Still Learning After All These Years: Dynamic Information Acquisition in Banking*

Matthew Botsch† Victoria Vanasco‡

December 30, 2016

Abstract

We explore the importance of information transmission in lending relationships between banks and firms. We develop a general methodology to test for bank learning, requiring the construction of a proxy variable from the future that is correlated with borrower quality but unobservable by the lender, so it cannot be used directly to price loans. Using data on syndicated loans originated between 1987 and 2003, we find strong evidence of dynamic information acquisition. Bank learning particularly benefits higher-quality borrowers, who receive lower interest rates on subsequent loans.

Keywords: banks, information acquisition, relationship lending, learning, syndicated loans.
JEL Classifications: D83, G21.

*We are indebted to the excellent research assistance of Mory Elsaify and Christopher McConnell. In addition, we would like to thank Vladimir Asriyan, Gabriel Chodorow-Reich, Darrell Duffie, Peter Koudkijs, Ross Levine, Doron Levit, Ulrike Malmendier, Christine Parlour, José Luis Peydró, Joshua Rauh, David Romer, Michel Serafinelli, Amit Seru, James Wilcox, Xiaoyu Xia, and seminar participants and discussants at Berkeley economics, Berkeley Haas, Stanford GSB, Universitat Pompeu Fabra, Santa Clara University, Wesleyan University, EconCon 2013, and the EuroFIT Research Workshop on Corporate Loans 2014 for helpful conversations and comments. The usual disclaimer applies.

†Bowdoin College, email: mbotsch@bowdoin.edu
‡Stanford University, email: vvanasco@stanford.edu.

1 Introduction

Modern economies rely on well-functioning financial markets to ensure that firms can raise funds and invest in profitable projects. Informational asymmetries between borrowers and lenders, however, hinder the functioning of credit markets –for example, a good entrepreneur will have difficulty raising funds if she cannot credibly communicate the quality of her project to investors.[1] A lengthy academic literature argues that banks are uniquely situated to ameliorate information frictions by centralizing the role of information acquisition in credit markets (Fama 1985; James 1987; Williamson 1987; Sharpe 1990), and are therefore essential to economic development. This is consistent with evidence of banks being one of the most important sources of firms' external financing (e.g. Demirgüç-Kunt et al. (2012); Degryse et al. (2015)). In view of this, understanding the determinants of bank lending is central for the study of firms access to finance and, more broady, economic growth.[2]

Banks perform exhaustive due diligence to obtain both hard and soft information about potential borrowers before making their lending decisions.[3] Banks, however, not only obtain private information at the loan screening stage (e.g., Mester et al. (2007); Agarwal and Hauswald (2010); Norden and Weber (2010)), but also through monitoring and servicing previously-issued loans to the same borrower and through time (e.g., Degryse et al. (2009); Ioannidou and Ongena (2010)). There is an extensive banking literature that has highlighted the importance of firm-bank relationships for a firm's access to finance (Petersen and Rajan (1995), Bharath et al. (2009), Goel and Zemel (2015a), Karolyi (2017), Kysucky and Norden (2015)). The main result

[1]Akerlof (1970) shows that under many scenarios, the presence of adverse selection can cause markets to break down: only bad entrepreneurs are able to raise funds (if bad projects have positive net present value (NPV)), or no one is able to raise funds (if bad projects have negative NPV).

[2]For the study of the relation between access to finance and growth see Levine (1997); Beck and Levine (2004); Beck and Demirguc-Kunt (2006).

[3]"Hard" information refers to information that is quantifiable and/or verifiable, while "soft" information includes perceptions, opinions, rumors, feelings, and other values which are harder to quantify or verify.

from this literature is that by establishing relationships with their banks, firms can improve their access to credit through lower rates or larger loan amounts. The result is robust to controlling for quantifiable measures of firm creditworthiness (e.g. Z-score, distance to default, leverage, ROA) and appears to be stronger for less transparent firms. These results combined provide powerful evidence of banks' ability to acquire valuable information through the development of firm-bank relationships.

Many questions, however, remain unanswered, such as: What do banks learn while establishing firm-bank relationships? Do different quality firms benefit differently from establishing relationships? In principle, firms with weaker fundamentals should benefit less from bank learning than those with stronger fundamentals. In this paper, we address these questions by developing a methodology that allows us to test for bank learning in a more direct manner. We construct a variable from the future that is correlated with a firm's creditworthiness and that is unobservable by the bank in real time, so it cannot be used directly to price loans. We refer to this variable as a "proxy," since it proxies for information about firm quality that banks may want to learn. We find that the loading on this proxy in the banks' pricing equations increases over relationship time, as we would expect if banks were learning about factors correlated with the firm's creditworthiness.

Our results present strong evidence that banks acquire relevant information about borrowers over relationship time. We interpret this as banks acquiring hard-to-document information that results from qualitative analysis that are based on ideas, opinions, rumors, feedback, or anecdotes that cannot be easily transmitted or verified. This information allows the bank to better understand the underlying factors that drive our proxy variable, which is correlated with firm creditworthiness. As a result, our methodology allows us to understand what type of information banks acquire – since our proxy captures something that banks are learning about. In addition, our proxy allows us to characterize the cross-sectional effect of learning on loan terms, which we show depends on the (ex-ante) unobserved creditworthiness of the borrower.

Our data come from the syndicated loan market, a unique setting where consortia of banks supply loans to large, corporate borrowers. We merge Reuters LPC DealScan data on syndicated loans with the financial characteristics of borrowing firms from the Compustat-CRSP Merged database. We construct a panel of approximately 5,794 lender-borrower pairs (relationships) that take out on average 3.4 loans between 1987 and 2003. We restrict the sample to borrowers who are still in existence in 2008, so we are able to observe *ex post* firm performance. Since the borrowers in this market tend to be large and transparent, private bank information probably matters less than in other commonly-studied contexts, such as small business and consumer lending. However, we show that learning matters even for large public borrowers that file detailed financial statements on a quarterly frequency. The average firm in our sample has over $12 billion in inflation-adjusted (year 2000) assets. The fact that banks appear to rely on the information they acquire over time, even for these borrowers, reinforces the importance of financial intermediaries as delegated information acquires.

Our methodology to test for bank learning is inspired by Farber and Gibbons (1996), who test for public learning by employers in labor markets. Farber and Gibbons (1996) show that wages are increasingly predicted by a worker's Armed Forces Qualifying Test (AFQT) score as a worker's tenure with a firm increases, even though the employer never observes this score. Since the AFQT score is presumably correlated with worker aptitude or ability, this supports the idea that employers learn about worker quality over time. As in their paper, our methodology is designed to detect information generation as the relationship between banks and a firm intensifies. We test for learning using a proxy variable that is correlated with borrower creditworthiness (which we also refer to as quality) but cannot be observed by banks in real time. If banks' pricing decisions are increasingly predicted by our unobservable proxy over time, then it must be that banks are acquiring information contained in our proxy through some other source.

This basic methodology is well-known in labor economics but has received relatively little attention in financial economics. Our innovation is to propose a general way to construct such proxies that can be applied in a wide set of economic environments.

To find a variable that is correlated with borrower creditworthiness but not observed by banks in real time, we travel back to the future – specifically, to five years after the last loan in our sample is originated. Our proxy is the firm's differential stock market response to a large adverse shock, the Lehman Brothers bankruptcy filing in 2008. This strategy relies on the stock price adjustment in the months after the shock containing relevant information about the firm's performance during the crisis, actual or expected. From an ex-ante perspective, if banks could observe the future, they would prefer to lend to those firms that will perform relatively better in adverse times. Thus, our proxy captures information that banks would like to have to evaluate a firm's creditworthiness; that is, its relative performance in tale events. Finally, as in Farber and Gibbons (1996), we guarantee that this proxy is not correlated with omitted, publicly-observable firm characteristics by orthogonalizing it to the bank's information set at relationship time zero. Thus, we have an orthogonalized proxy that is correlated with the firm's creditworthiness and uncorrelated with the bank's information set used to price the first loan between a bank and the firm in our sample.

The essence of our methodology is to run a regression of loan-level outcome variables (the spread over LIBOR or the quantity of credit extended) on both observable firm characteristics and on our orthogonalized proxy variable. We find that banks act as though they increasingly price on the proxy over relationship time, even though they cannot be directly observing the proxy since it is constructed with data from far into the future. If banks are increasingly pricing on our proxy, it must be that the proxy is effectively measuring unobservables (to the econometrician) that banks observe over relationship time, and that affects a firm's access to finance. We take this as evidence that banks are learning about the information contained in the proxy through some other source. Iyer et al. (2014) recently use a similar methodology to test for information acquisition at the loan screening stage. To our knowledge, we are the first to test directly for *dynamic* bank learning, i.e. over relationship time.

We test for and rule out other possible violations to our identification strategy. Although we remove the influence of any omitted borrower characteristics at relationship time zero, one might be concerned that the proxy is correlated with (public) omitted borrower characteristics that are time-varying. This is a valid concern, in particular since our proxy is constructed with information from the future. To address this, we run a separate regression in which we interact the proxy with calendar time, as opposed to relationship time. We find that the proxy interacted with calendar time is not significant; that is, it is not the passing of time – or the approaching future – that drives our main result.

Another important concern is the possibility of reverse causality. De Mitri et al. (2010)and Bolton et al. (2016) suggest that firms in longer relationships had improved access to bank loans during the credit crunch, enabling them to perform better during the crisis. However, we do not find strong evidence that relationship length during our sample (through year-end 2003) helps predict firm performance in 2008. To further address this concern, we construct our proxy so that it is orthogonal to firms' access to credit during the financial crisis.

Our identification strategy relies on two assumptions: first that the stock market response of our firms to the Lehman shock contained valuable information about creditworthiness, and second that banks did not have access to all of the information contained in the proxy at the beginning of our sample. Our second assumption partially rests on timing: because the proxy is from the future, banks could not have direct access to the proxy variable itself (although some of the information contained in the proxy might already be public knowledge). An implication of this identification strategy is that banks should incorporate all relevant information that was public prior to the loan origination date. We test this implication by running a series of placebo tests, using firms' stock market responses on arbitrary dates in the past rather than the future. Once we have orthogonalized these placebo proxies to initial loan terms, banks do not increasingly price on them, suggesting that any relevant information has already been incorporated into the first loan's terms.

We proceed to study the effect of bank learning on the cross-section of firms. We investigate the dynamics of loan pricing over time for high-quality versus low-quality borrowers and find that the benefits of bank learning affect high-quality firms the most. Why do so many borrowers form relationships with one lender, despite the risk of monopolistic lock-in pricing or hold-up problems (Goel and Zemel (2015a))? Depending on the importance of borrower switching costs and lender financing costs, banks might find it desirable to charge good borrowers less or to charge bad borrowers more as they learn borrowers' types. To distinguish between these two effects, we separate our proxy into its positive and negative components and estimate learning coefficients separately for "good" versus "bad" types. The coefficient on good types is strongly negative and significant, while the coefficient on bad types is statistically indistinguishable from zero. This suggests that pricing benefits accrue to good borrowers who are hard to distinguish from bad borrowers using public information. Both types are initially pooled together and pay high interest rates. As banks learn about which borrowers are more creditworthy, the good types benefit from cheaper subsequent loans, while the bad types continue to pay the high initial interest rate. We also investigate the cross-sectional dynamics of loan quantities and find that learning matters more in the bottom half of the distribution, suggesting that some of these borrowers face credit contraints.

Our setting does not let us make strong statements about whether bank learning is public (i.e., shared by all banks) or private (excludable to the originating lender or consortium of lenders). We see very few loans originated by multiple lead banks to the same borrower in the same calendar year, and since our proxy only varies across firms, the only source of relationship-level variation comes from the orthogonalization procedure and the initial loan terms. Moreover, information embodied in loan terms could easily spread across the entire market once the terms are revealed in the weekly *Gold Sheets* publication or a DealScan database update. We accordingly focus on tests of "public," market-wide learning in the majority of our specifications. However, we run two tests inspired by the relationship banking literature to test for any knock-on private learning. First, we interact our learning variable with a relationship loan indicator following the definition in Bharath et al. (2009). Second, we interact our learning variable with common indicators for borrower transparency (the relationship literature suggests that benefits accrue more strongly to less transparent borrowers). We do not find any strong evidence of dynamic, private bank learning – in addition to marketwide learning – in this setting.

Related Literature. Our paper is motivated by the extensive literature on relationship lending, that has established that (i) there is something special about bank lending; and (ii) longer bank-firm relationships are correlated with cheaper access to credit. Berger and Udell (1995) study credit lines to small firms and find that borrowers with longer bank relationships pay lower interest rates and are less likely to pledge collateral. Similar results are found by Bharath et al. (2009); Degryse and Van Cayseele (2000); Degryse et al. (2015); Ioannidou and Ongena (2010); Sufi (2007); Karolyi (2017); Norden and Weber (2010); Slovin et al. (1993). Petersen and Rajan (1995) find that firm-bank relationships improve the availability of credit to the firm, but they do not find a strong impact on pricing. Our results reconcile these findings: while longer firm-bank relationships are associated with larger loan amounts for all types of firms, loan spreads are reduced over time only for high quality borrowers. In addition, event studies have found evidence of bank loans being harder to substitute with other forms of credit. Slovin et al. (1993) examine the stock price of borrowing firms after the announcement of the failure of their main bank, Continental Illinois. They find that Continental borrowers incurred negative abnormal returns of 4.2% on average. If bank loans were indistinguishable from corporate bonds, borrowers could access funds directly from the market when their bank disappeared. Similarly, if banks were perfectly substitutable, the failure of one lender should have no impact on borrowers' stock prices. Slovin et al. conclude that Continental had private information about the borrowers unavailable to the rest of the market. Gibson (1997) reaches a similar conclusion by studying the effect of Japanese banks' health on borrowing firms.

Our results contribute to the existing literature on private information in credit markets. Sufi (2007), Liberti and Mian (2009), and Agarwal and Hauswald (2010) find that smaller distance (both hierarchical or geographical) facilitate the transmission of relevant soft information about borrowers' default likelihood. There is also substantial evidence on the presence of securitizers' private information in the mortgage market in the years leading to the financial crisis (Keys et al. 2010, Agarwal and Hauswald 2010, Jiang et al. 2014, Botsch 2015, and Rajan et al. 2015). Consistent with these findings, Iyer et al. (2014) document the importance of soft information in predicting default in new online credit markets. Bolton et al. (2016), using detailed credit register information for Italian banks around the Lehman Brothers default, find that *relationship* banks charged a higher spread before the crisis, but offered more favorable lending terms in response to the crisis while suffering fewer defaults. They interpret this as the result of the informational advantage of relationship banking. These papers find evidence of private or soft information in banking. Our paper extends these results by providing evidence of banks continuously acquiring private information about their borrowers by establishing firm-bank relationships. In contrast to these papers, our paper focuses on the presence of dynamic private information acquisition in banking.

In addition, we can shed some light on the discussion of relationship lending and bank competition. As Sharpe (1990) and Petersen and Rajan (1995) argue, the benefits from relationship lending are tamed by the presence of bank competition. Rajan (1992) presents a model where the benefits from bank financing are counteracted by the bank's increasing bargaining power over the firm. However, a bank's ability to extract surplus from its borrowers is limited by the presence of bank competition. If the monopoly over private information dominates the competitive forces, we should observe banks charging higher spreads to their borrowers as their relationship intensifies. In this context, how do firms benefit from developing firm-bank relationships? Do banks extract all the surplus generated by the acquisition of private information? Schenone (2010) addresses these question by studying loan pricing before and after events that release public information about borrower quality (IPOs) and thus level the playing field. She finds that relationship banks exploit their informational advantage by charging higher interest rates than those that would prevail were all banks symmetrically informed. Our findings, however, suggest that firms also obtain some of the surplus. In particular, we find that relatively better firms see their interest rates decrease over relationship time.

The rest of the paper is laid out as follows. Section 2 presents a simple model to provide theoretical foundations for our empirical exercise. In Section 3 we discuss the nature of our dataset and the construction of the proxy variable. Section 4 presents the main empirical results and several robustness checks. In Section 5, we explore two extensions: whether there is any heterogeneity in relationship benefits across loan types, and what type of borrowers benefit the most from relationship lending. Section 6 concludes.

2 Framework for the Empirical Strategy

2.1 A Simple Borrower-Lender Model

In this section we present a simple model of firm borrowing to discuss the determinants of loan agreements and the role of information in credit markets. Consider the problem of a risk-neutral bank that needs to decide whether to make a given loan, and how much to charge for it. Assume that the per-unit cost of funding for the bank is given by rate R_f. If a firm approaches a bank asking for loan of amount L, the bank will compare its expected return from making that loan to its cost of funding. Let π be the probability of the firm defaulting on the requested loan. We assume π is the firm's private information. The bank's beliefs about the firm's default probability are given by $E[\pi|I]$, where I is the information set of the bank at the time of making the loan. The bank determines the loan interest rate, R, and the percentage of the loan to be collateralized $c \in [0, 1]$ so that the following constraint is satisfied:

$$(1 - E[\pi|I]) R + E[\pi|I] c \geq R_f \tag{1}$$

The interest rate is set so that the expected return for making the loan exceed its cost of funding. Equation (1) describes a lower bound for the spread over its funding cost the bank can charge the firm: $S \equiv R - R_f \geq \frac{E[\pi|I]}{1-E[\pi|I]} (R_f - c) \equiv \underline{S}$. This simple model predicts that the minimum spread requested for a given loan increases with the expected default probability, $E[\pi|I]$, and the bank's funding rate, and that it decreases with the percentage of the loan being collateralized. All of these results are standard and very intuitive. We will assume that the bank's information set I at time t is given by a collection of observed signals about the firm's probability of default $s^t = \{s_0, ..., s_t\}$. In particular, if establishing a relationship with a firm allows the bank to observe private information about the firm's fundamentals, the bank should use this information to update its beliefs and recompute the required minimum spreads in subsequent loans. For example, if the bank learns good things about the firm, the lower bound on the spread should increase, and vice versa if bad news are received. Although the determination of the actual spread, S, could be a very complicated process, the behavior of the lower bound is sufficient to describe our main mechanism. For the actual spread to respond to changes in the bank's adjusting information set, it is only necessary that some of the surplus arising from the lending contract accrue to the firm.[4] Most importantly, this suggests that what we are able to measure is only a lower bound for learning, since some information could have affected the lower bound on spreads, but not the actual spread, which is what we observe.

2.2 Empirical Methodology

A linearized version of the model motivates the following reduced form pricing equation around the true default probability π[5]:

$$r \approx \alpha_0 + \alpha_1 E[\pi|I] + \gamma' w \tag{2}$$

where w denotes other loan characteristics. To describe our empirical strategy, we decompose the bank's information set about a given firm into three types of variables: $I_{fb,\tau} = \{x_{f,t_\tau}, z_{f,t_\tau}, s_{fb}^{\tau}\}$, where b denotes the bank, f denotes the firm, and τ denotes the length of the bank-firm relationship . The vector $x_{f,t}$ represents publicly-available characteristics of firm f at calendar time t that are observed by the bank but not by the econometrician (omitted variables). $z_{f,t}$ are public firm characteristics observed by both the bank and the econometrician (included variables). The set $s_{fb}^{\tau} = \{s_{fb,t_0}, s_{fb,t_1}, ..., s_{fb,t_\tau}\}$ represents the collection of private signals that only bank b observed during its relationship with firm *f*. The number of private signals is increasing in relationship length τ. For expositional purposes, suppose that firm characteristics $(x'_{f,t}, z'_{f,t})$ and other loan features $w_{l,fb,\tau}$ are time-invariant, so the "t" and "τ" subscripts may be suppressed.[6] We relax this assumption in the empirical section of the paper.

[4] Schenone (2010) shows that banks do, to some extent, exploit their monopoly power over their private information by charging higher interest rates than when information is symmetric across firms. What we need for our empirical strategy to succeed is that banks adjust their loan pricing on the arrival of private information, i.e., that the surplus arising from the relationship is somehow shared.

[5] For example, a first-order Taylor series expansion gives $r_{l,fb,\tau} = \frac{-\pi}{1-\pi} + \frac{1}{\pi(1-\pi)} p_{fb,\tau} + \log(R^F - c_{l,fb}) + o(p_{fb,\tau} - \pi)$.

[6] The "l" subscript on w counts if there are multiple loans between the same bank-firm pair at the same point in time.

The bank forms a forecast $p = E[\pi|I]$ of the firm's true, unobserved default probability π by conditioning on its information set $I = \{x, z, s\}$. Thus, $\{x, z, s\}$ should contain information about default probabilities that is relevant for the determination of loan prices, r, and other terms, w. As econometricians, we observe r, w, and z, and we would like to make inference about the link from π to r generated by the bank's private information, s. This channel represents banks acquiring private information about firm default probabilities relevant for loan pricing and used in their forecast model. The crucial identification problem is that there is a second channel from π to r via x that we need to control for. This information structure is depicted graphically in Figure 1.

Motivated by our linearized model (2), we could estimate the following population linear projection:[7]

$$\begin{aligned} E^*[r_{fb,\tau}|z_{f,t}, w_{fb,\tau}] &= \alpha_t + \alpha_i + \alpha_1 E^*[E[\pi|x_{f,t}, z_{f,t}, s^{\tau}_{fb}]|z_{f,t}, w_{fb,\tau}] + \gamma' w_{fb,\tau} \\ &= \alpha_t + \alpha_i + \alpha_1 E^*[\pi|z_{f,t}, w_{fb,\tau}] + \gamma' w_{fb,\tau} \\ &= \alpha_t + \alpha_i + \beta^{z\prime} z_{f,t} + \beta^{w\prime} w_{fb,\tau} \end{aligned} \tag{3}$$

The coefficient on w reflects both the substitutability between other loan characteristics and interest rate spreads (γ) and the correlation between w and omitted firm characteristics x and private signals s.[8] Similarly, the coefficient on z incorporates both direct and indirect pricing effects due to omitted variables.

What if an econometrician could include the true default probability π in a panel regression along with observable characteristics (z'_f, w_{fb})? At relationship time 0, there would be a positive loading on π because of omitted variable bias: the bank's internal model includes variables x_f that are relevant for forecasting default probabilities and setting loan spreads. As a relationship progresses, the bank observes additional signals

[7]See Appendix for detailed descriptions of the assumptions that are needed to make the empirical strategy valid

[8]In our empirical specifications we find that the second factor dominates. For example, loans with more collateral pay higher interest rates, presumably because these firms differ on omitted characteristics.

$s_{fb,t}$ that contain additional information about π not available in $\{x_f, z_f\}$. That is, the loading on π would increase over the course of the relationship due to private bank learning. This observation is at the heart of our empirical strategy, inspired by Farber and Gibbons (1996). These authors focus on learning and wage dynamics and show that time-invariant variables correlated with ability but unobserved by employers are increasingly correlated with wages as a worker's experience increases. In this paper, we instead focus on interest rate dynamics and show that time-invariant variables correlated by firms' fundamentals but unobserved by banks are increasingly correlated with interest rates as the bank-firm relationship increases.

The main drawback of this approach is that we do not observe the real probability of default estimated by the bank, π. As in Farber and Gibbons (1996), we address this issue by incorporating a variable b_f that proxies for the firm probability of default π, but that is not in the bank's information set I. In Figure 1, b_f is connected to π but not p. This variable is correlated with π but is not observed by banks, so it cannot be used to set loan prices. However, we expect that b_f is unconditionally correlated with publicly-observable variables $x_{f,t}$ that we omit from our pricing equation but the bank uses in its forecast model $p_{f,\tau} = E[\pi_{f,t}|x_{f,t}, z_{f,t}s^{\tau}_{fb}]$. To remove this dependency, we regress b on all observable firm characteristics and on the interest rate of the first loan in each relationship in our dataset, and take the residual. Conditioning on the latter ensures that the residual is orthogonal to all the information held by each bank at the start of each relationship in our sample, including x_{f,t_0}. Specifically, let:

$$b^*_{fb} = b_f - E^*\left[b_f|z_{f,t_0}, w_{fb,0}, r_{fb,0}\right] \tag{4}$$

This process removes the influence of any information the bank may have used to price its first loan to a firm from the original background variable, b_f. If information is non-excludable, we may remove the "b" subscripts and instead test for marketwide learning by estimating equation (4) on each firm's first syndicated loan in our sample. Consider adding b_f^* as a regressor to 3 with a slope that is allowed to vary over relationship time:

$$r_{f,\tau} = \alpha_t + \alpha_i + \beta' z_{f,t} + \gamma' w_{f,\tau} + \delta_\tau \cdot b_f^* + \varepsilon_{f,\tau} \tag{5}$$

We are interested in studying the evolution of the coefficient δ_τ calculated cross-sectionally across firms at the same market timeτ. By construction $\delta_0 = 0$. As banks receive additional signals s_f^τ, non-quantifiable information becomes increasingly important in their internal forecast model $E[\pi_{f,t}|x_{f,t}, z_{f,t} s_f^\tau]$. To the extent that b_f^* is correlated with these signals, the coefficient δ_τ should increase in magnitude with the number of signals and the length of the relationship τ.

3 Data

3.1 Sample Description

We use the DealScan database on syndicated loans from Reuters LPC (April 2012 vintage) to construct a panel of lender-borrower pairs ("relationships") observed repeatedly over time. DealScan provides data for approximately 176,000 contracts comprising 248,000 syndicated loans made between 1981 and 2012, but the coverage between 1981 and 1987 is extremely limited; more than 99% of loans in the database start in 1988 or later. Syndicated loans are between a single borrower and a syndicate of lenders. One lender acts as the lead arranger and negotiates contract terms for the entire group. Most of the lenders are large commercial banks, but many syndicates include non-bank financial companies. After the contract is agreed to, a lender referred to as the agent monitors the performance of the loan. The lead arranger and agent can be different members of the syndicate (but in our final sample, they are the same in 99% of cases). Lenders playing an active role in arranging loan terms have greater incentives to acquire borrower information than passive members of the syndicate. The lead lender in a syndicate is fairly persistent, even if other members change. 62 percent of lenders who served as a lead bank once serve as lead bank in every interaction with a given borrower. Each contract (or "package") can include multiple loans (or "facilities") made at the same time. A typical example is a borrower receiving both a term loan and a revolving line of credit.

Our exercise requires information on loan prices and firm financial characteristics that the bank might use to set interest rates. We obtain borrower financial data from Compustat using the link file created by Chava and Roberts (2008).[9] This requires the borrowers in our sample to file statements with the SEC, and reduces the sample size to 61,000 facilities. Our measure of loan price is the all-included drawn spread over LIBOR, which is the price including fees that a firm would pay if it drew upon 100% of its line of credit (for revolving loans) minus the spread over LIBOR including fees for term loans. Many of the rows in the DealScan tables contain missing values. Dropping loans without an all-in spread reduces our sample to 41,000 facilities. Since our proxy variable is constructed from market data, we further require that the borrowers appear in CRSP both in the year that the deal was consummated and over the six-year period 2003-2008.[10] This is a costly requirement and reduces our sample to 10,000 facilities. These data requirements restrict the sample to include only larger, more followed, and presumably more transparent firms. This should bias against finding any role for private bank learning. We exclude all loans with a start date after December 31, 2003 to ensure the unobservability of our 2008-based proxy variable (see below).

We follow DealScan borrower IDs to count every syndicated loan a borrwer takes out over time. There may be multiple observations at a particular moment in "market time" if a package contains multiple loan facilities (or occasionally, if a borrower takes out multiple packages on the same date). Since we care about the information set available

[9] We use the version of the link published on August 27, 2010, and made available on Wharton Research Data Services.

[10] We use the CRSP-Compustat Merged database to link between the two sources.

to banks at the time of the agreement, we order interactions by package date (variable "deal active date") rather than by each facility's specific start date. The market time counter begins at zero and starts running after the first interaction we observe with no missing loan terms. This counts loans that the same borrower takes out with different syndicates and lead banks. The average borrower in our sample takes out loans with 3 distinct lead banks. However, we follow Bharath et al. (2009) and use a broad definition of "lead bank" that counts, on average, 1.5 lead banks per syndicate, so this implies limited switching – on average, only two distinct syndicates per borrower.

Our final dataset has 8,673 facilities and 5,989 relationships between 1,992 unique borrowing firms and 807 unique lead banks. The deal active dates span the years 1987 to 2003. The average lead bank-firm relationship lasts 3.4 interactions (approximately five years in calendar time), and 10% of relationships last 7 or more interactions (approximately nine or more years). We report other summary statistics about the final "Lehman" sample of loans and relationships in Table 1. We also compare our sample to the more common sample of all public firms, removing the requirement that borrowers remain public through 2008. This requirement biases our sample toward larger, more transparent, and less risky borrowing firms. All nominal variables are deflated using the quarterly GDP implicit price deflator to constant 2000 dollars.

3.2 Observable Firm Characteristics

Our model requires that we condition on a subset of financial characteristics used by the bank in setting loan prices, $z_{f,t}$. Ideally these variables would be inclusive, so we would not have to worry about correlation between omitted variables $x_{f,t}$ and our proxy variable b_f (see the discussion below). We focus on a set of variables suggested by the corporate bankruptcy literature, since measures that are useful for predicting bankruptcy or default on debt obligations should also be relevant for loan pricing.

The first measure is Altman's Z score, denoted by Z. Altman (1968) investigated the determinants of corporate bankruptcy for a sample of 33 manufacturing firms that filed for bankruptcy between 1946-1965 and 33 firms still in existence in 1966 based on random stratified matching by industry and size. He uses discriminant analysis to estimate the following index:

$$Z = (1.2 \cdot WC + 1.4 \cdot RE + 3.3 \cdot EBIT + 0.6 \cdot MVE + .999 \cdot S)/AT \tag{6}$$

where WC is working capital, RE is retained earnings, EBIT is earnings before interest and taxes, MVE is market value of equity, S is sales, and AT is total assets.[11] Altman concludes that "firms having a Z score of greater than 2.99 clearly fall into the 'non-bankrupt' sector, while those firms having a Z below 1.81 are all bankrupt" (p. 606). So lower values of Z indicate an increased likelihood of bankruptcy. We Winsorize the top and bottom 0.5% of Z-score observations using the sample of all DealScan firms for which we have data over the years 1985-2012.

The second measure is the distance to default, denoted by NPD. This measure comes from the observation in Merton (1973, 1974) that the Black and Scholes (1973) options pricing model may also be used to calculate the market value of assets in place, by viewing the observed equity price as a call option on the unobserved market value of the entire firm. Once the market value of assets in place V_A has been estimated, a firm's probability of default T periods into the future is the probability that the value of its assets will drift below the "strike" price–the book value of liabilities. Since the Merton model assumes that V_A follows a geometric Brownian motion with deterministic drift μ and volatility σ_A, this probability is given by

$$P(V_{A,t+T} \leq L_t | V_{A,t}) = \Phi\left(-\frac{log(V_{A,t}/L_t) + (\mu + \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}}\right)$$

[11] There is an error in the placement of a decimal point in the original 1968 paper. The correct formula is given in subsequent papers–e.g., Altman (1984).

To calculate this exact probability, one must solve the Black-Scholes equations for V_A and σ_A. Rather than using a numerical solver, we use the "naive" alternative proposed by Bharath and Shumway (2004, 2008). This naive probability of default uses simple rules of thumb for variables in the formula above: L_t is the book value of debt in current liabilities plus one-half the book value of long-term debt; V_A is the sum of market value of equity plus book value of liabilities; equity volatility σ_E is the annualized standard deviation of the previous year's daily stock returns; debt volatility $\sigma_L = .05 + .25 \cdot \sigma_E$; and total firm volatility is the weighted sum of σ_E and σ_L. We solve for the naive probability of default for firm f at time t, $NPD_{f,t}$ for a one-year time horizon. In all tables and regressions, we truncate the probability of default to take values in the range $[0.001, 0.999]$.

3.3 The Bank Information Proxy

A good background variable b_f cannot be in the bank's information set at any time and it must be correlated with the firm's unobservable quality. Our candidate background variable is the differential response of the firms in our sample to a large negative aggregate shock: the onset of the financial crisis and the collapse of Lehman Brothers in September 2008. Specifically, we consider the idiosyncratic component of firms' stock returns in the three months around the Lehman Brothers bankruptcy. By using equity market data from five years after the last loan in our sample was made, we guarantee that the proxy cannot have been observed by banks in real time. Lehman's bankruptcy filing was a "shock" in the sense that it was not foreseen by market participants.[12]

Our identification strategy requires that idiosyncratic stock returns around the Lehman filing were partially driven by firms' latent quality. Suppose that during booms it is hard to differentiate good firms from bad firms, while during busts lemons are eas-

[12]When Bear Stearns failed six months earlier, the Fed and the Treasury avoided a regular bankruptcy process and arranged its sale to JP Morgan Chase, precisely to ameliorate turmoil in financial markets.

ier to identify. Those firms that perform relatively better during crises are spotted as high-quality firms, and investors should incorporate this information into the stock price. Moreover, the returns to identifying lemons might be greater in crisis states of the world; in booms all firms do well, while in busts only good firms do well. The forward-looking nature of stock market prices is well-suited to capture any new information about firm quality revealed during the financial crisis. Of course, a component of firms' stock returns during this period undoubtedly reflects subprime-crisis-specific exposure. To the extent that subprime exposure is industry-specific, we can remove this influence with industry fixed effects. Our identifying assumption is that at least part of firms' idiosyncratic returns are due to underlying firm characteristics that were revealed after Lehman, and not to subprime-crisis-specific risk exposure. We do not interpret loadings on the proxy as changes in the perceived probability of a Lehman-style crisis occurring, as we find it implausible that this risk was priced in loans made a decade or more in advance.

We construct b_f as follows. We compute the cumulative abnormal return (CAR) of each firm in a [-21, +42] day window centered around the collapse of Lehman:[13]

$$b_f := \sum_{s=-21}^{+42} \left(R_{f,s} - R^F\right) - \hat{\beta}_f' \left(R_{factor,s}\right)$$

where $R_{f,s}$ and $R_{factor,s}$ denote the daily returns on a firm's stock and the four Fama and French (1993) - Carhart (1997) factors at time s, R^F denotes the risk-free rate, and $s = 0$ on September 15, 2008. The factor betas are estimated from time-series regressions of daily excess stock returns over 2003-2007:

$$R_{f,t} - R^F = \alpha_f + \beta_f' \left(R_{factor,t}\right) + \varepsilon_{f,t}$$

With each firm's CAR in hand, we construct the final **bank information proxy** using

[13] Starting on August 14 and ending on November 12.

equation (4). We define market time 0 as the time of the first syndicated loan for each firm in our sample. We note that for many firms, it is likely that the first observation in our sample is not their first loan. Orthogonalization removes the influence both of omitted variables and of any bank learning that may have occurred prior to the beginning of our sample period. To the extent that learning is diminishing over time, the inclusion of more-mature relationships will bias our estimates toward zero.

The orthogonalization guarantees that b_f^* is uncorrelated with relevant omitted firm characteristics used in loan pricing at the beginning of the sample, x_{f,t_0}. However, an identification problem would arise if innovations in unobservable firm characteristics, $x_{f,t}$, reveals information about changes in default probabilities. That is, since b_f^* is from the future, the proxy could simply be picking up future innovations in a firm's default probability that are correlated with subsequent movements in omitted, publicly available variables. While we cannot completely rule this explanation out, we run a battery of robustness checks to rule out specific violations, including: unobserved time-invarint firm characteristcs, bank characteristics, match characteristics, time-varying firm characteristics, and time-varying coefficients.

4 Results

In this section, we proceed to test whether banks learn about customers as evidenced by an increasing loading on b_f over market time. We discuss and rule out several alternate explanations that might explain our findings, including reverse causality, firm-specific and lender-specific omitted variables, match quality, and time-varying loan pricing coefficients. Our results are consistent with the model described previously. We find robust evidence that banks learn about unobserved firm characteristics while in a relationship.

4.1 Orthogonaliztion

The LPM coefficients from equation (4) are presented in the first column of Table 2. Note first that the all-in-spread at market time zero is negatively correlated with the Lehman proxy, even after controlling for observable firm characteristics. A firm paying an additional 100 basis points on its first loan in our dataset is expected to experience an additional 1.7 percentage point negative CAR in the three-month window around Lehman. This indicates that initial loan prices contain omitted information that is correlated in the correct direction with the proxy variable. The bank information proxy b^*_{fb} is simply the residual from this regression.[14]

We also include controls for extensive and intensive margins of credit access in 2006-09. This is to address any reverse causality running from better credit access during our sample period to better credit access and superior market performance during the financial crisis. We include a dummy variable for whether each borrower received any syndicated loans over 2006-09, and if the firm received credit, the number of previous loans the borrower received from that same lead bank in the entire LPC DealScan database. If the borrower received loan from multiple lead banks over that time period, we take the maximum relationship length as our measure of relationship intensity. The results indicate that access to bank credit during this time period stronlgy predicts idiosyncratic market performance: firms receiving a syndicated loan over 2006-09 outperformed those that did not by 4 percentage points during the Lehman time period. Relationship intensity appears not to have mattered. By taking the residuals, we remove the influence of both variables in all subsequent regressions.

[14]If the borrower received multiple loans at market time 0, either due to multiple packages or multiple facilities in a single package, we include all loans in the regression. Each borrwoing firm's bank information proxy is then the average of the residuals: $b^*_f = 1/L \sum_{l=1}^{L} b^*_{l,fb}$.

4.2 Baseline Results

We begin the main part of our analysis graphically. Figure 2 plots the all-in spread versus relationship time for high- and low-quality firms according to the orthogonalized bank information proxy b_f^*. "High-quality" firms have values of the proxy in the top third of our sample, and "low-quality" firms are in the bottom third. There appears to be an overall relationship discount for both groups of firms: as the length of the relationship progresses, the interest rate falls. This is suggestive that firms benefit from remaining with the same bank, as is sometimes found in the previous literature. However, it could be driven by selection on observables and is not evidence of private bank learning *per se*.

The average all-in spread is the same for both high- and low-quality firms at relationship time zero. This is due to our orthogonalization procedure removing any correlation between the proxy and unobserved factors that the bank used to set $r_{l,f,0}$. Bank learning is evidenced by the gap that opens up between the two groups of firms over the course of a relationship. Firms with top-third values of b_f^* are consistently receiving lower interest rates than are firms with bottom-third values of b_f^*.[15] That is, banks appear to be learning about b_f^* and using it to price firms' subsequent loans. Of course, banks cannot be observing or pricing on b_f^* by construction. This suggests that the banks are learning about factors correlated with b_f^*, namely, firm quality.

Turning to a regression framework, we estimate pricing equations to assess the extent of bank learning while also controlling for time-varying firm covariates. We first report the results from estimating a standard pricing equation of the all-in drawn spread on firm characteristics, loan characteristics, and on relationship time:

$$r_{l,f,\tau} = \alpha_t + \alpha_i + \beta' z_{f,t} + \gamma' w_{l,f,\tau} + \varphi \cdot \tau + u_{l,f,\tau} \tag{7}$$

[15] The difference is economically significant, averaging about 15 basis points between time periods 2 and 6, and widening to a maximum of 73 basis points in time period 9.

where each observation is given by a loan l to firm f at market time τ. We control for year t and two-digit SIC industry i fixed effects. We report heteroskedasticity-robust standard errors that are clustered to allow for the presence of autocorrelation within borrower over time. Results are presented in the second column of Table 2. Larger predicted probabilities of default (lower Z score and higher NPD) are significantly associated with higher loan spreads. In this full regression framework, longer time in the market does not have a statistically significant correlation with loan rates.[16] Secured loans carry a 66 - 13 = 53 basis point higher spread than unsecured loans (the omitted group is loans where the collaterialization status is unreported), a seemingly counterintuitive result. This suggests that other loan characteristics also reflect omitted borrower characteristics: if secured loans tend to be lower quality on average, then they will also carry higher interest rates. Larger borrowers, as measured by the log of total assets, pay lower interest rates, and revolving loans are 60 basis points cheaper than term loans. Finally, loan maturity is insignificant.

The main result from this regression is a non-result: having an established presence in the syndicated loan market does not appear to lower the cost of credit for the average firm after controlling for relevant pricing characteristics, including firm size and borrower quality (as measured by Z-score and NPD). We proceed to test whether this non-result might be due to differential pricing, as lenders learn about borrower quality.

In our baseline learning specification, we add the private information proxy b_f^* to the previous regression. By construction the proxy variable can have no effect on loan prices at market time zero. The test is whether the loading varies over market time and whether "better" firms receive a discount. The coefficient of interest is δ_τ in the

[16] Lim and Minton (2012) also fail to find a relationship discount in the syndicated loan market; see their Table IV panel A.

following specification:

$$r_{l,f,\tau} = \alpha_t + \alpha_i + \beta' z_{f,t} + \gamma' w_{l,f,\tau} + \delta_0 \cdot b_f^* + \delta_\tau \cdot \left(b_f^* \times \tau\right) + \varphi \cdot \tau + u_{l,f,\tau} \quad (8)$$

Estimates are presented in the third column of Table 2. First note that the inclusion of our proxy variable does not materially affect any of the results obtained in the baseline case. This is because the proxy is orthogonalized to all loan characteristics at relationship time zero. Second, the coefficient on the proxy variable interacted with relationship time has a highly significant effect on the pricing of a firm's loans. Consider a one standard deviation increase in the proxy, an increase in the CAR of 0.36 log units (i.e., 36 percentage points). Holding other firm and loan features constant, this firm would benefit from a reduction in its interest rate on bank loans of -5.708×0.36 = -2.06 basis points per renewal. This is of the same order of magnitude as the effect shown in Figure 2. On an average sized loan ($360 million in constant 2000 dollars), this would result in annual savings of $74,000 per year. Since the average maturity of a loan in our sample is just over four years, the total savings from on the first loan renewal is $328,000. The savings increases with time in the market: on the third renewal it would be nearly one million dollars. Put another way, a one S.D. increase in the proxy has the same benefit per renewal on loan prices as a 1.8 percentage point decrease in the firm's default probability.[17]

4.3 Robustness Checks

4.3.1 Unobserved Borrower, Lender, or Relationship Heterogeneity.

Table 3 presents a series of robustness checks for omitted, time-invariant factors. For ease of reference, we replicate the baseline result from the previous table in column 1.

An omitted, time-invariant borrower characteristic that has a fixed effect on loan

[17]Berger and Udell (1995) find pricing effects for small-business borrowers that are one order of magnitude larger: a 48 basis-point discount for a ten-year banking relationship.

pricing would be removed by our orthogonalization procedure. However, suppose there is an omitted characteristic whose correlation with the loan price is increasing with relationship length. We control for any such time-invariant borrower characteristic via borrower fixed effects, rather than industry fixed effects, in column 2 of Table 3. Our learning coefficient actually increases in magnitude and is significant at the 1% level, suggesting that our private learning result cannot be due to unobserved borrower heterogeneity. Note that the proxy variable itself only varies across borrowers and is thus absorbed by the fixed effects.

Another possibility is that our results are driven by unobserved heterogeneity among lenders. For example, suppose that there is heterogeneity across lenders in the cost of funds and price-setting: some low cost-of-funds lenders are able to offer a particularly big relationship discount to repeat customers, and these same lenders also attract high-quality borrowers. This lender pricing effect would be correlated with the interaction between relationship time and our proxy variable, so we might simply be confusing learning for lender heterogeneity. Lender fixed effects would capture any such effect. We test for and reject the presence of such concerns with the inclusion of lender fixed effects in column 3. Our point estimate is essentially unchanged from the baseline column.[18]

What if some firm outcomes partly depend on the match quality with their bank? That is, suppose that a high-quality bank-firm match produces a surplus that is shared between the borrower and lender via declining interest rates. Since the Lehman proxy measures the firm's future financial performance, it might also be capturing that a firm is in a good match with its primary lender. So we would expect to see higher quality matches producing lower interest rates over time. Match quality is relationship-specific (i.e., it varies by the Cartesian product of lenders$\times$borrowers), so this is not captured by either lender or borrower fixed effects. We address this concern in Table 3, column

[18]We include lender fixed effects by running WLS on a relationship panel dataset, with each loan's characteristics replicated across all lead bank members of the syndicate.

4, by including relationship-specific fixed effects.[19] Within the same relationship, the bank information proxy becomes increasingly relevant for loan prices over time. This indicates that the importance of the proxy interaction variable is not driven by match heterogeneity either.

4.3.2 Forecast Window-Length Effect

A potential confound is that the proxy variable is taken from financial market data in a specific year in the future. It might be the case that lenders are forecasting some factor correlated with b_f^*, such as project returns over the lifetime of the loan, and that these forecasts become more correlated with the proxy as the origination year $t \to 2008$. To be confounding, such an effect would have to manifest as an interaction between the proxy and calendar time. If there were something special merely about time until 2008, it would be picked up by the calendar year fixed effects.

To control for such a forecast window-length effect, we estimate the following regression:

$$\begin{aligned} r_{l,f,\tau} &= \alpha_t + \alpha_i + \beta' z_{f,t} + \gamma' w_{l,f,\tau} + \delta_0 b_f^* + \delta_\tau \cdot \left(b_f^* \times \tau\right) \\ &+ \delta_t \cdot \left(b_f^* \times YearsUntil2008\right) + u_{l,f,\tau} \end{aligned} \tag{9}$$

where $YearsUntil2008 := 2008 - t$ is a time trend counting down from the origination year t. Under the forecast window-length hypothesis, $\delta_t > 0$ – the Lehman proxy is more relevant (more negative) when $YearsUntil2008$ is smaller. Moreover, this specification removes *any* correlation between the proxy and loan rates that linearly depends on calendar time instead of market time. However, we do not find a statistically significant or economically meaningful calendar time trend in the proxy coefficient (Table 4, column 1). Moreover, including this additional interaction barely affects the bank learning interaction coefficient, which remains a highly significant factor in bank pricing equation.

[19] As in column 3, we run WLS on a relationship panel dataset.

4.3.3 Omitted Time-Varying Firm Variables

Since both observed and unobserved firm characteristics $(z'_{f,t}, x'_{f,t})$ are also time-varying, our orthogonalization procedure might not fully eliminate omitted variable bias. Intuitively, the non-orthogonalized background variable contains information about both the total default probability and omitted firm characteristics in 2008. The orthogonalization procedure removes the influence of omitted variables at relationship time 0 but leaves information about total default probability. If subsequent values of omitted factors x contain information new information about the default probability, this will show up as a correlation with the orthogonalized private information proxy. As the innovations accumulate, the correlation will increase in magnitude. This will exhibit as omitted variable bias in our regressions – we would mistake banks pricing on publicly-observable variables for private learning.

To partially address this concern, we run a "kitchen-sink" style regression in Table 4, column 2, in which we include a bevy of additional time-varying firm-level characteristics suggested by the loan pricing literature: market-to-book ratio of assets, profitability ratio, book leverage, the log interest coverage ratio, tangibility, and dummy variables for S&P long-term debt rating categories (including unrated as a category). We additionally include the log facility amount as an explanatory variable, since loan quantities might contain additional information about observable but omitted borrower characteristics. The net impact of including all of these additional controls is that the learning proxy falls from -5.7 to -4.7 but remains highly significant. These additional control variables have the same economic impact on the coefficient as including lender fixed effects in the previous table. While there could be additional time-varying public characteristics that are both relevant and omitted, it seems unlikely that these will reduce the coefficient much further.

4.3.4 Time-varying coefficients

Suppose that the coefficients of the loan pricing equation are varying over time. This could be because banks increasingly rely on computerized credit scoring models. Alternately, banks might be learning about the correct pricing model based on observables, rather than about unobservable borrower characteristics. In either case, the bank learning coefficient might be picking up instability in the other regression coefficients rather than information acquisition.

We address this concern in Table 3 column 4 by including interaction variables between market time τ and *all* of the control variables, including the previously-mentioned "kitchen-sink" variables. This has barely an effect on the learning coefficient, which remains highly economically and statistically significant at -4.5.

4.3.5 Reverse Causality

Several authors (Ivashina and Scharfstein 2010, Santos 2011, and Chodorow-Reich 2014) have hypothesized that bank health affected firm performance post-2008. However, De Mitri et al. (2010) show that firms in longer relationships were less affected by the Lehman shock and had easier access to funds during the credit crunch, enabling them to better weather the shock. This suggests an alternate reason that we might observe a correlation among relationship length, the Lehman proxy, and loan rates. Under this scenario the causality is reversed and runs from interest rates to the Lehman proxy rather than vice versa. Firms in longer relationships might receive larger loans and lower interest rates for reasons unrelated to bank learning, driving their superior performance to the Lehman shock in 2008.

We addressed this concern in our initial orthogonalization procedure (Table 2, column 1), by including measures for syndicated credit access and relationship intensit over the 2006-09 period. Since the proxy is orthogonal to borrower credit access in 2008, reverse causality cannot be driving any of our results.

4.3.6 Placebo Tests

An assumption underlying all of our results is that the abnormal returns of firms to the large aggregate Lehman shock contain important information about latent firm quality. In addition, since our event is from the future, banks cannot directly observe it and price on it. This suggests two natural placebo tests: choose an event where information is revealed but has already been observed by banks and potentially incorporated into loan prices, or choose a random date when it is unlikely any information was revealed.

For the first placebo test, we construct firms' cumulative abnormal returns in a three-month window around the collapse of Continental Illinois. The general creditors of Continental Illinois National Bank and Trust Company were bailed out by the FDIC on May 18, 1984. This was the largest bank failure in US history prior to the global financial crisis of 2007-08. This event is of similar magnitude to the Lehman failure, but it occurs five years before our sample rather than five years after. Since it is from the past, we expect to find that banks have already incorporated any relevant information that was revealed into loan prices. After orthogonalizing the Continental Illinois proxy to the first loan in our sample, we should not observe banks learning about the proxy over time.

For the second placebo test, we choose several arbitrary dates from the past: March 25, 1983; November 12, 1984; and July 2, 1986.[20] For each of these event dates, we estimate a four-factor Fama-French-Carhart model for each firm using the previous five years of daily stock returns (ending on December 31 of the previous year), then construct the cumulative abnormal return over a [-21, +42] trading-day event window centered at the event date.

In Figure 3, we display the estimated coefficient and 95% confidence interval on the proxy×relationship time interaction variable for each of these placebo tests and various sets of fixed effects (all specifications include the same set of control variables

[20]The first two dates were chosen for non-economic reasons: the authors' birthdays are March 24 and 25, 1983, and the second date is one of the authors' spouse's birthday. The third date is random.

as in Table 2 column 3, including year and industry fixed effects, and all standard errors are clustered by borrower). The first three columns report the Lehman proxy results and 95% confidence intervals for reference. By comparison, the coefficient on the Continental Illinois interaction variable is the wrong sign (positive rather than negative) and estimated very imprecisely – all three confidence bands include zero. Our three randomly-chosen dates perform even worse. Only one of the 12 specifications produces a negative point estimate that is also statistically significant at standard levels. By contrast, all three of the Lehman proxy specifications are the correct sign and statistically different from zero.

4.3.7 Other Possible Explanations

In this subsection we discuss several other possible explanations for our results.

Functional form misspecification. Suppose the true pricing equation is a non-linear function of firm characteristics z, and that the proxy variable is correlated with this non-linear function. Controlling for z in a linear fashion is misspecified and does not remove the relevant correlation. However, any spurious relationship between b_f^* and $r_{f,\tau}$ should be constant over time. This does not explain our result that the loading on the proxy increases with relationship time.

Selection / survivorship bias. Suppose that banks screen on omitted but publicly-observable firm characteristics x, so that only the best firms have long-term relationships. In the extreme case, imagine that there are two firms, G and B. Firm G stays in a long-term relationship with its bank and pays a low interest rate because it is high quality, while firm B switches banks every period and pays a high interest rate because it is low quality. This would create a negative correlation between relationship length and interest rate spreads in our data. However, we control for relationship length and find that the interaction between relationship length and the proxy variable also matters.

As a further check, we plot the average value of firm characteristics over relationship time in Figure 4. The top left panel suggests that there may indeed be some selection on unobservables. The average value of the Lehman proxy becomes less negative (better) over relationship time, and differs from its original value in 5 of 10 observations at a 95% confidence level. However, this selection is essentially eliminated after we orthogonalize the proxy to the time zero interest rate (top right panel). Firms in long-term relationships do not appear to be consistently higher quality as measured by higher Z-scores or lower default probabilities (bottom two panels).

4.4 The Impact of Learning on Loan Sizes

Loans are multidimensional contracts, so banks could use their private information to adjust the loan's characteristics along a variety of margins. We consider the other major dimension – quantity – in Table 5. The dependent variable is now the log of loan size, rather than the all-in-drawn spread over LIBOR. (Recall that the proxy is already orthogonalized to the log of initial loan size in Table 2 column 1.) We consider regressions on all loans in our sample, and broken down into the two major categories of term loans vs. revolvers. The coefficient on the Information Proxy×Market Time interaction is statistically insignificant for all loans and for term loans, but it is positive and significant for revolvers. That is, higher quality borrowers do not seem to receive better access to credit via term loans, but they do seem to differentially receive larger credit lines. Finally, our results confirm previous evidence that banks grant more credit to firms that have been in the market longer – the coefficient on market time ranges between 75 basis points (for credit lines) and 3.3 percentage points (for term loans).

5 Extensions

In this section, we conduct several extensions to further understand the implications of bank learning. In particular, our focus is to understand whether and how firms benefit from the surplus generated by relationship lending.

5.1 Bank Learning and the Distribution of Interest Rates

Who benefits from longer relationships? Consider a stylized model where there are two types of borrowers, good and bad, who are observationally equivalent based on publicly-available information. In an initial loan contract, the bad type imitates the good and the loans are pooled together. Subsequently, the bank observes private signals that let it distinguish between good types and bad types. Assuming it is individually rational for the bank to extend correctly-priced loans to both types, we expect the bank to update its beliefs about which borrowers are good and charge them less. Whether the bank also charges bad types more depends on the nature of interbank competition, whether borrowers face switching costs, and what signal is sent by observing a borrower change lenders.

We offer two competing hypotheses:

H0 As banks acquire private information about borrowers, they charge good borrowers less.

H1 As banks acquire private information about borrowers, they charge bad borrowers more.

H1 would be plausible if there were an overall relationship discount for all repeat borrowers, for which we have failed to find consistent evidence. Alternately, borrowers might face a stigma from switching lenders, setting a positive upper bound on the existing lender's ability to raise interest rates as observes signals about borrower type.

We begin by estimating quantile regressions to explore how the private information proxy affects the entire distribution of interest rates, rather than simply its conditional mean. Table 6 presents estimates of the marginal effects of each explanatory variable on the 5th, 25th, median, 75th, and 95th quantiles of the all-in-drawn spread. These estimates are otherwise analogous to the OLS results presented in Table 2, column 3.

We find a monotonic relationship between the bank info proxy interaction coefficient and the quantile. This indicates that bank learning does not play much of a role for borrowers already receiving unusually low interest rates: most of the effect of private information is on borrowers receiving high interest rates. The marginal effect of Z score and Merton default probability on loan price are larger in the higher quantiles of the interest rate distribution. However, the rate of change is much larger for private learning. For example, the marginal effect of the default probability on interest rates is about five times as large at the 95th percentile as at the 5th percentile; the marginal effect of private learning is almost fifty times as large.

We also note that the coefficient on relationship time becomes increasingly negative as we move up the quantiles. Although it remains statistically insignificant overall, this is suggestive that if there is an overall relationship discount, it is likely most important for repeat borrowers at the top of the interest rate distribution.

Next, we decompose the orthogonalized private info proxy into its positive and negative parts:

$$b = b^+ + (-b^-)$$

We relax the implicit restriction that the coefficients on the positive and negative components are the same and estimate both the levels and the interactions separately. Since positive values of the proxy indicate positive abnormal stock returns in 2008, the coefficient on $b^+ \times \tau$ maps to the good type borrowers (H0). Similarly, the coefficient on $(-b^-) \times \tau$ maps to bad type borrowers (H1).

Our main results are presented in Table 7 column 1. Both components are signed so that their coefficients may be interpreted as before. The coefficient on good types is -7.9, somewhat larger than the magnitude of the restricted coefficient in Table 2 column 3, and highly significant. Conversely, the coefficient on bad types is half as large, -4.0, but remains statistically signficant. This presents evidence in favor of H0 and against H1. It appears that good types and bad types are initially pooled together and pay high interest rates. As banks learn about which borrowers are more creditworthy, the good types benefit more from cheaper subsequent loans, while the bad types continue to pay the high initial interest rate.

Column 2 indicates that the differential loan pricing behavior is robust to the inclusion of borrower fixed effects. If anything, the difference in dynamic loan pricing between good and bad types is exacerbated when we control for additional time-invariante borrower characteristics.

Columns 3 and 4 repeat the exercise using loan size rather than price as the dependent variable. Interestingly, we find that bank learning is more important for negative-proxy bad types than for positive-proxy good types, although the quantity results are again less robust than the price results. The learning coefficient on bad types is positive 0.05 and signficant in column 4, in the borrower fixed effect specification. This would indicate that the quantity of credit asymmetrically falls more for bad borrowers than it risese for good borrowers as banks learn each borrower's type.

5.2 Bank Learning and Lending Relationships

5.2.1 Public vs. Private Learning

Table 8 presents results in which we try to distinguish between public and private learning. We test for a knock-on benefit if the borrower takes out a "relationship loan" with the same bank as previously, measured either by relationship length or by a relationship loan dummy variable, following the 5-year definition used by Bharath et al. (2009). A more negative learning coefficient for within-relationship lending would be suggestive of private, as opposed to marketwide, bank learning.

The results are inconclusive at best. This may be due to survivorship bias in our sample. We do not find the well-established result that relationship lonans carry a discount (column 3); but in column 5, we re-estimate this equation on the more-common "all-public firms" sample and find a (static) relationship loan discount of 6 basis points.

5.2.2 Borrower Transparency

The relationship lending literature also suggests that relationships matter more for less transparent borrowers. As an additional test, we test for differential speeds of learning for less versus more transparent borrowers. We use three common measures of transparency: size, % of tangible assets, and whether or no the borrower has an S&P long-term debt rating. We classify all firms as "opaque" or "tranparent" based on whether they are below or above the sample median (unrated or rated) at market time 0, and we hold these classifications fixed. We then interact the learning coefficient with an indicator for borrower transparency.

We can reject the hypothesis that banks learn more quickly about more-opaque syndicated loan borrowers. If anything, banks appear to learn more quickly about transparent borrowers – however, none of the three triple-interaction coefficients in Table 9 are significant.

6 Conclusions

We began this paper by posing the question, "Do banks learn by lending?" Our answer is a resounding yes. We first verified that borrowers inside longer relationships unconditionally pay cheaper loan spreads. We then tested whether this reduction in spreads is driven by banks learning about firm fundamentals using a methodology adapted from Farber and Gibbons (1996). We constructed a proxy for firm fundamentals that is orthogonal to the bank's information set, based on the differential response of the firms to the collapse of Lehman Brothers in September 2008. We argue that this contains relevant information about firm's tail risk, which is precisely what lenders care about when pricing loans in this market. We showed that our proxy is increasingly relevant for loan prices as a relationship progresses. Finally, we investigated the dynamics of loan pricing and showed that learning matters more for good types than for bad types.

Our research suggests three main takeaways and one caveat. First, delegated information acquisition matters even in a market with large, transparent borrowers, such as the syndicated loan market. Second, some information acquisition is dynamic and occurs via repeated interaction – banks continue to learn, even after many years of doing business with the same borrower. Third, the bank shares some of its surplus with high quality borrowers via lower interest rates over time. The caveat is that we cannot confidently determine whether learning is public or private in this setting.

In future research, we plan to further investigate what it is that banks are learning about. Possible candidates include: firm-specific characteristic, such as the value of assets in place, or the effectiveness of the firm's corporate governance structure; the top management's character and ability; and the membership and the activeness of the firm's board. We hope to exploit variation in CEOs and board membership across firms to disentangle these possible explanations.

References

Agarwal, S. and Hauswald, R. (2010). Distance and private information in lending. *Review of Financial Studies*, 23(7):2757–2788.

Agrawal, A. K., Catalini, C., and Goldfarb, A. (2011). The Geography of Crowdfunding. NBER Working Papers 16820, National Bureau of Economic Research, Inc.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Allen, F. (1990). The Market for information and the Origin of Financial Intermediation. *Journal of Financial Intermediation*, 30:3–30.

Allen, F. (1999). Diversity of Opinion and Financing of New Technologies. *Journal of Financial Intermediation*, 8(1-2):68–89.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4):589–609.

Altman, E. I. (1984). A further empirical investigation of the bankruptcy cost question. *The Journal of Finance*, 39(4):1067–1089.

Asea, P. K. and Blomberg, B. (1998). Lending cycles. *Journal of Econometrics*, 83:89–128.

Beck, T. and Demirguc-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & Finance*, 30(11):2931–2943.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2006). Bank concentration, competition, and crises: First results. *Journal of Banking & Finance*, 30(5):1581–1603.

Beck, T. and Levine, R. (2004). Stock markets, banks, and growth: Panel evidence. *Journal of Banking & Finance*, 28(3):423–442.

Berger, A. N. and Udell, G. F. (1995). Relationship Lending and Lines of Credit in Small Firm Finance. *Journal of Business*, 68(3):351–381.

Berger, A. N. and Udell, G. F. (2005). A More Complete Conceptual Framework for Financing of Small and Medium Enterprises.

Bharath, S. and Shumway, T. (2004). Forecasting Default with the KMV-Merton Model. *AFA 2006 Boston Meetings Paper*.

Bharath, S. T., Dahiya, S., Saunders, A., and Srinivasan, A. (2009). Lending relationships and loan contract terms. *Review of Financial Studies*, 24(4):1141–1203.

Bharath, S. T. and Shumway, T. (2008). Forecasting Default with the Merton Distance to Default Model. *Review of Financial Studies*, 21(3):1339–1369.

Bharath, S. T., Sunder, J., and Sunder, S. V. (2008). Accounting Quality and Debt Contracting. *The Accounting Review*, 83(1):1–28.

Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, pages 637–654.

Bolton, P., Freixas, X., and Gambacorta, L. (2016). Relationship and transaction lending in a crisis. *Review of Financial Studies*, page hhw041.

Boot, A. W. A. and Thakor, A. V. (1993). Self-interested Bank Regulation. *The American Economic Review*, 83(2):206–212.

Botsch, M. J. (2015). Distance, asymmetric information, and mortgage securitization.

Boyd, J. and Prescott, E. C. (1985). Financial Intermediary-Coalitions.

Brock, W. A. and Manski, C. F. (2008). Competitive Lending with Partial Knowledge of Loan Repayment.

Butler, A. W. (2008). Distance Still Matters: Evidence from Municipal Bond Underwriting. *Review of Financial Studies*, 21(2):763–784.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2006). Robust Inference with Multi-way Clustering.

Campbell, T. S. and Kracaw, W. A. (2012). Information Production, Market Signalling, and the Theory of Financial Intermediation. *The Journal of Finance*, 35(4):863–882.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1):57–82.

Chodorow-Reich, G. (2014). The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis. *The Quarterly Journal of Economics*, 129(1):1–59.

Claessens, Stijn and Laeven, L. (2003). What Drives Bank Competition ? Some International Evidence. *Journal of Money, Credit and Banking*, 36(3):563–583.

De Mitri, S., Gobbi, G., and Sette, E. (2010). Relationship Lending in a Financial Turmoil. *Bank of Italy Temi di Discussione (Working Paper) No*, 772.

Degryse, H., Ioannidou, V., and Ongena, S. (2015). Bank-firm relationships: a review of the implications for firms and banks in normal and crisis times. In *The Economics of Interfirm Networks*, pages 177–189. Springer.

Degryse, H., Kim, M., and Ongena, S. (2009). *Microeconometrics of banking: methods, applications, and results*. Oxford University Press, USA.

Degryse, H. and Van Cayseele, P. (2000). Relationship lending within a bank-based system: Evidence from european small business data. *Journal of financial Intermediation*, 9(1):90–109.

Demirgüç-Kunt, A., Feyen, E., and Levine, R. (2012). The evolving importance of banks and securities markets. *The World Bank Economic Review*, page lhs022.

Diamond, D. W. (1984). Financial Intermediarion and Delegated Monitoring. *Economic Analysis*, 51(3):393–414.

Diamond, D. W. (1997). Liquidity, Banks, and Markets. *Journal of Political Economy*, 105(5):928.

Diamond, D. W. and Dybvig, P. H. (1983). Bank Runs, Deposit Insurance, and Liquidity. *Journal of Political Economy*, 91(3):401.

Diamond, D. W. and Rajan, R. G. (2001). Banks and Liquidity. *American Economic Association*, 91(2).

Duffie, D., Garleanu, N., and Pedersen, L. H. (2005). Over-the-Counter Markets. *Econometrica*, 73(6):1815–1847.

Fama, E. F. (1985). What's different about banks? *Journal of Monetary Economics*, 15(1):29 – 39.

Fama, E. F. and French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1):3–56.

Farber, H. S. and Gibbons, R. (1996). Learning and Wage Dynamics. *The Quaterly Journal of Economics*, 111(4):1007–1047.

Geanakoplos, J. (2003). Promises Promises.

Gibson, M. S. (1997). More Evidence on the Link Between Bank Health and Investment in Japan. *Journal of the Japanese And International Economies*, 11(3):296–310.

Goel, M. and Zemel, M. (2015a). Friendship matters less on a rainy day: Firm outcomes and relationship bank health. *Working paper*.

Goel, M. and Zemel, M. (2015b). Friendship matters less on a rainy day: Firm outcomes and relationship bank health. *Available at SSRN 2693413*.

Golosov, M., Tsyvinski, A., and Farhi, E. (2011). A Theory of Liquidity and Regulation of Financial Intermediation. *Review of Economic Studies*, 76(3):973–992.

Gorton, G. and Metrick, A. (2010). Regulating the Shadow Banking System.

Gorton, G. and Metrick, A. (2011). Securitized banking and the run on repo. *Journal of Financial Economics*, pages 1–27.

Gorton, G. and Pennacchi, G. (1990a). Bank and Loan Sales: Marketing Non-Marketable Assets.

Gorton, G. and Pennacchi, G. (1990b). Financial Intermediaries and Liquidity Creation. *The Journal of Finance*, 45(1):49–71.

Gorton, G. and Winton, A. (2002). Financial Intermediation.

Holmstrom, B. and Tirole, J. (1997). Financial Intermediation, Loanable Funds, and The Real Sector. *Quarterly Journal of Economics*, 112(3):663–691.

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

Ivashina, V. and Scharfstein, D. (2010). Bank lending during the financial crisis of 2008. *Journal of Financial Economics*, 97(3):319–338.

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. (2014). Screening peers softly: Inferring the quality of small borrowers. Revision requested at Management Science. 2014.

James, C. (1987). Some evidence on the uniqueness of bank loans. *Journal of Financial Economics*, 19(2):217 – 235.

Jiang, J., Li, Z., and Lin, C. (2014). Financing difficulties of smes from its financing source in china. *Journal of Service Science and Management*, (7):196–200.

Karolyi, G. A. and Taboada, A. G. (2015). Regulatory arbitrage and cross-border bank acquisitions. *The Journal of Finance*, 70(6):2395–2450.

Karolyi, S. A. (2017). Personal lending relationships. *forthocoming at The Journal of Finance*.

Keys, B., Mukherjee, T., Seru, A., and Vig, V. (2010). Did Securitization Lead to Lax Screening? Evidence from Subprime Loans. *The Quarterly Journal of Economics*, 125(1):307–362.

Kysucky, V. and Norden, L. (2015). The benefits of relationship lending in a cross-country context: A meta-analysis. *Management Science*, 62(1):90–110.

Leland, H. E. and Pyle, D. H. (1977). Informational Asymmetries, Financial Structure, and Financial Intermediation. *The Journal of Finance*, 32(2).

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

Liberti, J. M. and Mian, A. R. (2009). Estimating the effect of hierarchies on information use. *Review of Financial Studies*, 22(10):4057–4090.

Lim, J. and Minton, B. A. (2012). Syndicated Loan Spreads and the Composition of the Syndicate.

Merton, R. (1973). Theory of Rational Option Pricing. *The Bell Journal of Economics and Management*, 4(1):141–183.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29(2):449–470.

Mester, L. J., Nakamura, L. I., and Renault, M. (2007). Transactions accounts and loan monitoring. *Review of Financial Studies*, 20(3):529–556.

Morrison, A. D. and Wilhelm, W. J. (2007). Investment of Banking: Past, Present, and Future. *Journal of Applied Corporate Finance*, 19(1):42–54.

Norden, L. and Weber, M. (2010). Credit line usage, checking account activity, and default risk of bank borrowers. *Review of Financial Studies*, 23(10):3665–3699.

Petersen, M. A. and Rajan, R. G. (1995). The Effect of Credit Market Competition on Lending Relationships. *The Quaterly Journal of Economics1*, 110(2):407–443.

Rajan, R. G. (1992). Insiders and outsiders: The choice between informed and arm's-length debt. *The Journal of Finance*, 47(4):1367–1400.

Rajan, U., Seru, A., and Vig, V. (2015). The failure of models that predict failure: Distance, incentives, and defaults. *Journal of Financial Economics*, 115(2):237 – 260.

Santos, J. a. C. (2011). Bank Corporate Loan Pricing Following the Subprime Crisis. *Review of Financial Studies*, 24(6):1916–1943.

Schenone, C. (2010). Lending Relationships and Information Rents: Do Banks Exploit Their Information Advantages? *Review of Financial Studies*, 23(3):1149–1199.

Sharpe, S. (1990). Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *Journal of Finance*, 45(4):1069–87.

Slovin, M. B., Sushka, M. E., and Polonchek, J. A. (1993). The Value of Bank Durability: Borrowers as Bank Stakeholders. *The Journal of Finance*, 48(1):247–266.

Sufi, A. (2007). Information asymmetry and financing arrangements: Evidence from syndicated loans. *The Journal of Finance*, 62(2):629–668.

Tirole, J. and Holmstrom, B. (2010). Private and Public Supply of Liquidity. *Journal of Political Economy*, 106(1):1–40.

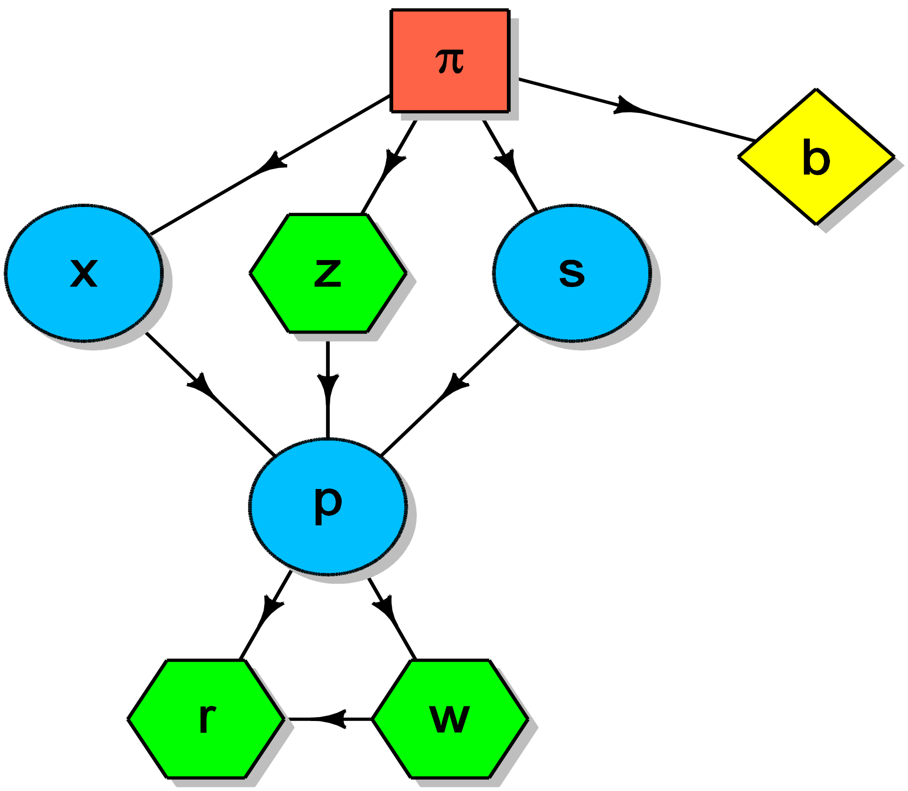
William H. Beaver (1966). Financial Ratios as Predictors a of Failure. *Journal of Accounting Research*, 4:71–111.

Williamson, S. (1987). Costly Monitoring , Loan Contracts , and Equilibrium Credit Rationing. *The Quarterly Journal of Economics*, 102(1):135–146.

Wilson, R. (1968). The Theory of Syndicates. *Econometrica*, 36(1):119–132.

Figure 1.

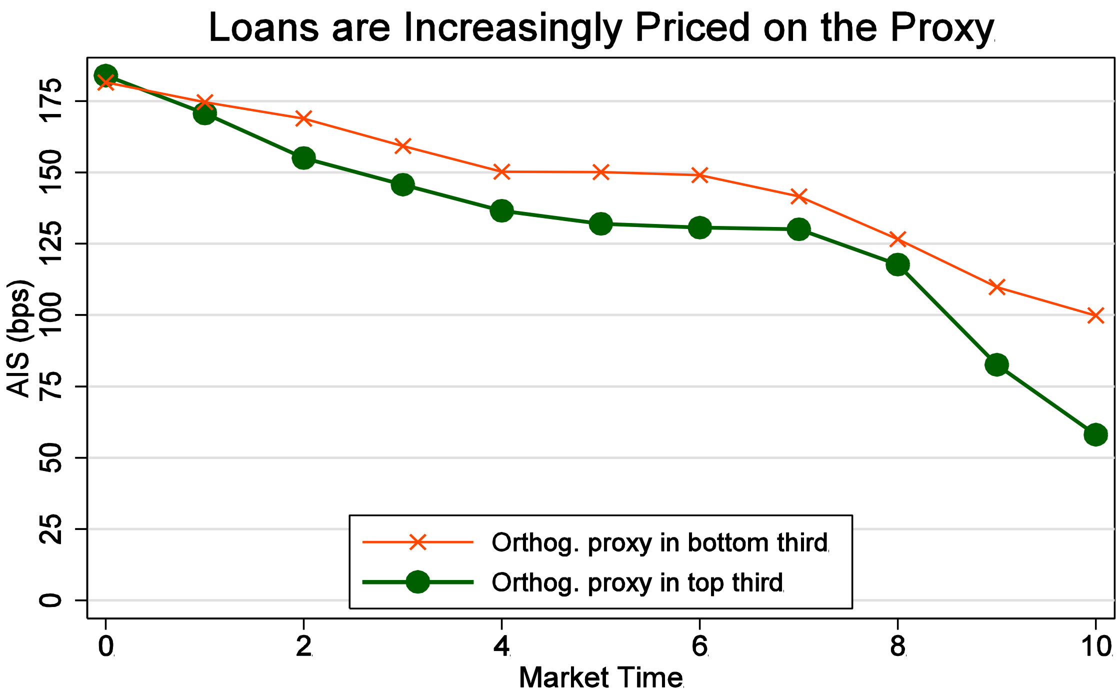
Overview of the Information Structure



Hexagons (in green) indicate the variables that we observe along with the bank.
Circles (in blue) indicate variables that only the bank sees.
Diamonds (in yellow) indicate variables that only we see, and not the bank.
Squares (in red) indicate variables that no one sees.

Arrows indicate correlations between variables.

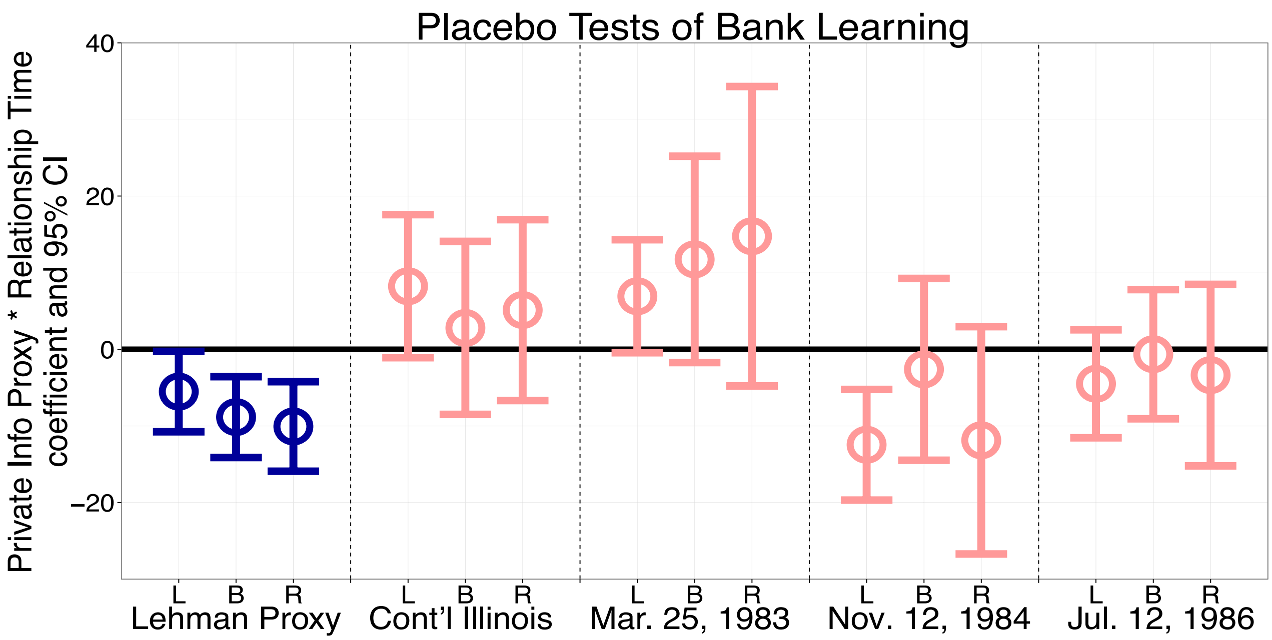
Figure 2.



Moving avererage [-1,+1] filter applied to each series.

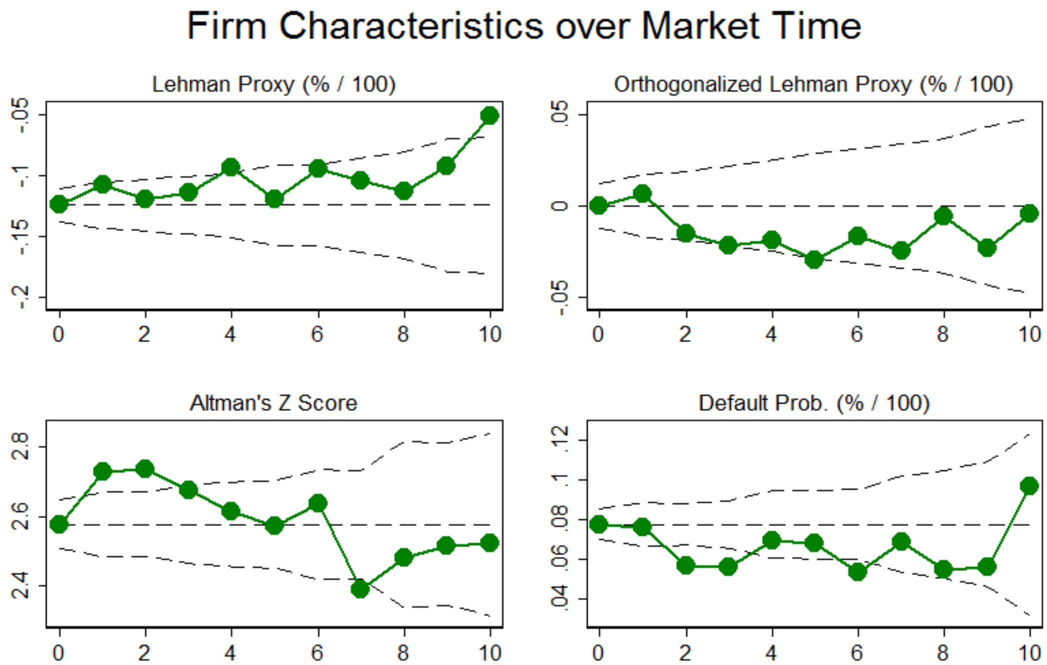
Source: LPC DealScan / authors' calculations.

Figure 3.



Fixed Effects Included in Specification
L-Lender
B-Borrower
R-Relationship

Figure 4.



Source: LPC DealScan / authors' calculations.

Panels show 95% CI for population mean centered at initial value.

Table 1: Summary Statistics

Table reports summary statistics of syndicated loan facilities originated between 1987 and 2003. "Lehman Sample" borrowers were publicly traded both at time of loan and in 2008. "All public firms" is Lehman Sample + firms that were publicly traded at time of loan but not in 2008.

Panel A: Facility Characteristics	**Lehman Sample N = 8,673**	**All public firms N = 23,594**	
	Mean (SD)	Mean (SD)	Difference
All-in spread (bps)	152.3	196.9	-44.5
	(125.71)	(140.11)	
Loan Size (constant 2000 $m)	361.6	265.3	96.4
	(837.91)	(722.38)	
Maturity (months)	40.9	42.3	-1.4
	(28.37)	(28.40)	
Fraction revolver	0.756	0.711	0.045
	(0.43)	(0.45)	
Fraction collateralized	0.378	0.504	-0.125
	(0.48)	(0.50)	
Fraction not collateralized	0.198	0.141	0.056
	(0.40)	(0.35)	
Syndicate size	8.7	7.1	1.6
	(10.0)	(9.2)	
Fraction relationship loans [1]	0.578	0.544	0.034
	(0.49)	(0.50)	
Market Time	2.7	2.2	0.5
	(3.5)	(3.1)	

Panel B: Borrowing Firm Characteristics	**Lehman Sample N = 6,112**	**Public Sample N = 16,133**	
	Mean (SD)	Mean (SD)	***Difference***
Total assets (constant 2000 $b)	6.38	4.11	2.27
	(20.8)	(16.4)	
Tangible assets (% of total)	36.4	34.8	1.6
	(23.8)	(23.8)	
Fraction with LT debt rating from S&P	0.54	0.42	0.12
	(0.5)	(0.5)	
Average Q [2]	1.53	1.44	0.10
	(1.53)	(1.48)	
ROA (%)	3.34	0.72	2.62
	(11.52)	(15.51)	
Z score	2.65	2.44	0.21
	(1.74)	(1.82)	
Naïve Probability of Default (% / 100)	0.064	0.105	-0.041
	(0.19)	(0.25)	
Three-month CAR around Lehman (% / 100)	-0.098	*N.A.*	
	(0.36)		

Notes.

In Panel A, each observation is a facility; in Panel B, a borrower-month.

[1] Relationship Loan = 1 if the borrower has received credit from the same lead bank at least once in the previous five years, and 0 if the borrower has received credit in the previous five years, but only from different lead banks. 2,128 (7,027) facilities are not classified in the Lehman (Public) Sample and are excluded from analysis.

[2] Q = (E + P + D) / A, where E is market value of common equity, P is liquidating value of preferred stock, D is book value of long-term debt plus current liabilities net of (current assets less inventories), and A is book value of total assets.

Table 2: Do Banks Learn?

OLS regressions of syndicated loan interest rates on proxy for bank learning. Sample is facilities originated between 1987 and 2003 to borrowers that were publicly traded both at time of loan and in 2008.

Dependent variable:	*CAR (% / 100)*	*Interest Rate spread over LIBOR (in bps)*	
Time period:	Mkt. Time 0	All	All
	(1)	(2)	(3)
All-in Spread at Mkt. Time 0	-0.000176**		
	(0.0001)		
log(Loan Amount) at Mkt. Time 0	-0.0139**		
	(0.006)		
1{Access to credit between 2006 and 2009}	0.0399**		
	(0.018)		
Relationship intensity between 2006 and 2009	-0.0047		
	(0.0032)		
Market Time		0.714	0.632
		(0.74)	(0.69)
Lehman Proxy × Market Time			-5.708***
			(2.20)
Lehman Proxy			8.087
			(6.06)
Borrower's Z score	0.0229***	-12.87***	-12.83***
	(0.00)	(1.05)	(1.05)
Naïve Probability of Default	0.011	117.6***	117.5***
	(0.033)	(10.90)	(10.70)
log(Total Assets)	0.0144**	-24.76***	-24.90***
	(0.006)	(1.09)	(1.09)
1{loan is secured}	-0.0524***	66.14***	65.45***
	(0.017)	(4.00)	(3.97)
1{loan is not secured}	0.0128	-13.05***	-13.06***
	(0.020)	(2.84)	(2.81)
Loan Maturity (months)	-3.81E-04	-0.024	-0.024
	(0.000)	(0.051)	(0.051)
1{revolver loan}	0.0408**	-58.18***	-58.20***
	(0.016)	(3.56)	(3.51)
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	2,871	8,673	8,673
R-squared	0.2	0.55	0.55

Standard errors clustered by borrower in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Notes.

The "Lehman proxy" is the 3-month cumulative abnormal return from a Fama-French-Carhart four-factor model in a [-21, +42] day window centered on the Lehman Bros. bankruptcy of 9/15/2008, orthogonalizec to borrower and loan characteristics at market time 0.

Column 1 reports a cross-sectional regression of each borrower's CAR on borrower and loan characteristics at the time of each borrower's first syndicated loan in our sample (market time 0). The residuals from this regression are the "Lehman proxy" in columns 2 and 3.

Colums 2 and 3 report panel regressions of syndicated loan interest spreads over LIBOR on the Lehman prc and contemporaneous borrower and loan characteristics from Compustat and CRSP. "Market time" count number of times each borrower has received credit in the syndicated loan market, starting at 0.

Table 3: Fixed Effect Regressions

Fixed-effect regressions of syndicated loan interest rates on proxy for bank learning. Sample is borrower-lead bank interactions from facilities originated between 1987 and 2003 to borrowers that were publicly traded both at time of loan and in 2008, weighted so each facility receies equal weight.

Dependent variable:	*Interest Rate spread over LIBOR (in bps)*			
	(1)	(2)	(3)	(4)
Market Time	0.632	2.819**	0.856*	-0.00642
	(0.690)	(1.100)	(0.520)	(1.170)
Lehman Proxy × Market Time	-5.708***	-6.246**	-4.879**	-6.193**
	(2.20)	(2.62)	(1.90)	(2.56)
Lehman Proxy	8.087	*Absorbed by FE*	4.333	*Absorbed by FE*
	(6.06)		(5.35)	
Borrower's Z score	-12.83***	-14.18***	-10.17***	-11.19***
	(1.05)	(1.80)	(0.87)	(2.28)
Naïve Probability of Default	117.5***	91.31***	113.4***	80.28***
	(10.7)	(10.7)	(9.5)	(11.1)
log(Total Assets)	-24.90***	-26.53***	-21.45***	-25.71***
	(1.09)	(3.74)	(1.07)	(6.53)
1{loan is secured}	65.45***	40.74***	58.37***	23.55***
	(3.97)	(3.99)	(3.32)	(5.12)
1{loan is not secured}	-13.06***	-4.374*	-12.30***	-5.332
	(2.81)	(2.57)	(2.44)	(3.33)
Loan Maturity (months)	-0.0244	-0.135***	0.056	-0.0277
	(0.051)	(0.050)	(0.057)	(0.055)
1{revolver loan}	-58.20***	-42.96***	-53.77***	-29.96***
	(3.510)	(2.960)	(3.670)	(2.790)
Year FE	YES	YES	YES	YES
Industry FE	YES			
Borrower FE		YES		
Lender FE			YES	
Relationship FE				YES
Number of facilities	8,673	8,673	8,614	8,614
R-squared	0.55	0.79	0.67	0.92

Standard errors clustered by borrower (cols. 1-2), lender (col. 3) and both (col. 4) in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Notes.

The "Lehman proxy" is the 3-month cumulative abnormal return from a Fama-French-Carhart four-factor model in a [-21, +42] day window centered on the Lehman Bros. bankruptcy of 9/15/2008, orthogonalized to borrower and loan characteristics at market time 0.

"Market time" counts number of times each borrower has received credit in the syndicated loan market, starting at 0.

Each observation is a borrower-lead bank interaction. Facilities with multiple lead banks appear multiple times in the sample but are downweighted to receive equal weight. Columns 3 and 4 omit 44 facilities for which we cannot identify a lead bank.

Table 4: Additional Robustness Tests

OLS regressions of syndicated loan interest rates on proxy for bank learning. Sample is facilities originated between 1987 and 2003 to borrowers that were publicly traded both at time of loan and in 2008.

Dependent variable:	*Interest Rate spread over LIBOR (in bps)*		
	(1)	(2)	(3)
Market Time	0.632	0.0978	-1.586
	(0.688)	(0.587)	(6.648)
Lehman Proxy ×	-5.675**	-4.745***	-4.477**
Market Time	(2.293)	(1.797)	(1.783)
Lehman Proxy	7.056	6.776	5.939
	(15.170)	(6.147)	(6.024)
Leman Proxy × time trend	0.0921		
(=2008 - Origination Year)	(1.169)		
Current ratio		-0.583	0.292
		(0.993)	(1.122)
M/B ratio of assets		3.014**	3.294**
		(1.345)	(1.530)
Profitability (% of sales)		-0.231	-0.482*
		(0.224)	(0.268)
Book leverage (% of assets)		0.308**	0.690***
		(0.121)	(0.139)
Log coverage ratio		-6.156***	-3.439
		(2.186)	(2.416)
Tangible Assets (% of assets)		-0.428***	-0.531***
		(0.099)	(0.114)
Log(Loan Amount)		-10.78***	-11.28***
		(1.504)	(1.918)
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Control variables	YES	YES	YES
S&P LT Debt Rating FE		YES	YES
Control variables × mkt. time			YES
Observations	8673	7701	7701
R-squared	0.553	0.598	0.575

Standard errors clustered by borrower in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Notes.

The "Lehman proxy" is the 3-month cumulative abnormal return from a Fama-French-Carhart four-factor model in a [-21, +42] day window centered on the Lehman Bros. bankruptcy of 9/15/2008, orthogonalized to borrower and loan characteristics at market time 0.

"Market time" counts number of times each borrower has received credit in the syndicated loan market, starting at 0.

Column 1 controls for the interaction between the proxy and the number of years until 2008.

Column 2 controls for additional borrower characteristics in both the first and second stages.

Column 3 controls for time-varying coefficients in all borrower / loan characteristics

Control variables are Z score, default probability, log(size), secured and revolver indicators, and loan maturity.

Table 5: Does Learning Affect Loan Size?

OLS regressions of log syndicated loan amount on proxy for bank learning. Sample is facilities originated between 1987 and 2003 to borrowers that were publicly traded both at time of loan and in 2008.

Dependent variable:		*Log(Loan Amount)*	
Sample:	All	Term Loans	Lines of Credit
	(1)	(2)	(3)
Market Time	0.0135***	0.0331***	0.00747**
	(0.004)	(0.009)	(0.004)
Lehman Proxy × Market Time	0.0127	-0.032	0.0285**
	(0.012)	(0.029)	(0.012)
Lehman Proxy	-0.0442	0.144	-0.107**
	(0.043)	(0.100)	(0.046)
Borrower's Z score	0.0914***	0.116***	0.0826***
	(0.008)	(0.019)	(0.008)
Naïve Probability of Default	-0.415***	-0.664***	-0.263***
	(0.057)	(0.110)	(0.067)
log(Total Assets)	0.676***	0.635***	0.687***
	(0.007)	(0.017)	(0.008)
1{loan is secured}	0.142***	0.375***	0.0247
	(0.029)	(0.065)	(0.031)
1{loan is not secured}	0.294***	0.547***	0.225***
	(0.031)	(0.110)	(0.030)
Loan Maturity (months)	0.00554***	0.0009	0.00869***
	(0.000)	(0.0008)	(0.001)
1{revolver loan}	0.521***		
	(0.028)		
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	8,673	2,120	6,553
R-squared	0.67	0.59	0.71

Standard errors clustered by borrower in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Note.

The "Lehman proxy" is the 3-month cumulative abnormal return from a Fama-French-Carhart four-factor model in a [-21, +42] day window centered on the Lehman Bros. bankruptcy of 9/15/2008, orthogonalized to borrower and loan characteristics at market time 0.

"Market time" counts number of times each borrower has received credit in the syndicated loan market, starting at 0.

Table 6: Bank Learning and the Distribution of Interest Rates

Quantile regression of syndicated loan facility interest rates on proxy for bank learning. Sample is loans originated between 1987 and 2003 to borrowers that were publicly traded both at time of loan and in 2008.

Dependent variable:	*Interest Rate spread over LIBOR (in bps)*				
Quantile:	0.05	0.25	0.50	0.75	0.95
Market Time	0.566*	0.239	-0.00623	0.447	0.806
	(0.300)	(0.350)	(0.420)	(0.710)	(1.050)
Private Info Proxy	**-0.297**	**-1.903****	**-3.968****	**-5.917****	**-8.756****
× Relationship Time	**(0.990)**	**(0.950)**	**(1.880)**	**(2.380)**	**(3.810)**
Private Info Proxy	-3.74	-1.504	4.863	7.088	16.44
	(3.390)	(4.720)	(5.660)	(7.100)	(13.700)
Borrower's Z score	-7.743***	-9.787***	-10.67***	-13.05***	-15.67***
	(0.910)	(0.870)	(0.940)	(1.200)	(1.850)
Naïve Probability of Default	27.79***	80.60***	104.5***	149.9***	219.3***
	(6.270)	(12.700)	(12.800)	(15.500)	(29.200)
Relationship Time	0.566*	0.239	-0.00623	0.447	0.806
	(0.300)	(0.350)	(0.420)	(0.710)	(1.050)
log(Total Assets)	-13.67***	-18.26***	-21.41***	-26.75***	-31.29***
	(0.940)	(0.990)	(1.040)	(1.260)	(1.800)
1{loan is secured}	30.47***	62.12***	82.31***	83.22***	64.54***
	(2.540)	(3.690)	(4.780)	(5.310)	(7.700)
1{loan is not secured}	1.69	-1.781	-3.036	-12.87***	-40.90***
	(1.570)	(1.850)	(2.040)	(3.310)	(6.480)
Loan Maturity (months)	-0.013	0.0274	0.0311	0.00651	2.94E-09
	(0.043)	(0.032)	(0.039)	(0.060)	(0.110)
1{revolver loan}	-13.08***	-28.77***	-49.97***	-65.82***	-124.5***
	(2.820)	(3.310)	(3.850)	(5.040)	(12.000)
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES
Observations	8,673	8,673	8,673	8,673	8,673
Pseudo R^2	0.19	0.34	0.42	0.42	0.36

Standard errors clustered by borrower in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Note.

The "Lehman proxy" is the 3-month cumulative abnormal return from a Fama-French-Carhart four-factor model in a [-21, +42] day window centered on the Lehman Bros. bankruptcy of 9/15/2008, orthogonalized to borrower and loan characteristics at market time 0.

"Market time" counts number of times each borrower has received credit in the syndicated loan market, starting at 0.

Table 7: Does the Market Charge Good Borrowers Less or Bad Borrowers More Over Time?

OLS regression of syndicated loan facility interest rates on proxy for bank learning, separated into positve and negative parts. Sample is loans originated between 1987 and 2003 to borrowers that were publicly traded both at time of loan and in 2008.

Dependent variable:	*Interest Rate over LIBOR (bps)*		*Log(Loan Amount)*	
	(1)	(2)	(3)	(4)
Market time	1.049**	3.569***	0.0173***	0.009
	(0.450)	(0.680)	(0.005)	(0.008)
(Lehman Proxy)$^{+}$	**-7.888*** **	**-10.78*** **	**-0.00781**	**-0.0131**
× Market Time	**(2.030)**	**(2.340)**	**(0.024)**	**(0.029)**
-(Lehman Proxy)$^{-}$	**-3.971** **	**-2.867**	**0.0269**	**0.0539** **
× Market Time	**(1.650)**	**(1.890)**	**(0.020)**	**(0.023)**
(Lehman Proxy)$^{+}$	21.87***	*Absorbed by*	-0.11	*Absorbed by*
	(7.490)	*FE*	(0.090)	*FE*
-(Lehman Proxy)$^{-}$	-3.034	*Absorbed by*	0.00789	*Absorbed by*
	(6.410)	*FE*	(0.077)	*FE*
Borrower's Z score	-12.76***	-14.28***	0.0908***	0.100***
	(0.650)	(1.100)	(0.008)	(0.014)
Naïve Probability of Default	117.5***	91.40***	-0.413***	-0.257***
	(4.750)	(5.420)	(0.057)	(0.067)
Market Time	1.049**	3.569***	0.0173***	0.00938
	(0.450)	(0.680)	(0.005)	(0.008)
log(Total Assets)	-24.71***	-26.43***	0.674***	0.485***
	(0.610)	(2.180)	(0.007)	(0.027)
1{loan is secured}	65.33***	40.76***	0.144***	0.149***
	(2.400)	(2.590)	(0.029)	(0.032)
1{loan is not secured}	-13.10***	-4.542*	0.293***	0.186***
	(2.560)	(2.550)	(0.031)	(0.032)
Loan Maturity (months)	-0.023	-0.132***	0.00554***	0.00469***
	(0.035)	(0.036)	(0.000)	(0.000)
1{revolver loan}	-58.22***	-42.92***	0.522***	0.402***
	(2.310)	(2.130)	(0.028)	(0.026)
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Borrower FE		YES		YES
F test: equality of learning coeffs	1.558	4.821	0.845	2.263
Prob > F	[0.212]	[0.028]	[0.358]	[0.133]
F test: joint sig. of learning coeffs	18.07	18.4	0.963	2.851
Prob > F	[0.000]	[0.000]	[0.382]	[0.058]
Observations	8,673	8,673	8,673	8,673
R-squared	0.55	0.79	0.67	0.84

Standard errors clustered by borrower in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Note.

The "Lehman proxy" is the 3-month cumulative abnormal return from a Fama-French-Carhart four-factor model in a [-21, +42] day window centered on the Lehman Bros. bankruptcy of 9/15/2008, orthogonalized to borrower and loan characteristics at market time 0.

"Market time" counts number of times each borrower has received credit in the syndicated loan market, starting at 0.

The proxy is separated into its positive and negative parts:

$(\text{Lehman Proxy})^{+} \equiv |\text{Lehman Proxy}| \times 1\{\text{Lehman Proxy} > 0\};$

$(\text{Lehman Proxy})^{-} \equiv |\text{Lehman Proxy}| \times 1\{\text{Lehman Proxy} < 0\}.$

Table 8: Private Learning Tests

OLS regressions of syndicated loan interest rates on proxies for public and private learning. "Public Sample" is facilities originated between 1987 and 2003 to borrowers that were publicly traded. "Lehman Sample" excludes borrowers not also appearing in CRSP in 2008. "RT" excludes loans that cannot be classified as relationship or transaction.

Dependent variable:	*Interest Rate spread over LIBOR (in bps)*				
Sample:	Lehman	Lehman	Lehman RT	Lehman RT	Public RT
	(1)	(2)	(3)	(4)	(5)
Market Time	0.825	0.84	0.707	0.707	1.687***
	(0.840)	(0.840)	(0.730)	(0.730)	(0.460)
Lehman Proxy ×	-5.564**	-4.432*	-6.365**	-6.217**	
Market Time	(2.200)	(2.440)	(2.490)	(2.670)	
Lehman Proxy	7.443	6.645	11.5	11.46	
	(6.020)	(6.060)	(8.130)	(8.080)	
Relationship Time of Lead	-0.962	-1.032			
Bank in Syndicate	(1.170)	(1.140)			
Proxy × Mkt. Time ×		**-0.455**			
Relationship Time		**(0.490)**			
1{Relationship Loan}			-0.428	-0.442	-6.360***
			(2.900)	(2.900)	(2.010)
Proxy × Mkt. Time ×				**-0.228**	
1{Relationship Loan}				**(2.250)**	
Borrower's Z score	-12.80***	-12.81***	-13.19***	-13.19***	-13.78***
	(1.050)	(1.050)	(1.190)	(1.190)	(0.810)
Naïve Probability of Default	119.4***	119.2***	125.5***	125.5***	106.0***
	(10.700)	(10.600)	(12.600)	(12.500)	(5.730)
log(Total Assets)	-24.84***	-24.84***	-24.90***	-24.90***	-27.48***
	(1.090)	(1.090)	(1.240)	(1.240)	(0.830)
1{loan is secured}	64.85***	64.91***	66.15***	66.16***	68.40***
	(3.960)	(3.970)	(4.700)	(4.700)	(2.900)
1{loan is not secured}	-13.27***	-13.31***	-10.02***	-10.03***	-15.36***
	(2.830)	(2.830)	(3.080)	(3.090)	(2.380)
Loan Maturity (months)	-0.0175	-0.0175	-0.012	-0.0118	-0.314***
	(0.050)	(0.050)	(0.057)	(0.057)	(0.042)
1{revolver loan}	-58.06***	-58.01***	-58.62***	-58.61***	-61.11***
	(3.510)	(3.510)	(4.090)	(4.080)	(2.480)
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES
Observations	8,614	8,614	6,545	6,545	16,567
R-squared	0.55	0.55	0.56	0.56	0.52

Standard errors clustered by borrower in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Notes.

The "Lehman proxy" is the 3-month cumulative abnormal return from a Fama-French-Carhart four-factor model in a [-21, +42] day window centered on the Lehman Bros. bankruptcy of 9/15/2008, orthogonalize to borrower and loan characteristics at market time 0.

"Market time" counts number of times each borrower has received credit in the syndicated loan market, starting at 0. "Relationship time" countes number of times each borrower has received credit from the lead bank in the syndicate, starting at 0.

Relationship Loan Indicator = 1 if the borrower has received credit from the same lead bank at least once in the previous five years, and 0 if the borrower has received credit in the previous five years, but only from different lead banks. Loans to borrowes that did not receive credit in the previous five years are excluded.

Table 9: Transparency and Learning

OLS regression of syndicated loan facility interest rates on proxy for public learning, interacted with indicator for borrower transparency. Sample is loans originated between 1987 and 2003 to borrowers that were publicly traded both at time of loan and in 2008.

Dependent variable:	*Interest Rate spread over LIBOR (in bps)*		
Transparency Measure:	Size > 50th pctl.	Tangibility > 50th pctl.	Has LT Debt Rating
	(1)	(2)	(3)
Market Time	-0.415	0.243	0.494
	(0.880)	(0.790)	(0.910)
Market Time ×	**1.137**	**0.66**	**0.0549**
1{Transparent Borrower}	**(1.100)**	**(1.090)**	**(1.100)**
Lehman Proxy × Market Time	-4.585**	-2.838	-3.177
	(1.990)	(2.160)	(1.980)
Lehman Proxy × Market Time	**-2.067**	**-3.975**	**-3.613**
× 1{Transparent Borrower}	**(2.870)**	**(2.920)**	**(2.370)**
1{Transparent Borrower}	-14.51***	5.15	13.21**
	(5.610)	(5.610)	(6.100)
Lehman Proxy	8.646	8.156	6.00
	(6.240)	(6.150)	(6.200)
Borrower's Z score	-13.92***	-14.17***	-11.08***
	(1.250)	(1.290)	(1.180)
Naïve Probability of Default	116.0***	115.2***	102.2***
	(11.300)	(11.100)	(11.000)
log(Total Assets)	-21.98***	-24.25***	-22.09***
	(1.510)	(1.160)	(1.420)
1{loan is secured}	64.14***	63.72***	53.30***
	(4.090)	(4.060)	(3.940)
1{loan is not secured}	-12.16***	-12.00***	-11.14***
	(2.870)	(2.830)	(2.680)
Loan Maturity (months)	-1.17E-02	-0.0259	-0.127**
	(0.053)	(0.053)	(0.052)
1{revolver loan}	-58.82***	-58.68***	-55.52***
	(3.730)	(3.660)	(3.450)
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Additional controls		Tangibility	Credit rating DVs
F test: sum of learning coefficients	5.78	6.58	10.02
Prob > F	0.016	0.010	0.002
Observations	7,912	7,908	7,912
R-squared	0.54	0.55	0.58

Standard errors clustered by borrower in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Notes.

The "Lehman proxy" is the 3-month cumulative abnormal return from a Fama-French-Carhart four-factor model in a [-21, +42] day window centered on the Lehman Bros. bankruptcy of 9/15/2008, orthogonalized to borrower and loan characteristics at market time 0.

"Market time" counts number of times each borrower has received credit in the syndicated loan market, starting at 0.

We classify borrowers as "transparent" based on their characteristics at market time 0. In column 1, transparent borrowers have assets above the time 0 median; in column 2, transparent borrowers have % tangible assets above the time 0 median; and in column 3, transparent borrowers are those with a long-term debt rating from S&P.

Appendix

In our empirical model, we assume that the fth firm's default probability at time t follows an error-components structure that may depend on the macroeconomic environment m_t, industry-i-specific shocks v_i and idiosyncratic firm shocks $\xi_{f,t}$: $\tilde{\pi}_{f,t} := \eta_f + \tilde{\xi}_{f,t} = \eta_f + \alpha_m' m_t + v_i + \xi_{f,t}$. We allow for arbitrary forms of cross-sectional and time-series correlation in the m_t and v_i components. These are nuisance parameters that may be removed by including time and industry fixed effects in our model, leaving two firm-specific components:

$$\pi_{f,t} := \eta_f + \xi_{f,t}$$

The parameter of interest to the bank as well as the econometrician is η_f, which we assume the bank does not know. We call this component a firm's latent *quality*. The following assumptions motivate our empirical strategy:

Assumption 1: There is a stationary distribution $F\left(\eta_f, \xi_{f,t}, x_{f,t}, z_{f,t}, b_f, s^{\tau}_{fb}, m_t, v_i\right)$ known by all bankers; i.e. bankers have symmetric information about the underlying distributions.

assumption 2: Our dataset contains a time-invariant, background firm characteristic b_f that is correlated with η_f but has no direct effect on the probability of default: $E(\pi_{f,t}|\eta_f, b_f) = E(\pi_{f,t}|\eta_f)$.

assumption 3: Non-interest contract features are conditionally uninformative about default probabilities: $E\left[\pi_{f,t}|x_{f,t}, z_{f,t}, s^{\tau}_{fb}, w_{l,fb,\tau}\right] = E\left[\pi_{f,t}|x_{f,t}, z_{f,t}, s^{\tau}_{fb}\right]$.

Assumption 4: Firm characteristics $(x'_{f,t}, z'_{f,t})$ are not informative about the idiosyncratic component of default probabilities: $E\left[\xi_{f,t}|x_{f,t}, z_{f,t}\right] = 0$.

assumption 5: Default probabilities $\{\pi_{f,t} : t = 1, ..., T\}$ are cross-sectionally independent draws from a conditional distribution $G\left(\pi_{f,t}|\eta_f, x_{f,t}, z_{f,t}\right)$; i.e., shocks are conditionally i.i.d. across firms.

Unlike Farber and Gibbons, we assume that the information held by banks about firm quality is asymmetric. All banks know the distribution $F\left(\eta_f, \xi_{f,t}, x_{f,t}, z_{f,t}, b_f, s_{fb}^{\tau}, m_t, v_i\right)$, and the conditional distribution $G\left(\pi_{f,t}|\eta_f, x_{f,t}, z_{f,t}\right)$, all observe $\{x_{f,t}, z_{f,t}\}$ and whether a firm has defaulted or not, but they differ on their observed set of signals s_{fb}^{τ} as well as the number of signals (the length of the relationship) τ. The claim that we test in this paper is that access to these private signals allows the inside bank to price loans to firm f better than outside banks with a less-established relationship.

Imagine a panel dataset covering a cohort of firms entering the market for bank loans and taking out one-period loans from initially identical, perfectly competitive banks. The data reveal some firm and loan characteristics relevant for loan pricing ($z_{f,t}$ and $w_{l,fb,\tau}$, respectively) when the loan is applied for at the beginning of each period, but omits some firm characteristics $x_{f,t}$ relied on by the banks. Motivated by our linearized model (2), and given Assumptions 1-5, we could estimate the following population linear projection:

$$\begin{aligned}
E^*[r_{l,fb,\tau}|z_{f,t}, w_{l,fb,\tau}] &= \alpha_t + \alpha_i + \alpha_1 E^*[E[\pi|x_{f,t}, z_{f,t}, s_{fb}^{\tau}]|z_{f,t}, w_{l,fb,\tau}] + \gamma' w_{l,fb,\tau} \\
&= \alpha_t + \alpha_i + \alpha_1 E^*[\pi|z_{f,t}, w_{l,fb,\tau}] + \gamma' w_{l,fb,\tau} \\
&= \alpha_t + \alpha_i + \beta^{z\prime} z_{f,t} + \beta^{w\prime} w_{l,fb,\tau}
\end{aligned}$$

We use Assumption 3 to apply the Law of Iterated Linear Projections. The coefficient on w reflects both the substitutability between other loan characteristics and interest rate spreads (γ) and the correlation between w and omitted firm characteristics x and private signals s.[21] Similarly, the coefficient on z incorporates both direct and indirect pricing effects due to omitted variables.

[21] In our empirical specifications we find that the second factor dominates. For example, loans with more collateral pay higher interest rates, presumably because these firms differ on omitted characteristics.