NBER WORKING PAPER SERIES

RELATIONSHIP LENDING AND THE GREAT DEPRESSION

Jon Cohen
Kinda Cheryl Hachem
Gary Richardson

Working Paper 22891
http://www.nber.org/papers/w22891

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2016, Revised July 2018

We thank Joseph Johnson, Padma Sharma, and Daniel Tracht for excellent research assistance. We also thank Ben Bernanke, Gabe Chodorow-Reich, Jonathan Dingel, Joe Haubrich, Erik Hurst, Chris Koch, Fabrizio Perri, and Johannes Wieland as well as seminar and conference participants at Chicago Booth, UW-Madison, FRB Cleveland, NBER SI 2016 Monetary Economics, and the 2018 Texas Monetary Conference for helpful comments. Kinda Hachem gratefully acknowledges financial support from Chicago Booth. Part of this research was conducted while Gary Richardson was at the Federal Reserve Bank of Richmond. The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Richmond, the Federal Reserve System, or the National Bureau of Economic Research.

Relationship Lending and the Great Depression
Jon Cohen, Kinda Cheryl Hachem, and Gary Richardson
NBER Working Paper No. 22891
December 2016, Revised July 2018
JEL No. E44,G01,G21,L14,N22

ABSTRACT

The collapse of long-term lending relationships amplified the Great Depression and perturbed recovery. We demonstrate this by developing a new measure of lending relationships that can be calculated from widely available data at any level of aggregation. Our approach exploits differences in the responsiveness of loan rates to bank funding costs and is supported by historical evidence and theoretical arguments. The new measure reveals that bank suspensions' marginal impact on economic activity in the early 1930s was higher in more relationship-intensive areas. Furthermore, the rebuilding of destroyed relationships helps explain regional variation in economic performance during the 1937-38 recession.

Jon Cohen
Department of Economics
University of Toronto
Max Gluskin House
150 St. George Street, 322
Toronto, Ontario M5S 3G7
Canada
jon.cohen@utoronto.ca

Kinda Cheryl Hachem
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
kinda.hachem@chicagobooth.edu

Gary Richardson
Department of Economics
University of California, Irvine
3155 Social Sciences Plaza
Irvine, CA 92697-5100
and NBER
garyr@uci.edu

1 Introduction

The Great Depression was the longest and deepest downturn in U.S. history. Scholars have long debated how the collapse of commercial banking in the early 1930s contributed to the contraction of commerce and industry. The pioneering work of Bernanke (1983) has since spawned a large literature that seeks to estimate the size and significance of this credit channel. The assumption that underlies almost all of these papers is that bank suspensions had real effects because they destroyed – or at the very least prevented the immediate redeployment of – some critical input into the bank lending process that impacts the cost of credit. Absent this assumption, it is difficult to explain why bank suspensions would trigger such a deep and protracted decline in economic activity. But what exactly was this critical input and why was it so difficult to redeploy? Despite the abundant scholarship on the Great Depression, data limitations have impeded attempts to answer this question.

In this paper, we demonstrate that the critical input impaired by bank suspensions was the soft information that banks had acquired about the quality of their borrowers during multi-period lending relationships. This information cannot be directly observed with data surviving from the Great Depression. The type of microeconomic data on bank-firm interactions used in modern analyses of relationship lending are also unavailable for the 1920s and 1930s. We overcome this lacuna in the evidence by developing a new measure of continuing lending relationships that can be constructed from data aggregated at any level or frequency, enabling comparison to the time-series and panel data typically analyzed by macroeconomists. We then use our measure to show that disruptions to continuing relationships as a result of bank suspensions in the early 1930s contributed to both the severity of the Great Depression and the uneven pace of recovery later in the decade.

Our measure is based on the idea that loan rates charged in continuing relationships are less responsive to changes in bank funding costs than are those charged on other loans. We start by comparing average interest rates on two types of bank loans: one that historical narratives describe as relational (commercial loans) and one that they describe as transactional (brokers' loans). We find that the average interest rate on the relational loans was less responsive to changes in bank funding costs than was the average interest rate on the transactional loans. We present a model based on Hachem (2011) to help understand why this was the case. We choose this model because its assumptions are consistent with the institutional details of the period we study, namely the short-term and uncollateralized nature of commercial loans, the potential sequencing of these loans as part of longer-term lending relationships, and the competitive nature of bank lending.

At the heart of the model are the asymmetric information problems inherent in financial intermediation. The first is adverse selection, which exists because borrowers have private information about their ability to operate certain projects. The second is moral hazard, where higher loan rates increase a borrower's incentive to undertake projects with high default risk. Moral hazard generates a threshold loan rate above which borrowers shift their unobservable efforts towards riskier projects. Heterogeneous ability implies that this threshold depends on the borrower's type. A continuing relationship lender is able to use information gleaned about his borrower's type during previous interactions to determine the threshold and decide whether or not to pass an increase in bank funding costs through to the borrower. New relationship lenders and/or transactional lenders cannot condition this decision on the borrower's type because they lack information about it.

The model predicts that the incentive compatible contract offered by a continuing relationship lender eliminates risk-shifting by reducing the pass-through from bank funding costs to loan rates for some borrowers. From this, we derive that relational loan rates are on average less elastic with respect to bank funding costs than are transactional loan rates. The difference is driven by continuing rather than new relationships. Moreover, the elasticity of the weighted average of all loan rates in a region is decreasing in the degree to which that region's banks are engaged in continuing relationships. Our new measure of continuing lending relationships is based on this last prediction.

We use data from the consolidated balance sheets and income statements of national banks in the 1920s to implement our measure for the eve of the Great Depression. These data – reported by the Comptroller of the Currency at a semi-annual frequency for Federal Reserve districts, states, and reserve cities – allow us to infer the weighted average loan rate at each point in time in each of 82 locations. We then calculate the elasticity of the loan rate with respect to the discount rate for each location. Since the twelve Federal Reserve Banks at the time had latitude to operate largely independent discount windows, the marginal cost of funds for banks varied across districts (Richardson and Troost (2009)). To control for location-specific differences in rates of return and their responsiveness, we also calculate the elasticity of securities returns with respect to the discount rate in each location and net it out from the loan rate elasticity. The remaining cross-sectional differences in loan rate elasticity indicate differences in the nature and intensity of lending relationships on the eve of the Depression. Regions with relatively less elastic loan rates are those where the theory predicts bank portfolios contained a higher proportion of information-intensive continuing relationship loans. We demonstrate the veracity of our measure by showing its correlation to other measures that historical sources have argued should be proxies for relationship lending, such as bank size and location and the structure of local industry. That being said, the economic content of our measure exceeds these proxies: our measure is a statistically significant predictor of the real effects of bank suspensions even when the aforementioned proxies are included as controls.

Our first set of results using our new measure establishes the importance of relationship lending for understanding the real effects of banking distress in the early 1930s. We estimate that bank suspensions alone explain one-eighth of the national economic contraction observed during the Great Depression. Equally important, we are able to show that the marginal impact of bank suspensions on economic activity was much greater in areas with more continuing relationships, other things the same. In other words, the destruction of continuing relationships amplified the severity of the Depression. Although Calomiris and Mason (2003), Richardson and Troost (2009), Ziebarth (2013), and Carlson and Rose (2015) all find evidence that the collapse of commercial banks adversely affected the availability of credit and thus the activities of firms during this period, none, due to data limitations, manages to isolate the role of relationship lending in propagating the real effects of banking sector distress.[1]

Our second set of results focuses on the recovery. When business is dependent on the soft information that banks acquire over time, it is not enough for bank suspensions to cease and new banks to open. These banks also need time to interact with firms, learn about them, and rebuild continuing relationships. With this in mind, we show that cross-sectional differences in the rebuilding of lending relationships destroyed by the bank suspensions of the early 1930s contributed to cross-sectional differences in economic performance during the 1937-38 recession. Specifically, areas that were rebuilding the types of continuing relationships which existed in the 1920s fared better during the 1937-38 recession than otherwise similar areas that failed to rebuild. In contrast, areas that were rebuilding weaker relationships fared worse. The effects of the Great Depression thus appear to have persisted long after the end of the crisis phase and, because of the time required to rebuild destroyed relationships, the banking panics of the early 1930s are also likely to have contributed to what some believe was a surprisingly muted recovery in the wake of such a catastrophic collapse.[2]

In keeping with Bernanke (1983) and others, we focus on the Great Depression. No other crisis comes close in terms of the dimensions of the downturn and the largely unfettered nature of the banking panics. However, our findings are of interest beyond the historical context. Relationship lending continues to be a principal source of working capital in many parts of the world, facilitating the allocation of scarce financial resources to the full range of credit-worthy borrowers. The soft information on which these relationships depend is intrinsically difficult to transfer from one bank to another. Our results suggest that, all else

[1] In contrast to the mainstream view, Cole and Ohanian (1999, 2004, 2007) argue that the collapse of the U.S. banking system in the early 1930s had few real economic effects, citing weak (unconditional) correlation between bank failures and economic outcomes, even at the state level. On this point, our results suggest that the correlation becomes much more compelling when conditioned on cross-sectional differences in the nature and intensity of lending relationships.

[2] Rajan and Ramcharan (2016) also observe lingering effects of the Depression, although, unlike us, they focus on banking concentration post-WWII, not on economic activity in the late 1930s and the role of relationship lending.

constant, the flow of funds is more likely to be disrupted by failures of lenders who have accumulated knowledge of their borrowers through continuing relationships than by failures of those in the early stages of new relationships or of those who extend credit transactionally. This is a valuable lesson for policy-makers as they debate the scope for bank bailouts and the parameters of government safety nets. Our method is also useful for analysis of modern economies when loan-level data are either unavailable or available only with long lags.

The rest of the paper proceeds as follows. Section 2 provides historical background on the lending activities of Depression-era commercial banks to motivate our new measure of continuing lending relationships. Section 3 presents theoretical foundations for this measure and Section 4 describes its empirical implementation for the 1920s. Sections 5 and 6 present the key empirical results, with Section 5 using cross-sectional differences in continuing relationships in the 1920s to pinpoint the real effects of banking distress in the early 1930s and Section 6 demonstrating how relationship rebuilding in the aftermath of the Depression affected economic performance during the 1937-38 recession. Section 7 concludes.

2 Historical Background

We start by reviewing the lending activities of Depression-era commercial banks. The annual report of the Comptroller of the Currency provides information on U.S. bank balance sheets in the 1920s and 1930s in a format that is consistent across geographic locations. Up until 1928, the Comptroller divided bank loans into four categories: (i) real estate loans, (ii) loans on financial securities, (iii) uncollateralized loans, and (iv) loans collateralized by personal security. Based on the Comptroller's description of personal security, the last category consists primarily of loans secured by difficult-to-evaluate collateral such as goods in the process of production and distribution, warehouse receipts, and, in the case of farm loans, future crops. The sum of (iii) and (iv) constitutes commercial lending and is the largest category overall, accounting for an average of 67% of loans by national banks during the 1920s. On the eve of the Great Depression, almost 50% of national bank loans were uncollateralized, with another 10% collateralized by personal security.

Category (ii), which consisted largely of loans to brokers, was the second largest category, accounting for an average of 28% of national bank loans during the 1920s. Real estate loans were the smallest category, accounting for only 5% of lending by national banks over the same period. Brokers' loans were short-term (e.g., call loans or 90-day time loans) and facilitated the purchase of stocks and bonds. Historical sources are unequivocal about the lack of relationship lending in brokers' loans. These loans were "usually made on an impersonal basis with the borrower and lender dealing through agents" and "not connected with established customer relationships" (Board of Governors 1943, p. 425). Thomas (1935) further notes that in the U.S. the agents made little inquiry into the borrower's credit-worthiness, in contrast to practices in London. Brokers' loans in the U.S. were thus purely transactional, with no decisions made on the basis of soft information gathered in previous periods. The transactional nature of brokers' loans is also emphasized by Currie (1931).

In contrast, commercial loans, which provided working capital to merchants and manufacturers, were a prime venue for relationship building (e.g., Miller (1927), Foulke and Prochnow (1939)). The maturity of bank loans to commercial clients typically ranged from one to six months and never exceeded one year. To be eligible for use as collateral at a Federal Reserve Bank's discount window, commercial loans had to mature in 90 days or less, unless collateralized by agricultural commodities, in which case they had to mature in 180 days or less. As noted earlier, commercial loans were either uncollateralized or collateralized by goods in the process of production and distribution. Such goods were difficult to value, costly to repossess, and if liquidated, could be sold only with a long delay and/or at a considerable loss. For all intents and purposes, then, commercial loans were an unsecured form of lending. As a result, these loans tended to be relationship-based, in that firms borrowed repeatedly from the same bank, often also obtaining deposit, payment, and other intermediation services from this bank. This approach allowed banks to incorporate information acquired during the initial loan period into future credit terms, including the possibility of discontinuing the lending relationship (Langston (1921)).

The commercial lending activities of Depression-era banks thus involved the production and use of information. Such information is "soft" because it is acquired over time through repeated personal interactions between banks and their clients. Hard information instead involves easily obtained facts about borrowers that can be discovered upfront. By and large, businesses most dependent on banks for commercial loans were small with little or no public record. While these firms knew their intrinsic quality, banks did not and could only learn it over time by cultivating a repeated relationship. A survey from the Department of Commerce in 1929 determined that 86% of small manufacturing firms depended on banks for working capital (Department of Commerce 1935, p. 65-66). In response to a survey by the National Industrial Conference Board in 1932, 72% of retailers also reported regularly borrowing working capital from banks (National Industrial Conference Board 1932, p. 62). The Census of Retail Trade in 1929 indicates that half of all retail establishments sold $12,000 or less of goods annually (Census 1930, p. 15).

Within the set of relationship loans, it is then important to separate loans made as part of a continuing relationship from first-time relationship loans. Soft information on borrower quality is accumulated over time through repeated interactions and is hence embodied in continuing relationships, not new ones. Bank failures, such as those that ravaged the U.S. financial system in the early 1930s, destroy this knowledge, disrupt the flow of funds to borrowers and, because it takes time to re-establish these relationships, also impact the pace of recovery. We therefore need to isolate continuing relationships in order to accurately estimate the real effects of financial crises.

The available data on loan quantities in the 1920s and 1930s do not distinguish between stages of a commercial lending relationship (i.e., the Comptroller's report does not divide commercial loans in this way). To overcome this limitation, we propose to do with price what we cannot do with quantity. Specifically, we argue that interest income can be used to construct a measure that distinguishes not only between relationship and transactional loans but also between new and continuing relationships.

We first provide some motivating evidence on the information content of interest rates using the distinction between relationship and transactional loans. In the decades before the Great Depression, the Federal Reserve tracked interest rates on several types of bank loans in New York City. Figure 1 displays monthly average interest rates on new 90-day brokers' loans (red dots) and loans to commercial clients (tan dots). The black line represents the discount rate of the Federal Reserve Bank of New York.[3] From the foundation of the Fed through the 1930s, the discount window of each Federal Reserve Bank was operated in a manner that made the discount rate the marginal cost of funding a loan in that district.[4] Figure 1 shows that interest rates on brokers' loans rose swiftly and substantially when the discount rate rose. Interest rates on commercial loans moved in the same direction, but slowly and slightly. Even the Federal Reserve noted that "because of a variety of considerations arising out of customer relationships," rates on commercial loans behaved differently than rates on brokers' loans, which were arm's length and impersonal (Board of Governors 1943, p. 426). The Federal Reserve did not elaborate on these considerations (we will do so in Section 3) but Figure 1 clearly shows that commercial loan rates and interest rates on brokers' loans differed in their responsiveness to bank funding costs despite having similar maturities.

Appendix A provides a more formal treatment of the data in Figure 1 and reaches the same conclusion. To provide additional support, Appendix A also considers the responsive-

[3] Discount rate changes on or before the 15th day of a month are assigned to that month. Rate changes after the 15th day of a month are assigned to the next month.

[4] The Fed expected member banks to discount commercial loans to accommodate seasonal and cyclical peaks in demands for credit, one of the principle motivations for the creation of the Federal Reserve System. The Fed also needed banks to discount loans because they were its primary source of earned income. From the early 1920s through the early 1930s, the Fed was not allowed to hold government bonds, corporate securities, derivatives, or equities so, to cover its operating costs, the Fed encouraged member banks to use the discount window. Banks wanting to expand their commercial loan book thus knew they could always raise the necessary funds by discounting those loans at their district Fed, making the discount rate the relevant cost of funds for commercial loans. In principle, loans on financial securities were not discountable at the Fed. In practice, however, the Fed expressed concern about banks using discount loans to invest in financial securities, suggesting that the discount rate was also a relevant cost of funds for brokers' loans and security holdings by banks more generally.

ness of commercial paper rates to changes in the discount rate. Like commercial loans from banks, commercial paper financed goods in the process of production and sale, in contrast to brokers' loans which financed the purchase of securities. For the period we consider, the Fed kept track of interest rates on commercial paper maturing in four to six months in New York City. This maturity is similar to that of commercial loans. However, unlike the recipients of commercial loans from banks, borrowers in the commercial paper market were large, well-known firms with good credit ratings and a public history of repayment. There was little information left to be acquired about them and hence no need for learning through relationship lending. Investors buying commercial paper and banks making brokers' loans thus shared the feature that neither used soft information acquired as part of a lending relationship. We find in Appendix A that commercial paper rates were more responsive than commercial loan rates to changes in the discount rate. This confirms that the responsiveness of interest rates to changes in bank funding costs in Figure 1 is determined by the nature of the loan (i.e., relational versus transactional), not by the purpose of the loan (i.e., financing working capital versus buying securities).

It follows from the above discussion that the responsiveness of interest rates can be used to distinguish relationship loans from transactional ones. Why do such differences in responsiveness exist? Whatever model one uses to think about this, it has to be consistent with the institutional features of the 1920s and 1930s, namely that bank credit was extended on a short-term basis and lending relationships were effectively uncollateralized.[5] It must also incorporate the fact that the banking sector in the U.S., both nationally and locally, was much less concentrated and therefore much more competitive than it is today. In the next section, we present a model that embraces these institutional features. The model makes the following key predictions. First, the average interest rate on relationship loans will be less responsive to changes in bank funding costs than the average interest rate on

[5] Prior to the 1930s, real estate loans also had short maturities, commonly 3 to 5 years. While refinancing was obviously necessary for most real estate loans because of their short maturities, mortgages were often refinanced with different lenders, mitigating the extent to which these could have been relationship-based. Moreover, as discussed above, real estate loans constituted only a small fraction of national banks' lending.

transactional loans, consistent with the evidence in Figure 1 and Appendix A. Second, within the set of relationship loans, interest rates on loans made as part of a continuing relationship will be less responsive to changes in bank funding costs than the rates on loans extended to borrowers at the beginning of a relationship. Third, the responsiveness of the weighted average of all loan rates in a region to bank funding costs will be decreasing in the degree to which that region's banks are engaged in continuing relationships. As will become clear, this last prediction is important for our empirical analysis later on because it permits us to use geographical areas as the unit of observation to estimate how bank suspensions impacted the real economy through the destruction of continuing relationships.

3 New Indicator of Continuing Relationships: Theory

By definition, a continuing relationship involves prior interactions between the same lender and borrower pair. This is a widely used measure of relationship strength in modern analyses of relationship lending.[6] However, it requires transaction-level data that are not available for a representative sample of banks and firms in the 1920s and 1930s. We must therefore find another way to detect the presence of continuing lending relationships.[7]

We propose a measure that exploits changes in loan pricing over the course of a relationship. Our measure is based on responsiveness over time, not on levels at a given point in time. This distinction is important. Petersen and Rajan (1994) argue that lending relationships have little effect on the price of credit. However, their result is about levels in a single cross-section, not about changes over time. We are instead looking at how relationships affect the responsiveness of loan rates to changes in bank funding costs over time. Based on the evidence in Section 2, the effects that we are interested in are potentially quite large.[8]

[6] For example, Berger and Udell (1995), Ongena and Smith (2001), Elsas (2005), Chodorow-Reich (2014).

[7] In other contexts where data are a constraint, the use of historical narratives has proved fruitful. See Romer and Romer (1989, 2004, 2015). While it may be possible to gain some insight into lending practices through the narrative approach, we lack a source which provides narratives of local conditions in the Depression-era U.S. that is consistent across both time and space.

[8] Recent analyses of the transmission of the Lehman shock to loan rates in Italy also suggest that rela-

To help interpret the evidence and provide insight into the Federal Reserve's statement that "considerations arising out of customer relationships" led interest rates on commercial loans to behave differently than those on brokers' loans, we sketch a simple model of relationship lending based on Hachem (2011). We choose this model because its assumptions are consistent with the historical features of national bank lending described above: the short-term and uncollateralized nature of commercial loans, the potential sequencing of these loans as part of longer-term lending relationships, and the competitive nature of bank lending. At the heart of the model are the asymmetric information problems inherent in financial intermediation. There is moral hazard in that higher loan rates increase a borrower's incentive to undertake projects with high default risk. There is also adverse selection in that each borrower has private information about his ability to operate certain projects. The threshold loan rate above which a borrower engages in risk-shifting thus depends on his type. A continuing relationship lender is able to use information gleaned about the borrower's type during previous interactions to decide whether or not to pass through an increase in bank funding costs. New relationship lenders and/or transactional lenders cannot condition this decision on the borrower's type because they lack information about it. The model environment and equilibrium loan rates are discussed in Sections 3.1 and 3.2. We then derive a prediction about loan rate elasticities (Section 3.3) that can be used in empirical work.

3.1 Environment

There are two stages and three periods. All agents are risk neutral. There is a continuum of firm types, denoted by ω and distributed uniformly over the unit interval. Types are private information. In each stage, firm ω can undertake a production project that generates output θ_1 with probability $p(\omega)$, where $p'(\omega) > 0$. For simplicity, we can consider $p(\omega)$ linear. The project fails with probability $1 - p(\omega)$, in which case zero output is generated. Output is independently distributed across firms and stages.

tionships affect the pricing of credit. See Gambacorta and Mistrulli (2014), Gobbi and Sette (2015), and Bolton et al (2016) for examples.

The firm's project requires one unit of capital input each time it is operated. Capital is available to a mass of ex ante identical banks at an exogenous policy rate r. Firms are not endowed with capital, nor can they store capital or output across stages. The credit contracts that transfer capital from banks to firms are uncollateralized and mature at the end of the stage in which they are signed.

In the first stage, banks are perfectly competitive and offer the same zero-profit interest rate R_1^*. The interest rate decision is subject to a risk-shifting problem. In particular, instead of putting capital into production, the firm can unobservably invest in a speculative project which produces output $\theta_2 > \theta_1$ with probability $q < p(0)$ and nothing otherwise. Assume $q\theta_2 = p(0)\theta_1$ to reduce notation. Banks can detect the presence of output so firms with positive output repay their contracts. However, banks cannot detect the exact value of output so credit contracts cannot be made contingent on realized output.

The loan rate that makes firm ω indifferent between the two projects is:

$$\overline{R}(\omega) = \frac{p(\omega)\theta_1 - q\theta_2}{p(\omega) - q}$$

It is straightforward to show $\overline{R}'(\omega) > 0$. Firm ω undertakes the production project in the first stage if and only if the interest rate in the first stage does not exceed $\overline{R}(\omega)$. All banks know that this is the best response of a type ω firm. Firms select banks randomly in the first stage then decide which projects to undertake. All agents play the first-stage game in the first period.

At the beginning of the second period, banks and firms face an exogenous separation probability s. Separated firms become new firms, drawing new types from the uniform distribution and playing the first-stage game again. Firms that are not separated continue to the second stage and have their types discovered by their first-stage bank ("insider"). The insider learns this information by virtue of having interacted with the firm during the first stage. Other banks ("outsiders") can only observe that this firm is not a new firm.[9] Each

[9] Allowing outsiders to also observe whether or not the firm repaid its insider in the first stage does not,

insider then decides whether to continue the lending relationship and extend another unit of capital to the firm. If the insider wants to keep the firm, the interest rate cannot exceed what outsiders would charge, otherwise the firm will move to one of these outsiders. All banks have equilibrium beliefs so outsiders know that they are being adversely selected.

Interest rate choices in the second stage are still subject to the risk-shifting problem, meaning that firm ω undertakes the production project in the second stage if and only if the interest rate it is charged in the second stage does not exceed $\overline{R}(\omega)$. Outsiders are perfectly competitive and offer the same zero-profit interest rate to firms that are endogenously separated from their insiders. This rate differs from the zero-profit interest rate R_1^* offered to new (i.e., exogenously separated) firms because banks understand the adverse selection problem. Insiders have an informational advantage over outsiders so the interest rate offered by an insider in the second stage needs not generate zero profits for the insider. Instead, the insider seeks to maximize his profit, subject to the risk-shifting problem and the ability of the firm to move to an outsider.

The game ends at the end of the second stage. To produce the same first-stage interest rate R_1^* in the first and second periods, there is a third period where agents exogenously separated at the beginning of the second period play the second-stage game with probability $1-s$, after having played the first-stage game during the second period.

3.2 Equilibrium Loan Rates

Figure 2 depicts the equilibrium division of firm types between insiders and outsiders in the second stage, along with the interest rates they are charged. We refer the reader to Hachem (2011) for a detailed proof and only describe here the results we will build on in Section 3.3.

The equilibrium for a given policy rate r can be described with reference to two cutoff types. The first, $\widetilde{\omega}$, is implicitly defined by $p(\widetilde{\omega})\overline{R}(\widetilde{\omega}) \equiv r$ and represents the type on which the insider expects to break even by charging the reservation rate $\overline{R}(\cdot)$. The second,

in equilibrium, result in outsiders conditioning on credit history (Hachem (2011)). Accordingly, we do not introduce notation for credit history here.

$\widehat{\omega}$, is implicitly defined by $q\overline{R}(\widehat{\omega}) \equiv r$ and represents the lowest type that will choose the production project if charged $\frac{r}{q}$. Types below $\widetilde{\omega}$ move to outsiders and are charged a pooled interest rate $\frac{r}{q}$, prompting them to undertake the speculative project since $\widetilde{\omega} < \widehat{\omega}$. Types above $\widetilde{\omega}$ stay with their insiders and are charged interest rates that lead them to choose the production project. Notice that types between $\widetilde{\omega}$ and $\widehat{\omega}$ are getting policy-invariant interest rates (i.e., $\overline{R}(\cdot)$ does not depend on r). If these types were instead charged $\frac{r}{q}$, they would undertake the speculative project. The relationship lender is therefore using his information to mitigate the risk-shifting problem, incentivizing higher repayment rates by not passing through increases in r to some of the borrowers he retains.

For a given r, we can integrate over the type space in Figure 2 to get the average interest rate charged in the second stage:

$$R_2^* = \int_0^{\widetilde{\omega}} \frac{r}{q} d\omega + \int_{\widetilde{\omega}}^{\widehat{\omega}} \overline{R}(\omega)\, d\omega + \int_{\widehat{\omega}}^{1} \frac{r}{q} d\omega$$

To get the first-stage interest rate R_1^*, define $R_1^* \equiv \overline{R}(\xi)$ so that ξ denotes the firm type that is exactly indifferent between the production project and the speculative project when charged R_1^*. The zero-profit condition for lenders with new firms is then:

$$\underbrace{\left[q\xi + \int_{\xi}^{1} p(x)\, dx\right] \overline{R}(\xi) - r}_{\text{expected profit from first stage}} + \underbrace{\beta(1-s)\left[\int_{\widetilde{\omega}}^{\widehat{\omega}} \left[p(x)\,\overline{R}(x) - r\right] dx + \int_{\widehat{\omega}}^{1} \left(p(x)\frac{r}{q} - r\right) dx\right]}_{\text{expected present discounted continuation value}} = 0$$

where $\beta \in (0,1)$ is the lender's discount factor. The lender's continuation value comes from playing the second-stage game with the firm and earning profit on types above $\widetilde{\omega}$.

The model so far has assumed that all banks have the potential to become continuing relationship lenders in the second stage. As an alternative, we can imagine that fraction φ of the economy follows the model just described while fraction $1 - \varphi$ is characterized by transactional lending. We define a transactional loan to be a one-period contract where the firm and the bank exogenously separate with probability 1 at the end of the period.

A transactional loan is therefore similar to a first-stage loan, except that the transactional lender has a continuation value of zero from the firm. Formally, we can write the interest rate on transactional loans as $R^*_{TL} \equiv \overline{R}(\eta)$, where the zero-profit condition for the lender implies:

$$\left[q\eta + \int_{\eta}^{1} p(x)\,dx\right]\overline{R}(\eta) - r = 0$$

Comparing to the products discussed in Section 2, the transactional loans modeled here are more like brokers' loans than commercial paper in that information is not acquired even though an information asymmetry exists between the borrower and lender. In the commercial paper market, information was not acquired because there was no such asymmetry. We choose to model transactional loans more like brokers' loans because (i) these were actual loans made by banks whereas commercial paper was debt traded on the open market and (ii) any transactional lending outside of brokers' loans would have involved an information asymmetry given the small and private nature of bank-dependent borrowers. We refer the reader to Hachem (2011) for a discussion of interest rates when there is no asymmetric information.

3.3 Elasticity Prediction

The following proposition shows that the average interest rate on second-stage relationship loans is less elastic with respect to the policy rate than either the interest rate on first-stage relationship loans or the interest rate on transactional loans:

Proposition 1 *Let* $e_i \equiv \frac{dR_i^*}{dr}\frac{r}{R_i^*}$ *denote the elasticity of the interest rate* R_i^* *with respect to the policy rate* r*. The elasticities satisfy* $e_2 < \min\{e_1, e_{TL}\}$*.*

Proof. *See Appendix B.* ■

Consider now the weighted average loan rate in the economy during the second period, which is the main period in the model:

$$\mathcal{R} \equiv \varphi \left[sR_1^* + (1-s) R_2^* \right] + (1-\varphi) R_{TL}^*$$

Assuming the exogenous separation rate s and the transactional fraction $1-\varphi$ do not vary in a first-order way with the policy rate r, we can derive:

$$\frac{d\mathcal{R}}{dr}\frac{r}{\mathcal{R}} \approx \frac{\varphi s R_1^*}{\mathcal{R}} e_1 + \frac{\varphi (1-s) R_2^*}{\mathcal{R}} e_2 + \frac{(1-\varphi) R_{TL}^*}{\mathcal{R}} e_{TL} \tag{1}$$

where $\frac{d\mathcal{R}}{dr}\frac{r}{\mathcal{R}}$ is the elasticity of $\mathcal{R}$ with respect to the policy rate, $\frac{\varphi s R_1^*}{\mathcal{R}}$ is the fraction of interest income that comes from first-stage relationship loans, $\frac{\varphi(1-s)R_2^*}{\mathcal{R}}$ is the fraction from second-stage relationship loans, and $\frac{(1-\varphi)R_{TL}^*}{\mathcal{R}}$ is the fraction from transactional lending.

Taken together with Proposition 1, equation (1) implies that the weighted average loan rate in the economy will be less elastic with respect to the policy rate when banks are more heavily engaged in continuing relationships, as measured by a higher fraction of interest income coming from second-stage loans. For our purposes, this means that areas with less elastic loan rates are areas where continuing relationships are more substantial. It is this crucial insight that we exploit next in Section 4.

4 Empirical Implementation for the 1920s

We have demonstrated that the elasticity of the weighted average loan rate with respect to bank funding costs will be lower in an area where continuing relationships are more substantial. There is little micro data on how Depression-era banks adjusted loan rates charged to repeat customers. A survey by Ford (1928) of country banking practices in Northern Texas is the only study we found on the issue. Interestingly, Ford did find that relationship lenders often maintained a constant loan rate for borrowers they wanted to keep, even in times when the discount rate was changing.

Section 4.1 explains how we translate our theoretical prediction into an empirical indicator of continuing relationships using the data available for the 1920s and 1930s in the U.S., namely the balance sheets and income statements of national banks aggregated by geographic region. Data sources are described in more detail in Section 4.2 and summary statistics for our computed indicator are presented in Section 4.3.

4.1 Methodology

Let β_i^ℓ denote the elasticity of loan returns with respect to the discount rate during the pre-Depression era in location i. As discussed in Section 2, the discount rate is the relevant bank funding cost for the period we consider. For each i, we estimate β_i^ℓ by running the regression:

$$\log\left(ReturnOnLoans_{i,t}\right) = \alpha_i^\ell + \beta_i^\ell \log\left(DiscountRate_{i,t}\right) + \varepsilon_{i,t}^\ell$$

where t denotes time, specifically six-month periods from 1923 to 1929 inclusive as this is the frequency for which commercial bank data are available in the 1920s. The dependent variable, $ReturnOnLoans_{i,t}$, is calculated by dividing the interest earnings of banks on loans in location i during date t by the stock of bank loans in location i at the end of date t.

The intercepts α_i^ℓ in the above regressions control for any heterogeneity across locations that was constant over time. To also control against the possibility that there was time-varying heterogeneity unrelated to relationship lending in the 1920s, we will consider the estimate of β_i^ℓ relative to the elasticity of other returns in location i. The identifying assumption is that time-varying heterogeneity unrelated to relationship lending affects the return elasticity of all interest-earning assets in a given location in roughly the same way. To this end, we estimate the elasticity of securities returns with respect to the discount rate in each location i during the pre-Depression era by running the regression:

$$\log\left(ReturnOnSecurities_{i,t}\right) = \alpha_i^s + \beta_i^s \log\left(DiscountRate_{i,t}\right) + \varepsilon_{i,t}^s$$

The variable $ReturnOnSecurities_{i,t}$ is calculated by dividing the interest earned by banks on securities in location i during date t by the securities holdings of banks in location i at the end of date t. The sample period for the estimation of β_i^s is the same as for β_i^ℓ.

We then define the net elasticity for location i in the 1920s as:

$$NE20_i \equiv \beta_i^s - \beta_i^\ell$$

If $NE20_i > NE20_j$ for two locations i and j, it means that loan rates respond less in location i than in location j, relative to interest rates on other financial products, when the discount rate changes. We thus take $NE20_i > NE20_j$ to indicate the presence of more continuing relationships in location i than in location j on the eve of the Great Depression. When running regressions later on, we will also present specifications where β_i^ℓ and β_i^s are included as separate regressors rather than restricting their coefficients to sum to zero by including only $NE20_i$.

4.2 Data Sources

Historical data on commercial bank assets come from the annual reports of the Comptroller of the Currency. These reports tabulate the balance sheets of commercial banks aggregated by Federal Reserve district, state, and major municipalities (principally financial centers then termed reserve cities) for June and December of each year. The Comptroller's balance sheet data are reported separately for banks with national charters and banks with state charters. In contrast, income statement data, described in more detail below, are only tabulated for nationally chartered banks. Since the income statements are necessary to impute loan rates, we limit our sample to national banks.

The earning assets of national banks include loans, government bonds, and other financial securities. Until 1928, national bank loans were classified according to the four categories discussed in Section 2. Starting in 1929, the Comptroller reduced the number of categories to three, combining uncollateralized loans and loans collateralized by personal security into one category called "all other loans" and refining somewhat the cutoffs between the three resulting categories in line with the Fed's 1929 direct action campaign. Despite these refinements, the "all other loans" category remains a valid source of information on the sum of unsecured loans and loans on personal collateral. First, for the U.S. as a whole, the fraction of "all other loans" in June 1929 is roughly of the same magnitude as the fraction of loans secured by either no collateral or personal collateral in June 1928. Second, the cross-sectional correlation between the fraction of "all other loans" in June 1929 and the fraction of loans secured by either no collateral or personal collateral in June 1928 is 0.75, large enough to suggest that the two definitions are reasonably similar.

The Comptroller's annual reports also published earnings and expense tables for nationally chartered banks aggregated at the district, state, and city level. One table starts in 1919 and contains data for the months of January through June for each year. Another table starts in 1925 and contains data for the months of July through December for each year. These income statements are a valuable input into the construction of our price-based indicator. To the best of our knowledge, the literature has made little use of them.

Starting in 1926, the Comptroller reported earnings on loans separately from earnings on government bonds and other securities. Prior to 1926, when the Comptroller did not separate earnings by asset class, we can construct estimates. We first multiply the stock of securities held on bank balance sheets by market yields to estimate earnings on publicly-traded securities. Market yields appear in *Banking and Monetary Statistics, 1914 to 1941.* This tome recapitulates information previously published in annual and monthly reports of the Federal Reserve Board and Fed District Banks as well as trade publications such as the *Commercial and Financial Chronicle* and the *Wall Street Journal.* We also estimate interest earned from balances at other banks and interest earned on Fed securities. We then subtract these estimates from the total interest income reported in the Comptroller's report to get an estimate of loan income. This procedure delivers reasonable predictions when applied to 1926 to 1929, a period where we can compare against actuals.[10]

[10] For loans, the correlation between predicted returns and actual returns is 0.95. For securities, it is 0.83.

Dividing income by the stock of assets at the end of the period, we obtain the average loan returns and average securities returns that we need to run the elasticity regressions explained in Section 4.1. Federal Reserve discount rates are then obtained from *Banking and Monetary Statistics, 1914 to 1941.* As the discount rate in each Fed district often changed over the course of the six-month-periods that make up the time units in our panels, we calculate the average time-weighted rate in effect for each period.[11]

The geographic disaggregation in the Comptroller's report allows us to construct complete time series on assets and income for 82 locations: 33 reserve cities, 31 states (net of reserve cities) each fully contained within a single Fed district, 12 parts of states (net of reserve cities) each fully contained within a single Fed district, and 6 district remainders. We focus on locations that are fully contained in a single district because discount rates varied across Fed districts and we need to use the appropriate discount rate when running our elasticity regressions. The issue arises because some states are split between two districts (e.g., only part of Pennsylvania is in the Philadelphia Fed district; the rest is in the Cleveland Fed district). For a district that has only one split state, we isolate the part of this state contained in the district by subtracting from district-level data the relevant city-level data as well as state-level data for states fully contained in the district. We can then subtract the isolated part of this state from the state's total to ascertain the part of the state contained in another district. As long as this other district does not have more than two split states, we can repeat the process to back out any additional splits in the other district. For six districts, where there are simply too many split states to be fully identified by this iterative procedure, we define district remainders.

4.3 Net Elasticity Estimates

In this section, we present our net elasticity estimates and report their correlation with observables. Our goal is to show that these estimates actually do capture variations in the

[11] The weighted rate is a summation over the product of the interest rate and the number of days it was in effect, divided by the total number of days in the period.

nature and intensity of relationship lending and do not, instead, represent spurious patterns in the data or other potentially confounding factors.

Figure 3 plots the distributions of β_i^ℓ, β_i^s, and $NE20_i$, as estimated from the data. Overall, the distribution of the net elasticities is roughly normal with mean slightly greater than zero. Since we are interested primarily in the relative ranking of locations (e.g., whether $NE20_i$ is above or below $NE20_j$, as explained in Section 4.1), we do not infer much from the mean of the distribution. If anything, there may be a downward bias in our calculations because of the approximation used to separate interest income in the Comptroller's report into income from loans and income from securities pre-1926. As discussed in Section 4.2, the approximation uses market yields to estimate the amount of interest that banks earned on the various components of their securities portfolios. It is an approximation because market yields are averages over different issues (e.g., the market yield on municipal bonds is an average over multiple municipalities). Accordingly, these yields are likely smoother than the yields in any individual location, understating fluctuations in securities income and overstating, by subtraction, fluctuations in loan income.[12] As we have no reason to believe that the precision of our approximations varies systematically across locations, particularly in the early 1920s when institutional investing was less advanced, the ordinal ranking of net elasticities remains informative. The main takeaway from Figure 3 is that there is significant dispersion in net elasticity across locations, which is exactly what we need for our analysis.

In Figure 4, we plot the net elasticity, $NE20_i$, against the average fraction of national bank loans made on little to no collateral in location i during the 1920s. Recall from Section 2 that observers in the 1920s and 1930s viewed uncollateralized loans as the ones most closely associated with relationship building. We would therefore expect an area with more continuing relationships, as measured by its net elasticity, to also have more uncollateralized lending. This positive correlation shows up in Figure 4.

[12] If we were to dispense with the approximation and calculate net elasticities using only actuals from 1926 to 1937 (stopping in 1929 would be too short), we would find that all net elasticities are positive. However, because we want to separate the 1920s from the 1930s, we cannot use only actuals from 1926 to 1937.

Table 1 reports additional cross-sectional correlations between $NE20_i$ and a variety of economic indicators from the 1920 population census. We also report correlations with banking variables in 1920. We can see from Table 1 that net elasticity tended to be higher in areas that were more rural and/or had more small banks, characteristics that arise frequently in historical anecdotes about relationship lenders (e.g., Ford (1928)). This also makes sense theoretically. In contrast to big banks in urban centers, small rural banks were not relied upon to be liquidity providers to other financial institutions in emergencies or on short notice, making them less likely to have to suddenly sever relationships with non-financial borrowers for reasons unrelated to the borrower's health. This can be mapped into Section 3 as a lower probability of exogenous separation s in the middle of a potential relationship. More relationships with policy-invariant loan rates would then be fostered in these areas, suggesting that we should indeed observe a higher net elasticity. We will control for the characteristics in Table 1 in later regressions to isolate the role of continuing relationships over and above bank size and location and the structure of local industry.

As derived in Section 3, our price-based approach to measuring continuing relationships rests on the idea that differences in loan rates are driven by differences in lending practices, not by differences in funding practices. As a final confirmation, then, we would like to show that the patterns we find in loan rates are not driven by patterns in deposit rates. This is relevant to consider as Neumark and Sharpe (1992) have argued that deposit rates are slower to rise in concentrated markets. One may therefore wonder whether loan rates are stickier in some areas because market concentration makes deposit rates in those areas stickier, not because there are more continuing relationships. We already have some evidence that refutes this: Table 1 shows that locations with more concentrated banking markets, as measured by a lower number of banks per capita, had lower net elasticities and hence loan rates that were less, not more, sticky. However, we can also test the relevance of deposit rates more formally by running Granger causality tests for each location in our sample. Figure 5 plots the results. The median p-value for the null hypothesis that deposit rates did not cause loan rates in the 1920s is 0.11. In other words, for more than half of the locations in our sample, we accept the hypothesis that loan rates were not Granger-caused by deposit rates. For locations where we reject this hypothesis, we also tend to reject the opposite hypothesis that deposit rates were not Granger-caused by loan rates. It can also be seen from Figure 5 that there is a mildly positive cross-sectional correlation between the p-values of the Granger tests and our net elasticity estimates. This suggests that loan rates were less likely to be Granger-caused by deposit rates in precisely those locations where our price-based indicator finds the most evidence of continuing lending relationships.

5 The Effects of Banking Distress in the Early 1930s

We now use our new net elasticity measure to show that cross-sectional differences in continuing lending relationships during the 1920s played a fundamental role in determining the real effects of banking distress in the early 1930s.

5.1 Empirical Specification

We measure economic outcomes using retail sales. Unlike construction contracts and business failures, which are respectively leading and lagging indicators of economic activity, retail sales are generally viewed as a contemporaneous indicator. In addition, few other measures of economic activity are readily available at the same frequency and level of disaggregation. We obtain retail sales for each location in our sample by aggregating the appropriate counties in the Census of Business. This census is available for the years 1929, 1933, 1935, and 1939. A subsample of the firms from the 1935 census was also surveyed in 1937. The 1937 survey reports what the 1935 results would have been had they been based on the same subsample, making it possible to scale up the 1937 results and thus approximate total retail sales for 1937. Although we focus primarily on the 1933 survey in this section, other years are considered later in the paper.

Our first step is to define an indicator of banking distress that provides an accurate representation of its nature in the early 1930s. An example will elucidate what needs to be done. Suppose bank suspensions in location i amount to 10% of deposits. This could have been caused by the suspension of one bank with a 10% market share or by the suspension of ten banks each with a 1% market share. Although the size of the banking shock is the same in both cases – 10% of deposits – the suspension of many small banks more closely approximates the nature of the banking panics documented by Friedman and Schwartz (1963) and Wicker (1996) for the early 1930s. Accordingly, two separate indicators are needed to describe accurately distress during this period: one that captures the dispersion of distress across banks and another that controls for the size of the shock.

We measure the dispersion of banking distress in location i by the fraction of banks suspended in the early 1930s:

$$SuspNum32_i = \sum_{t=1930}^{1932} \frac{number\_of\_suspended\_national\_banks_{i,t}}{number\_of\_national\_banks_{i,1929}}$$

The size of the shock is then measured using the share of deposits in suspended banks:

$$SuspVal32_i = \sum_{t=1930}^{1932} \frac{deposits\_in\_suspended\_national\_banks_{i,t}}{deposits\_in\_national\_banks_{i,1929}}$$

We will use $SuspNum32_i$ as our main indicator of banking distress in each location i, with $SuspVal32_i$ as a control.

Define $Sales_{i,t}$ as retail sales in location i in year t. Using the 82 locations discussed at the end of Section 4.2, we estimate:

$$\begin{aligned}\frac{Sales_{i,1933}}{Sales_{i,1929}} &= \gamma_0 + \gamma_1 NE20_i + \gamma_2 SuspNum32_i + \gamma_3 NE20_i * SuspNum32_i \\ &\quad + \gamma_4 SuspVal32_i + \gamma_5 NE20_i * SuspVal32_i + \Gamma X_i + \upsilon_i\end{aligned}$$

where X_i is a vector of controls. For reference, the cross-sectional correlation between $NE20_i$

and $SuspNum32_i$ is 0.17 while that between $NE20_i$ and $SuspVal32_i$ is 0.21.

If banking panics had real effects because they destroyed soft yet valuable information embodied in continuing lending relationships, we should find $\gamma_3 < 0$. That is, the negative effect of bank suspensions on economic activity should have been more pronounced in locations with more continuing relationships, as measured by higher values of $NE20_i$, than in comparable locations with fewer such relationships. Intuitively, firms in continuing relationships lose a critical source of working capital when their banks are suddenly suspended. If the soft information accumulated over the course of a relationship is either unimportant or easy to redeploy, then the credit challenges faced by these firms should be more or less the same as the credit challenges faced by firms in otherwise similar locations, in which case we should find relatively little difference in their performance. If, on the other hand, soft information does matter and is difficult to redeploy across lenders, then it will be harder and/or take longer for these firms to regain access to credit on the terms they had prior to suspension, leading to a more noticeable impact on their operations.

5.2 Regression Results

The regression results are reported in Table 2. In columns (1) to (4), the control vector X_i includes district fixed effects and an indicator variable for reserve cities. In column (5), we add demographic and economic controls from the 1920 population census as well as controls for banking market structure.

The first column of Table 2 reports baseline results (i.e., estimation of the regression in Section 5.1 by ordinary least squares with simple controls). The estimates of γ_2 and γ_3 are negative and statistically significant, whereas the estimates of γ_1, γ_4, and γ_5 are not statistically different from zero. The magnitudes of γ_2 and γ_3 suggest that suspending 10% of national banks in the early 1930s would have led to a 3.77% decline in retail sales between 1929 and 1933 in locations where net elasticity in the 1920s was one standard deviation above the mean. This is almost double a 2.16% decline in comparable locations where net elasticity was at the mean.[13] The marginal impact of banking distress on retail sales during the Great Depression was therefore more severe in areas with more continuing relationships. This is an important result and, as we will see below, it is robust to many alternative specifications and additional controls.

In the second column, we separate the net elasticity $NE20_i$ into its components: the elasticity of loan returns β_i^ℓ and the elasticity of securities returns β_i^s. The second column can therefore be interpreted as the regression in the first column without the restriction that the coefficients on β_i^ℓ and β_i^s sum to zero. With β_i^s as a separate regressor, β_i^ℓ is the relevant price-based indicator of continuing relationships. As explained in Section 3, lower loan rate elasticity indicates more continuing relationships. Therefore, to confirm the baseline result that, on the margin, banking distress had a more severe negative effect on retail sales in areas with more continuing relationships, we should find a negative and statistically significant coefficient on $SuspNum32_i$ and a positive and statistically significant coefficient on the interaction between β_i^ℓ and $SuspNum32_i$. This is exactly what we do find, as can be seen in the second column of Table 2.

In the third column, we rerun the baseline results using purged indicators of banking distress. Specifically, we attempt to remove from $SuspNum32_i$ and $SuspVal32_i$ bank suspensions that could plausibly have been caused by bad loans made in the 1920s. From the perspective of destroying continuing relationships, it does not matter why exactly banks were suspended as long as the suspensions were not driven by the relationships themselves. The low correlation reported in Section 5.1 between bank suspensions and our measure of continuing relationships already suggests that banking distress was not more common in relationship lending areas but, as an additional exercise, we can redo the baseline regression expunging suspensions of delinquent loans to focus more sharply on the unpredictable component of the banking distress (e.g., panics) that hit in the 1930s. To this end, we define an indicator of pre-Depression loan losses for each location i:

[13] The mean of the $NE20$ distribution in Figure 3 is 0.045 and the standard deviation is 0.494. So we get -2.16=(-0.201-0.327*0.045)*10 and -3.77=(-0.201-0.327*(0.045+0.494))*10 with γ_2=-0.201 and γ_3=-0.327.

$$avgLLpre_i = \frac{1}{4} \sum_{t=1926}^{1929} \frac{loan\_losses_{i,t}}{deposits_{i,t}}$$

Data on loan losses come from the Comptroller's report and are aggregated across all types of lending. The cross-sectional correlation between $NE20_i$ and $avgLLpre_i$ is 0.29, which is also low. We run Tobit regressions of $SuspNum32_i$ and $SuspVal32_i$ on $avgLLpre_i$, including as controls $NE20_i$, district fixed effects, and the indicator for reserve cities.[14] Using the estimated coefficients on $avgLLpre_i$ from the Tobit regressions, we obtain predictions for the banking distress indicators when all controls are set to zero. We then subtract these predictions from the actuals, and re-censor so that all negative values are recorded as zero, to generate the purged indicators of banking distress. The results using these indicators are shown in the third column of Table 2 and confirm that the marginal effect of banking distress on retail sales was more negative in areas with more continuing relationships.

In the fourth column, we instrument the original (unpurged) indicators of banking distress. One may worry that lower demand for retail goods drives both a drop in retail sales and an increase in bank suspensions as firm profitability falls and defaults rise. Retail sales involve tradeable goods so district fixed effects should already soak up many of the demand-side determinants. To this point, the Federal Reserve Act of 1913 explicitly drew district boundaries around large regional markets to accord with "the convenience and customary course of business" in 1914, including trade, transportation, and communication links. In contrast, the cost of production, which is strongly influenced by the cost and availability of credit, was determined locally during the 1920s and 1930s. This is confirmed by a number of surveys conducted at the time, including the Bureau of Foreign and Domestic Commerce's *Consumer Debt Study* and the *Survey of Reports of Credit and Capital Difficulties* compiled by the Bureau of the Census. Nevertheless, it is still useful to show that our results are robust to instrumenting the indicators of banking distress. Calomiris and Mason (2003)

[14]The coefficient on $avgLLpre_i$ is positive and statistically significant in both Tobit regressions, suggesting that some of the bank suspensions in the early 1930s were predictable based on loan losses in the late 1920s. The coefficient on $NE20_i$ is not statistically significant in either regression.

have argued that the following variables, all measured in 1929, are acceptable instruments: logged banking assets, real estate owned by banks as a fraction of non-cash banking assets, and bank capital as a fraction of banking assets. Accordingly, we check the robustness of our baseline results to the use of these instruments, adding cash-to-deposit and loan-to-deposit ratios in 1929 to improve the first-stage estimation.[15] The fourth column of Table 2 shows that, once banking distress is instrumented, the only significant predictor of the change in retail sales from 1929 to 1933 is the interaction between $SuspNum32_i$ and $NE20_i$ and, as before, the coefficient is negative. In an alternative specification, we removed the loan-to-deposit ratio as an instrument and replaced it with the ratio of demand deposits to banking assets in 1929. The results were very similar.

In the fifth column, we add demographic and economic variables from the 1920 population census to the vector of controls in the baseline regression (see Table 1 for a list of these variables). We also control for banking concentration – specifically, the log of banks per capita in 1920 or $LogBankpc_i$ as defined in Table 1 – as well as its interactions with the banking distress indicators.[16] There are two noteworthy results. First, the coefficient on the interaction between $SuspNum32_i$ and $NE20_i$ remains negative and statistically significant, confirming that the marginal effect of banking distress on retail sales was more negative in areas with more continuing relationships. In an alternative specification, we replaced $NE20_i$ with the average fraction of national bank loans made on little to no collateral in location i during the 1920s. The interaction term was not negative or statistically significant, underscoring the importance of extracting continuing relationships from the broader pool of potential relationship loans. Second, the coefficients on $SuspVal32_i$ and its interactions are now also statistically significant. The coefficient on $SuspVal32_i$ is negative, meaning that a

[15] The identifying assumption of Calomiris and Mason (2003) is that shocks in the 1930s were not just a continuation of shocks in the 1920s. The two ratios we add are informative about the ability of banks in different areas to withstand deposit withdrawals (positive for the former, negative for the latter) and should therefore have predictive power for the amount of banking distress, at least in the early stages of the crisis.

[16] We exclude the bank size variable in Table 1 ($LogBankSize_i$) from the vector of controls because its correlation with $LogBankpc_i$ is -0.83. Including the remaining variables ($DepRatio_i$ and $DDratio_i$) as controls does not change the results.

larger banking shock leads to a bigger decline in retail sales. In contrast, the coefficient on $SuspVal32_i * NE20_i$ is positive, as is the coefficient on $SuspVal32_i * LogBankpc_i$.

Increasing $SuspVal32_i$ without also increasing $SuspNum32_i$ is akin to considering the effects of the suspension of larger banks. In one important respect the suspension of large banks in the early 1930s is likely to have been less damaging to the economy than suspension of their smaller competitors. Large banks during this period were liquidated quite rapidly, both in absolute terms and relative to smaller banks, meaning that depositors in large banks were able to gain access to their funds much more quickly than depositors in small banks. Receivership data for 1931 and 1932 show that depositors in national banks with more than \$6 million in deposits on the date of failure received an average of 4.2 cents per month for each dollar of deposits during the initial year of liquidation. Depositors in banks that failed with \$2-6 million in deposits received only 3.0 cents per month while depositors in banks that failed with less than \$2 million in deposits received 2.0 cents per month. In practice, suspended deposits in large banks could be redeemed even more quickly than the receivership data suggest because clearinghouses often provided advances to depositors of failed members, which would typically be the largest failures in the municipality. Given these efforts by receivers and clearinghouses to ensure that depositors of large suspended banks still had at least partial access to their funds, the suspension of a large bank would have created space in the business landscape for surviving banks to increase their deposits and thus their lending activities. The coefficients in the fifth column of Table 2 suggest that this reallocation was most valuable in less concentrated banking markets (as captured by more banks per capita) and in areas where business was more dependent on continuing relationships (as measured by higher net elasticity in the 1920s).

5.3 Discussion

The main takeaway from Table 2 is that banking panics had a more pronounced negative effect on retail sales in locations with more continuing relationships, other things the same.

Across all specifications, the coefficient on the interaction between $SuspNum32_i$ and $NE20_i$ was negative and statistically significant. We also found some evidence that efforts by receivers and clearinghouses to ensure a rapid release of funds to depositors of large suspended banks had a more pronounced positive effect on retail sales in locations with more continuing relationships. While this was observed most clearly in the last column of Table 2, with a positive and statistically significant coefficient on the interaction between $SuspVal32_i$ and $NE20_i$, the positive coefficient is also visible in other columns, albeit without the same statistical significance.

In Table 3, we repeat the analysis using the retail sales ratio in 1935, rather than the retail sales ratio in 1933, as the dependent variable. The coefficient on the interaction between $SuspNum32_i$ and $NE20_i$ is still negative but no longer statistically significant, except in the last column which is the regression with census controls. The interaction between $SuspVal32_i$ and $NE20_i$ is also still positive and statistically significant in this column. In Table 4, we use the retail sales ratio in 1937 as the dependent variable and find that all statistical significance disappears. In other words, bank suspensions from the early 1930s no longer had a direct effect on retail sales by 1937. We emphasize that this is only a statement about direct effects. We will turn to indirect effects in Section 6.

The estimated coefficients can also be used to make statements about the aggregate implications of banking distress in the early 1930s. Specifically, we can predict what retail sales would have been had there been no bank suspensions (i.e., $SuspNum32_i = SuspVal32_i = 0$ for all locations i) then compare this to the fitted values when all regressors are as observed in the data. Depending on which column in Table 2 is used to generate the comparison, we find that total retail trade in the U.S. would have been 1% to 4% higher in 1933 had there been no banking distress. If we set only $SuspNum32_i = 0$ to try to isolate the impact of the initial distress without the mitigating effect from the rapid liquidation of large suspended banks, we find that total retail trade would have been 3% to 9% higher in 1933 had there been no banking distress. In both counterfactuals, the upper bounds come from the instrumental variables regression (fourth column) while the lower bounds come from the regression with census controls (fifth column). That being said, the regression with census controls delivers more sizeable estimates for 1935 than it does for 1933. In particular, it predicts that total retail trade would have been 3% higher in 1935 had there been no banking distress and 5% higher had there also been no mitigating effect.

Are the aggregate effects just derived economically significant? This is a question about both the magnitude of the effects we found and the importance of retail trade for the broader economy. Geographically disaggregated measures of total economic activity (e.g., GNP) are not available for this period, but we can perform a back-of-the-envelope calculation to translate the total retail sales decline into an aggregate GNP effect. The "multiplier" we use is the coefficient from a simple regression for the period 1920-1929 of GNP growth on retail sales growth, with the latter instrumented by its one period lag. We start in 1920 because that is when the Federal Reserve began publishing its monthly index of retail sales. We seasonally adjust this index using Census software then take annual averages to match the frequency of the GNP estimates available from Romer (1989) and Balke and Gordon (1989). We end in 1929 to ensure that our results are not distorted by the Great Depression, the 1937-38 recession, or WWII. We also eliminate the post-WWII period since the ratio of retail sales to GNP declines markedly after the war. The multipliers from various specifications are reported in the first row of Table 5. Overall, we find that a 4% drop in retail sales is consistent with a 6-7% decline in nominal GNP and a 3-4% fall in real GNP, which is about one-eighth of the economic contraction experienced during the Great Depression. This would seem to leave little doubt about the economic significance of the results.

6 Relationship Rebuilding and the 1937-38 Recession

In this section, we show that cross-sectional differences in the rebuilding of lending relationships destroyed during the Great Depression are important for understanding cross-sectional differences in economic performance later in the decade. Specifically, areas that appear to have been rebuilding relationships in the mold of those destroyed by the banking panics of the early 1930s fared better than others during the 1937-38 recession. This finding is important for two reasons. First, it sheds light on the heterogeneous nature of recovery from the Depression. Second, it strengthens the interpretation of our previous results by serving as evidence that what we measured to be continuing relationships on the eve of the Great Depression do indeed have value in a period where they are not forcibly destroyed.

6.1 Empirical Specification

The spirit of the analysis is similar to Section 5 in that we want to see how continuing relationships affect performance in a crisis. The crisis we considered in Section 5 was the Great Depression. The crisis we are considering here is the 1937-38 recession. That being said, there are two differences between these events that will affect both the regressions we run and the intuition behind the results: the magnitude of the crisis and the precision with which we can detect continuing relationships on its eve.

Consider first the magnitude. The Great Depression involved bank runs by depositors and swaths of bank suspensions. In contrast, the 1937-38 recession involved an increase in bank funding costs that did not spiral out of control, possibly benefitting from the introduction of deposit insurance in 1934. Friedman and Schwartz (1963) attribute higher bank funding costs in the late 1930s to the Fed's decision to begin doubling reserve requirements in August 1936. Hanes (2006) and Calomiris et al (2011) instead assign a more important role to the Treasury's decision to sterilize gold inflows starting in December 1936. The bottom line is the same in both cases: contractionary monetary policy was introduced in the U.S. in 1936. The model in Section 3 predicts that continuing relationships will mitigate policy tightening by extending policy-invariant loan rates to some borrowers, thus enabling them to continue operations uninterrupted.[17] Of course, the ability of continuing relationships

[17]By making loan rates less responsive to changes in bank funding costs, continuing relationships also

to absorb a cost-of-funds shock is conditional on bank survival. If the banks are instead destroyed, as they were during the runs of the early 1930s, the continuing relationships are also destroyed and, as we saw in Section 5, locations dependent on those relationships suffer greatly. Whether continuing relationships absorb or amplify potential crises is, therefore, sensitive to the nature of the shock.

A second difference between the Great Depression and the 1937-38 recession is the precision with which we can detect continuing relationships on the eve of the crisis. We were able to construct our price-based indicator of continuing relationships on the eve of the Great Depression using data from the 1920s. This is much more difficult to do for the 1937-38 recession for the obvious reason that early years of the decade were scarred by the Depression, giving us only three years (1934-36) of recovery to work with, too few to produce reliable estimates of the elasticity of asset returns using the semi-annual data published by the Comptroller.[18] If we were to estimate elasticities using data from 1930 to 1936, we would find that the cross-sectional correlation between net elasticity in the 1920s and the change in net elasticity from the 1920s to the 1930s is -0.90, which only reinforces the point that the Great Depression destroyed continuing relationships in areas where such relationships existed. We therefore need a proxy to measure the extent to which continuing relationships had been (re)built by 1936.

We select as our proxy the interaction between two variables prior to the 1937-38 recession. The first variable is $NE20_i$, our price-based indicator of continuing relationships in the 1920s. The idea is that locations where continuing relationships had previously flourished are likely to be the ones where they will again thrive, on the assumption that the preferences of banks and the needs of firms were, for the most part, stable over time. We will return to this point when discussing control variables in Section 6.2. The second variable, $CollLite30_i$, is the average fraction of bank loans in location i that were either uncollater-

make firm project choice, and hence production, less responsive to those changes. See Hachem (2011) for a formal treatment of how relationship lending affects the output response to monetary policy.

[18] Note also that the first real test for the majority of post-Depression lenders may have been the 1937-38 recession itself.

alized or collateralized by personal security between 1934 and 1936. We know from Section 2 that observers in the 1920s and 1930s viewed uncollateralized loans as the most natural loans for relationship building. It then stands to reason that locations with high values of $NE20_i$ and $CollLite30_i$ are the most likely candidates for relationship rebuilding: they had a history of continuing relationships and they were still making the types of loans that were an input into such relationships.

To study whether continuing relationships affected the severity of the 1937-38 recession, we run the regression:

$$\frac{Sales_{i,1939}}{Sales_{i,1937}} = \lambda_0 + \lambda_1 NE20_i + \lambda_2 CollLite30_i + \lambda_3 NE20_i * CollLite30_i + \Pi Z_i + \zeta_i$$

where Z_i is a vector of controls described in more detail below. If continuing relationships absorbed the contractionary monetary policy shocks that others have argued occurred in the U.S. in 1936, we should find $\lambda_3 > 0$. That is, the late 1930s recession should have been less pronounced in those locations where we would expect to find the greatest concentration of rebuilt lending relationships.

6.2 Results and Discussion

The regression results are reported in Table 6.[19] In column (1), the control vector Z_i includes district fixed effects and an indicator variable for reserve cities. In column (2), we add as a control the excess reserve ratio in location i on the eve of the 1937-38 recession and interact it with the main regressors. This ratio, $ExcessRR_i$, is calculated as the ratio of excess reserves to demand deposits, both measured in June 1936. In column (3), we instrument the excess reserve ratio and its interactions to address potential endogeneity concerns.

There are two reasons why banks in some areas may have been holding high excess reserve

[19] Recall from Section 4.2 that uncollateralized lending for the 1930s is based on the "all other loans" category in the Comptroller's report. The Comptroller separated "all other loans" from "loans on securities" for cities and states but not districts so we cannot use any of the split states or district remainders in the analysis here. As a result, the number of locations is less than 82.

ratios in June 1936. The first is backward-looking and is the one that motivates inclusion of $ExcessRR_i$ in the vector of controls: the banking panics of the early 1930s may have prompted bankers in those areas to hold precautionary reserve cushions. The problem is that heterogeneous changes in liquidity management from the 1920s to the 1930s confound the ability of continuing relationships in the pre-Depression era to predict the rise of similar relationships once the Depression had subsided. Controlling for excess reserve ratios in June 1936 helps control for the effect of the Great Depression on banker preferences, making $NE20_i * CollLite30_i$ a better predictor of continuing relationships on the eve of the 1937-38 recession.[20]

The second possible reason for high excess reserve ratios in June 1936 is forward-looking: bankers in those areas may have opted for large reserve cushions in anticipation of poor economic conditions ahead. This introduces an endogeneity concern so we use instrumental variables to isolate only the backward-looking component of excess reserve ratios in June 1936. Changes in liquidity management from the 1920s to the 1930s are more likely to have occurred in locations where banks were burned by liquidity problems in the early 1930s, motivating the following two instruments for $ExcessRR_i$. The first instrument is the ratio of cash to demand deposits in location i in 1929. Areas where banks held less cash relative to demand deposits on the eve of the Great Depression would have been more vulnerable once depositors began to panic and withdraw funds. The second instrument is the purged indicator of banking distress in the early 1930s based on the number of suspended banks in location i (see Section 5.2). This captures the extent to which banks with otherwise good assets succumbed to the panics.

The third column of Table 6 presents the results. We find that λ_3 (the coefficient on $NE20_i * CollLite30_i$) is positive and statistically significant while the coefficient on the interaction between $NE20_i * CollLite30_i$ and $ExcessRR_i$ is negative and statistically signifi-

[20] The seminal work of Diamond and Dybvig (1983) expounds the fundamental fragility of banks vis-à-vis liquidity so it seems reasonable to focus specifically on the preference for liquidity when considering the effect of the Depression on banker preferences.

cant. The magnitudes of the estimated coefficients imply that economic performance during the late 1930s recession was increasing in $NE20_i * CollLite30_i$ for any location i with an excess reserve ratio below 8.4%. For comparison, the mean of $ExcessRR_i$ across locations was 7.0% with a standard deviation of 3.9%.

We would argue that the Great Depression did not give rise to a preference for liquidity among banks in locations where our two instruments predict low excess reserve ratios in June 1936. Accordingly, any relationships rebuilt in these locations are the ones most likely to have been cut from the same cloth as the continuing relationships that existed in the 1920s. The positive coefficient on $NE20_i * CollLite30_i$ for these locations substantiates this claim in that the 1937-38 recession was less pronounced in areas where more continuing relationships had been built, consistent with the shock-absorbing nature of these relationships in Section 3. The negative coefficient on $NE20_i * CollLite30_i$ for locations where our instruments predict high excess reserve ratios then provides evidence that the recession was more pronounced when attempts at relationship rebuilding had been made but banks were less dedicated to maintaining them than they had been in the 1920s. Bankers chastened by the Great Depression into holding large excess reserves may also have been inclined to preemptively discontinue even good lending relationships in order to expand their liquidity buffers at the first sign of trouble. Such behavior would cause economic activity to contract for the very same reason that it did during the Depression: the soft information about borrowers cannot be easily redeployed.

We also find that the coefficient on $NE20_i$ in the third column of Table 6 is negative and statistically significant while the coefficient on the interaction between $NE20_i$ and $ExcessRR_i$ is positive and statistically significant. At the average value of $ExcessRR_i$, the magnitudes of the estimated coefficients imply that economic performance during the late 1930s recession was decreasing in $NE20_i$ for any location i with $CollLite30_i$ below 0.56. For comparison, the mean of $CollLite30_i$ across locations was 0.55 with a standard deviation of 0.12. Section 5 established that bank suspensions in the early 1930s had a more contractionary effect on economic activity in locations with more continuing relationships, other things equal. It would therefore appear that these locations stayed weak into the late 1930s unless they continued making the types of loans that observers argued built relationship capital (i.e., uncollateralized loans). The interaction between $NE20_i$ and $CollLite30_i$ is therefore important for capturing relationship rebuilding and, as discussed above, relationship rebuilding in the aftermath of the Great Depression helped mitigate the 1937-38 recession when the relationships reproduced those that had existed in the 1920s.

7 Conclusion

We proposed a novel measure of continuing lending relationships that resolves the data limitations of the 1920s and 1930s and pinpoints the non-monetary effects of banking distress in a way that the existing literature on the Great Depression has been unable to do. Our measure is based on the idea that longer relationships involve loan rates that are less responsive to changes in bank funding costs. We presented both historical and theoretical arguments in support of our measure then implemented it to study the transmission of bank suspensions to economic activity through the disruption of long-term lending relationships.

We showed that the marginal impact of bank suspensions on economic activity in the early 1930s was more negative in areas that had more continuing relationships on the eve of the Great Depression. We also showed that relationship lending played an important role during the recovery period. In particular, areas that rebuilt the types of continuing relationships that they embraced in the 1920s fared better during the 1937-38 recession than otherwise similar areas that did not rebuild. In contrast, areas that rebuilt relationships but, based on their reserve holdings, appear to have been less dedicated to these relationships, fared worse than otherwise similar areas that failed to rebuild.

In keeping with Bernanke (1983) and others, we have focused on the Great Depression. Given the dimensions of the downturn and the scope of the banking panics, this episode remains ground zero for studying the real effects of financial crises. However, the value of our methodology transcends resolution of data limitations for the 1920s and 1930s. Policy-makers working in real time, often in crisis situations, usually have only aggregate data to inform them about what is happening at more disaggregated levels. Our method permits the extraction of more detailed information from these aggregates. Our findings also transcend the historical context of the Great Depression. In many parts of the developing world, where relationship loans constitute a major part of bank assets, bank failures are likely to have serious and potentially long-lasting deleterious effects on the real economy. Fortified with this knowledge, policy-makers can mitigate the damage if they act swiftly and strongly enough to shore up the banks and salvage the soft yet crucial information embodied in their continuing lending relationships.

References

Balke, N. and R. Gordon. 1989. "The Estimation of Prewar Gross National Product: Methodology and New Evidence." *Journal of Political Economy*, 97(1), pp. 38-92.

Berger, A. and G. Udell. 1995. "Relationship Lending and Lines of Credit in Small Firm Finance." *Journal of Business*, 68(3), pp. 351-382.

Bernanke, B. 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review*, 73(3), pp. 257-276.

Board of Governors. 1943. *Banking and Monetary Statistics*.

Bolton, P., X. Freixas, L. Gambacorta, and P. Mistrulli. 2016. "Relationship and Transaction Lending in a Crisis." *Review of Financial Studies*, 29(10), pp. 2643-2676.

Calomiris, C. and J. Mason. 2003. "Consequences of Bank Distress During the Great Depression." *American Economic Review*, 93(3), pp. 937-947.

Calomiris, C., J. Mason, and D. Wheelock. 2011. "Did Doubling Reserve Requirements Cause the Recession of 1937-1938? A Microeconomic Approach." NBER Working Paper No. 16688.

Carlson, M. and J. Rose. 2015. "Credit Availability and the Collapse of the Banking Sector in the 1930s." *Journal of Money, Credit, and Banking*, 47(7), pp. 1239-1271.

Chodorow-Reich, G. 2014. "The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008-9 Financial Crisis." *Quarterly Journal of Economics*, 129(1), pp. 1-59.

Cole, H. and L. Ohanian. 1999. "The Great Depression in the United States from a Neoclassical Perspective." *Federal Reserve Bank of Minneapolis Quarterly Review*, 23(1), pp. 2-24.

Cole, H. and L. Ohanian. 2004. "New Deal Policies and the Persistence of the Great Depression: A General Equilibrium Analysis." *Journal of Political Economy*, 112(4), pp. 779-816.

Cole, H. and L. Ohanian. 2007. "A Second Look at the U.S. Great Depression from a Neoclassical Perspective." In Kehoe, T. and E. Prescott (eds.), *Great Depressions of the Twentieth Century*, Federal Reserve Bank of Minneapolis, pp. 21-57.

Currie, L. 1931. "The Decline of the Commercial Loan." *Quarterly Journal of Economics*, 45(4), pp. 698-709.

Diamond, D. and P. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy*, 91(3), pp. 401-419.

Elsas, R. 2005. "Empirical Determinants of Relationship Lending." *Journal of Financial Intermediation*, 14(1), pp. 32-57.

Ford, A.W. 1928. "Country Banking Practices in Northern Texas." *Journal of Business*, 1(4), pp. 468-484.

Foulke, R. and H. Prochnow. 1939. *Practical Bank Credit*. New York: Prentice-Hall.

Friedman, M. and A. Schwartz. 1963. *A Monetary History of the United States, 1867-1960.* Princeton, NJ: Princeton University Press.

Gambacorta, L. and P. Mistrulli. 2014. "Bank Heterogeneity and Interest Rate Setting: What Lessons Have We Learned since Lehman Brothers." *Journal of Money, Credit and Banking*, 46(4), pp. 753-778.

Gobbi, G. and E. Sette. 2015. "Relationship Lending During a Financial Crisis." *Journal of the European Economic Association*, 13(3), pp. 453-481.

Hachem, K. 2011. "Relationship Lending and the Transmission of Monetary Policy." *Journal of Monetary Economics*, 58(6-8), pp. 590-600.

Hanes, C. 2006. "The Liquidity Trap and U.S. Interest Rates in the 1930s." *Journal of Money, Credit, and Banking*, 38(1), pp. 163-194.

Langston, L. 1921. *Practical Bank Operation*. New York: The Ronald Press Company.

Miller, M. 1927. *Bank Loans on Statement and Character*. New York: The Ronald Press Company.

Neumark, D. and S. Sharpe. 1992. "Market Structure and the Nature of Price Rigidity: Evidence from the Market for Consumer Deposits." *Quarterly Journal of Economics*, 107(2), pp. 657-680.

Ongena, S. and D. Smith. 2001. "The Duration of Bank Relationships." *Journal of Financial Economics*, 61(3), pp. 449-475.

Petersen, M. and R. Rajan. 1994. "The Benefits of Lending Relationships: Evidence from Small Business Data." *Journal of Finance*, 49(1), pp. 3-37.

Rajan, R. and R. Ramcharan. 2016. "Crises and the Development of Economic Institutions: Some Microeconomic Evidence." *American Economic Review: Papers and Proceedings*, 106(5), pp. 524-527.

Richardson, G. and W. Troost. 2009. "Monetary Intervention Mitigated Banking Panics During the Great Depression: Quasi-Experimental Evidence from the Federal Reserve District Border in Mississippi, 1929 to 1933." *Journal of Political Economy*, 117(6), pp. 1031-1073.

Romer, C. 1989. "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869-1908." *Journal of Political Economy*, 97(1), pp. 1-37.

Romer, C. and D. Romer. 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." *NBER Macroeconomics Annual*, Vol. 4, pp. 121-170.

Romer, C. and D. Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review*, 94(4), pp. 1055-1084.

Romer, C. and D. Romer. 2015. "New Evidence on the Impact of Financial Crises in Advanced Countries." NBER Working Paper No. 21021.

Thomas, W. 1935. "Use of Credit in Security Speculation." *American Economic Review*, 25(1), pp. 21-30.

Wicker, E. 1996. *The Banking Panics of the Great Depression*. Cambridge: Cambridge University Press.

Ziebarth, N. 2013. "Identifying the Effects of Bank Failures from a Natural Experiment in Mississippi during the Great Depression." *American Economic Journal: Macroeconomics*, 5(1), pp. 81-101.

Table 1:
Cross-Sectional Correlations

		Correlation with NE20
LogArea	log(area in square miles)	0.460
Urban	% population urban in 1920	-0.543
Nwnp	% population with native white parents in 1920	0.325
Age1844	% population aged 18 to 44 in 1920	-0.435
School1620	% of aged 16 to 20 in school in 1920	0.329
LogMfgEst	log(number of mfg establishments in 1920)	-0.084
LogMfgSize	log(workers per mfg establishment in 1920)	-0.297
MfgWork	mfg workers as % of population in 1920	-0.390
LogMfgVa	log(value added per mfg establishment in 1920)	-0.314
LogFarms	log(number of farms in 1920)	0.491
Acres	farm land as % of area in 1920	0.291
LogAvgAcre	log(farm acres / number of farms in 1920)	0.344
LogAvgCrop	log(crop value / number of farms in 1920)	0.186
LogAvgValue	log(value of farm land, equip, etc / no. of farms in 1920)	-0.124
OwnerOp	% of farms owner-operated in 1920	-0.080
HomeOwnClr	home ownership rate in 1920	0.457
LogBankpc	log(number of national banks / population in millions in 1920)	0.479
LogBankSize	log(banking assets / number of banks in 1920)	-0.511
DepRatio	demand and time deposits / banking assets in 1920	0.276
DDratio	demand deposits / demand and time deposits in 1920	-0.302

Notes: All variables are from the 1920 population census, except the last four which are from the Comptroller's annual report in 1920.

Table 2:
Retail Sales Ratio, 1933 / 1929

	Baseline	Comp. Elast.	Purged Susp.	IV Susp.	Extra Controls
	(1)	(2)	(3)	(4)	(5)
NE20	0.00777		0.0131	0.0252	0.0164
	(0.0174)		(0.0175)	(0.0304)	(0.0195)
SuspNum32	-0.201**	-0.257***	-0.201***	-0.326	0.0717
	(0.0668)	(0.0654)	(0.0610)	(0.225)	(0.391)
SuspNum32 × NE20	-0.327*		-0.832*	-0.554*	-0.624**
	(0.162)		(0.458)	(0.292)	(0.280)
SuspVal32	0.141	0.197	0.0881	0.393	-0.563**
	(0.0971)	(0.139)	(0.0990)	(0.338)	(0.186)
SuspVal32 × NE20	0.330		1.032	0.133	0.714**
	(0.269)		(0.658)	(0.657)	(0.254)
β^{ℓ}		-0.0210			
		(0.0414)			
SuspNum32 × β^{ℓ}		0.433*			
		(0.229)			
SuspVal32 × β^{ℓ}		-0.493			
		(0.957)			
β^{s}		0.00025			
		(0.0172)			
SuspNum32 × β^{s}		-0.301			
		(0.194)			
SuspVal32 × β^{s}		0.318			
		(0.241)			
LogBankpc					0.0108
					(0.0352)
SuspNum32 × LogBankpc					-0.0424
					(0.114)
SuspVal32 × LogBankpc					0.153**
					(0.0505)
Observations	82	82	82	82	81
District and RC Dummies	✓	✓	✓	✓	✓
Census Controls	×	×	×	×	✓
R-squared	0.496	0.500	0.498		0.781
J-Statistic (p-value)				0.830	
AR Wald Test (p-value)				0.000	

Notes: All columns include district fixed effects and a dummy variable for reserve cities. Column (1) is the baseline regression in Section 5.1. Column (2) splits NE20 into the component elasticities defined in Section 4.1. Column (3) purges SuspNum32 and SuspVal32 of bank suspensions that could plausibly be attributed to bad loans made in the 1920s. Column (4) instruments SuspNum32 and SuspVal32 following Calomiris and Mason (2003). Column (5) adds controls from Table 1 to the baseline regression, specifically all the variables from the 1920 population census and LogBankpc. Standard errors clustered at the district level are in parentheses. ***p<0.01, **p<0.05, *p<0.1

Table 3:
Retail Sales Ratio, 1935 / 1929

	Baseline	Comp. Elast.	Purged Susp.	IV Susp.	Extra Controls
	(1)	(2)	(3)	(4)	(5)
NE20	-0.0198		-0.00819	-0.0104	-0.0189
	(0.0129)		(0.0138)	(0.0409)	(0.0245)
SuspNum32	-0.291**	-0.383*	-0.366**	-0.307	0.593
	(0.117)	(0.176)	(0.130)	(0.357)	(0.430)
SuspNum32 × NE20	-0.227		-0.913	-0.267	-0.531**
	(0.296)		(0.816)	(0.583)	(0.211)
SuspVal32	0.250	0.206	0.131	0.447	-1.312*
	(0.216)	(0.233)	(0.208)	(0.530)	(0.614)
SuspVal32 × NE20	0.192		1.378	-0.140	0.706**
	(0.396)		(1.180)	(0.705)	(0.314)
β^{ℓ}		0.0223			
		(0.0584)			
SuspNum32 × β^{ℓ}		0.177			
		(0.299)			
SuspVal32 × β^{ℓ}		0.489			
		(1.065)			
β^{s}		-0.0181			
		(0.0182)			
SuspNum32 × β^{s}		-0.116			
		(0.292)			
SuspVal32 × β^{s}		0.0754			
		(0.344)			
LogBankpc					0.0111
					(0.0399)
SuspNum32 × LogBankpc					-0.225*
					(0.117)
SuspVal32 × LogBankpc					0.359**
					(0.153)
Observations	82	82	82	82	81
District and RC Dummies	✓	✓	✓	✓	✓
Census Controls	×	×	×	×	✓
R-squared	0.474	0.485	0.489		0.755
J-Statistic (p-value)				0.737	
AR Wald Test (p-value)				0.017	

Notes: All columns include district fixed effects and a dummy variable for reserve cities. Column (1) is the baseline regression in Section 5.1. Column (2) splits NE20 into the component elasticities defined in Section 4.1. Column (3) purges SuspNum32 and SuspVal32 of bank suspensions that could plausibly be attributed to bad loans made in the 1920s. Column (4) instruments SuspNum32 and SuspVal32 following Calomiris and Mason (2003). Column (5) adds controls from Table 1 to the baseline regression, specifically all the variables from the 1920 population census and LogBankpc. Standard errors clustered at the district level are in parentheses. ***p<0.01, **p<0.05, *p<0.1

Table 4:
Retail Sales Ratio, 1937 / 1929

	Baseline	Comp. Elast.	Purged Susp.	IV Susp.	Extra Controls
	(1)	(2)	(3)	(4)	(5)
NE20	-0.431		-0.384	-4.134	0.280
	(0.613)		(0.470)	(4.980)	(1.124)
SuspNum32	-0.817	-0.973	-0.586	22.61	-1.496
	(0.833)	(1.397)	(0.777)	(28.81)	(18.35)
SuspNum32 × NE20	7.359		11.23	55.30	17.12
	(9.418)		(11.13)	(64.47)	(19.20)
SuspVal32	0.582	7.020	0.0781	-19.47	-36.55
	(1.074)	(6.194)	(1.457)	(40.48)	(36.54)
SuspVal32 × NE20	-7.963		-7.669	8.671	-32.74
	(8.744)		(9.094)	(77.52)	(28.51)
β^ℓ		3.184			
		(2.773)			
SuspNum32 × β^ℓ		-13.40			
		(13.94)			
SuspVal32 × β^ℓ		-12.92			
		(13.69)			
β^s		0.431			
		(0.361)			
SuspNum32 × β^s		12.59			
		(13.95)			
SuspVal32 × β^s		-16.92			
		(15.71)			
LogBankpc					-2.425
					(3.273)
SuspNum32 × LogBankpc					-0.184
					(5.273)
SuspVal32 × LogBankpc					9.874
					(9.852)
Observations	82	82	82	82	81
District and RC Dummies	✓	✓	✓	✓	✓
Census Controls	×	×	×	×	✓
R-squared	0.060	0.076	0.059		0.258
J-Statistic (p-value)				0.964	
AR Wald Test (p-value)				0.918	

Notes: All columns include district fixed effects and a dummy variable for reserve cities. Column (1) is the baseline regression in Section 5.1. Column (2) splits NE20 into the component elasticities defined in Section 4.1. Column (3) purges SuspNum32 and SuspVal32 of bank suspensions that could plausibly be attributed to bad loans made in the 1920s. Column (4) instruments SuspNum32 and SuspVal32 following Calomiris and Mason (2003). Column (5) adds controls from Table 1 to the baseline regression, specifically all the variables from the 1920 population census and LogBankpc. Standard errors clustered at the district level are in parentheses. ***p<0.01, **p<0.05, *p<0.1

Table 5:
GNP Multiplier for Retail Sales

	Nominal GNP Growth		Real GNP Growth	
	Romer	Balke-Gordon	Romer	Balke-Gordon
RetailGrowth	1.690***	1.565***	0.696***	0.896***
	(0.242)	(0.299)	(0.163)	(0.242)
Constant	-0.0102	-0.00644	0.0278***	0.0246***
	(0.00998)	(0.0115)	(0.00512)	(0.0056)

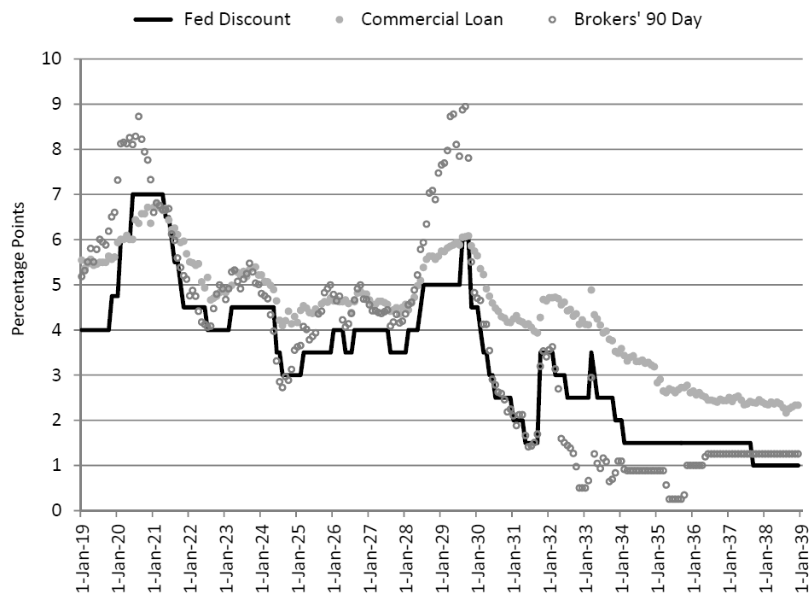
Notes: Data are annual for the period 1920-1929. RetailGrowth is instrumented using its one period lag. Robust standard errors are in parentheses. ***p<0.01, **p<0.05, *p<0.1

Table 6:
Retail Sales Ratio, 1939 / 1937

	OLS	OLS	IV
	(1)	(2)	(3)
NE20	0.612	-0.777	-5.256**
	(0.387)	(1.363)	(2.361)
CollLite30	-0.114	-0.562	-1.695
	(0.436)	(0.499)	(1.582)
NE20 × CollLite30	-0.890	1.421	9.155**
	(0.595)	(2.321)	(4.024)
ExcessRR		-3.170	-14.12
		(5.587)	(17.14)
NE20 × ExcessRR		15.60	62.88**
		(11.21)	(26.77)
CollLite30 × ExcessRR		6.274	24.77
		(8.082)	(26.88)
NE20 × CollLite30 × ExcessRR		-26.00	-109.0***
		(18.90)	(42.32)
Observations	64	63	63
District and RC Dummies	✓	✓	✓
R-squared	0.619	0.653	
J-Statistic (p-value)			0.328
AR Wald Test (p-value)			0.000

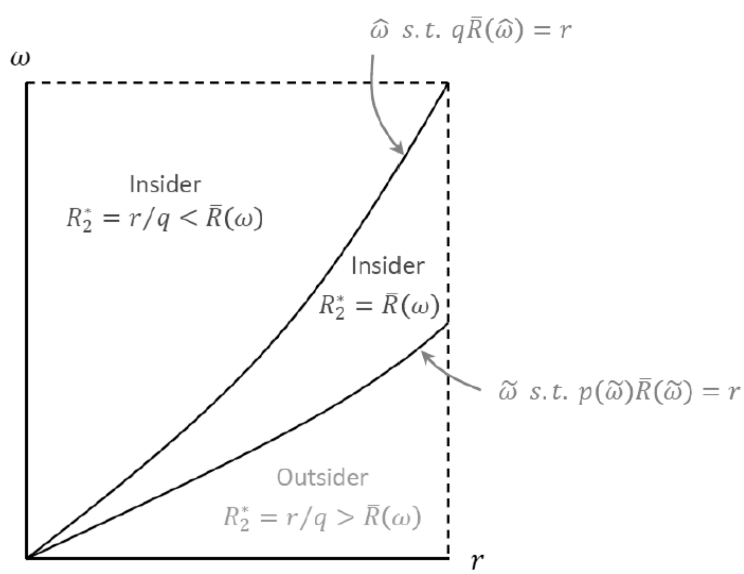
Notes: Column (3) instruments ExcessRR to isolate the component that can be explained by differences in liquidity management as a result of the Great Depression. Standard errors clustered at the district level are in parentheses. ***p<0.01, **p<0.05, *p<0.1

Figure 1:
Illustrative Example from NYC Rates



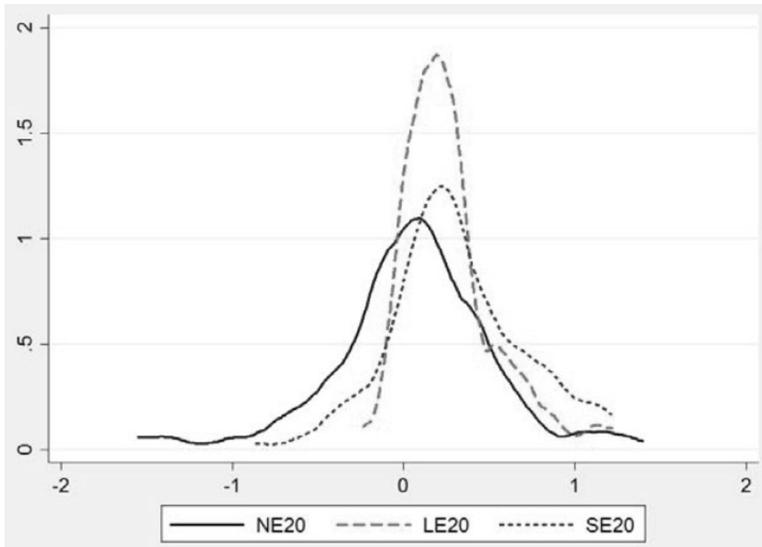
Notes: Data is from *Banking and Monetary Statistics*

Figure 2:
Equilibrium Loan Rates in the Second Stage



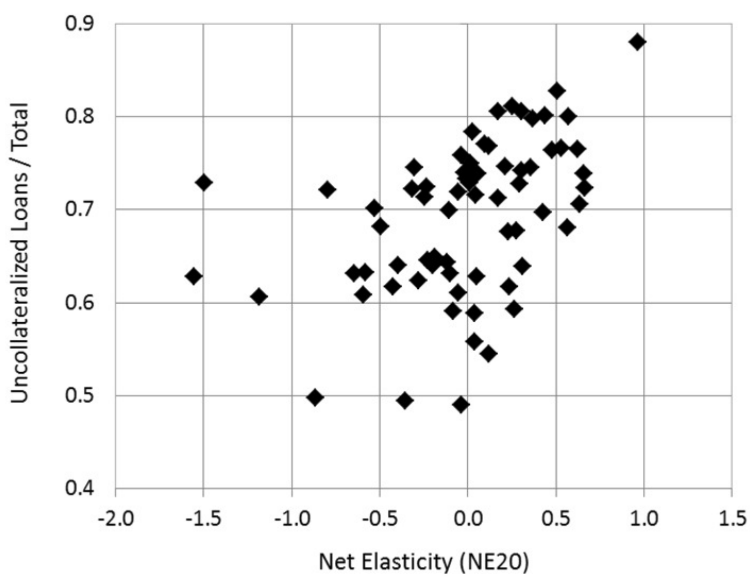
Notes: The policy rate is r and firm types are denoted by ω. The interest rate that makes ω indifferent towards risk-shifting is $\overline{R}(\omega)$. The information sets of insiders and outsiders are as defined in Section 3.

Figure 3:
Distribution of Elasticity Estimates in the 1920s



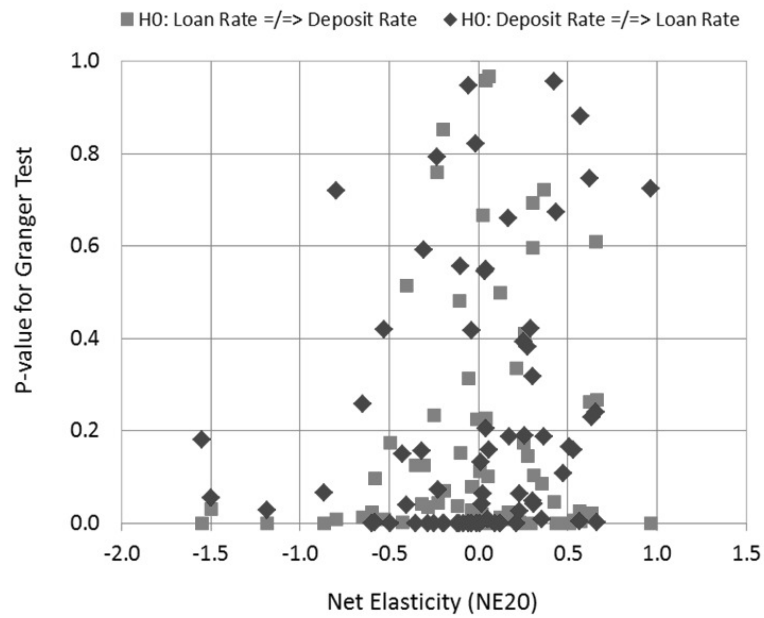
Notes: LE20 is the elasticity of loan returns (β_i^ℓ), SE20 is the elasticity of securities returns (β_i^s), and NE20 is the net elasticity ($NE20_i = \beta_i^s - \beta_i^\ell$).

Figure 4:
Net Elasticities by Location in the 1920s



Notes: Each point represents a location. Uncollateralized loans include loans collateralized by personal security.

Figure 5:
Granger Causality Test Results



Notes: Each point represents a location. Tests are based on semi-annual data from 1925 to 1929 inclusive.

Appendix A – Supplement to Figure 1

This appendix provides a more extensive treatment of the data presented in Figure 1. Recall the three monthly series plotted in Figure 1 from January 1919 to December 1938: the average interest rate on 90-day brokers' loans in New York City, the average interest rate on one to six month commercial loans in New York City, and the average discount rate set by the Federal Reserve Bank of New York.

We start by regressing changes in loan rates on changes in the discount rate. Figure A1 presents a scatterplot of the data. The horizontal axis indicates the change in the Fed's discount rate. The vertical axis indicates the change in the loan rate. The red dots plot changes in the brokers' loan rate against changes in the discount rate. The tan dots plot changes in the commercial loan rate against changes in the discount rate. The average response for each type of loan is indicated by the fitted regression line. On average, brokers' loans rates responded substantially within a month to changes in the discount rate whereas commercial loan rates responded much less.

The slopes of the regression lines in Figure A1 are reported along with robust standard errors in column (1) of Table A1. The coefficients imply that, when the Fed increased the discount rate by 100 basis points, the commercial loan rate rose on average by 18 basis points while the brokers' loan rate rose on average by 79 basis points in that month. These responses differ significantly in statistical terms. The row labelled "Test B=C" reports the χ^2 value of the hypothesis test for the equality of the two coefficients. The test rejects the null hypothesis of equality at the 1% level.

The rest of Table A1 demonstrates the robustness of this result. Column (2) adds month fixed effects to the regression to control for any seasonal patterns in the data. Column (3) excludes observations where the NY Fed raised or lowered the discount rate by more than 50 basis points. There are only 8 such observations in our sample but one may be concerned that they are outliers. Column (4) addresses concerns about endogenous changes in the discount rate by restricting the sample to months where the NY Fed's discount rate decision was dominated by factors exogenous to the New York economy. These months include April 1924 through February 1925 (when the NY Fed lowered rates to help Britain return to the gold standard at its pre-war parity), February 1928 through September 1929 (when the NY Fed kept rates constant because its requests to raise rates to control stock market speculation were denied by the Federal Reserve Board), and October 1931 through January 1932 (when the NY Fed raised rates to stem gold outflows from the U.S. and defend the gold standard). Column (5) returns to the full sample and drops the observations for May through September 1931 to control for the possibility that banks in New York City colluded to keep all loan rates constant during the international financial crisis following the collapse of Creditanstalt in Austria in May 1931. Column (6) instead drops the observations for the years before 1920 and the years from 1932 to 1939 to control for direct intervention in credit markets by the federal government via the War Finance Corporation, Reconstruction Finance Corporation, and Federal Reserve Banks in the years during and immediately after World War I and again from 1932 through 1939. Column (7) includes only observations for the prosperous years known as the Roaring 20s (1922 to 1928) to control for the economic turmoil and extreme policies during the downturn after World War I and the Great Depression. Together, the columns in Table A1 reveal a consistent message. Brokers' loan rates responded three to six times as much as commercial loan rates to changes in the discount rate.

Next, we examine the dynamic response of interest rates to Fed policies. Interest rates often respond to policy with a lag. The lags were longer in the past, when the Fed was less experienced, less centralized, and less transparent and the public had to infer Reserve Bank policies from limited information and short histories. We address this issue by regressing changes in loan rates on the contemporaneous change in the discount rate, six months of lagged changes in the discount rate, and month fixed effects. Setting the coefficients on the lagged changes to zero would return the specification in column (2) of Table A1.

The results are reported in Table A2. For each row, columns (3) to (9) indicate how a change in the discount rate in column (2) affects the interest rate in column (1) over various horizons. The first and third rows present the results for brokers' loans and commercial loans in New York City, which are the two interest rates discussed so far. The coefficient in column (3) of the first row implies that the 90-day brokers' loan rate rose on average 78 basis points in months when the discount rate rose 100 basis points. Columns (4) to (9) show no statistically significant change in the brokers' loan rate in subsequent months. The impact of discount rate changes on brokers' loan rates was thus swift and substantial. The response of commercial loan rates was different. The coefficient in column (3) of the third row implies that commercial loan rates in New York City rose on average 18 basis points in months when the discount rate rose 100 basis points. Columns (4) to (9) show that the commercial loan rate continued to rise for at least two and perhaps as long as five months after the change in the discount rate. However, the cumulative response remained well below the response of the interest rate on brokers' loans. Cumulative responses are plotted in Figure A2 to illustrate this point.

The remaining rows in Table A2 explore the impact of discount rate changes on other interest rates, namely the commercial paper rate in New York City (second row), the commercial loan rate in Northern and Eastern cities (fourth row), and the commercial loan rate in Southern and Western cities (fifth row). As discussed in Section 2, commercial paper was used for the same purposes as commercial loans from banks and was of similar maturity. The difference is that borrowers in commercial paper markets were larger, older, and better known, leaving little need for extra information acquisition by the lender. We are also able to consider commercial loan rates outside of New York City. In the decades before the Great Depression, the Fed tracked a sample of these rates and reported an average series for Northern and Eastern cities (with the sample drawn principally from Chicago, Boston, Philadelphia, Pittsburgh, Buffalo, Cleveland, and Detroit) and an average series for Southern and Western cities (with the sample drawn principally from San Francisco, Los Angeles, St. Louis, Dallas, Minneapolis, Kansas City, New Orleans, Seattle, Atlanta, Baltimore, and Richmond).[21] The data averaged for these loan rates span several Federal Reserve Districts so we report results for the policy rate to which the loan rates appear most responsive (Boston Fed for Northern and Eastern cities; Richmond Fed for Southern and Western cities).

The commercial loan rates (NYC, Northern/Eastern, and Southern/Western) respond similarly to changes in the discount rate. They rise slightly in the initial month and continue to rise slowly for the next four to six months. Their cumulative six-month response averages about 44 basis points and, at every month, the null hypothesis that the three series are equally responsive cannot be rejected. The open market rates (brokers' loans and commercial paper) also behave like each other. They respond substantially in the initial month, less in the next month, and not at all after a few months. Their cumulative six-month response averages 83 basis points. The null hypothesis that their cumulative response equals that of the three commercial loan rates is rejected at the 1% significance level. The null hypotheses that their cumulative response equals 100 basis points in the second through sixth month cannot be rejected at that level.

These results appear robust to the issues addressed in Table A1. Hypothesis tests recover similar results, for example, when we limit the regression to the years 1922 through 1929. Overall, it is clear that in response to changes in the discount rate, interest rates on bank loans to commercial customers changed less than interest rates on loans that did not occur in the context of a lending relationship.

[21] Board of Governors 1943, p. 427

Table A1:
Interest Rate Responses to Discount Rate Changes, Contemporaneous

	All	All	±0.5	Exogenous	No Collude	No Govt	Roaring 20s
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Variable							
Δ Commercial (C)	0.18	0.18	0.19	0.18	0.18	0.14	0.14
	(0.04)	(0.04)	(0.04)	(0.04)	(0.39)	(0.04)	(0.08)
Δ Brokers' (B)	0.79	0.78	0.53	0.74	0.78	0.86	0.53
	(0.14)	(0.13)	(0.19)	(0.06)	(0.13)	(0.19)	(0.10)
Test B=C (χ^2)	26.5	27.8	10.4	53.6	27.0	13.3	14.4
Monthly Indicators	×	✓	✓	✓	✓	✓	✓
Observations	239	239	231	34	234	96	84

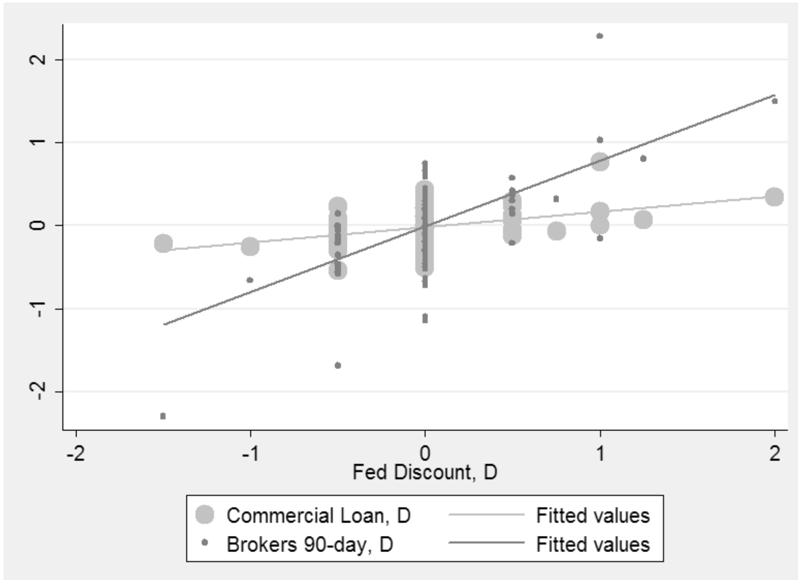
Notes: Robust standard errors in parenthesis. Row "Test B=C" indicates χ^2 statistic from test that coefficient of regression for brokers' loans equals coefficient of regression for commercial loans.

Table A2:
Interest Rate Responses to Discount Rate Changes, Allowing Lags

Dependent Variable (Δ Interest Rate)	Discount Rate (Fed District)	Initial Month	1st Mo. After	2nd Mo. After	3rd Mo. After	4th Mo. After	5th Mo. After	6th Mo. After
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Brokers' Loan, 90-day	New York	0.781***	0.108	0.041	-0.078	-0.039	-0.022	0.010
Commercial Paper	New York	0.507***	0.260***	0.067**	0.022	0.001	-0.013	0.003
Comm. Loan, NYC	New York	0.177***	0.118**	0.072**	0.028	-0.011	0.078***	-0.036
Comm. Loan, N/E Cities	Boston	0.093***	0.098***	0.070***	0.082***	0.093***	0.059**	-0.017
Comm. Loan, S/W Cities	Richmond	0.086***	0.072***	0.059***	0.057***	0.117***	0.030	0.019

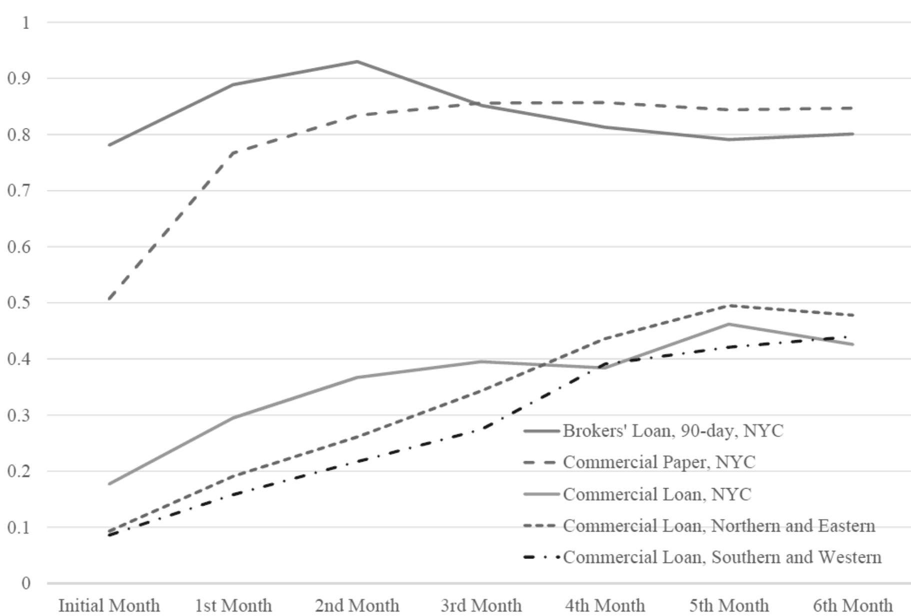
Notes: ***p<0.01, **p<0.05, *p<0.1

Figure A1:
Scatterplot of Contemporaneous Interest Rate Changes



Notes: Scatterplot of loan rate changes on discount rate changes. All changes are in percentage points. Fitted values based on estimates in column (1) of Table A1.

Figure A2:
Cumulative Response of Interest Rates to Changes in Discount Rate



Notes: Shock is a 1 percentage point increase in discount rate in initial month. Cumulative responses, in percentage points, are generated using the estimates in Table A2.

Appendix B – Proof of Proposition 1

This appendix proves Proposition 1 in the main text.

The elasticities of the relationship lending interest rates are:

$$e_1 = \frac{\left[q\xi + \int_\xi^1 p(x)\,dx\right]\overline{R}'(\xi) + \beta(1-s)\frac{\overline{R}'(\xi)}{\overline{R}(\xi)}\int_{\widetilde{\omega}}^{\widehat{\omega}} p(x)\overline{R}(x)\,dx}{\left[q\xi + \int_\xi^1 p(x)\,dx\right]\overline{R}'(\xi) - [p(\xi)-q]\overline{R}(\xi)}$$

and:

$$e_2 = \frac{1-\widehat{\omega}+\widetilde{\omega}+\left[1-\frac{q}{p(\widetilde{\omega})}\right]\frac{r}{p'(\widetilde{\omega})\overline{R}(\widetilde{\omega})+p(\widetilde{\omega})\overline{R}'(\widetilde{\omega})}}{1-\widehat{\omega}+\widetilde{\omega}+\frac{1}{\overline{R}(\widehat{\omega})}\int_{\widetilde{\omega}}^{\widehat{\omega}}\overline{R}(\omega)\,d\omega}$$

while the elasticity of the transactional interest rate is:

$$e_{TL} = \frac{\left[q\eta + \int_\eta^1 p(x)\,dx\right]\overline{R}'(\eta)}{\left[q\eta + \int_\eta^1 p(x)\,dx\right]\overline{R}'(\eta) - [p(\eta)-q]\overline{R}(\eta)} \equiv f(\eta)$$

The denominators of e_1 and e_{TL} are positive under minimal conditions on $p(\cdot)$ (see the online appendix of Hachem (2011) for a formal proof). Accordingly, $e_1 > f(\xi)$ and it will suffice to show $e_2 < f(z)$ for any $z \in (0,1)$ satisfying:

$$\left[qz + \int_z^1 p(x)\,dx\right]\overline{R}(z) \leq r \tag{B.1}$$

The definitions of ξ and η in Section 3 imply that (B.1) holds with strict equality at $z=\eta$ and strict inequality at $z=\xi$. Use the expressions for e_2 and $f(z)$ to rewrite $e_2 < f(z)$ as:

$$\begin{aligned}
&\left[qz + \int_z^1 p(x)\,dx\right]\overline{R}'(z)\left[\frac{[p(\widetilde{\omega})-q]\overline{R}(\widetilde{\omega})}{p'(\widetilde{\omega})\overline{R}(\widetilde{\omega})+p(\widetilde{\omega})\overline{R}'(\widetilde{\omega})} - \frac{1}{\overline{R}(\widehat{\omega})}\int_{\widetilde{\omega}}^{\widehat{\omega}}\overline{R}(\omega)\,d\omega\right] \\
< \;& [p(z)-q]\overline{R}(z)\left[1-\widehat{\omega}+\widetilde{\omega}+\frac{[p(\widetilde{\omega})-q]\overline{R}(\widetilde{\omega})}{p'(\widetilde{\omega})\overline{R}(\widetilde{\omega})+p(\widetilde{\omega})\overline{R}'(\widetilde{\omega})}\right]
\end{aligned} \tag{B.2}$$

Linearity of $p(\omega)$ implies $p(\omega) = p(0) + [p(1)-p(0)]\,\omega$ so we can write (B.2) as:

$$\begin{aligned}
&\frac{[p(\widetilde{\omega})-q]\overline{R}(\widetilde{\omega})}{p'(\widetilde{\omega})\overline{R}(\widetilde{\omega})+p(\widetilde{\omega})\overline{R}'(\widetilde{\omega})} - \frac{1}{\overline{R}(\widehat{\omega})}\int_{\widetilde{\omega}}^{\widehat{\omega}}\overline{R}(\omega)\,d\omega \\
< \;& \frac{[p(z)-q]z}{p(0)-q}\frac{p(z)-q}{qz+\int_z^1 p(x)\,dx}\left[1-\widehat{\omega}+\widetilde{\omega}+\frac{[p(\widetilde{\omega})-q]\overline{R}(\widetilde{\omega})}{p'(\widetilde{\omega})\overline{R}(\widetilde{\omega})+p(\widetilde{\omega})\overline{R}'(\widetilde{\omega})}\right]
\end{aligned}$$

where we have also used $p(0)\,\theta_1 = q\theta_2$. From (B.1):

$$\frac{p(z) - q}{qz + \int_z^1 p(x)\,dx} \geq \frac{[p(1) - p(0)]\,\theta_1 z}{r}$$

so a sufficient condition for (B.2) is:

$$\begin{aligned} & \frac{[p(\widetilde{\omega}) - q]\,\overline{R}(\widetilde{\omega})}{p'(\widetilde{\omega})\,\overline{R}(\widetilde{\omega}) + p(\widetilde{\omega})\,\overline{R}'(\widetilde{\omega})} - \frac{1}{\overline{R}(\widehat{\omega})}\int_{\widetilde{\omega}}^{\widehat{\omega}} \overline{R}(\omega)\,d\omega \\ < \;& \frac{[p(z) - q]\,z^2}{p(0) - q}\frac{[p(1) - p(0)]\,\theta_1}{r}\left[1 - \widehat{\omega} + \widetilde{\omega} + \frac{[p(\widetilde{\omega}) - q]\,\overline{R}(\widetilde{\omega})}{p'(\widetilde{\omega})\,\overline{R}(\widetilde{\omega}) + p(\widetilde{\omega})\,\overline{R}'(\widetilde{\omega})}\right] \end{aligned}$$

Going through the algebra, we can rewrite the sufficient condition as:

$$\begin{aligned} & \frac{p(\widetilde{\omega})\,\widetilde{\omega}^2}{[p(1) - p(0)]\,\widetilde{\omega} + p(\widetilde{\omega})\,\frac{p(0)-q}{p(\widetilde{\omega})-q}} - \frac{q(\widehat{\omega} - \widetilde{\omega})\,\widetilde{\omega}}{p(\widetilde{\omega}) - q} \\ < \;& \frac{[p(z) - q]\,z^2}{p(0) - q}\left[1 - \widehat{\omega} + \widetilde{\omega} + \frac{[p(\widetilde{\omega}) - q]\,\overline{R}(\widetilde{\omega})}{p'(\widetilde{\omega})\,\overline{R}(\widetilde{\omega}) + p(\widetilde{\omega})\,\overline{R}'(\widetilde{\omega})}\right] + \frac{\int_{\widetilde{\omega}}^{\widehat{\omega}}\left[\overline{R}(\omega) - \overline{R}(\widetilde{\omega})\right]d\omega}{[p(1) - p(0)]\,\frac{\theta_1}{q}} \end{aligned}$$

The right-hand side is positive so it will be enough for the left-hand side to be negative. What we want to show is therefore:

$$\frac{p(\widetilde{\omega})\,\widetilde{\omega}}{[p(1) - p(0)]\,\widetilde{\omega} + p(\widetilde{\omega})\,\frac{p(0)-q}{p(\widetilde{\omega})-q}} < \frac{q(\widehat{\omega} - \widetilde{\omega})}{p(\widetilde{\omega}) - q} \tag{B.3}$$

Use $p(\widetilde{\omega})\,\overline{R}(\widetilde{\omega}) = q\overline{R}(\widehat{\omega})$ to isolate:

$$\widehat{\omega} = \frac{[p(0) - q]\,p(\widetilde{\omega})\,\widetilde{\omega}}{q\,[p(\widetilde{\omega}) - q] - [p(1) - p(0)]\,p(\widetilde{\omega})\,\widetilde{\omega}}$$

We can then rewrite (B.3) as:

$$\left[(p(\widetilde{\omega}))^2 + q\,[p(\widetilde{\omega}) - q]\right][p(1) - p(0)]\,\widetilde{\omega} > p(\widetilde{\omega})\,q\,[p(\widetilde{\omega}) - p(0)]$$

This simplifies to $p(\widetilde{\omega}) > q$, which is true. ■