced with a meager, flat rate, conditional benefit (Arbeitslosengeld II). Owner-occupied housing of reasonable size is treated as a protected asset as before, but the overall limit for *other* protected assets is significantly lower (~150 Euro times age). This somewhat odd (Kaiser, 2018) differential treatment of assets means that a mortgage provides an opportunity to protect assets from the government by shifting them into owner-occupied housing. Consequently, for those at risk of unemployment, incentives for seeking a mortgage increased with the Hartz-IV reforms. For banks, however, there is now much more reason to worry about default among those with risk of long-term unemployment.

[18] Interest rate may not be more than twice the comparable market rate in relative terms, and not more than 12 percentage points in absolute terms (BGH, Urt. vom 13. März 1990 - XI ZR 252/8).

Overall, the reforms had two potential effects on financial markets: (i) they made it more difficult for some people to qualify for a loan, and (ii) they increased the risk of default among some borrowers who did qualify. The logic is illustrated in Figure 7, where the solid line is the pre-reform distribution of default-risk, and the dashed line is the post-reform distribution. The share with observed risk above the threshold for approval increases, and the distribution of those below the threshold "flattens" (becomes more dispersed), reflecting a more right-skewed default-risk distribution. The implications are a reduction in the number of loans granted among higher risks and an increase in the interest rate spread of loans that are granted. We have data that can illuminate the former effect.

Figure 7: Distribution of default risk before and after the Hartz-IV reforms in 2005



Unfortunately, we do not have data on declined mortgage applications, but we have data on homeownership rates that allow us to shed some light on the fortunes of the unemployed compared to the employed (or poor versus rich) before and after the reform. We do so in three steps. First, we track and compare homeownership rates of the employed vs. the unemployed over time. The data show a sharp drop in homeownership among the unemployed that coincides with the cuts of unemployment benefits.

Second, to gauge the effects of reforms on relinquishing ownership we track homeownership rates by employment status for those that were homeowners before the reform. The data indicate that, for this subsample, ownership rates did not differently drop among either the unemployed or employed. We infer from this that the unemployed rarely relinquished housing assets and that the drop in homeownership rates for the unemployed must be due to lower rates of home acquisition after the reform, presumably because of more difficult access to credit. It may seem surprising that the reform is not associated with more widespread selloff among the unemployed, but as we noted, the Hartz-IV reforms gave people incentives to hold onto their housing wealth, which is exempt from the requirement to spend down personal savings before drawing benefits ("Schonvermögen"). Those who were already homeowners before the reform were undoubtedly also in a stronger financial situation than those who were not, and therefore less likely to default. The point for our purposes is simply that the drop in homeownership among the unemployed must be because people are less likely to obtain mortgages after the reform.

Third, among the employed we find that homeownership rates among the poor and the rich diverged after the reform. That is especially true for those who were not already owners before the reform. Since the poor are at much higher risk of becoming unemployed, the obvious explanation is that mortgage lenders increasingly avoided bad risks after the reform.

The following four tables display the evidence just summarized. Each table shows the difference in homeownership between two groups, before and after the reform (“diff-in-diff”). The first is a comparison of homeownership rates by employment status. The most authoritative data for this information is the “Sample Survey of Income and Expenditure” (EVS), which is based on about 60,000 respondents and conducted every 5 years (we have data for 1993, 1998, 2003, 2008, 2013, and 2018). We do not have access to the micro-level data, but the Federal Statistical Office publishes—or provided us with—aggregate data on homeownership[19] for all households and for the unemployed. This allows us to compare homeownership rates among the “Unemployed” (treated group) versus the “Employed”[20] (control group), before and after 2005. Table 2 display the results of a difference-in-difference test of this comparison. The results show that homeownership rates among the employed did not significantly change while they sharply declined among the unemployed. The difference-in-difference estimate is about -12.7 percentage points for homeownership and statistically significant.[21]

[19] Data are published for (i) “households with house- or land property” and (ii) “households with land property,” among other breakdowns. We report as “homeownership” item (i) minus (ii).

[20] The “Unemployed” are households where the main earner is unemployed. The “Employed” are defined as ‘all households’ minus the “Unemployed.” For presentational ease, we refer to this group as the “employed” even though it technically is the group of “Not Unemployed” households.

[21] We re-estimated the models in Table 2 but added control variables, namely the unemployment rate and/or (linear of factorial) time. The difference-in-difference estimate remains statistically significant.

Table 2: Homeownership (EVS)

Homeownership	Pre-2005	Post-2005	Difference
Employed	0.423	0.455	0.0316 (0.023)
Unemployed	0.222	0.126	-0.0959** (0.023)
Difference	0.201** (0.023)	0.329** (0.023)	-0.127** (0.033)

Note: Standard errors in parentheses. N=12.[22]
* $p < 0.05$, ** $p < 0.01$

The German Socio-Economic Panel (GSOEP) (Liebig et al., 2019) is another source that allows us to track homeownership rates among the unemployed versus the employed.[23] The EVS and the GSOEP use somewhat different definitions of homeownership[24] and employment status,[25] and they cover different time-periods.[26] So, while estimates from the two sources are not directly

[22] We only have aggregate data and for the years listed above, so the number of observations in Table 2 is 12.

[23] Again, strictly speaking the "not unemployed" since it includes pensioners and other people not in the labor force.

[24] The SOEP provides information on home ownership and how the property has been acquired (inherited vs. purchased). We construct a binary home ownership variable that equals one for homeowners that have purchased their home and zero for those that do not own a home. About 25% of home ownership is the result of inheritance, and we drop these cases from the analysis. Substantive results are similar when we include inherited homeowners into the analysis.

[25] Our unit of analysis is the household, but we have person-level information that allows us to code the employment status of the household head and her/his partner. We code unemployment as unemployment of either the head or her/his partner, or both.

[26] The SOEP survey started out in 1984 with a sample that was representative for West Germany. Since then, refreshment samples have been periodically added to keep the survey representative. We make use of all samples that cover the 2000s (samples A to F, with F starting in 2000) and apply cross-sectional weights.

comparable, we do expect that they reveal similar patterns. Table 3, which displays difference-in-difference estimates—comparing the unemployed with the employed before and after 2005—shows that this is, indeed, the case: homeownership rates dropped markedly among the unemployed, comparing 2000-2004 with 2005-2010, and the difference-in-difference estimate is about -7.5 percentage points.

Table 3: Homeownership (GSOEP)

Homeownership	**Pre-2005**	**Post-2005**	**Difference**
Employed	0.402	0.384	-0.0182** (0.005)
Unemployed	0.256	0.162	-0.0936** (0.014)
Difference	0.146** (0.013)	0.221** (0.013)	-0.0754** (0.015)

Note: Standard errors (clustered at household-level) in parentheses. N=100,667
* $p < 0.05$, ** $p < 0.01$

The second step in our three-pronged approach compares the development of homeownership for the subsample of respondents who already were homeowners before the reform (during 2000-2004), using the GSOEP data. We again use difference-in-difference estimates even though, by construction, there are no differences between the unemployed and employed before the reform since the sample is restricted to those that are homeowners before the reform. Table 4 shows that there are no meaningful differences after the reform, either – the estimated difference-in-difference is essentially zero, and not statistically significant. This suggests that the unemployed in the GSOEP sample are not disproportionally relinquishing their homes after the reform and that the divergence in ownership rates between the unemployed and employed documented above is driven by the inability of the unemployed to secure mortgages credit after the reform.

Table 4: Homeownership (GSOEP), conditional on being homeowner pre-reform

Homeownership	Pre-2005	Post-2005	Difference
Employed	1	0.963	
Unemployed	1	0.968	
Difference	0	0.005 (0.012)	0.005 (0.012)

Note: Standard errors (clustered at household-level) in parentheses. N=39,170
* $p < 0.05$, ** $p < 0.01$

It could still be the case, however, that the unemployed simply decide that they cannot afford a mortgage after the reform. That is consistent with the model, but not speaking to the role of lenders. Therefore, in the third step, we compare changes in homeownership among the employed only, comparing rich and poor employed respondents. The Hartz reforms made lower income groups more likely to default—because they are at higher risk of unemployment[27]—and we therefore expect homeownership rates among the employed poor and the employed rich to diverge after the reform. This is what the data show, with a statistically significant difference-in-

[27] To distinguish between low and high risks, we divide people by income. Although income is only one factor affecting default risks, those with lower incomes are expected to experience a higher increase in the risk of default after the reform for two reasons. First, they are at higher risk of unemployment – simply because income and unemployment risk are negatively correlated – and the lowering of long-term unemployment benefits makes them worse default risks. Second, private assets (k in our model) become more important when unemployment benefits are lower (lower b can be offset with higher k), and people with lower income generally have fewer private assets. Moreover, by reducing protected assets (other than home equity), the reform made people with lower savings higher default risks. Therefore, we expect that access to mortgage credit becomes more difficult for lower income households after the reform.

difference estimate of about -3.2 percentage points (Table 5). As for the unemployed, the effect is mostly due to a relative drop in homeownership among those poor who were not already owners before the reform, suggesting that they faced tighter access to credit after the reform.

Table 5: Homeownership (GSOEP), rich vs. poor employed

Homeownership	Pre-2005	Post-2005	Difference
Rich	0.577	0.583	0.00590 (0.009)
Poor	0.249	0.223	-0.0259** (0.008)
Difference	0.328** (0.014)	0.360** (0.017)	-0.0318** (0.012)

Note: Standard errors (clustered at household-level) in parentheses. N=55,798
* $p < 0.05$, ** $p < 0.01$

Overall, the patterns in both datasets (the EVS and the GSOEP) are consistent with the hypothesis that access to credit worsened for the unemployed as well as those at higher risk of unemployment, after the Hartz-reforms lowered unemployment benefits. Different data-sources and our three-pronged empirical approach support this conclusion, but since we do not have data on mortgage applications (and rejections), we cannot be certain that lending decisions drove the results. Future research will have to (dis)confirm that interpretation.[28]

[28] We would have liked to test the hypothesis that access to, and conditions of, mortgage credit vary as a function of income support generosity, using cross-national data. But because regulatory frameworks of financial markets vary greatly (even within the EU) and because there is very little data, we can only offer a very preliminary test. In appendix E, we show that the spread of interest rates (the coefficient of variation, the Gini coefficient, and p90/p10 ratios)—a measure for the inequality in access to credit—cross-nationally correlates, in the predicted direction, with

4 Conclusion

Financialization of advanced economies has made ability to access credit markets, and the terms of such access, increasingly important for individual welfare and economic inequality. Creditworthiness affects who can purchase a home and who can move between work and family and between work and further education, and it also affects the interest rate spread and therefore the dispersion of discretionary income. More plentiful information strengthens this relationship and empowers lenders to differentiate between high- and low-risk groups, raising interest rates for low-income groups or excluding them from credit markets altogether. The combination of financialization and Big Data is therefore a double whammy for the poor: like everyone else, they increasingly depend on borrowing to smooth income and acquire assets, but they are increasingly identified as bad risks and face worse access to, and terms of, borrowing.

Yet these inequalizing effects are strongly conditioned by the regulatory regime and by the welfare state. Where the state assumes some of the risks of lending—for example by acting as a backstop in mortgage markets—or where the social protection system is generous, the effects of financialization and Big Data are muted. Our evidence from the housing market backs up these claims. While our evidence is based on isolated cases, they suggest a consistent (possibly causal) story across very diverse contexts, consistent with more descriptive evidence, that lends credence to our model. We believe the model captures an important consequence of the massive shifts towards financialization and individualized data, which has been mostly neglected in the CPE literature and points well beyond credit markets to social insurance and the welfare state.

---

two measures of income replacement generosity; one measure of public subsidies for home ownership; and the home ownership rate.

An important question for future research is whether the improved capacity of markets to differentiate between risk groups will lead to a weakening of the regulatory regime and possibly also the welfare state. A major difficulty in keeping together progressive coalitions is that the underlying risk distribution is strongly right-skewed, which means that the median in the distribution—who is likely to be political influential—is someone who would benefit from greater differentiation. Is it too pessimistic to suppose that the rising importance of credit combined with better information about the shape of the distribution will lead to intensified calls for the state to step back?

References

Acharya, V. V., Richardson, M., Nieuwerburgh, S. van, & White, L. J. (2011). *Guaranteed to Fail: Fannie Mae, Freddie Mac, and the Debacle of Mortgage Finance*. Princeton University Press.

Ahlquist, J. S., & Ansell, B. W. (2017). Taking Credit: Redistribution and Borrowing in an Age of Economic Polarization. *World Politics*, *69*(4), 640–675. https://doi.org/10.1017/S0043887117000089

Ansell, B. (2014). The Political Economy of Ownership: Housing Markets and the Welfare State. *American Political Science Review*, *108*(2), 383–402. https://doi.org/10.1017/S0003055414000045

Arent, S., & Nagl, W. (2013). Unemployment Compensation and Wages: Evidence from the German Hartz Reforms. *Jahrbücher Für Nationalökonomie Und Statistik / Journal of Economics and Statistics*, *233*(4), 450–466.

Blackwell, T., & Kohl, S. (2018). The origins of national housing finance systems: A comparative investigation into historical variations in mortgage finance regimes. *Review of International Political Economy*, *25*(1), 49–74. https://doi.org/10.1080/09692290.2017.1403358

Blackwell, T., & Kohl, S. (2019). Historicizing housing typologies: Beyond welfare state regimes and varieties of residential capitalism. *Housing Studies*, *34*(2), 298–318. https://doi.org/10.1080/02673037.2018.1487037

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, *17*(2), 372–404.

DeFusco, A. A., & Paciorek, A. (2017). The Interest Rate Elasticity of Mortgage Demand: Evidence from Bunching at the Conforming Loan Limit. *American Economic Journal: Economic Policy*, *9*(1), 210–240. https://doi.org/10.1257/pol.20140108

Eichengreen, B. (2008). Origins and responses to the current crisis. *CESifo Forum*, *9*, 6–11.

Fernandez, R., & Aalbers, M. B. (2016). Financialization and housing: Between globalization and Varieties of Capitalism. *Competition & Change*, *20*(2), 71–88.

Fligstein, N., & Goldstein, A. (2015). The emergence of a finance culture in American households, 1989–2007. *Socio-Economic Review*, *13*(3), 575–601.

Foote, C., Loewenstein, L., & Willen, P. (2019). *Technological Innovation in Mortgage Underwriting and the Growth in Credit: 1985–2015* (Federal Reserve Bank of Boston Research Department Working Papers) [Federal Reserve Bank of Boston Research Department Working Papers]. Federal Reserve Bank of Boston. https://doi.org/10.29412/res.wp.2019.11

Froud, J., Johal, S., Montgomerie, J., & Williams, K. (2010). Escaping the tyranny of earned income? The failure of finance as social innovation. *New Political Economy*, *15*(1), 147–164.

Gelman, A., & Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456. https://doi.org/10.1080/07350015.2017.1366909

Godechot, O. (2020). Financialization and the Increase in Inequality. In P. Mader, D. Mertens, & N. van der Zwan (Eds.), *The Routledge International Handbook of Financialization* (pp. 413–424). Routledge. https://doi.org/10.4324/9781315142876-37

Goering, J., & Wienk, R. (2018). *Mortgage lending, racial discrimination and federal policy*. Routledge.

Goodman, L. S., Landy, B., Ashworth, R., & Yang, L. (2014). A Look at Freddie Mac's Loan-Level Credit Performance Data. *The Journal of Structured Finance*, *19*(4), 52–61. https://doi.org/10.3905/jsf.2014.19.4.052

Gross, D. B., & Souleles, N. S. (2002). Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data. *The Quarterly Journal of Economics*, *117*(1), 149–185. https://doi.org/10.1162/003355302753399472

Hacker, J. S. (2008). *The Great Risk Shift: The New Economic Insecurity and the Decline of the American Dream* (2nd edition). Oxford University Press.

Hacker, J. S., Rehm, P., & Schlesinger, M. (2013). The Insecure American: Economic Experiences and Policy Attitudes amid the Great Recession. *Perspectives on Politics*, *11*(1), 23–49. https://doi.org/10.1017/S1537592712003647

Hall, R. E. (1988). Intertemporal Substitution in Consumption. *Journal of Political Economy*, *96*(2), 339–357. https://doi.org/10.1086/261539

Hariri, J. G., Jensen, A. S., & Lassen, D. D. (2020). Middle Class Without a Net: Savings, Financial Fragility, and Preferences Over Social Insurance. *Comparative Political Studies*, *53*(6), 892–922. https://doi.org/10.1177/0010414019879718

Häusermann, S., Kurer, T., & Schwander, H. (2015). High-skilled outsiders? Labor market vulnerability, education and welfare state preferences. *Socio-Economic Review*, *13*(2), 235–258. https://doi.org/10.1093/ser/mwu026

Hausman, C., & Rapson, D. S. (2018). Regression Discontinuity in Time: Considerations for Empirical Applications. *Annual Review of Resource Economics*, *10*(1), 533–552. https://doi.org/10.1146/annurev-resource-121517-033306

Hurley, M., & Adebayo, J. (2017). Credit scoring in the era of big data. *Yale Journal of Law and Technology*, *18*(1), 5.

Iversen, T., & Soskice, D. (2019). *Democracy and Prosperity. Reinventing Capitalism through a Turbulent Century*. Princeton University Press.

Kaiser, T. (2018, November 28). Schonvermögen: Darf ein Hartz-IV-Empfänger reich sein? *Die Welt*. https://www.welt.de/wirtschaft/article184579080/Schonvermoegen-Darf-ein-Hartz-IV-Empfaenger-reich-sein.html

Lazarus, J. (2020). Financial Literacy Education: A Questionable Answer to the Financialization of Everyday Life. In P. Mader, D. Mertens, & N. van der Zwan (Eds.), *The Routledge International Handbook of Financialization* (pp. 390–399). Routledge. https://doi.org/10.4324/9781315142876-35

Liebig, S., Schupp, J., Goebel, J., Richter, D., Schröder, C., Bartels, C., Fedorets, A., Franken, A., Giesselmann, M., Grabka, M., Jacobsen, J., Kara, S., Krause, P., Kröger, H., Kroh, M., Metzing, M., Nebelin, J., Schacht, D., Schmelzer, P., … Deutsches Institut Für Wirtschaftsforschung (DIW Berlin). (2019). *Socio-Economic Panel (SOEP), data from 1984-2017 Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2017* (Version v34) [SAS,SPSS,Stata,SPSS,SAS]. SOEP Socio-Economic Panel Study. https://doi.org/10.5684/SOEP.V34

Palier, B. (2010). *A Long Goodbye to Bismarck? The Politics of Welfare Reforms in Continental Europe*. Amsterdam University Press.

Pei, Z., Lee, D. S., Card, D., & Weber, A. (2021). Local Polynomial Order in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*. https://doi.org/10.1080/07350015.2021.1920961

Rueda, D. (2007). *Social Democracy Inside Out: Partisanship and Labor Market Policy in Advanced Industrialized Democracies*. Oxford University Press.

Schwartz, H., & Seabrooke, L. (2008). Varieties of Residential Capitalism in the International Political Economy: Old Welfare States and the New Politics of Housing. *Comparative European Politics*, *6*(3), 237–261. https://doi.org/10.1057/cep.2008.10

VantageScore. (2020). *The Dynamic Relationship Between a Credit Score and Risk*. https://vantagescore.com/pdfs/Credit-Scores-and-Risk-Relationship-WP-FINAL_2020-10-29-021228.pdf

Wiedemann, A. (2021). *Indebted Societies: Credit and Welfare in Rich Democracies*. Cambridge University Press.

Wood, J. D. G. (2019). Mortgage Credit: Denmark's Financial Capacity Building Regime. *New Political Economy*, *24*(6), 833–850. https://doi.org/10.1080/13563467.2018.1545755

Online appendices

Appendix A: The model

We assume that the individual *i*'s time horizon is equal to the term of any loan, so that the interest rate on the loan is proportional to the total interest that has to be paid back (in addition to the principal). The loan amount is L_i and the interest rate is r_i, where the money is used to pay for housing, daycare and other services, or time off work for education and retraining that are part of an anticipated career trajectory. There is also a risk, p_i, of "catastrophic" loss of income and the non-loan private funds available for consumption in this case is k_i, which is income from selling assets or bringing forward long-term pension savings, etc. The (von Neumann–Morgenstern) expected utility of individual *i* with income Y_i is now defined as:

(A1) $$U_i = \left[\ln\left(Y_i - L_i \cdot (1+r_i)\right) + \alpha \cdot \ln\left(L_i\right)\right] \cdot (1-p_i) + \ln(k_i) \cdot p_i \,,$$

where α is the demand for credit; which we here assume to be common.

The model uses a log function to capture a standard concave utility function (u'>0 & u''<0) in a simple and tractable form. Note that if the catastrophic life event is triggered, we have assumed that the individual cannot afford to pay back the loan and will default. We will endogenize the default decision below.

From the perspective of the lender, we assume the competitive rate in a market with no default risk is $\bar{r}$. But in determining the interest rate for borrower *i* the lender adjusts for *i*'s risk of default. If the lender has full information about *i*'s risk type, and if there are many other borrowers with the same risk-profile, the lender will break even when:

(A2) $(1-p_i)\cdot(1+r_i)\cdot L_i - p_i \cdot L_i = (1+\overline{r})\cdot L_i$,[29]

which implies that:

(A3) $r_i = \dfrac{\overline{r}+2p_i}{1-p_i}$.

The more likely *i* is to default, the higher the interest rate charged to that individual.

The optimal loan requested by individual *i* is found by setting the first-order condition of (1) equal to 0, which yields:

(A4) $L_i^* = \frac{1}{2}\cdot\alpha_i\cdot\dfrac{Y_i}{1+r_i}$.

The lower the interest rate, the greater the demand for credit, another standard result.

Discretionary income, D_i, is:

(A5) $D_i = Y_i - L_i \cdot r_i$.

Inserting the optimal loan amount (A4) at the break-even interest rate (A3) we find that:

(A6)
$$\begin{aligned} D_i &= Y_i - \tfrac{1}{2}\cdot\alpha\cdot\frac{Y_i}{1+\dfrac{\overline{r}+2p_i}{1-p_i}}\cdot\frac{\overline{r}+2p_i}{1-p_i} \\ &= Y_i\cdot\left(1-\tfrac{1}{2}\cdot\alpha\cdot\frac{\overline{r}+2p_i}{1+\overline{r}+p_i}\right). \end{aligned}$$

[29] Strictly speaking this equation applies to groups of borrowers with the same risk-profile, not individuals. So we should use means for each group – 1, 2, 3, … N -- and use the subscripts *i*=1, *i*=2, *i*=3, …., *i*=N. If the equation literally referred to an individual *i* the lender could no longer behave in a risk-neutral manner, as we have assumed. But since the meaning is clear, we save the complication in notation.

If the lender has no information about risk type, it will have to set an average interest rate that is proportional in equilibrium to the amount of defaulted loans among all borrowers (which is always observed), so the break-even condition is now:

(A7) $\sum\left((1-p_i)\cdot L_i\right)\cdot(1+\bar{\bar{r}})-\sum p_i\cdot L_i=\sum(1+\bar{r})\cdot L_i=(1+\bar{r})\cdot\sum L_i$,

which implies that:

(A8) $\bar{\bar{r}}=\dfrac{(1+\bar{r})\cdot\bar{L}+\overline{pL}}{\overline{(1-p)\cdot L}}-1$,

where $\bar{\bar{r}}$ is the interest rate charged to any borrower.[30]

Lenders can learn about individual risks through credit history. With Bayesian updating the "observed" risk is a weighted function of a prior and the signal. If $p_l^o=[p_{\min}^o, p_{\max}^o]$ is observed risk of individual *i* by lender *l*, we can write:

(A9) $p_l^o=\iota\cdot p_i^s+(1-\iota)\cdot\bar{p}$,

where p_i^s is a noisy signal drawn from a distribution that is centered on the individual's true risk, p_i, and $\bar{p}$ is the mean among all borrowers, which is the prior. The parameter ι is a measure of the "precision" of the signal, which equals the information about *i* available to the lender. With no information ($\iota=0$) *i* only observes the population mean, $p_i^o=\bar{p}$, and the range is therefore

[30] Note that since the individual loan amount depends on income, if p_i is (negatively) related to income the average loan amount among those in who end up in the bad state is not the same as among those who stay in the good state. Hence $\sum p_i\cdot L_i\neq\bar{p}\cdot\bar{L}$.

zero. At the other extreme, with complete information, $p_i^o = p_i$, the range equals the difference between those with the lowest and highest risk.

In the next iteration of the model, "catastrophic" loss of income does not necessarily lead to default. Instead, we assume that if assets that be used in the bad state of the world are at or below a certain threshold, T_i the borrower will default; otherwise not:

(A10) $$\text{If } \begin{Bmatrix} k_i \le T_i \text{ then default} \\ k_i > T_i \text{ then do not default} \end{Bmatrix}.$$

If the lender cannot observe either risk of income loss, p_i, or individual thresholds, T_i we find the break-even common (average) interest rate to be:

(A11) $$\begin{aligned} &\sum(1-p_i)\cdot(1+r_i)\cdot L_i + \sum p_i \cdot p_{(k_i>T_i)}\cdot(1+r_i)\cdot L_i - \sum p_i \cdot p_{(k_i<T_i)}\cdot L_i = (1+\bar{r})\cdot\sum L_i \\ &(1+\bar{\bar{r}})\cdot\sum(1-p_i)\cdot L_i + (1+\bar{\bar{r}})\cdot\sum p_i\cdot p_{(k_i>T_i)}\cdot L_i - \sum p_i\cdot p_{(k_i<T_i)}\cdot L_i = (1+\bar{r})\cdot\sum L_i \\ &\bar{\bar{r}} = \frac{(1+\bar{r})\cdot\sum L_i + \sum p_i\cdot p_{(k_i<T_i)}\cdot L_i}{\sum(1-p_i)\cdot L_i + \sum p_i\cdot(1-p_{(k_i>T_i)})\cdot L_i} - 1 \\ &\bar{\bar{r}} = \frac{(1+\bar{r})\cdot\bar{L} + \overline{p\cdot p_{(k<T)}\cdot L}}{\overline{(1-p)\cdot L} + \overline{p\cdot(1-p_{(k<T)})\cdot L}} - 1 \end{aligned}$$

(The expected repayment on the LHS is the probability of being in the good state times the (certain) repayment (first term), plus the probability of being in the bad state and getting paid (second term) and not paid (third term)).

If the lender knows p_i and k_i, but not individual subjective thresholds for defaults, the break-even interest rate offered to each individual is:

$$(1-p_i)\cdot(1+r_i)\cdot L_i + p_i \cdot p_{(k_i>T)} \cdot (1+r_i)\cdot L_i - p_i \cdot p_{(k_i<T)} \cdot L_i = (1+\overline{r})\cdot L_i$$

$$(1+r_i)\cdot((1-p_i) + p_i \cdot p_{(k_i>T)}) = (1+\overline{r}) + p_i \cdot p_{(k_i<T)}$$

$$\text{(A12)} \quad r_i = \frac{1+\overline{r} + p_i \cdot p_{(k_i<T)}}{1-p_i + p_i \cdot p_{(k_i>T)}} - 1$$

$$r_i = \frac{\overline{r} + 2\cdot p_i \cdot p_{(k_i<T)}}{1-p_i \cdot p_{(k_i<T)}}$$

Note that if everyone defaults after falling into the bad state, the result converges to equation (A3) where higher risk exposure means a higher interest rate. But those who are likely to service their debt in the bad state will be rewarded with a lower interest rate, and that makes differences in k_i a source of inequality even in the good state.

When the state transfers income to those in the bad state, we assume that the benefit, b_i , is paid for by a flat-rate tax of all income earners (i.e., those in the good state). Specifically, the benefit is:

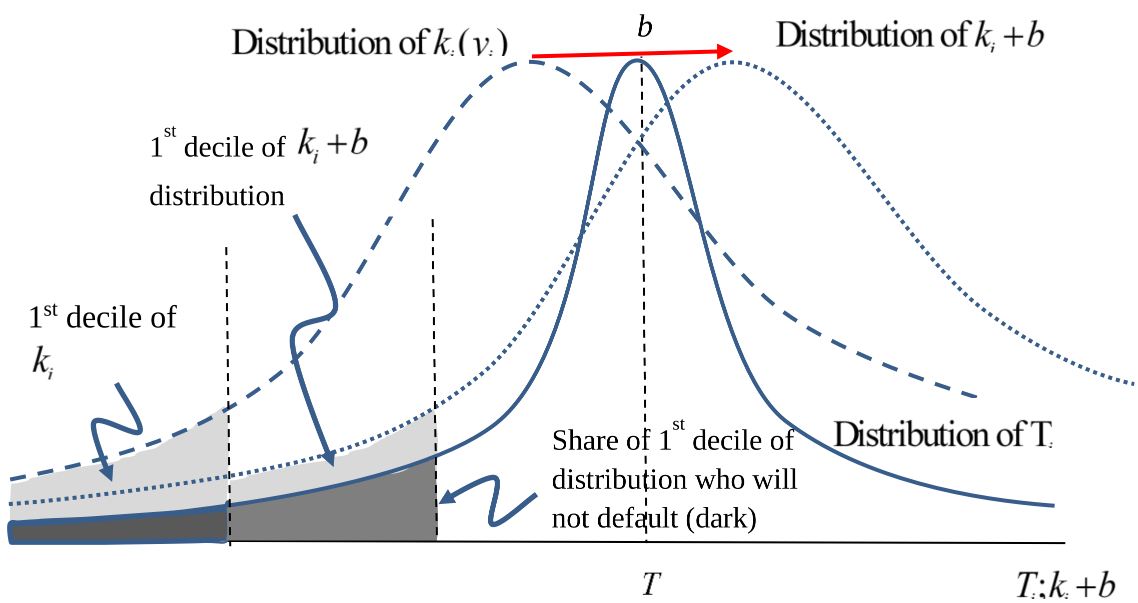
$$\text{(A13)} \quad b_i = b = \frac{t \cdot \sum Y_i}{n} = \frac{t \cdot \overline{y}}{n/N} = t \cdot \frac{\overline{y}}{\overline{p}} ,$$

where *n* is the number of people in that bad state, *N* is the total population, and $\overline{p}$ is the mean probability of falling into the bad state.

Appendix B: The effect of a flat-rate benefit on the distribution of default risks

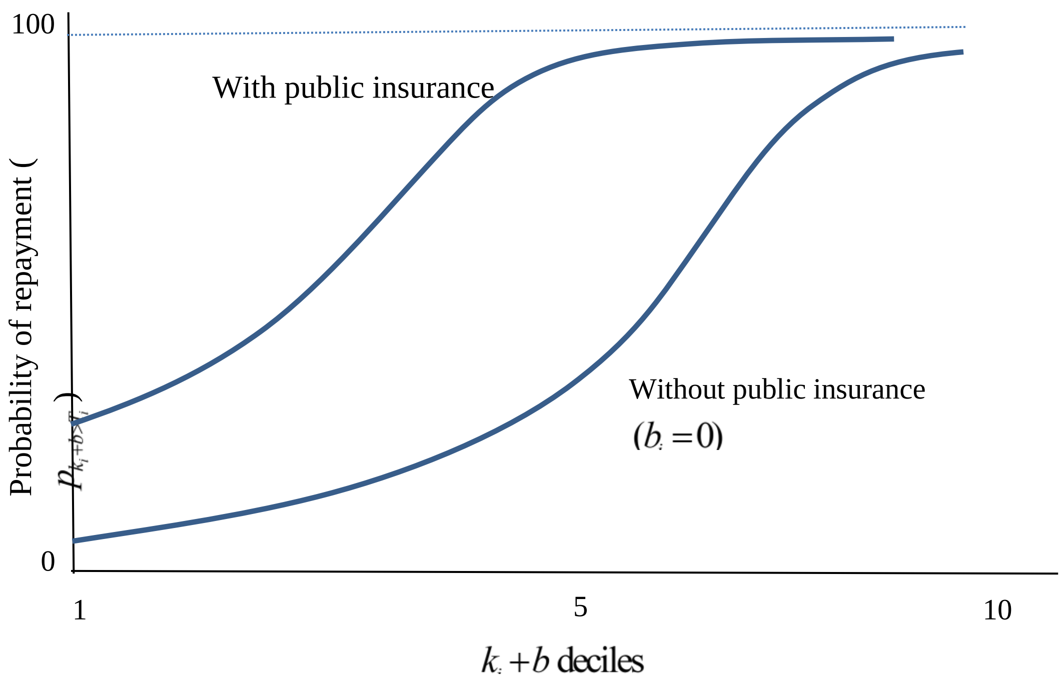
Figure B8 compares the entire distribution of income in the bad state, k_i, with the distribution of default thresholds, T_i. The k_i-distribution before government transfers is the dashed curve, while the distribution after government transfers is the dotted curve. The effect is to raise the income of everyone in the bad state by the amount b.

Figure B8: The effect of public spending on the relative location of the distribution of income in the bad state relative to T_i



As public transfers shift the income distribution ($k_i + b$) to the right, the probability of those in the bottom decile servicing their debt increases. In the example in Figure B8, the combined dark and very dark area under the T_i-distribution is the share of the bottom decile of the distribution who do not default (because they are above the threshold). In the example, this share is about 60 percent of those in the bottom decile for the $k_i + b$-distribution, compared with about 20 percent for the k_i-distribution. At the high end of the k-distribution, on the other hand, the effect of the subsidy is to only slightly decrease the default from about 10 percent to about 5 percent (because we are now at the “thin” tail of the T-distribution). Using this logic Figure B9 shows the relationship between income deciles and probability of not defaulting, contingent on whether the state redistributes resources to those in the bad state or not.

Figure B9: The relationship between income in bad state and probability of repayment



We can see that the variance of the distribution with public insurance is lower than the distribution without. In our example, if we measure the variance as d9/d1 ratios, it falls from about 4 (80 in the top decile; 20 in the bottom) to about 1.5 (90 in the top; 60 in the bottom).

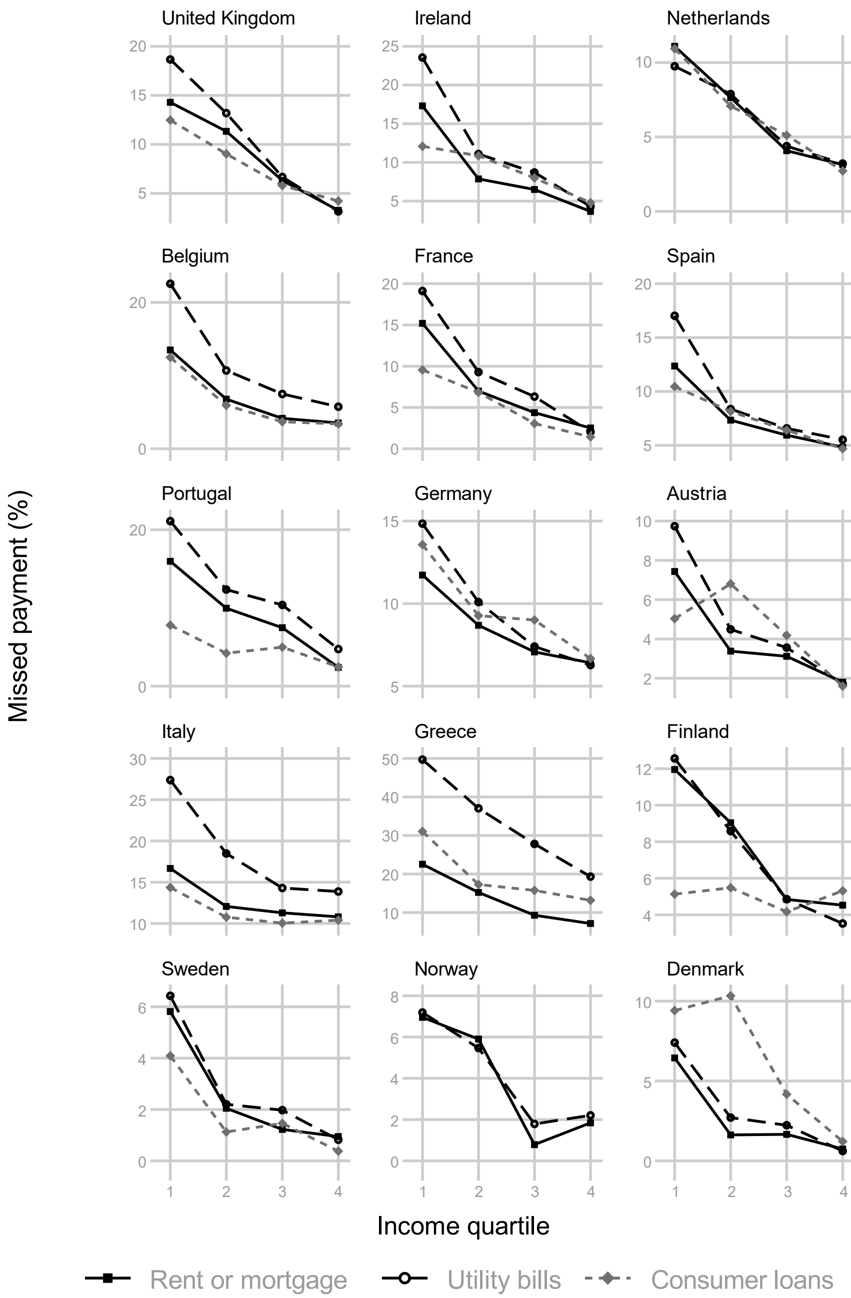
Appendix C: The relationship between income and default risk

The assumption that default risk is negatively related to incomes can be tested with various existing survey data. For example, the European Quality of Life Survey includes the following set of questions: "Has your household been in arrears at any time during the past 12 months, that is, unable to pay as scheduled any of the following? [Q60]":

- Rent or mortgage payments for accommodation [Q60a]
- Utility bills, such as electricity, water, gas [Q60b]
- Payments related to consumer loans, including credit card overdrafts (to buy electrical appliances, a car, furniture, etc.) [Q60c]

This allows us to explore the relationship between being behind in paying ones rent/mortgage, utility bills, and consumer loans, on the one hand, and income, on the other. As expected, there is generally a clear income gradient to being in arrears (see Figure C10).

Figure C10: The relationship between income and being in arrears



Note: Values are averaged across all available years. Y-axis varies by country.
Source: European Quality of Life Survey Integrated Data File, 2003–2016 (http://doi.org/10.5255/UKDA-SN-7348-3, last accessed June 3, 2021 [https://perma.cc/2JLS-CAHT]). Averages across all available years.

Appendix D: Regression discontinuity estimates

This appendix presents a variety of robustness checks. In particular:

- Table D6: Different bandwidth selection procedures
- Table D7: Different kernel functions
- Table D8: Covariate adjusted estimates
- Table D9: Different specification of the running variable (months, quarters, trimesters, half-years, years), using the year-month-FICO2d-level data
- Table D10: Different specification of the running variable (months, quarters, trimesters, half-years, years) with the outcome variable measured at the respective unit level (months, quarters, trimesters, half-years, years)
- Table D11: Sensitivity to observations near the cutoff (donut hole approach)
- Table D12: Placebo outcomes
- Table D13: Placebo cutoffs
- Table D14: Masspoints

Table D6: Different bandwidth selection procedures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
BW type	mserd	msetwo	msesum	msecomb1	msecomb2	cerrd	certwo	cersum	cercomb1	cercomb2
RD estimate	1.003^{***}	1.233^{***}	0.762^{***}	1.003^{***}	1.003^{***}	1.146^{***}	1.274^{***}	0.888^{***}	1.146^{***}	1.146^{***}
	(0.0976)	(0.104)	(0.0887)	(0.0976)	(0.0976)	(0.108)	(0.0994)	(0.0908)	(0.108)	(0.108)
Robust 95% CI	[.901 ; 1.283]	[1.113 ; 1.556]	[.685 ; 1.012]	[.901 ; 1.283]	[.901 ; 1.283]	[.984 ; 1.428]	[1.133 ; 1.548]	[.776 ; 1.117]	[.984 ; 1.428]	[.984 ; 1.428]
Kernel	Triangular									
Poly. (p)	1									
Bias (q)	2									
N	5025	5025	5025	5025	5025	5025	5025	5025	5025	5025
N (l)	2147	2147	2147	2147	2147	2147	2147	2147	2147	2147
N (r)	2878	2878	2878	2878	2878	2878	2878	2878	2878	2878
Eff. N (l)	340	320	460	340	340	280	260	380	280	280
Eff. N (r)	360	220	480	360	360	300	200	400	300	300
BW est. (l)	17.15	16.33	23.48	17.15	17.15	14.26	13.58	19.53	14.26	14.26
BW est. (r)	17.15	10.85	23.48	17.15	17.15	14.26	9.03	19.53	14.26	14.26
BW bias (l)	28.13	16.67	29.80	28.13	28.13	28.13	16.67	29.80	28.13	28.13
BW bias (r)	28.13	20.21	29.80	28.13	28.13	28.13	20.21	29.80	28.13	28.13

Note: Standard errors in parentheses (clustered at the FICO-2d level). Estimates adjusted for mass points in the running variable. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

- mserd: one common MSE-optimal bandwidth selector for the RD treatment effect estimator.
- msetwo: two different MSE-optimal bandwidth selectors (below and above the cutoff) for the RD treatment effect estimator.
- msesum: one common MSE-optimal bandwidth selector for the sum of regression estimates (as opposed to difference thereof).
- msecomb1: for min(mserd,msesum).
- msecomb2: for median(msetwo,mserd,msesum), for each side of the cutoff separately.
- cerrd: one common CER-optimal bandwidth selector for the RD treatment effect estimator.
- certwo: two different CER-optimal bandwidth selectors (below and above the cutoff) for the RD treatment effect estimator.
- cersum: one common CER-optimal bandwidth selector for the sum of regression estimates (as opposed to difference thereof).
- cercomb1: for min(cerrd,cersum).
- cercomb2: for median(certwo,cerrd,cersum), for each side of the cutoff separately.

Table D7: Different kernel functions

	(1)	(2)	(3)
Kernel function	Triangular	Epanechnikov	Uniform
RD estimate	1.003***	1.245***	1.290***
	(0.0976)	(0.100)	(0.0945)
Robust 95% CI	[.901 ; 1.283]	[1.148 ; 1.572]	[1.195 ; 1.6]
BW type	mserd	mserd	mserd
Kernel	Triangular	Epanechnikov	Uniform
Order Loc. Poly. (p)	1	1	1
Order bias (q)	2	2	2
N	5025	5025	5025
N (l)	2147	2147	2147
N (r)	2878	2878	2878
Eff. N (l)	340	160	120
Eff. N (r)	360	180	140
BW est. (l)	17.15	8.93	6.91
BW est. (r)	17.15	8.93	6.91
BW bias (l)	28.13	18.54	16.65
BW bias (r)	28.13	18.54	16.65

Note: Standard errors in parentheses (clustered at the FICO-2d level)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimates adjusted for mass points in the running variable.

Table D8: Covariate adjusted estimates

	(1)	(2)	(3)	(4)	(5)
Covariate(s)	None	Month dummies	Lagged DV	Average interest rate	Month dummies + lagged DV + average interest rate
RD estimate	1.003^{***} (0.0976)	1.510^{***} (0.107)	0.962^{***} (0.0844)	1.313^{***} (0.0998)	0.656^{***} (0.0349)
Robust 95% CI	[.901 ; 1.283]	[1.442 ; 1.888]	[.809 ; 1.154]	[1.249 ; 1.664]	[.618 ; .824]
BW type	mserd	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular
Order Loc. Poly. (p)	1	1	1	1	1
Order bias (q)	2	2	2	2	2
N	5025	5025	5001	5025	5001
N (l)	2147	2147	2125	2147	2125
N (r)	2878	2878	2876	2878	2876
Eff. N (l)	340	220	120	240	280
Eff. N (r)	360	240	140	260	300
BW est. (l)	17.15	11.98	7.00	12.82	14.01
BW est. (r)	17.15	11.98	7.00	12.82	14.01
BW bias (l)	28.13	23.36	15.08	23.99	26.68
BW bias (r)	28.13	23.36	15.08	23.99	26.68

Note: Standard errors in parentheses (clustered at the FICO-2d level)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimates adjusted for mass points in the running variable.
Covariate-adjusted estimates. Additional covariates included: 0 (model 1), 11 (model 2), 1 (models 3 and 4), and 13 (model 5).

Table D9: Different specification of the running variable (months, quarters, trimesters, half-years, years), using the year-month-FICO2d-level data

	(1)	(2)	(3)	(4)	(5)
Unit of running variable	Month	Quarter	Trimester	Biannual	Annual
RD estimate	1.003^{***} (0.0976)	1.228^{***} (0.101)	1.178^{***} (0.103)	0.933^{***} (0.0918)	0.704^{***} (0.104)
Robust 95% CI	[.901 ; 1.283]	[1.062 ; 1.496]	[1.046 ; 1.471]	[.92 ; 1.293]	[.334 ; .753]
BW type	mserd	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular
Order Loc. Poly. (p)	1	1	1	1	1
Order bias (q)	2	2	2	2	2
N	5025	5025	5025	5025	5025
N (l)	2147	2147	2147	2147	2147
N (r)	2878	2878	2878	2878	2878
Eff. N (l)	340	180	160	240	960
Eff. N (r)	360	240	240	360	1198
BW est. (l)	17.15	3.16	2.83	2.72	4.14
BW est. (r)	17.15	3.16	2.83	2.72	4.14
BW bias (l)	28.13	4.24	3.68	4.13	4.42
BW bias (r)	28.13	4.24	3.68	4.13	4.42

Note: Standard errors in parentheses (clustered at the FICO-2d level)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimates adjusted for mass points in the running variable.
The data is at the month-year-FICO-2d level (as in the previous analyses), but the running variable is recoded to quarters, trimesters, half-years, and years, respectively. Note that this changes the interpretation of the bandwidth estimates.

Table D10: Different specification of the running variable (months, quarters, trimesters, half-years, years), with the outcome variable measured at the respective unit level (months, quarters, trimesters, half-years, years)

	(1)	(2)	(3)	(4)	(4)
Unit of measurement of outcome & running variable	Month	Quarter	Trimester	Biannual	Annual
RD estimate	1.003*** (0.0976)	2.128*** (0.0504)	1.401*** (0.0689)	0.514*** (0.0707)	0.658*** (0.0843)
Robust 95% CI	[.901 ; 1.283]	[2.212 ; 2.551]	[1.309 ; 1.618]	[.273 ; .535]	[.205 ; .54]
BW type	mserd	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular
Order Loc. Poly. (p)	1	1	1	1	1
Order bias (q)	2	2	2	2	2
N	5025	1680	1260	840	420
N (l)	2147	720	540	360	180
N (r)	2878	960	720	480	240
Eff. N (l)	340	40	40	80	40
Eff. N (r)	360	60	60	100	60
BW est. (l)	17.15	2.09	2.05	4.10	2.56
BW est. (r)	17.15	2.09	2.05	4.10	2.56
BW bias (l)	28.13	4.44	4.89	4.21	3.76
BW bias (r)	28.13	4.44	4.89	4.21	3.76

Note: Standard errors in parentheses (clustered at the FICO-2d level, but not in model 3 due to non-convergence)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimates adjusted for mass points in the running variable.
The dataset is constructed at the year-month, year-quarter, year-trimester, year-biannual, and year-level (all at FICO-2d). Note that this changes the overall number of observations (N).

Table D11: Sensitivity to observations near the cutoff (donut hole approach)

	(1)	(2)	(3)	(4)
Donut hole (months)	0	+/-1 = 2 months	+/-2 = 4 months	+/-3 = 6 months
RD estimate	1.003^{***} (0.0976)	0.895^{***} (0.0897)	1.020^{***} (0.0485)	1.208^{***} (0.101)
Robust 95% CI	[.901 ; 1.283]	[.833 ; 1.16]	[1.158 ; 1.387]	[1.29 ; 1.882]
BW type	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Triangular	Triangular
Order Loc. Poly. (p)	1	1	1	1
Order bias (q)	2	2	2	2
N	5025	5005	4965	4925
N (l)	2147	2147	2127	2107
N (r)	2878	2858	2838	2818
Eff. N (l)	340	400	140	100
Eff. N (r)	360	400	140	100
BW est. (l)	17.15	20.12	8.73	7.98
BW est. (r)	17.15	20.12	8.73	7.98
BW bias (l)	28.13	29.79	17.34	17.43
BW bias (r)	28.13	29.79	17.34	17.43

Note: Standard errors in parentheses (clustered at the FICO-2d level)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimates adjusted for mass points in the running variable.
Excluded are 0 (model 1), 1 (model 2), 2 (model 3), or 3 (model 4) months to the left and right of the cutoff.

Table D12: Placebo outcomes

	(1)	(2)	(3)	(4)
Outcome	Gini of interest rates	Loan-to-value	Avg. interest rate	Avg. FICO score
RD estimate	1.003***	-1.344	-0.144**	-0.348
	(0.0976)	(0.895)	(0.0514)	(18.74)
Robust 95% CI	[.901 ; 1.283]	[-2.819 ; .573]	[-.161 ; .051]	[-37.038 ; 36.343]
BW type	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Triangular	Triangular
Order Loc. Poly. (p)	1	1	1	1
Order bias (q)	2	2	2	2
N	5025	5025	5025	5025
N (l)	2147	2147	2147	2147
N (r)	2878	2878	2878	2878
Eff. N (l)	340	760	40	2147
Eff. N (r)	360	779	60	2858
BW est. (l)	17.15	38.45	2.91	143.00
BW est. (r)	17.15	38.45	2.91	143.00
BW bias (l)	28.13	24.96	5.17	31.27
BW bias (r)	28.13	24.96	5.17	31.27

Note: Standard errors in parentheses (clustered at the FICO-2d level)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimates adjusted for mass points in the running variable.
Columns (2) to (4) are placebo outcomes. Note that the robust 95% CI in model (3) includes 0.

Table D13: Placebo cutoffs

	(1)	(2)	(3)	(4)	(5)
Cutoff	1-2008	6-2007	1-2007	6-2006	1-2006
RD estimate	1.003*** (0.0976)	0.0419 (0.0372)	0.181*** (0.0499)	-0.0970*** (0.0217)	0.114*** (0.0290)
Robust 95% CI	[.901 ; 1.283]	[.029 ; .217]	[.012 ; .219]	[-.123 ; -.018]	[.049 ; .167]
BW type	mserd	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular
Order Loc. Poly. (p)	1	1	1	1	1
Order bias (q)	2	2	2	2	2
N	5025	2147	2147	2147	2147
N (l)	2147	2027	1907	1787	1667
N (r)	2878	120	240	360	480
Eff. N (l)	340	80	100	120	220
Eff. N (r)	360	100	120	140	240
BW est. (l)	17.15	4.98	5.41	6.55	11.19
BW est. (r)	17.15	4.98	5.41	6.55	11.19
BW bias (l)	28.13	10.26	10.86	6.81	9.19
BW bias (r)	28.13	10.26	10.86	6.81	9.19

Note: Standard errors in parentheses (clustered at the FICO-2d level)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimates adjusted for mass points in the running variable.
The placebo cut-offs are every 6 months prior to the discontinuity (1-2008), up 1-2006. Only observations to the left of the discontinuity are part of the sample, to avoid contamination. Note the small size of the coefficients in models (2) to (5).

Table D14: Masspoints

	(1)	(2)
Masspoints	Adjusted	Ignored
RD estimate	1.003***	1.228***
	(0.0976)	(0.102)
Robust 95% CI	[.901 ; 1.283]	[1.103 ; 1.525]
BW type	mserd	mserd
Kernel	Triangular	Triangular
Order Loc. Poly. (p)	1	1
Order bias (q)	2	2
N	5025	5025
N (l)	2147	2147
N (r)	2878	2878
Eff. N (l)	340	180
Eff. N (r)	360	200
BW est. (l)	17.15	9.56
BW est. (r)	17.15	9.56
BW bias (l)	28.13	22.32
BW bias (r)	28.13	22.32

Note: Standard errors in parentheses (clustered at the FICO-2d level)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Model 1: Estimates adjusted for mass points in the running variable.
Model 2: Estimates not adjusted for mass points in the running variable.

Appendix E: Cross-country correlation matrix

Table F15: Correlation matrix

		(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
(1)	CoV of interest rates [a)]	1					
(2)	Gini of interest rates	0.982*	1				
(3)	P90/P10 interest rates	0.861*	0.868*	1			
(4)	Income replacement [b)] (% of GDP)	-0.453*	-0.400	-0.529*	1		
(5)	Income replacement [b)] (PPS per capita)	-0.516*	-0.458*	-0.584*	0.792*	1	
(6)	Homeownership subsidies [c)]	-0.535*	-0.573*	-0.516*	0.640*	0.494	1
(7)	Homeownership rate [a)]	0.462*	0.430*	0.401*	-0.665*	-0.683*	-0.681*

Note: N=17 for all variables (BEL, CYP, DEU, ESP, EST, FRA, GRC, HUN, IRL, ITA, LUX, LVA, MLT, NLD, PRT, SVK, SVN, but N=12 for variable (6) (No data for BEL, GRC, ITA, SVK, SVN). * $p < 0.1$.
Variables are averaged across available years since 1994. Variable (6) refers to around 2015, and variable (7) to around 2010–2014.
Sources:
a) Interest rate and homeownership data are from the Household Finance and Consumption Survey (HFCS), waves 1 and 2. https://www.ecb.europa.eu/pub/economic-research/research-networks/html/researcher_hfcn.en.html (last accessed June 3, 2021 [https://perma.cc/TA4R-XWVZ]).
b) Income replacement = social protection expenditures for sickness/healthcare + disability + survivor + unemployment. Source: http://ec.europa.eu/eurostat/cache/metadata/en/spr_esms.htm (last accessed June 3, 2021 [https://perma.cc/6B3S-TB8C]) / http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=spr_exp_sum (last accessed June 3, 2021 [https://perma.cc/YRW9-JKHT]).
c) Average of three measures of public support for homeownership: (i) Public spending on grants and financial support to home buyers [PH2.1]; (ii) Forgone tax revenue due to tax relief for access to homeownership [PH2.2]; (iii) Spending on housing allowances by type of housing-related costs covered [PH3.1]. Source: http://www.oecd.org/social/affordable-housing-database.htm (last accessed June 3, 2021 [https://perma.cc/YQA3-RFXS]). Some imputation.