Hard Information and Local Lending Markets: Evidence from Invoice Digitalization*

Yunyue Chen[1,2]

[1]Ghent University [2]NYU Stern School of Business

August 11, 2026

Abstract

Does hard information substitute for soft information in small business lending? I exploit the staggered provincial rollout of China's Golden Tax Phase III reform. The reform made transaction records verifiable in underwriting. Confidential data cover approximately 154,000 loan originations from multiple banks between 2012 and 2022. Banks revise assessed borrower income and repayment capacity upward, approve credit lines about 20 percent larger, reduce rates and spreads by roughly 40 basis points, and extend maturities. The central evidence comes from distance: distance schedules in quantity and price flatten, gains rise with borrower distance, and banks lend about 15 percent farther. Effects are strongest where preexisting verifiable information is scarce. At the expanded lending frontier, comparable internal repayment capacity scores, longer travel time, and higher loan rates are consistent with residual contact and monitoring costs. The credit response is also stronger in more competitive markets, consistent with invoice verification weakening the proprietary value of soft information accumulated through lending relationships. Hard information therefore substitutes for soft information in bank screening, weakens local informational advantages, and makes small business lending less local.

Keywords: Banking, Small Business Lending, Financial Technology, Information Asymmetry, Digital Infrastructure

JEL Codes: G21, D82, O33

*I am grateful to Yakov Amihud, Quirin Fleckenstein, Ziying Fan, Michael Frömmel, Yiqing Lü, Jingyuan Mo, Leentje Moortgat, Cecilia Parlatore, Xuyuanda Qi, Tianyue Ruan, Anthony Saunders, and Alexi Savov for helpful comments and suggestions. All errors are my own. Formerly circulated as: *Invoice Digitization and Credit Availability: Hard Information Effects in Small Business Lending.*

1 Introduction

Soft information makes small business lending possible and keeps it local. Opaque firms often lack audited accounts and standardized records, so banks learn about them through repeated interactions, local contact, and site visits (Petersen and Rajan, 1994; Berger and Udell, 1995). The resulting knowledge is soft information, which is specific to the borrower and difficult to transmit through bank hierarchies (Stein, 2002; Liberti and Petersen, 2019). Its precision declines with distance, limiting the scale and geographic reach of small business credit (Petersen and Rajan, 2002; Agarwal and Hauswald, 2010). Hard information, by contrast, can be verified, standardized, and transmitted within bank organizations, allowing banks to screen opaque firms with less reliance on soft information (Fama, 1985; Liberti and Petersen, 2019). This raises two central questions in academic and policy circles: Does hard information substitute soft information in loan origination? And does this substitution benefit the borrower historically underestimated by banks? My study show that hard information improves precision of information set, brings better loan terms, reduces the sensitivity of loan terms to distance, and leads banks to lend farther away. Where prior borrower information is scarce, hard information matters more. These results show that hard information substitutes for soft information in bank screening and makes small business lending less local.

China's Golden Tax Phase III reform (GT3) provides a setting in which transaction records become hard information for bank underwriting. Rolled out across provinces between 2013 and 2016, GT3 integrated invoice recording and authentication into a unified national tax administration system. The resulting digital records linked counterparties, amounts, and dates and could be authenticated through the tax system. When firms submitted invoices, tax filings, contracts, receivables documents, and related materials in loan applications, banks could use the authenticated records to assess reported revenue and transmit the supporting evidence through internal approval layers. Public descriptions of bank loan products document the use of invoices, tax filings, and related records in small business

credit review. Designed to modernize tax administration, GT3 increased the evidentiary value of documents banks already collected.

I match provincial implementation dates to approximately 154,000 confidential small business loan originations from nine banks between 2012 and 2022. The records connect banks' internal assessments of annual income and repayment capacity to the contract terms approved at origination and distance between borrower and lender. I assign exposure using the implementation date in the borrower's registered province. GT3 was rolled out province by province under a centrally coordinated national deployment plan designed to standardize tax administration and data processing. The credit line event study shows no evidence of differential movement before implementation and a clear increase afterward. Cohort stacked estimates yield same conclusions. This dataset allows me to trace the hard information shock from banks' internal assessments into contract decisions and the geographic reach of lending.

Loan contracts improve materially after GT3. In the preferred specifications, banks approve credit lines about 20% larger, reduce interest rates and spreads by roughly 40 basis points, and extend loan maturities. The geographic distribution of these gains reveals whether hard information substitutes for distance sensitive soft information.

Before GT3, loan terms deteriorate with distance. Banks approve smaller credit lines and charge higher rates and spreads as borrowers become more distant. All three distance schedules flatten after implementation. Moving from the 25th to the 75th percentile of distance increases the GT3 credit line response by about 6% and produces an additional reduction of roughly 12 basis points in the interest spread. Banks also lend about 15% farther after the reform. Credit quantity, price, and realized lending reach therefore move together. This joint pattern provides the paper's central evidence that hard information substitutes for distance sensitive soft information in bank screening. By reducing the informational content of distance, hard information changes the geography of small business lending.

The internal loan files show how this substitution occurs. Banks' assessed annual income increases by about 21%, while internal repayment capacity scores improve by 6.6 percentile

points. The increase in assessed income is larger for more distant borrowers, where soft information was initially less precise. Provincial tax collections also rise after implementation, consistent with more reported activity entering the formal reporting and verification system. The reform changes banks' assessment of borrower revenue and repayment capacity before contract terms are approved. These revisions show that banks place greater weight on verifiable revenue evidence precisely where soft information was previously weakest.

The remaining role of distance helps locate the substitution. At the expanded lending frontier, borrowers have comparable internal repayment capacity scores, while driving time and loan rates increase. This pattern is consistent with residual contact and monitoring costs outside the screening margin affected by hard information. The boundary evidence therefore reinforces the interpretation that hard information substitutes for soft information in screening.

Hard information should matter most where preexisting verifiable information is scarce. Credit line responses are larger in consumer facing industries, where transactions previously generated fewer records verified by third parties than interfirm trade (Pomeranz, 2015; Naritomi, 2019). Banks with lower industry specialization before GT3 also respond more strongly. These banks begin with less accumulated sector specific screening expertise, making verified transaction records more valuable at the margin (Stein, 2002; Blickle et al., 2026). After GT3, bank lending also covers a broader set of industries. Listed firms provide a useful benchmark. Audited accounts and public disclosures already supply lenders with extensive standardized information, and listed firms' borrowing shows no response to GT3 (Healy and Palepu, 2001; Minnis, 2011; Minnis and Sutherland, 2017). Across industries, banks, and disclosure environments, the effects of hard information are largest where the preexisting stock of verifiable information is smallest.

The substitution also has a market dimension. Soft information built through a lending relationship gives the incumbent bank an advantage because competitors cannot readily verify or reuse it (Broecker, 1990; Hauswald and Marquez, 2003). GT3 adds an external

verification standard whose credibility does not depend on any single lending relationship. The credit line response is stronger in less concentrated local banking markets, consistent with competition allowing more of the improvement in verifiability to be reflected in approved credit and with invoice verification reducing the proprietary value of soft information accumulated through lending relationships. Market structure therefore shapes how gains from verifiable hard information are transmitted into approved credit.

This paper makes three contributions. First, it provides direct causal evidence that hard information substitutes for soft information in bank screening and changes the geography of small business lending. Relationship lending allows banks to evaluate opaque firms through repeated interaction and borrower specific knowledge (Petersen and Rajan, 1994; Berger and Udell, 1995; Bharath et al., 2011). Reliance on this knowledge also constrains the reach of lending (Granja et al., 2022). Soft information is difficult to communicate through bank hierarchies (Stein, 2002; Liberti and Mian, 2009), and its precision declines as the distance between borrower and lender increases (Degryse and Ongena, 2005; Agarwal and Hauswald, 2010). Research on credit scoring shows that standardized information can support lending at greater scale and distance (Berger and Frame, 2007; DeYoung et al., 2008). Recent theory conceptualizes technological change as an expansion in the span of hard information available to lenders (He et al., 2024). The open question is whether hard information causes banks to place less weight on soft information in small business lending. GT3 supplies this variation, while the confidential loan files reveal banks' internal assessments, contract decisions, and realized lending distance. Banks revise assessed income and repayment capacity, improve terms more for distant borrowers, flatten the distance schedules in credit quantity and price, and extend lending farther away. The common movement in internal assessments, loan quantity, pricing, and realized lending distance reveals a change in how banks screen opaque firms. By substituting for soft information which is distance sensitive, hard information changes contract design and decreases credit inequality based on the geography.

Second, the paper connects tax digitalization to bank underwriting by showing how tax authenticated records become hard information for credit allocation. Research on VAT paper trails and third party reporting shows how transaction records strengthen tax reporting and enforcement (Pomeranz, 2015; Naritomi, 2019). Electronic invoicing increases the standardization and verifiability of business transactions (Bellon et al., 2022). A related banking literature shows that credible financial statements and transaction account information shape lender assessments and loan contracts (Norden and Weber, 2010; Minnis, 2011; Cassar et al., 2015). This paper traces a tax authentication shock into bank assessed income, repayment capacity, and origination decisions. Closest to this paper, Casado et al. (2026) show that an eInvoicing mandate reallocates credit toward firms already using invoice based finance and away from other firms as lenders redirect information production toward borrowers represented in invoice credit markets. Their analysis centers on credit reallocation according to firms' positions in invoice finance before the reform. This paper studies whether tax authenticated records enter ordinary small business underwriting and substitute for soft information where prior verifiable information is scarce. Here, the gains are largest where the preexisting stock of verifiable information is smallest. The comparison highlights a broader implication: the distributional effects of invoice digitalization depend on how verified records enter lender screening.

Third, the paper develops a market implication of external verification. Digital footprints, payment records, and platform data expand the information available to lenders (Berg et al., 2020; Hau et al., 2024; Minnis et al., 2024). In consumer credit, Ouyang (2026) shows that payment-flow data inform loan assessments and expand access, with larger effects for financially underserved consumers. Research on open banking studies how formal access rights to financial data affect screening and competition (Parlour et al., 2022; He et al., 2023). Agarwal and Hauswald (2010) show that proximity allows banks to produce proprietary soft information and carve out captive local markets: nearby borrowers receive greater access to credit and pay higher interest rates. GT3 works in the opposite direction. An invoice verifi-

cation system produces hard information outside any single lending relationship and reduces the extent to which credible borrower information remains proprietary to the incumbent bank. The stronger credit response in more competitive markets is consistent with external verification weakening this informational advantage and increasing the pass through of information improvements into approved credit. Public digital infrastructure can therefore produce hard information that weakens the informational basis of local lender market power and expands the geographic reach of small business credit.

2 Institutional Setting and Predictions

2.1 GT3 and the Production of Hard Information

Transaction records are useful in lending only to the extent that a bank can establish what they represent. Invoices identify the counterparties, amount, date, and taxable content of a transaction. VAT paper trails and third-party reporting make these records harder to manipulate because the same transaction can enter the records of more than one party (Pomeranz, 2015; Naritomi, 2019). Electronic invoicing further standardizes how such records are generated and retained (Bellon et al., 2022).

GT3 was a nationwide modernization of China's tax administration. The component relevant to this paper standardized tax administration, centralized tax data processing, and supported electronic invoice administration and nationwide invoice authenticity queries.[1] By placing the underlying invoice information within a standardized authentication system, GT3 increased the evidentiary value of borrower-supplied transaction records.

Existing credit-application channels allowed borrowers to supply invoices and related transaction documents to banks. Public descriptions of invoice-based and small business credit products from different banks of China document the use of invoices, contracts, re-

[1] See Shanghai Municipal Tax Service, State Taxation Administration, official description of GT3 (accessed July 12, 2026).

ceivables, delivery documents, and invoicing histories in credit review.[2] Appendix Table A.2 reports the product-level evidence.

Figure 1 summarizes this institutional channel. GT3 authenticates invoice records that borrowers supply through existing credit-application channels, and banks use these authenticated records in credit review.

Public authentication is the institutional source of the change, while hard information is the economic input whose use the paper studies. The following framework develops the implications for internal assessments, approved contracts, and lending across distance.

2.2 Hard Information in Bank Screening

A precision-weighted updating framework organizes the screening predictions. Let q_i denote borrower i's latent repayment capacity and let $\hat{q}_i^{pre}$ denote the bank's assessment before incorporating the new hard signal. The pre-GT3 information set combines existing hard information with soft information. Assume these components are conditionally independent, unbiased signals of q_i on a common scale, so precision is additive:

$$\tau_i^{pre} = \tau_{H0,i} + \tau_S(d_i), \qquad \tau_S'(d_i) < 0,$$

where $\tau_{H0,i}$ is the precision of existing hard information, $\tau_S(d_i)$ is the precision of soft information, and d_i is borrower to lender distance. Greater separation makes repeated interaction and the transmission of borrower-specific knowledge more difficult (Petersen and Rajan, 2002; Agarwal and Hauswald, 2010).

[2]See ICBC domestic invoice financing; Agricultural Bank of China domestic trade financing; Agricultural Bank of China "Zhang Ying Tong"; Ping An Bank supermarket supplier invoice loan; and Ping An Bank Speed Invoice Loan (accessed July 12, 2026).

Authenticated records supplied in a credit application provide an additional hard signal h_i with precision τ_H. The bank's updated assessment is

$$\begin{aligned} \hat{q}_i^{post} &= \frac{\tau_i^{pre}\hat{q}_i^{pre} + \tau_H h_i}{\tau_i^{pre} + \tau_H} = \hat{q}_i^{pre} + K_i\big(h_i - \hat{q}_i^{pre}\big), \\ K_i &= \frac{\tau_H}{\tau_{H0,i} + \tau_S(d_i) + \tau_H}. \end{aligned} \tag{1}$$

The share of total screening precision attributable to soft information therefore falls:

$$\omega_{S,i}^{pre} = \frac{\tau_S(d_i)}{\tau_{H0,i} + \tau_S(d_i)} > \omega_{S,i}^{post} = \frac{\tau_S(d_i)}{\tau_{H0,i} + \tau_S(d_i) + \tau_H}.$$

This decline in the screening weight placed on soft information is the sense in which hard information substitutes for it. Signal content determines the direction of the average revision. Holding the signal-prior gap fixed, the revision is larger where the pre-GT3 information set is less precise.

The updating framework yields three screening predictions.

Underwriting. Authenticated records inform bank assessed income and the internal repayment capacity score. Favorable revisions support larger credit lines, lower rates and spreads, and longer maturities. Together, the assessment and contract outcomes test whether hard information enters underwriting and changes lending decisions.

Distance. Lower screening precision can restrict quantity and raise an informational premium, producing smaller credit lines and higher rates for more distant borrowers (Stiglitz and Weiss, 1981; Petersen and Rajan, 2002; Agarwal and Hauswald, 2010). Because $\tau_S(d_i)$ falls with distance, Equation (1) implies that the weight on the new hard signal rises with borrower to lender distance. If authenticated records imply a favorable average revision, their greater weight at distance predicts flatter quantity and price schedules, larger contract responses for more distant borrowers, and an expansion in realized lending distance. A general expansion in credit supply can improve average contracts without naturally predicting simultaneous flattening across quantity and price. This joint flattening provides the paper's

signature test of hard information substituting for soft information in screening. Related theory shows how expanding the span of hard information changes the borrower characteristics that lenders can assess (He et al., 2024).

Information environments. The response should be strongest where the information used in screening was weakest before GT3. In B2C industries, transactions more often end with consumers rather than another reporting firm, leaving a thinner cross-firm paper trail (Pomeranz, 2015; Naritomi, 2019). A specialized bank, by contrast, has accumulated industry-specific screening knowledge, so its prior information set should be more precise (Blickle et al., 2026). Listed firms provide the opposite borrower benchmark because public disclosure makes their information environment richer (Healy and Palepu, 2001). More generally, verified financial statements are valuable in debt contracting (Minnis, 2011). The three comparisons locate distinct sources of pre-GT3 information: the availability of borrower transaction records, the bank's accumulated screening knowledge, and public disclosure.

Beyond screening, two predictions concern the market transmission of hard information and the boundary of substitution.

Market transmission. Invoice verification can reduce the informational advantage of proprietary borrower knowledge, while local competition shapes how improvements in screening reach borrowers through approved credit. Related models show how wider access to borrower information changes screening incentives and lender competition (Hauswald and Marquez, 2003; He et al., 2023). The empirical prediction is a larger credit-line response in more competitive local markets, a pattern that would be consistent with greater pass-through of the information improvement.

Boundary. A separating prediction follows because GT3 changes screening information while travel technology remains unchanged. It should attenuate the informational component of the distance penalty, but it cannot reduce travel for site visits, borrower contact, and monitoring (Vives and Ye, 2025). Banks should therefore lend farther and flatten the quantity and price schedules, yet loans requiring more travel should retain a geographic premium. A

general decline in geographic frictions would weaken both informational and travel penalties. The predicted combination of flatter distance schedules and a residual travel premium instead distinguishes screening substitution from that alternative.

3 Data and Empirical Design

3.1 Loan Data and Treatment Assignment

I use confidential data on approximately 154,000 small-business loan originations from several banks between 2012 and 2022. Each observation is a newly originated loan. The records link banks' internal borrower assessments to the contract terms approved at origination and to the borrower and lender locations used to measure lending geography.

For each origination, the data report the origination date, the borrower's registered location and industry, the originating bank, and the city of the originating branch. The underwriting file contains bank assessed income and the bank's internal repayment capacity score. Contract fields include the approved credit line, interest rate, interest spread, and maturity. Appendix Table A.4 reports variable construction, transformations, units, and merge levels.

Treatment assignment follows the staggered provincial rollout of GT3. Let $p(i)$ denote borrower i's registered province and $t(i)$ the loan origination date. The exact-date indicator $GT3_{p(i),t(i)}$ equals one when loan i is originated on or after the GT3 implementation date in province $p(i)$. The borrower's registered province and the province of the originating branch coincide in the loan sample, so each origination maps to a unique treatment province and implementation date. The provincial rollout calendar appears in Appendix Table A.1.

The sample observes assessments and approved terms among originated loans. It does not contain rejected applications and therefore does not identify approval probabilities, new credit access, or market-wide credit allocation.

3.2 Bank Assessments, Contracts, and Distance

The outcome variables map directly into the paper's economic questions. Internal assessments reveal whether hard information enters underwriting, contract terms reveal whether it changes approved lending decisions, and borrower to lender distance provides the geographic margin used to test substitution in screening.

I use two measures of the bank's assessment of borrower repayment capacity at origination. The first is $\ln(\textit{Income})$, the log of bank assessed annual borrower income recorded in the underwriting file. The second is the internal repayment capacity score, reported as Credit Score in the tables. Because the nine banks use different internal rating systems, I map their original scores into within-bank percentile ranks from 1 to 100, with higher values indicating stronger repayment capacity. The resulting percentile is a bank-specific internal underwriting measure placed on a common scale.

The contract variables describe the quantity, price, and maturity of approved credit. Credit Line is the limit approved at origination, and $\ln(\textit{Credit Line})$ is the log of the RMB amount expressed in thousands. Interest Rate is the contractual nominal annual rate; Interest Spread is the contractual rate minus the contemporaneous Loan Prime Rate; and Maturity is the contractual term measured in days.

The principal geographic measure is borrower to lender distance. I calculate Distance in kilometers between the centroid of the borrower's registered district or county and the centroid of the city of the originating branch, using WGS84 coordinates. The main specifications use $\ln(\textit{Distance})$. Because the loan data identify the city of the originating branch rather than its street address, the lender endpoint is measured at the branch-city level. Appendix Table A.4 reports the geographic construction and transformations.

Borrower to lender distance is the geographic margin along which both screening frictions and the physical costs of contact and monitoring can vary. The analysis therefore combines the baseline distance measure with routed driving distance and driving time. Driving dis-

tance provides an alternative measure of geographic separation, while driving time measures the travel burden near the expanded lending frontier.

3.3 Supporting Data

Provincial tax statistics from CNRDS measure reporting outcomes around the GT3 rollout. The unit of observation is province by tax category by taxpayer category by year, and the principal outcome is reported tax collections. GT3 exposure begins in the provincial rollout year, and the annual specifications include lagged macroeconomic controls. This series provides upstream evidence on whether reported tax collections change following GT3.

Listed-firm borrowing announcements from CSMAR provide a benchmark in which lenders already have access to extensive public disclosure and audited financial statements. The outcome is the disclosed upper bound of an announced credit line, transformed in the same way as the credit-line measure in the main loan sample. Treatment is assigned using the lending-branch province and the announcement date.

Local banking concentration from CNRDS captures market structure around each origination. I construct HHI_{ct} from bank branch shares within city c and year-month t, with higher values indicating a more concentrated local banking market. The competition analysis uses this measure to study how the pass-through of information improvements varies with local concentration.

Geographic routing data from Baidu Map provide routed driving distance and estimated driving time between the borrower and lender endpoints used for the baseline distance measure. Appendix C describes the expansion-boundary construction and reports the driving-distance analysis.

Finally, I construct a bank fiscal-year panel from publicly filed annual reports(CNRDS) to assess the aggregate scope of the loan-level response. The panel contains measures of funding costs, loan growth, total lending, asset composition, nonperforming loans, and reserve coverage. To assign provincial GT3 exposure, the bank-year sample is restricted to

banks whose domestic branch networks are confined to a single province. The corresponding estimates appear in Table 12. Appendix Table A.3 reports summary statistics for the listed-firm, tax, and bank-year samples.

3.4 Summary Statistics

Table 1 summarizes the main loan origination dataset. The median approved credit line is RMB 800,000, and the mean is approximately RMB 1.52 million. The average contractual interest rate is 7.29%, the average interest spread is 2.66 percentage points, and average maturity is approximately 507 days. Among loans with matched locations, mean borrower to lender distance is 27.2 kilometers, with an interquartile range from approximately 8.1 to 38.8 kilometers.

Observation counts differ because assessment fields and matched locations are not available for every origination. The contract variables contain approximately 154,000 observations, bank assessed income is available for approximately 126,000 loans, the internal repayment capacity score for approximately 101,000 loans, and borrower to lender distance for approximately 97,000 loans. Table 1 reports row-wise non-missing observations; each regression uses the corresponding non-missing estimation sample.

3.5 Empirical Design and Identification

Identification comes from differences in the timing of provincial GT3 implementation and within-province changes around the exact rollout date. Let $m(i)$ denote the year-month of origination, $b(i)$ the bank, $c(i)$ the originating city, and $j(i)$ the borrower industry. The baseline difference-in-differences specification is

$$Y_i = \beta GT3_{p(i),t(i)} + \alpha_{m(i)} + \mu_{b(i)} + \lambda_{c(i)} + \gamma_{j(i)} + X_i'\delta + \varepsilon_i, \tag{2}$$

where Y_i is the relevant assessment, contract, or geographic outcome. The preferred loan-level specifications include Year-Month, Bank, City, and Industry fixed effects. The vector X_i contains grouped borrower and contract controls entered as indicator sets with explicit missing categories. Controls that mechanically overlap with the dependent variable are omitted. Income-group fixed effects are omitted from bank-assessed-income regressions, Credit-score fixed effects from internal-repayment-capacity-score regressions, and Loan-term fixed effects from maturity regressions. Appendix B reports the complete control definitions and column progressions. Standard errors are clustered at the province-by-year-month level.

The paper's central specification asks whether GT3 changes the borrower to lender distance gradient:

$$\begin{aligned} Y_i = {} & \beta_1 GT3_{p(i),t(i)} + \beta_2 \ln(Distance_i) + \beta_3 \left[GT3_{p(i),t(i)} \times \ln(Distance_i)\right] \\ & + \alpha_{c(i)\times m(i)} + \mu_{b(i)} + \gamma_{j(i)} + X_i'\delta + \varepsilon_i. \end{aligned} \tag{3}$$

City × Year-Month fixed effects absorb local monthly lending conditions. Because $GT3_{p(i),t(i)}$ switches on at the exact implementation date, it continues to vary within the implementation city-month, and β_1 is identified from changes within that local monthly environment around rollout. The coefficient β_2 is the pre-GT3 distance gradient, $\beta_2+\beta_3$ is the post-GT3 gradient, and β_3 measures the change in slope. A flatter quantity schedule implies a less negative credit-line gradient, while flatter price schedules imply smaller positive rate and spread gradients.

The joint behavior of quantity and price is the signature test of substitution. A general expansion in lending can improve average contract terms without predicting that the credit-line, rate, and spread schedules all become less sensitive to distance. Simultaneous flattening across quantity and price is instead the pattern implied when hard information substitutes for soft information in screening. I complement the slope estimates by using $\ln(Distance)$ as the outcome in Equation (2), which tests whether banks originate loans farther from their lending locations after GT3.

The identifying assumption is that, conditional on the fixed effects and controls, loan outcomes would have followed parallel trends absent GT3. The identifying variation comes from the centrally coordinated staggered rollout described in Section 2. I examine pre-treatment dynamics using

$$Y_i = \sum_{k \neq -1} \beta_k 1\{\tau_{p(i),t(i)} = k\} + \alpha_{m(i)} + \mu_{b(i)} + \lambda_{c(i)} + \gamma_{j(i)} + X_i'\delta + \varepsilon_i, \tag{4}$$

where $\tau_{p(i),t(i)}$ is event time in quarters and the quarter immediately before implementation is omitted. The pre-GT3 coefficients describe differential pre-treatment dynamics, while the post-GT3 coefficients trace the timing of the response. Figure 2 reports the credit-line event study. Figure 3 places the realized credit-line estimate of 0.179 in the extreme right tail of the distribution generated by 500 random permutations of the GT3 indicator across loan observations.

I also estimate a stacked difference-in-differences design for the principal assessment, quantity, and pricing outcomes. Each stack is centered on an implementation cohort and spans 24 months before through 25 months after the cohort-specific implementation month. The adopting cohort is compared with not-yet-treated provinces in the same event window, and the regressions absorb cohort-specific fixed effects (Cengiz et al., 2019; Baker et al., 2022). Appendix B.3 reports the stacked credit-line event study and corresponding estimates for bank assessed income, credit quantity, and loan pricing. The stacked estimates yield the same qualitative conclusions as the baseline design.

The remaining comparative-static tests preserve the same treatment timing while examining distinct economic margins. Borrower information environments, bank specialization, and listed firms measure where the marginal value of hard information should differ; local concentration measures variation in market pass-through. Specifications using supporting data are estimated at the native frequency and unit of each dataset.

4 Hard Information and Bank Underwriting

4.1 Loan Contracts

GT3 changes the terms on which banks originate small business credit. Table 2 shows that banks approve larger credit lines, lower rates and spreads, and longer maturities after implementation. In the preferred specifications, credit lines increase by about 20%, interest rates decline by about 40 basis points, interest spreads decline by about 44 basis points, and maturities increase by approximately 49 days.

Figure 2 shows no systematic differential movement before GT3; the post-GT3 estimates are positive and become largest several quarters after rollout.

These estimates establish an economically important contract response. I next examine the reporting environment and the assessments recorded inside bank underwriting.

4.2 Reporting Environment and Bank Assessments

Section 2.1 describes how borrower-supplied authenticated records enter credit review. To locate the contract response within underwriting, I examine the reporting environment and banks' internal borrower assessments.

Table 3 shows that reported provincial tax collections increase by about 8% after implementation, documenting a change in aggregate reporting outcomes around the GT3 rollout.

The response is also visible inside bank underwriting. Table 3 shows that bank assessed income increases by about 21%, while the internal repayment capacity score rises by 6.6 percentile points.

These results show that hard information enters bank underwriting and changes the terms approved at origination. Section 5 tests whether this hard information substitutes for soft information in screening.

5 Hard Information Substitutes for Soft Information

Distance provides the paper's sharpest test of whether hard information substitutes for soft information in screening. Substitution predicts flatter distance penalties in credit quantity and price and an expansion in realized borrower to lender distance. Evidence at the expanded lending boundary then helps distinguish substitution in screening from the physical costs of contact and monitoring.

5.1 Distance Schedules in Credit Quantity and Price

Before GT3, borrower to lender distance is associated with both the quantity and price of approved credit. Table 4 and Figure 5 show that credit lines decline with distance, while interest rates and spreads rise. The fitted post-GT3 schedules are substantially flatter. The credit-line gradient moves from -0.123 to -0.070, while the rate and spread gradients each fall from about 0.19 to about 0.08. Approved credit quantity and price therefore become markedly less sensitive to distance.

Table 5 reports the interaction specification in Equation (3). City $\times$ Year-Month fixed effects absorb local monthly conditions, so the interaction compares within-city-month distance gradients before and after GT3. The interaction between GT3 and $\ln(\textit{Distance})$ is positive for the approved credit line and negative for the interest rate and spread. Moving from the 25th to the 75th percentile of the pre-GT3 borrower to lender distance distribution increases the credit-line response by about 6% and makes the decline in the interest spread about 12 basis points larger.

The response of bank assessed income also rises with distance, linking the contract-schedule changes to the internal assessment evidence in Section 4. Figure 6 plots the implied point estimates across the observed distance distribution. Appendix Table C.2 reports the corresponding pointwise estimates and linear-combination tests, while Appendix Table C.1 reports similar interaction estimates using routed driving distance.

Because a general credit expansion can improve average quantities and prices without reducing their distance sensitivity, the joint flattening across both margins provides the paper's signature test of substitution. The evidence shows that hard information substitutes for soft information in bank screening.

5.2 The Geographic Reach of Lending

The schedule rotation has a realized geographic counterpart. Table 6 shows that, among observed originations, banks originate loans approximately 15% farther from their lending locations after GT3. Together, the flatter schedules and the increase in average borrower to lender distance show that hard information changes both contract design and the geographic reach of lending, making small business lending less local. That expansion defines the lending boundary used to examine which geographic costs remain.

5.3 The Boundary of Substitution

GT3 changes screening information without directly altering the travel technology required for borrower contact and monitoring (Heider and Inderst, 2012; Heitz et al., 2026). This creates a separating prediction at the expanded lending boundary: internal assessments can remain comparable while driving times and loan rates rise at the outer margin.

I construct a city-specific expansion boundary using the post-GT3 lending frontier and the estimated average expansion in log distance reported in Table 6. Table 7 compares post-GT3 loans in adjacent bands just inside and outside this boundary. Appendix C.1 describes the cutoff and band construction in detail.

The comparison reveals no detectable difference in the internal repayment capacity score across the two bands. Loans in the outer band involve approximately 14.9 additional minutes of driving and carry rates about 16 basis points higher. Table 7 and Figure 4 report a local RD analysis around the same constructed expansion boundary. The local estimates show

higher rates and driving times on the outer side of the boundary, while the internal repayment capacity score shows no statistically detectable discontinuity.

The higher driving time and rates in the outer band are consistent with residual contact and monitoring costs outside the screening margin affected by hard information. Their coexistence with flatter quantity and price schedules and expanded geographic reach helps distinguish substitution in screening from a general decline in geographic frictions. The joint evidence shows that hard information substitutes for soft information in bank screening and makes small-business lending less local. The next section asks where the marginal value of hard information is greatest.

6 Where Hard Information Matters

Equation (1) predicts that hard information has greater marginal value when the information used in screening was weaker before GT3. I examine this prediction across borrower transaction-record environments, banks' accumulated industry knowledge, and borrowers with extensive public disclosure. Each comparison isolates a distinct source of pre-GT3 screening information.

6.1 Borrower Information Environments

Consumer-facing industries provide a borrower-side test of this prediction. Interfirm transactions are more likely to leave corresponding records at two reporting firms, whereas transactions with final consumers generate a thinner cross-firm paper trail (Pomeranz, 2015; Naritomi, 2019). B2C exposure therefore serves as a proxy for the borrower's pre-GT3 transaction-record environment.

Following the classification approach in this literature, I begin with a conservative definition of consumer-facing industries and progressively broaden it across three versions. Table 8

estimates the credit-line response under each classification.[3] Across all three classifications, the B2C point estimates are consistently larger. They range from 0.206 to 0.213, compared with 0.128 to 0.141 in the comparison industries. Appendix Table D.1 reports separate estimates for one-digit industries with at least 3,000 observations.

The consistently larger B2C estimates match the prediction that hard information is more valuable when borrowers' transactions leave thinner cross-firm paper trails.

6.2 Bank Specialization

Pre-GT3 screening information also varies on the lender side. Repeated lending within an industry can build borrower comparables, screening routines, and knowledge of industry-specific risks (Blickle et al., 2026; Berger et al., 2017). Bank specialization therefore captures accumulated industry screening knowledge. The pre-GT3 measure compares a bank's local industry credit-line shares with the corresponding province-quarter industry portfolio; Appendix D.2 reports the construction.

GT3 also changes the industry breadth of banks' local lending portfolios. Table 9 shows that bank specialization, Industry HHI, and the top-industry share all decline, while the number of active industries increases by about 1.4. Banks therefore lend across a broader set of local industries after GT3.

Pre-GT3 specialization also predicts the strength of the credit-line response. Table 9 shows a negative interaction between GT3 and Bank Specialization: banks with lower pre-GT3 specialization respond more strongly on the credit-line margin. Banks broaden their industry coverage after GT3, and the credit-line response is strongest among banks that entered the reform with less accumulated industry screening knowledge.

A final borrower-side benchmark comes from firms whose public disclosure already supplies extensive hard information.

[3]Version 1 includes industries F, H, O, and R; Version 2 adds P and Q; Version 3 additionally includes L and M. The remaining classified industries form the comparison group.

6.3 Borrowers with Abundant Hard Information

Listed firms provide an abundant information benchmark on the borrower side. Audited financial statements and regular public disclosure give lenders an extensive standardized information environment before GT3 (Healy and Palepu, 2001; Minnis, 2011). Table 10 uses listed-firm borrowing announcements, with the disclosed upper bound of an announced credit line as the outcome.

Announced credit lines show little response to GT3 across all three specifications. The fully controlled estimate is approximately 1% and statistically indistinguishable from zero. The weak listed-firm response provides a benchmark consistent with a lower marginal value of additional hard information when audited accounts and public disclosure already supply lenders with extensive information.

Borrower transaction records, lenders' accumulated industry knowledge, and public disclosure capture distinct sources of pre-GT3 screening information. Across these margins, the response is stronger where that information was weaker. The next section examines how local competition shapes the transmission of this information improvement into approved credit.

7 Market Consequences and Scope

The preceding section shows where hard information is most valuable. I now examine how local competition shapes its transmission into credit lines. I then ask whether the loan-level response extends to aggregate bank outcomes.

7.1 Competition and the Transmission of Hard Information

Hard information improves screening, whether this improvement translates into larger approved credit lines may depend on local competition. Banks accumulate soft information through lending relationships, competing lenders cannot easily reproduce this borrower-

specific knowledge. This gives the bank an informational advantage based on soft information (Sharpe, 1990). Invoice verification gives banks a shared basis for evaluating borrower-supplied records and can reduce the proprietary value of this advantage (Sutherland, 2018; Liberti et al., 2022). Models of information technology and credit-market competition therefore predict that market structure can shape how information improvements enter lending decisions (Hauswald and Marquez, 2003; He et al., 2023).

As described in Section 3.3, I measure local banking concentration using HHI_{ct}, constructed from bank branch shares within city c and year-month t. Higher HHI indicates a more concentrated and therefore less competitive local banking market. I interact HHI_{ct} with GT3 in the loan-level credit-line specification, retaining City, Year-Month, Bank, and Industry fixed effects and grouped borrower and contract controls. Because HHI is measured contemporaneously, I use the interaction to characterize how the GT3 response varies with local market structure and interpret the pattern as evidence on pass-through.

Table 11 shows that the interaction between GT3 and HHI is negative. The estimate implies a 0.133 log point, or approximately 14.2%, larger GT3 response when HHI is 0.10 lower. The stronger response in less concentrated markets links the credit line effect of GT3 to the ability of hard information to cross organizational boundaries. It provides further evidence that hard information, rather than soft information, drives the GT3 credit line response. I next turn to bank funding, aggregate lending, and asset quality.

7.2 Aggregate Bank Outcomes

To assess the aggregate effect of the bank, I use annual financial report data of the bank that described in Section 3.3. The sample is restricted to banks only have branches in a single province. This restriction assigns each bank to a single provincial GT3 rollout. Table 12 reports estimates for funding costs, aggregate lending, and asset quality. All specifications include bank and year fixed effects and lagged balance-sheet controls.

Funding costs, measured by interest expense relative to assets, remain unchanged after GT3. Aggregate bank lending does not increase: loan growth, the ratio of loans to assets, and log total loans remain unchanged. Asset quality does not deteriorate: the NPL ratio does not increase, and the reserve coverage ratio does not decline. The absence of aggregate credit expansion or deterioration in asset quality rejects a general relaxation of lending standards.

8 Conclusion

This paper asks whether hard information substitutes for soft information in small business lending. I study this question using the provincial rollout of China's Golden Tax Phase III reform and confidential loan originations linking banks' internal assessments, approved contracts, and borrower to lender distance. The results show that it does. Hard information reduces banks' reliance on soft information in screening, improves approved loan terms, and expands the distance over which banks lend. It therefore changes both contract design and the geography of small business lending.

GT3 turns borrower supplied transaction records into hard information that banks can verify and transmit through underwriting. Banks revise assessed income and repayment capacity upward before approving larger credit lines, lowering rates and spreads, and extending maturities. The central evidence comes from distance. The gains increase as borrowers become more distant, the distance schedules in credit quantity and price flatten, and banks extend lending farther away. The effects are also strongest where the preexisting stock of verifiable information is smallest. The common movement in internal assessments, contract terms, and realized lending distance shows that hard information replaces soft information in bank screening.

The persistence of distance after the reform helps locate the boundary of this substitution. At the expanded lending frontier, borrowers have comparable internal repayment capacity scores, while driving time and loan rates increase. The remaining premium is consistent with

contact and monitoring costs outside the screening margin affected by hard information. The market evidence points in the same direction. Soft information accumulated through lending relationships gives incumbent banks an informational advantage. The stronger credit response in more competitive markets is consistent with invoice verification weakening that advantage and allowing more of the information gain to pass into approved credit.

The broader implication is that public digital infrastructure can change the information technology of bank lending. By authenticating records firms already generate, public systems can alter what banks recognize about opaque borrowers and how far credit travels through existing banking relationships. By producing verifiable hard information, public infrastructure can weaken banks' reliance on soft information and make small business lending less local.

References

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

A. C. Baker, D. F. Larcker, and C. C. Y. Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395, 2022.

M. Bellon, E. Dabla-Norris, S. Khalid, and F. Lima. Digitalization to improve tax compliance: Evidence from VAT e-invoicing in Peru. *Journal of Public Economics*, 210:104661, 2022.

T. Berg, V. Burg, A. Gombović, and M. Puri. On the rise of FinTechs: Credit scoring using digital footprints. *Review of Financial Studies*, 33(7):2845–2897, 2020.

A. N. Berger and W. S. Frame. Small business credit scoring and credit availability. *Journal of Small Business Management*, 45(1):5–22, 2007.

A. N. Berger and G. F. Udell. Relationship lending and lines of credit in small firm finance. *Journal of Business*, 68(3):351–381, 1995.

P. G. Berger, M. Minnis, and A. Sutherland. Commercial lending concentration and bank expertise: Evidence from borrower financial statements. *Journal of Accounting and Economics*, 64(2–3):253–277, 2017.

S. T. Bharath, S. Dahiya, A. Saunders, and A. Srinivasan. Lending relationships and loan contract terms. *Review of Financial Studies*, 24(4):1141–1203, 2011.

K. Blickle, C. Parlatore, and A. Saunders. Specialization in banking. *Journal of Finance*, 81(3):1531–1572, 2026.

T. Broecker. Credit-worthiness tests and interbank competition. *Econometrica*, 58(2):429–452, 1990.

S. Calonico, M. D. Cattaneo, and R. Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

A. Casado, M. Giometti, J. E. Gutiérrez, D. Martínez-Miera, A. Matyunina, and T. Terracciano. Digitalization and credit markets: Evidence from e-Invoicing. Working Paper 2619, Banco de España, 2026.

G. Cassar, C. D. Ittner, and K. S. Cavalluzzo. Alternative information sources and information asymmetry reduction: Evidence from small business debt. *Journal of Accounting and Economics*, 59(2–3):242–263, 2015.

D. Cengiz, A. Dube, A. Lindner, and B. Zipperer. The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics*, 134(3):1405–1454, 2019.

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

R. DeYoung, D. Glennon, and P. Nigro. Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. *Journal of Financial Intermediation*, 17(1):113–143, 2008.

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

J. Granja, C. Leuz, and R. G. Rajan. Going the extra mile: Distant lending and credit cycles. *Journal of Finance*, 77(2):1259–1324, 2022.

H. Hau, Y. Huang, C. Lin, H. Shan, Z. Sheng, and L. Wei. FinTech credit and entrepreneurial growth. *Journal of Finance*, 79(5):3309–3359, 2024.

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

Z. He, J. Huang, and J. Zhou. Open Banking: Credit market competition when borrowers own the data. *Journal of Financial Economics*, 147(2):449–474, 2023.

Z. He, J. Huang, and C. Parlatore. Information span in credit market competition. NBER Working Paper 33141, National Bureau of Economic Research, 2024.

P. M. Healy and K. G. Palepu. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1–3):405–440, 2001.

F. Heider and R. Inderst. Loan prospecting. *Review of Financial Studies*, 25(8):2381–2415, 2012.

A. R. Heitz, C. Martin, and A. Ufier. Bank monitoring with on-site inspections. *Journal of Finance*, 81(2):687–737, Apr. 2026.

J. Liberti, J. Sturgess, and A. Sutherland. How voluntary information sharing systems form: Evidence from a U.S. commercial credit bureau. *Journal of Financial Economics*, 145(3): 827–849, 2022.

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

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

M. Minnis. The value of financial statement verification in debt financing: Evidence from private U.S. firms. *Journal of Accounting Research*, 49(2):457–506, 2011.

M. Minnis and A. Sutherland. Financial statements as monitoring mechanisms: Evidence from small commercial loans. *Journal of Accounting Research*, 55(1):197–233, 2017.

M. Minnis, A. G. Sutherland, and F. W. Vetter. Financial statements not required. *Journal of Accounting and Economics*, 78(2–3):101732, 2024.

J. Naritomi. Consumers as tax auditors. *American Economic Review*, 109(9):3031–3072, 2019.

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

S. Ouyang. Cashless payment and financial inclusion. *Journal of Financial Economics*, 180: 104277, 2026.

C. A. Parlour, U. Rajan, and H. Zhu. When FinTech competes for payment flows. *Review of Financial Studies*, 35(11):4985–5024, 2022.

M. A. Petersen and R. G. Rajan. The benefits of lending relationships: Evidence from small business data. *Journal of Finance*, 49(1):3–37, 1994.

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

D. Pomeranz. No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8):2539–2569, 2015.

S. A. Sharpe. Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *Journal of Finance*, 45(4):1069–1087, 1990.

J. C. Stein. Information production and capital allocation: Decentralized versus hierarchical firms. *Journal of Finance*, 57(5):1891–1921, 2002.

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

A. Sutherland. Does credit reporting lead to a decline in relationship lending? evidence from information sharing technology. *Journal of Accounting and Economics*, 66(1):123–141, 2018.

X. Vives and Z. Ye. Information technology and lender competition. *Journal of Financial Economics*, 163:103957, 2025.

Tables

General table notes: Unless stated otherwise, parentheses report t-statistics; significance is denoted by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. Variable construction is reported in Appendix Table A.4, and loan level fixed effects and control blocks are described in Appendix B.

Data and Sample

Table 1: Summary Statistics

	N	Mean	Std. Dev.	P25	Median	P75
Approved Credit Line (RMB)	154,233	1,521,540	2,050,424	330,000	800,000	1,890,000
Interest Rate	154,233	7.290	2.133	5.655	6.960	8.528
Interest Spread	154,233	2.655	2.015	1.250	2.213	3.700
Maturity	154,233	506.9	598.722	356	364	366
Bank Assessed Income (Log)	126,282	14.95	2.018	13.592	15.294	16.300
Internal Repayment Capacity Score	101,242	45.997	28.680	21.000	49.130	64.385
Borrower to Lender Distance (km)	96,883	27.2	25.035	8.073	22.351	38.772

Notes: This table reports summary statistics for the main sample of small business loan originations from nine banks between 2012 and 2022. Approved Credit Line is measured in RMB, Interest Rate and Interest Spread in percentage points, Maturity in days, and Borrower to Lender Distance in kilometers. Bank Assessed Income is reported in logs. Internal Repayment Capacity Score is a within-bank percentile rank from 1 to 100, with higher values indicating stronger repayment capacity. Each row uses observations with nonmissing values for the corresponding variable.

Table 2: GT3 and Loan Contract Terms

Panel A. Approved Credit Line and Interest Rate

	ln(*Approved Credit Line*)		Interest Rate	
	(1)	(2)	(3)	(4)
GT3	0.360***	0.179***	-0.315***	-0.404***
	(8.96)	(4.21)	(-3.22)	(-3.77)
Constant	6.464***	6.575***	7.482***	7.537***
	(241.48)	(268.03)	(119.35)	(116.28)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Control FE	No	Yes	No	Yes
Observations	154,182	154,170	154,233	154,221
R^2	0.384	0.543	0.579	0.670

Panel B. Interest Spread and Maturity

	Interest Spread		Maturity	
	(5)	(6)	(7)	(8)
GT3	-0.338***	-0.442***	110.939***	49.014**
	(-3.34)	(-3.95)	(5.40)	(2.22)
Constant	2.862***	2.926***	438.933***	476.854***
	(44.33)	(43.13)	(29.61)	(32.96)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Control FE	No	Yes	No	Yes
Observations	154,233	154,221	154,182	154,170
R^2	0.523	0.628	0.419	0.610

Notes: This table reports loan-level estimates for the main sample of small business originations. GT3 equals one when a loan is originated on or after the reform's rollout date in the borrower's registered province. Approved Credit Line is measured in logs, Interest Rate and Interest Spread in percentage points, and Maturity in days. Odd-numbered columns include Year-Month and Bank fixed effects. Even-numbered columns additionally include City, Industry, and Control fixed effects. The Control fixed effects and outcome-specific omissions are described in Appendix B. Standard errors are clustered by province × year-month.

Table 3: GT3, the Reporting Environment, and Bank Assessment

Panel A. Reporting Environment

	ln(*Tax Amount*)
GT3	0.076**
	(2.29)
Lagged ln(*GDP per capita*)	1.142***
	(7.08)
Lagged ln(*Population*)	1.328***
	(3.92)
Lagged ln(*Road infrastructure*)	−0.196**
	(−2.60)
Lagged secondary sector share	0.532
	(0.50)
Lagged tertiary sector share	0.212
	(0.15)
Lagged ln(*Trade*)	−0.100
	(−1.69)
Constant	−10.081***
	(−2.78)
Province FE	Yes
Tax Category × Year FE	Yes
Taxpayer Category × Year FE	Yes
Observations	44,239
R^2	0.759

Panel B. Bank Assessment

	Bank Assessed Income		Internal Repayment Capacity Score	
	(1)	(2)	(3)	(4)
GT3	0.132***	0.193***	2.987***	6.642***
	(2.84)	(3.31)	(2.75)	(4.99)
Constant	14.869***	14.831***	43.459***	40.354***
	(381.84)	(428.88)	(42.45)	(35.74)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	No	Yes
Industry FE	No	Yes	No	Yes
Control FE	No	Yes	No	Yes
Observations	126,282	126,268	101,240	101,228
R^2	0.337	0.480	0.090	0.151

Notes: Panel A reports annual estimates from the CNRDS tax panel, where an observation is a province × tax category × taxpayer category × year. GT3 equals one in and after the provincial rollout year, and the macroeconomic controls are lagged one year. Panel B reports loan-level estimates for the main small business sample. In this panel, GT3 is assigned using the exact rollout date in the borrower's registered province, Bank Assessed Income is measured in logs, and Internal Repayment Capacity Score is a within-bank percentile rank from 1 to 100. Columns (1) and (3) include Year-Month and Bank fixed effects. Columns (2) and (4) additionally include City, Industry, and Control fixed effects. The Control fixed effects and outcome-specific omissions are described in Appendix B. Loan-level standard errors are clustered by province × year-month.

Distance

Table 4: Pre and Post GT3 Distance Gradients

Dep. Var.:	ln(*Approved Credit Line*)		Interest Rate		Interest Spread	
	(1)	(2)	(3)	(4)	(5)	(6)
	Pre	Post	Pre	Post	Pre	Post
ln(*Distance*)	-0.123^{***}	-0.070^{***}	0.187^{***}	0.084^{***}	0.186^{***}	0.081^{***}
	(-22.90)	(-8.33)	(10.77)	(3.80)	(10.71)	(3.45)
Constant	6.582^{***}	6.930^{***}	8.248^{***}	6.578^{***}	2.919^{***}	2.302^{***}
	(404.18)	(293.72)	(157.08)	(105.85)	(55.54)	(35.23)
City ×						
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Control FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	48,609	45,354	48,609	45,354	48,609	45,354
R^2	0.458	0.544	0.465	0.758	0.583	0.755

Notes: This table reports loan-level estimates separately for loans originated before and after GT3. Borrower to Lender Distance is measured in kilometers and enters in logs. Approved Credit Line is measured in logs, and Interest Rate and Interest Spread are measured in percentage points. All columns include City × Year-Month, Bank, Industry, and Control fixed effects. Standard errors are clustered by province × year-month.

Table 5: GT3 and Borrower to Lender Distance Schedules

Dep. Var.:	ln(*Approved Credit Line*)	Interest Rate	Interest Spread	ln(*Income*)
	(1)	(2)	(3)	(4)
GT3	0.083	0.309**	0.332**	0.172
	(0.62)	(2.30)	(2.44)	(0.57)
ln(*Distance*)	−0.124***	0.190***	0.190***	−0.212***
	(−21.06)	(10.09)	(10.03)	(−12.15)
GT3 ×				
ln(*Distance*)	0.052***	−0.103***	−0.107***	0.155***
	(4.78)	(−3.39)	(−3.40)	(7.65)
Constant	6.715***	7.284***	2.451***	15.273***
	(101.98)	(94.02)	(31.80)	(96.83)
City ×				
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Control FE	Yes	Yes	Yes	Yes
Observations	93,993	93,993	93,993	86,026
R^2	0.533	0.683	0.668	0.430

Notes: This table reports loan-level estimates of Equation (3) for the main sample with observed Borrower to Lender Distance. GT3 equals one when a loan is originated on or after the reform rollout date in the borrower's registered province. Borrower to Lender Distance is measured in kilometers and enters in logs. Approved Credit Line and Bank Assessed Income are measured in logs, and Interest Rate and Interest Spread are measured in percentage points. All columns include City × Year-Month, Bank, Industry, and Control fixed effects. Income-group fixed effects are omitted in column (4). Standard errors are clustered by province × year-month.

Table 6: The Geographic Reach of Lending

Dep. Var.:	ln(*Borrower to Lender Distance*)
	(1)
GT3	0.135**
	(2.10)
Constant	2.855***
	(91.95)
City × Year-Month FE	Yes
Bank FE	Yes
Industry FE	Yes
Control FE	Yes
Observations	93,993
R^2	0.395

Notes: This table reports a loan-level estimate for the main sample with observed Borrower to Lender Distance. The dependent variable is the log distance in kilometers between the borrower and lender locations. GT3 equals one when a loan is originated on or after the reform rollout date in the borrower's registered province. The specification includes City × Year-Month, Bank, Industry, and Control fixed effects. Standard errors are clustered by province × year-month.

Table 7: Evidence at the Expansion Boundary

Panel A. Boundary Band Comparison

Dep. Var.:	Interest Rate (1)	Internal Repayment Capacity Score (2)	Driving Time (3)
Outer band	0.160**	0.744	14.922***
	(2.00)	(0.46)	(10.03)
City ×			
Year-Month FE	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Control FE	Yes	Yes	Yes
Observations	14,303	10,107	14,303
R^2	0.767	0.309	0.986

Panel B. Local Estimates at the Expansion Boundary

	Baseline h			Robustness $2h$		
Dep. Var.:	Interest Rate (1)	Internal Repayment Capacity Score (2)	Driving Time (3)	Interest Rate (4)	Internal Repayment Capacity Score (5)	Driving Time (6)
Local estimate	0.430***	−3.358	10.225***	0.313***	−2.791	10.327***
	(2.91)	(−1.03)	(2.62)	(3.15)	(−1.01)	(4.53)
City ×						
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Control FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,303	10,107	14,303	14,303	10,107	14,303

Notes: This table reports estimates around the city-specific lending expansion boundary defined in Appendix C.1. Panel A compares post GT3 loans in adjacent outer and inner bands. Panel B reports local linear estimates at the same boundary using bandwidths h and $2h$, a triangular kernel, and a mass point adjustment. Each outcome in Panel B is residualized using the fixed effects and controls reported in the table. Interest Rate is measured in percentage points, Internal Repayment Capacity Score is a within-bank percentile rank from 1 to 100, and Driving Time is measured in minutes. Credit-score fixed effects are omitted when Internal Repayment Capacity Score is the dependent variable.

Where Hard Information Matters

Table 8: Borrower Information Scarcity and Credit Line Effects

Dep. Var.:	ln(*Approved Credit Line*)					
	Version 1		Version 2		Version 3	
	B2B	B2C	B2B	B2C	B2B	B2C
GT3	0.135***	0.213***	0.141***	0.209***	0.128***	0.206***
	(3.69)	(3.59)	(3.82)	(3.54)	(3.66)	(3.51)
Constant	6.666***	6.552***	6.660***	6.558***	6.649***	6.578***
	(301.98)	(188.96)	(299.21)	(190.84)	(310.24)	(192.60)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Control FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	68,767	73,146	67,658	74,254	61,934	79,983
R^2	0.603	0.522	0.605	0.521	0.612	0.523

Notes: This table reports loan-level estimates separately for B2B and B2C industries under three classifications of consumer-facing activity. The dependent variable is the log of Approved Credit Line. Version 1 classifies industries F, H, O, and R as B2C; Version 2 adds P and Q; Version 3 adds L and M. The remaining classified industries form the B2B group. Observations with unclassified or missing one-digit industry codes are omitted. All columns include City, Year-Month, Bank, Industry, and Control fixed effects. Standard errors are clustered by province × year-month.

Table 9: Bank Industry Information and Local Credit Allocation

	Credit-Line Response	Local Industry Allocation			
Dep. Var.:	ln(*Approved Credit Line*)	Bank Specialization	Industry HHI	Top-Industry Share	Active Industries
	(1)	(2)	(3)	(4)	(5)
GT3	0.073**	−0.138**	−0.211***	−0.167***	1.372**
	(2.08)	(−2.28)	(−3.39)	(−3.17)	(2.46)
Bank Specialization	0.050				
	(1.53)				
GT3 ×					
Bank Specialization	−0.125**				
	(−2.11)				
Constant	6.781***	0.289***	0.730***	0.798***	3.101***
	(187.15)	(11.96)	(30.28)	(38.96)	(14.49)
City FE	Yes	–	–	–	–
Year-Month FE	Yes	–	–	–	–
Bank FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	–	–	–	–
Control FE	Yes	–	–	–	–
Province FE	–	Yes	Yes	Yes	Yes
Year-Quarter FE	–	Yes	Yes	Yes	Yes
Observation unit	Loan	Bank × Province × Year-Quarter	Bank × Province × Year-Quarter	Bank × Province × Year-Quarter	Bank × Province × Year-Quarter
Observations	117,945	1,020	1,020	1,020	1,020
R^2	0.525	0.393	0.470	0.428	0.493

Notes: Column (1) reports the loan-level credit-line response as a function of Bank Specialization measured before GT3. The specialization measure compares a bank's local industry credit-line shares with the corresponding province-quarter industry portfolio and is defined in Appendix D.2. Columns (2) through (5) report bank × province × year-quarter estimates of local industry allocation. Bank Specialization measures the distance between the bank's local industry portfolio and the province-quarter portfolio; Industry HHI and Top-Industry Share measure portfolio concentration; Active Industries counts one-digit industries receiving positive credit. The fixed effects for each specification are reported in the table. Standard errors in column (1) are clustered by province × year-month.

Table 10: Listed-Firm Credit Lines

Dep. Var.:	ln(*Approved Credit Line*)		
	(1)	(2)	(3)
GT3	−0.006	0.011	0.010
	(−0.25)	(0.52)	(0.50)
Constant	9.235***	9.221***	9.221***
	(493.76)	(512.34)	(519.93)
Firm FE	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes
Loan Type FE	No	Yes	Yes
Loan Purpose FE	No	Yes	Yes
Currency FE	No	Yes	Yes
Collateral FE	No	No	Yes
Syndicate FE	No	No	Yes
Observations	151,074	151,066	151,065
R^2	0.613	0.656	0.657

Notes: This table reports estimates from the listed-firm borrowing-announcement sample. The dependent variable is the log of the disclosed upper bound of the announced credit line. GT3 exposure is assigned using the lending-branch province and the announcement date. All columns include Firm, Year-Month, and Bank fixed effects. Column (2) adds Loan Type, Loan Purpose, and Currency fixed effects. Column (3) additionally includes Collateral and Syndicate fixed effects.

Market Transmission and Scope

Table 11: Local Competition and Credit Line Pass Through

Dep. Var.:	ln(*Approved Credit Line*)
	(1)
GT3	0.341***
	(3.81)
HHI	−0.496
	(−1.31)
GT3 × HHI	−1.327***
	(−3.37)
Constant	6.542***
	(129.27)
City FE	Yes
Year-Month FE	Yes
Bank FE	Yes
Industry FE	Yes
Control FE	Yes
Observations	140,253
R^2	0.517

Notes: This table reports a loan-level estimate of the relation between local banking market concentration and the GT3 credit-line response. The dependent variable is the log of Approved Credit Line. HHI is constructed from bank branch shares within each city and year-month; a higher value indicates greater local market concentration. The specification includes City, Year-Month, Bank, Industry, and Control fixed effects. Standard errors are clustered by province × year-month.

Table 12: Bank Funding Costs, Loan Quantity, and Asset Quality

Panel A. Funding Costs

Dep. Var.:	(1) Interest Expense / Assets
GT3	0.0009
	(1.22)
Bank FE	Yes
Year FE	Yes
Lagged Controls	Yes
Observations	1,492

Panel B. Loan Quantity

Dep. Var.:	(2) Loan Growth	(3) Loan-to-Asset Ratio	(4) Log Total Loans
GT3	0.017	−0.007	0.003
	(0.28)	(−0.53)	(0.17)
Bank FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Lagged Controls	Yes	Yes	Yes
Observations	1,494	1,494	1,494

Panel C. Asset Quality

Dep. Var.:	(5) NPL Ratio	(6) Reserve Coverage Ratio
GT3	0.209	32.893
	(1.49)	(0.96)
Bank FE	Yes	Yes
Year FE	Yes	Yes
Lagged Controls	Yes	Yes
Observations	1,477	1,328

Notes: This table reports bank-year estimates for banks whose domestic branch networks are confined to a single province. GT3 exposure is assigned using the provincial rollout date. Panel A reports funding costs, Panel B reports loan quantity, and Panel C reports asset quality. All specifications include Bank and Year fixed effects and the lagged balance sheet controls defined in Appendix Table A.4.

Figures

Information Channel and Timing

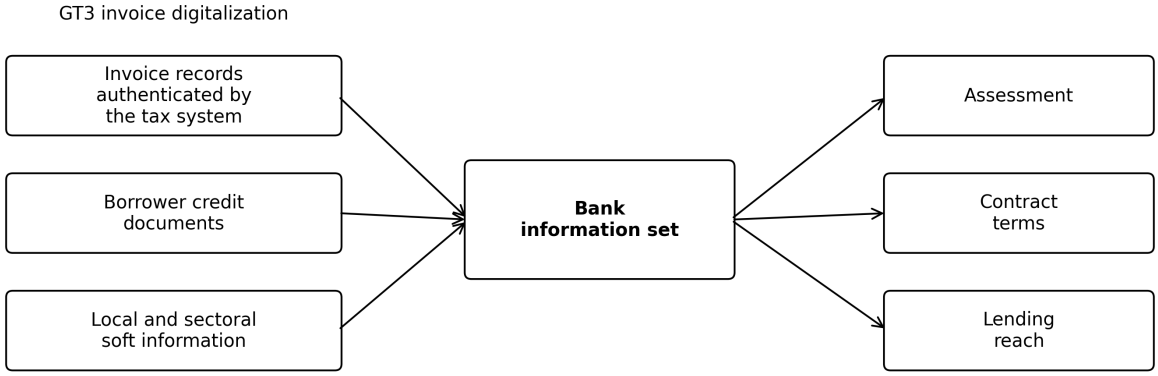


Figure 1: Invoice Digitalization and the Bank Information Set

Notes: Invoice digitalization enters credit files through borrower documents and expands the bank information set. Local and sectoral soft information enter through relationships, monitoring, branch knowledge, and industry familiarity.

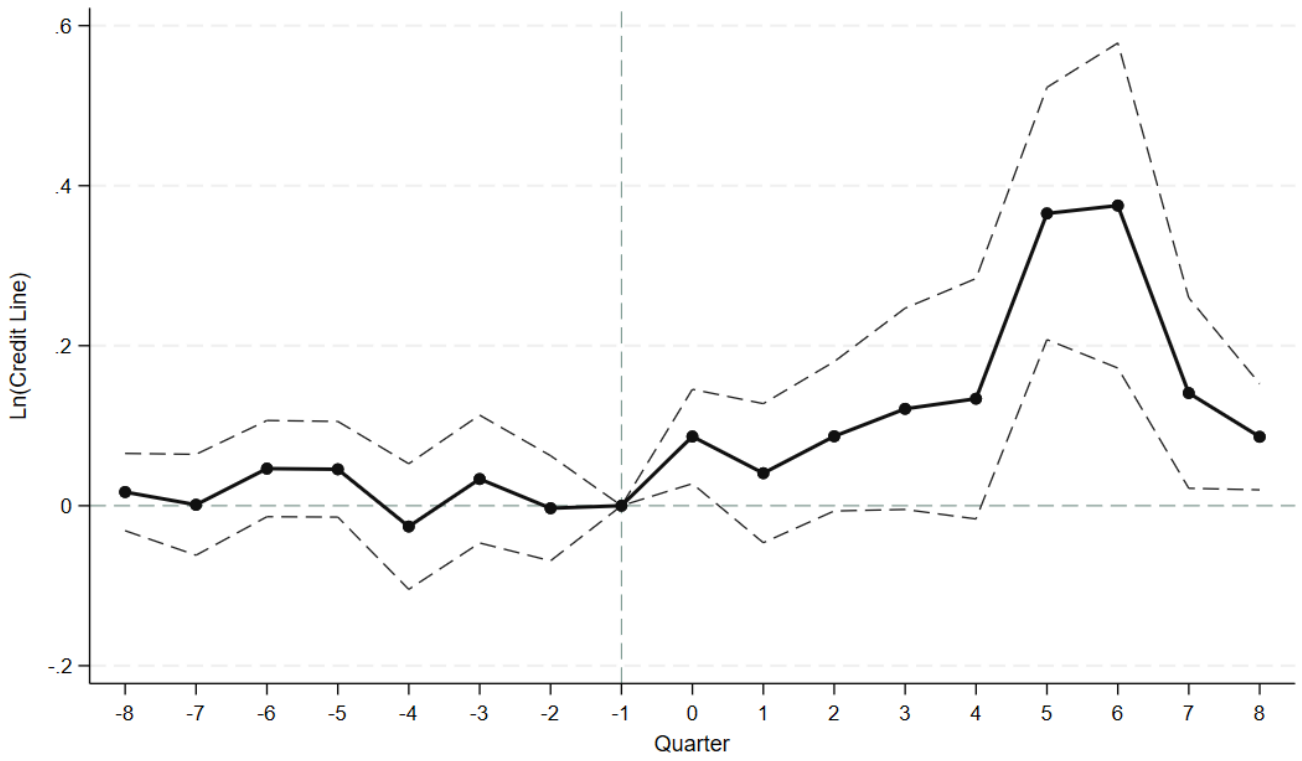


Figure 2: Event Study for ln(*Credit Line*)

Notes: Event time coefficients come from the preferred event study specification for ln(*Credit Line*), with the quarter immediately before GT3 rollout omitted.

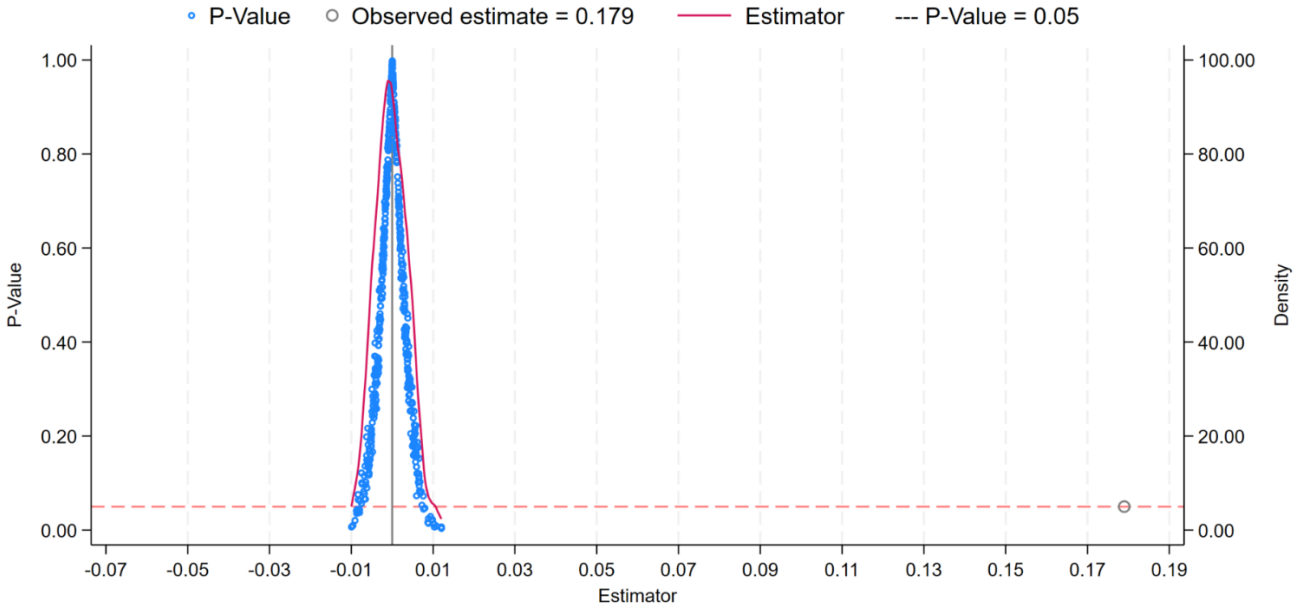


Figure 3: Placebo Distribution for ln(*Approved Credit Line*)

Notes: The figure reports 500 placebo estimates obtained by randomly permuting the GT3 indicator across loan observations in the estimation sample and re-estimating the baseline specification for ln(*Approved Credit Line*). Blue circles show the corresponding two-sided p values, the red curve shows the density of placebo coefficients, and the hollow marker denotes the realized estimate of 0.179 from Table 2. The dashed line marks $p = 0.05$. The permutations use random seed 123.

Lending Frontier

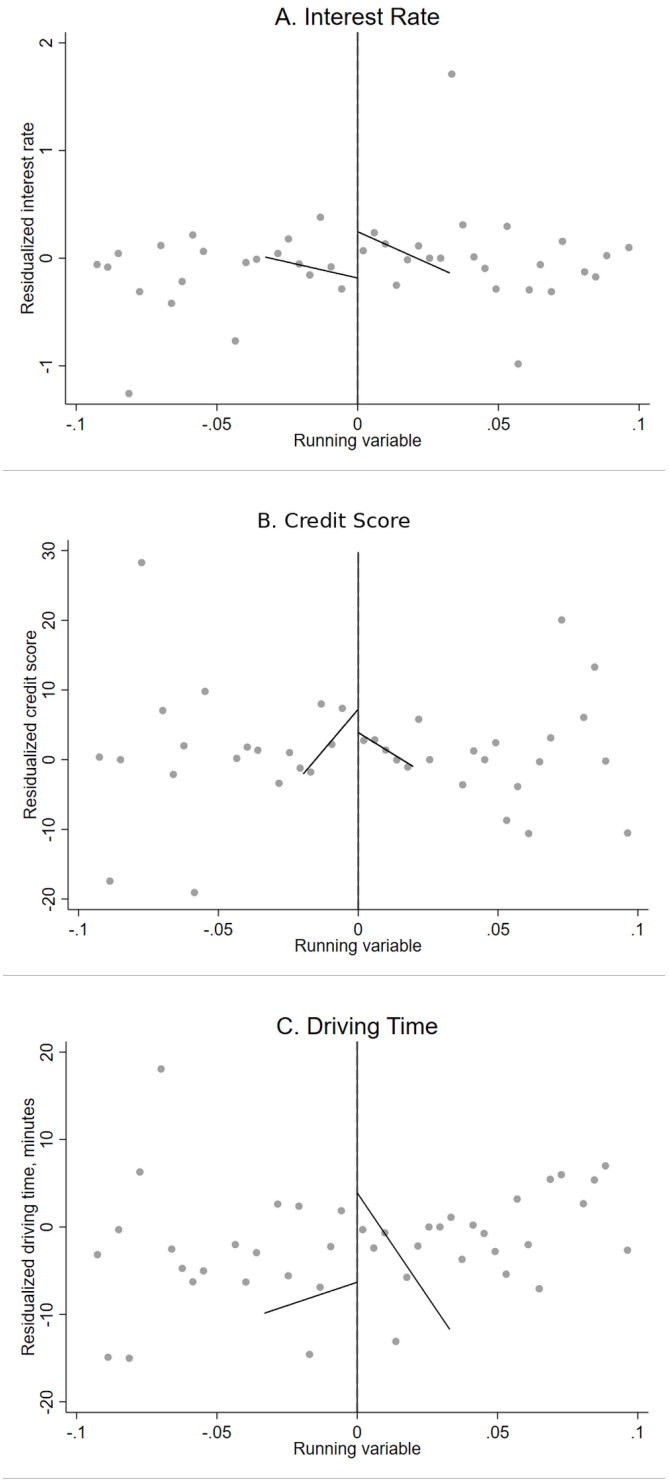


Figure 4: Local RD Plots at the GT3 Implied Expansion Boundary

Notes: Binned means of the residualized outcome are plotted against the boundary running variable. Panel A reports Interest Rate, Panel B reports Credit Score, and Panel C reports Driving Time.

Distance Gradients

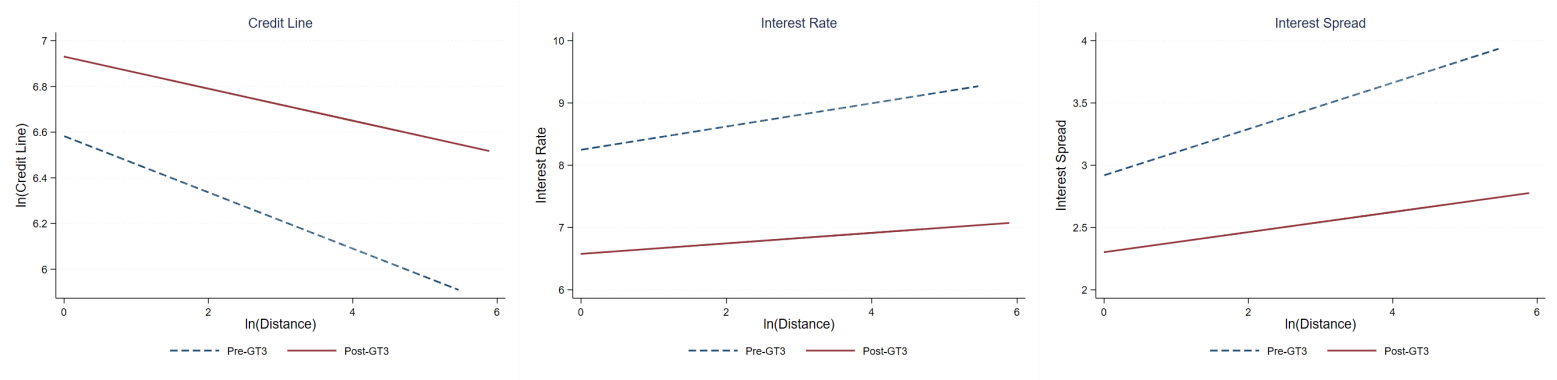


Figure 5: Pre- and Post-GT3 Distance Schedules

Notes: The three plots show fitted pre- and post-GT3 distance schedules for ln(*Credit Line*), Interest Rate, and Interest Spread.

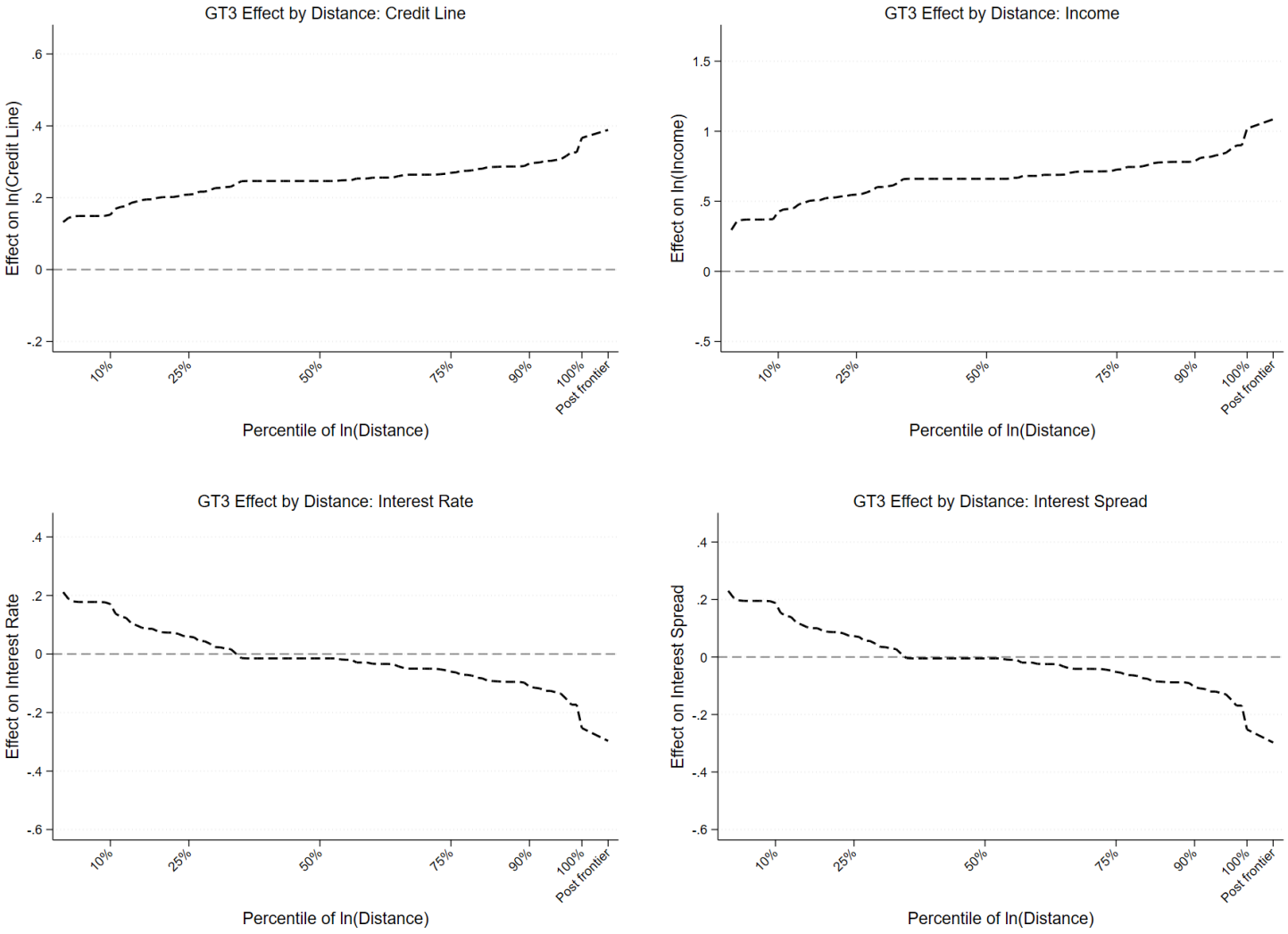


Figure 6: GT3 Effects Along the Distance Distribution

Notes: Each plot shows implied point estimates evaluated at support points of the relevant distance distribution. The four plots report ln(*Credit Line*), ln(*Income*), Interest Rate, and Interest Spread. The support points are the pre GT3 percentiles, the pre GT3 maximum, and the post GT3 frontier. The implied effect equals $\widehat{\beta}_{GT3} + x\widehat{\beta}_{GT3\times\ln(Distance)}$. The income plot uses the income sample, while the remaining plots use the common loan sample. Appendix Table C.2 reports the underlying pointwise estimates.

A Institutional Details, Data, and Variable Construction

A.1 GT3 Rollout

GT3 was introduced in stages across provincial administrative units. The loan-level analysis assigns treatment using the exact implementation date, while annual panels assign exposure beginning in the implementation year.

Table A.1: GT3 Rollout by Province

Province or Municipality	Implementation Date	Province or Municipality	Implementation Date
Chongqing Municipality	2013.02.22	Shanxi Province	2013.10.08
Shandong Province	2013.10.08	Guangdong Province	2015.01.08
Henan Province	2015.01.08	Inner Mongolia Autonomous Region	2015.01.08
Ningxia Hui Autonomous Region	2015.07.01	Tibet Autonomous Region	2015.09.01
Hebei Province	2015.09.08	Guizhou Province	2015.09.08
Yunnan Province	2015.09.08	Guangxi Zhuang Autonomous Region	2015.09.08
Hunan Province	2015.10.08	Qinghai Province	2015.10.08
Hainan Province	2015.10.08	Gansu Province	2015.10.08
Anhui Province	2016.01.08	Xinjiang Uygur Autonomous Region	2016.01.08
Sichuan Province	2016.01.08	Jilin Province	2016.01.08
Shanghai Municipality	2016.07.08	Fujian Province	2016.07.08
Jiangxi Province	2016.07.08	Liaoning Province	2016.07.08
Beijing Municipality	2016.08.08	Tianjin Municipality	2016.08.08
Heilongjiang Province	2016.08.08	Hubei Province	2016.08.08
Shaanxi Province	2016.08.08	Jiangsu Province	2016.10.08
Zhejiang Province	2016.10.08		

Notes: The table reports GT3 implementation dates by provincial administrative unit. Municipalities are treated as province-level units. Dates use the YYYY.MM.DD format.

A.2 Invoice Records in Small Business Lending

Public bank product descriptions document the use of invoices, contracts, receivables, delivery records, and invoicing histories in small business credit review. These materials enter investigation, approval, or credit-limit decisions.

Table A.2: Public Bank Product Evidence on Invoice and Transaction Records in Small Business Lending

Bank or product	Records used	Role in credit review
ICBC small business domestic invoice financing	Domestic transaction invoices, transaction contracts, receivables	Invoice backed receivables support short term financing, with receivables identified as the first repayment source.
Agricultural Bank of China domestic invoice financing	Purchase and sale contracts, invoices, transport documents	Borrower documents enter the investigation, review, approval, and disbursement workflow.
Agricultural Bank of China "Zhang Ying Tong"	Receivables, purchase and sale contracts, invoices, delivery or transfer documents	Receivables based financing uses invoices and related transaction documents to support small business credit approval.
Ping An Bank supermarket supplier invoice loan	Supplier invoice data	Supplier invoice records are transmitted electronically and used in rapid credit review for upstream suppliers.
Ping An Bank Speed Invoice Loan	Invoicing behavior and invoicing history	Enterprise invoicing behavior and at least 18 months of invoicing records enter internet credit approval for small business owners.

Notes: The table summarizes publicly available bank product descriptions and the stated role of invoice and transaction records in credit review.

A.3 Auxiliary Sample Summary Statistics

Table A.3: Auxiliary Sample Summary Statistics

Panel A. Listed-firm borrowing-announcement sample

	N	Mean	Std. Dev.	P25	Median	P75
Announced Credit Line (million RMB)	151,074	342.522	13,544.210	50	100	250

Panel B. Provincial tax panel

	N	Mean	Std. Dev.	P25	Median	P75
Accounting year	52,173	2010.124	5.612039	2005	2010	2015
GT3 rollout year	52,140	2015.29	.8871095	2015	2015	2016
GT3 exposure indicator	52,173	.2521611	.4342574	0	0	1
Tax Amount	44,257	445212.4	1631849	2272	28450	213823
ln(*Tax Amount*)	44,239	9.817983	3.32395	7.730614	10.25696	12.27417
Lagged ln(*GDP per capita*)	52,173	10.0312	.8303778	9.381264	10.15518	10.65495
Lagged ln(*Population*)	52,173	8.082321	.8617372	7.784473	8.243808	8.711443
Lagged ln(*Road infrastructure*)	52,173	5.082042	.9350047	4.691348	5.267858	5.700444
Lagged secondary sector share	52,173	.4265007	.0829062	.3821188	.4356856	.4825657
Lagged tertiary sector share	52,173	.4507732	.0883211	.3977692	.4326409	.487055
Lagged ln(*Trade*)	52,173	16.72243	1.902105	15.37783	16.73073	17.98333

Panel C. Bank annual report panel

	N	Mean	Std. Dev.	P25	Median	P75
Interest Expense / Assets	1,493	0.0208	0.0085	0.0171	0.0200	0.0233
Loan Growth	1,494	0.2493	2.0537	0.1141	0.1626	0.2227
Loan to Asset Ratio	1,494	0.5024	0.1606	0.4363	0.5092	0.5712
Log(Total Loans)	1,494	24.3890	1.1443	23.5962	24.3206	25.2078
NPL Ratio	1,479	1.7295	1.3987	0.9900	1.5000	1.9800
Reserve Coverage Ratio	1,331	296.9807	333.2781	174.8200	223.8500	324.4100
Log(Total Assets)	1,500	25.1083	1.1744	24.2798	25.0917	25.9433
Equity / Assets	1,500	0.0783	0.0203	0.0653	0.0771	0.0880
Liquid Assets / Assets	1,245	0.2374	0.1086	0.1549	0.2155	0.3026
Operating Expenses / Assets	1,488	0.0095	0.0038	0.0071	0.0090	0.0113
Deposits / Assets	1,493	0.7541	0.2589	0.6676	0.7570	0.8333

Notes: Panel A reports the listed-firm borrowing-announcement sample, Panel B the province × tax category × taxpayer category × year panel, and Panel C the bank fiscal-year panel. Statistics are computed over the non-missing observations for each variable, so observation counts differ across rows.

A.4 Variable Definitions and Construction

Table A.4: Variable Definitions and Construction

Variable	Construction	Unit or merge level
Panel A. Treatment and geography		
GT3	Indicator equal to one if an observation falls on or after the GT3 rollout date in the assigned province. Loan-level exposure uses the exact implementation date in the borrower's registered province. Annual panels assign exposure beginning in the implementation year. The listed-firm sample uses the lending-branch province and announcement date.	Province date or year
Borrower to Lender Distance	Geographic distance in kilometers between the WGS84 centroid of the borrower's registered district or county and the WGS84 centroid of the originating branch city.	Loan level
ln(*Distance*)	Natural logarithm of Borrower to Lender Distance.	Loan level
Driving Distance	Routed road distance between the same borrower and lender endpoints used for Borrower to Lender Distance.	Kilometers
Driving Time	Estimated routed travel time between the same borrower and lender endpoints.	Minutes
Panel B. Loan contract terms		
Approved Credit Line	Approved credit limit recorded at loan origination. In the listed-firm sample, this variable is the disclosed upper bound of the announced credit line.	RMB
ln(*Approved Credit Line*)	Natural logarithm of Approved Credit Line after dividing the RMB amount by 1,000. The same transformation is applied to the listed-firm announced upper bound.	Loan or announcement
Interest Rate	Contractual nominal annual interest rate.	Percentage points
Interest Spread	Interest Rate minus the contemporaneous Loan Prime Rate.	Percentage points
Maturity	Contractual loan maturity.	Days
Panel C. Bank assessment measures		
Bank Assessed Income	Annual borrower income recorded by the bank in the underwriting file.	RMB
ln(*Bank Assessed Income*)	Natural logarithm of Bank Assessed Income.	Loan level

Table A.4 continued

Variable	Construction	Unit or merge level
Internal Repayment Capacity Score	Bank-specific internal repayment-capacity assessment mapped into within-bank percentile ranks from 1 to 100. Higher values indicate stronger internally assessed repayment capacity. This measure is distinct from an external credit-bureau score. Regressions that condition on the score use the underlying bank-specific score categories.	Loan level
Panel D. Market structure and bank specialization		
Local Banking HHI	Herfindahl-Hirschman index of local banking concentration constructed from branch shares at the city × year-month level. Higher values indicate greater concentration.	City × year-month
Pre GT3 Bank Specialization	Leave-one-out, pre-GT3 measure of how far a bank's local industry credit-line shares deviate from the province-quarter industry portfolio. The construction is reported in Appendix D.2.	Loan level
Local Excess Specialization	One half of the sum of absolute deviations between a bank's industry credit shares within a bank province quarter and the corresponding province quarter industry credit shares.	Bank province quarter
Industry HHI	Herfindahl-Hirschman index of amount weighted industry credit line shares.	Bank province quarter
Top Industry Share	Largest amount weighted industry credit line share.	Bank province quarter
Active Industries	Number of one digit industries receiving positive credit.	Bank province quarter
Panel E. Tax and bank panel variables		
Tax Amount	Reported tax collections from CNRDS provincial tax statistics.	Province × tax category × taxpayer category × year
ln(*Tax Amount*)	Log transformed Tax Amount used as the reporting environment outcome.	Annual tax panel
Lagged ln(*GDP per capita*)	Natural logarithm of provincial GDP per capita, lagged one year.	Province year
Lagged ln(*Population*)	Natural logarithm of provincial population, lagged one year.	Province year
Lagged ln(*Road infrastructure*)	Natural logarithm of provincial road infrastructure, lagged one year.	Province year

Table A.4 continued

Variable	Construction	Unit or merge level
Lagged Secondary Sector Share	Secondary-sector share of provincial output, lagged one year.	Province year
Lagged Tertiary Sector Share	Tertiary-sector share of provincial output, lagged one year.	Province year
Lagged ln(*Trade*)	Natural logarithm of provincial trade, lagged one year.	Province year
Interest Expense / Assets	Interest expense divided by total assets.	Bank fiscal year
Loan Growth	Annual growth in total loans.	Bank fiscal year
Loan to Asset Ratio	Total loans divided by total assets.	Bank fiscal year
Log Total Loans	Log of total loans reported in bank annual statements.	Bank fiscal year
NPL Ratio	Nonperforming loans divided by total loans.	Bank fiscal year
Reserve Coverage Ratio	Loan loss reserves divided by nonperforming loans, reported from annual statements.	Bank fiscal year
Log Total Assets	Natural logarithm of total assets.	Bank fiscal year
Equity / Assets	Total equity divided by total assets.	Bank fiscal year
Liquid Assets / Assets	Liquid assets divided by total assets.	Bank fiscal year
Operating Expenses / Assets	Operating expenses divided by total assets.	Bank fiscal year
Deposits / Assets	Total deposits divided by total assets.	Bank fiscal year

B Empirical Design and Identification Checks

B.1 Fixed Effects and Control Variables

The loan-level specifications use Year-Month, City, Bank, and one-digit Industry fixed effects. Year-Month fixed effects absorb aggregate shocks common to loans originated in the same calendar month. City, Bank, and Industry fixed effects absorb time-invariant differences across originating-branch cities, banks, and one-digit GB/T 4754 (2017) industries. Specifications that include City $\times$ Year-Month fixed effects identify the relevant gradients within the same city and month. Each table reports the fixed effects included in that specification.

The Control fixed effects comprise Payment Method, Borrower Type, Age Group, Income Group, Leverage Quintile, Credit Score, and Loan Term fixed effects. Continuous borrower and contract characteristics enter as grouped indicators, and each set includes a separate category for missing values. Age Group fixed effects apply to personal-name borrowers. Leverage Quintile fixed effects apply to corporate-name borrowers. Credit Score fixed effects use bank-specific rating categories or score bins. Loan Term fixed effects use maturity groups of less than one year, one to three years, three to five years, and more than five years.

In the four-column specification progression, column (1) includes Year-Month and Bank fixed effects. Column (2) adds City fixed effects, column (3) adds Industry fixed effects, and column (4) adds the complete Control fixed-effect block. To avoid mechanically absorbing the dependent variable, Income Group fixed effects are omitted from Bank Assessed Income regressions, Credit Score fixed effects from Internal Repayment Capacity Score regressions, and Loan Term fixed effects from Maturity regressions.

B.2 Coefficient Stability Across Specifications

Table B.1: Coefficient Stability Across Specifications

	(1)	(2)	(3)	(4)
Panel A. Bank Assessed Income				
GT3	0.132***	0.210***	0.211***	0.193***
	(2.84)	(3.73)	(3.68)	(3.31)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Control FE	No	No	No	Yes
Observations	126,282	126,268	126,268	126,268
R^2	0.337	0.376	0.385	0.480
Panel B. Internal Repayment Capacity Score				
GT3	2.987***	1.880	3.233***	6.642***
	(2.75)	(1.56)	(2.61)	(4.99)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Control FE	No	No	No	Yes
Observations	101,240	101,229	101,228	101,228
R^2	0.090	0.118	0.120	0.151

Table B.1 continued

	(1)	(2)	(3)	(4)
Panel C. ln(*Approved Credit Line*)				
GT3	0.360***	0.349***	0.331***	0.179***
	(8.96)	(9.44)	(7.87)	(4.21)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Control FE	No	No	No	Yes
Observations	154,182	154,172	154,172	154,170
R^2	0.384	0.426	0.429	0.543
Panel D. Interest Rate				
GT3	-0.315***	-0.378***	-0.464***	-0.404***
	(-3.22)	(-4.06)	(-4.46)	(-3.77)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Control FE	No	No	No	Yes
Observations	154,233	154,223	154,223	154,221
R^2	0.579	0.617	0.619	0.670

Table B.1 continued

	(1)	(2)	(3)	(4)
Panel E. Interest Spread				
GT3	-0.338***	-0.400***	-0.493***	-0.442***
	(-3.34)	(-4.12)	(-4.53)	(-3.95)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Control FE	No	No	No	Yes
Observations	154,233	154,223	154,223	154,221
R^2	0.523	0.566	0.568	0.628
Panel F. Maturity				
GT3	110.939***	63.272***	75.501***	49.014**
	(5.40)	(2.84)	(3.19)	(2.22)
Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
City FE	No	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes
Control FE	No	No	No	Yes
Observations	154,182	154,172	154,172	154,170
R^2	0.419	0.452	0.452	0.610

Notes: The table reports the four-column fixed-effect progression described in Appendix B.1. The dependent variables are Bank Assessed Income, Internal Repayment Capacity Score, the log of Approved Credit Line, Interest Rate, Interest Spread, and Maturity. Panel A omits Income Group fixed effects, Panel B omits Credit Score fixed effects, and Panel F omits Loan Term fixed effects. Standard errors are clustered by province × year-month.

B.3 Stacked Difference in Differences

Each stack is centered on a GT3 rollout cohort and covers the 24 months before through the 25 months after the cohort-specific rollout month. The adopting cohort is compared with provinces that remain untreated within the same event window. The regressions successively add cohort-specific Year-Month and Bank fixed effects, cohort-specific City fixed effects, cohort-specific Industry fixed effects, and cohort-specific Control fixed effects.

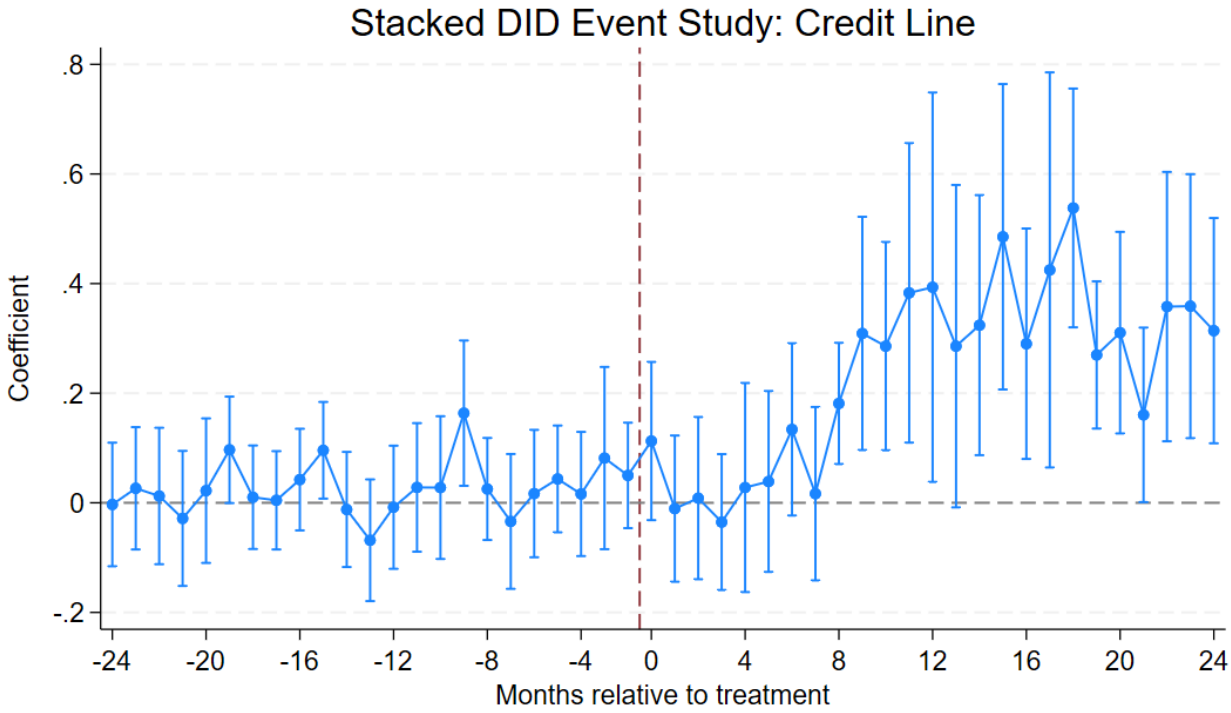


Figure B.1: Stacked Difference in Differences Event Study for Approved Credit Line

Notes: The figure plots stacked difference-in-differences event-study coefficients for ln(*Approved Credit Line*). Event time is measured in months relative to the GT3 rollout month, and the month immediately before implementation is the omitted period.

Table B.2: Stacked Difference in Differences Estimates

	(1)	(2)	(3)	(4)
Panel A. ln(*Approved Credit Line*)				
GT3	0.382***	0.180***	0.176***	0.108***
	(5.46)	(4.64)	(4.55)	(3.33)
Constant	6.286***	6.297***	6.298***	6.301***
	(438.92)	(1286.51)	(1295.08)	(1600.54)
Cohort × Year-Month FE	Yes	Yes	Yes	Yes
Cohort × Bank FE	Yes	Yes	Yes	Yes
Cohort × City FE	No	Yes	Yes	Yes
Cohort × Industry FE	No	No	Yes	Yes
Cohort × Control FE	No	No	No	Yes
Observations	401,141	400,760	400,759	400,721
R^2	0.292	0.396	0.399	0.535
Panel B. Interest Rate				
GT3	−0.399**	−0.325***	−0.329***	−0.188**
	(−2.51)	(−3.63)	(−3.68)	(−2.17)
Constant	8.757***	8.753***	8.753***	8.746***
	(253.33)	(760.94)	(765.90)	(763.13)
Cohort × Year-Month FE	Yes	Yes	Yes	Yes
Cohort × Bank FE	Yes	Yes	Yes	Yes
Cohort × City FE	No	Yes	Yes	Yes
Cohort × Industry FE	No	No	Yes	Yes
Cohort × Control FE	No	No	No	Yes
Observations	401,223	400,841	400,840	400,799
R^2	0.239	0.372	0.377	0.455

Table B.2 continued

	(1)	(2)	(3)	(4)
Panel C. Interest Spread				
GT3	−0.435***	−0.330***	−0.334***	−0.206**
	(−2.66)	(−3.69)	(−3.73)	(−2.34)
Constant	3.636***	3.631***	3.631***	3.624***
	(105.12)	(312.66)	(315.00)	(310.97)
Cohort × Year-Month FE	Yes	Yes	Yes	Yes
Cohort × Bank FE	Yes	Yes	Yes	Yes
Cohort × City FE	No	Yes	Yes	Yes
Cohort × Industry FE	No	No	Yes	Yes
Cohort × Control FE	No	No	No	Yes
Observations	401,223	400,841	400,840	400,799
R^2	0.355	0.467	0.471	0.538
Panel D. Bank Assessed Income				
GT3	0.124**	0.158***	0.154***	0.168***
	(2.53)	(2.70)	(2.67)	(2.79)
Constant	15.152***	15.149***	15.150***	15.149***
	(801.97)	(1427.18)	(1434.42)	(1427.13)
Cohort × Year-Month FE	Yes	Yes	Yes	Yes
Cohort × Bank FE	Yes	Yes	Yes	Yes
Cohort × City FE	No	Yes	Yes	Yes
Cohort × Industry FE	No	No	Yes	Yes
Cohort × Control FE	No	No	No	Yes
Observations	356,702	356,321	356,320	356,292
R^2	0.147	0.206	0.216	0.235

Notes: The table reports stacked difference-in-differences estimates under the cohort-specific fixed-effect progression described in Appendix B.3. Panel D, column (4), omits Income Group and Credit Score fixed effects.

C Distance Measures and Expansion Boundary

C.1 Distance and Expansion Boundary Construction

Borrower to lender distance is measured between the WGS84 centroid of the borrower's registered district or county and the WGS84 centroid of the originating branch city. The baseline measure is geographic distance. Routed driving distance and driving time use the same endpoints.

Let

$$\ell_i = \ln(Distance_i)$$

denote log distance for loan i, and let $\ell_c^{\max,post}$ denote the city-specific post-GT3 maximum of ℓ_i in city c. The boundary construction uses the average post-GT3 expansion in log distance reported in Table 6, column (1), which is 0.135. The two city-specific cutoffs are

$$\ell_c^{cut1} = \ell_c^{\max,post} - 0.135$$

and

$$\ell_c^{cut2} = \ell_c^{\max,post} - 2 \times 0.135.$$

The Inner Band is $\ell_i \in (\ell_c^{cut2}, \ell_c^{cut1}]$, and the Outer Band is $\ell_i \in (\ell_c^{cut1}, \ell_c^{\max,post}]$. The indicator for the Outer Band is used in Table 7, Panel A.

The local estimates in Table 7, Panel B, use the same constructed boundary. Each outcome is first residualized using the fixed effects and controls reported in that table. With $x_i^{band} = \ell_i - \ell_{c(i)}^{cut1}$, the local estimand is

$$\tau_Y = \lim_{x \downarrow 0} E[\widetilde{Y}_i \mid x_i^{band} = x] - \lim_{x \uparrow 0} E[\widetilde{Y}_i \mid x_i^{band} = x].$$

The implementation uses local linear fits, a triangular kernel, bandwidths h and $2h$, robust bias-corrected inference, and an adjustment for mass points in the running variable (Calonico et al., 2014).

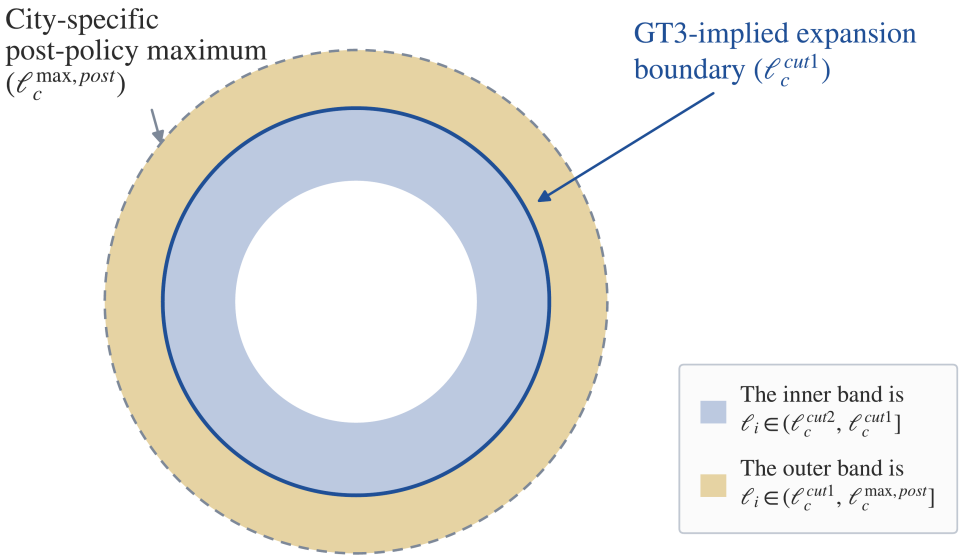


Figure C.1: Construction of the GT3 Implied Expansion Boundary

Notes: The figure illustrates the city-specific post-GT3 lending frontier, the constructed expansion boundary, and the adjacent Inner and Outer Bands.

C.2 Alternative Distance Measure

Table C.1: Alternative Distance Measure

Dep. Var.:	ln(*Approved Credit Line*) (1)	Interest Rate (2)	Interest Spread (3)	ln(*Bank Assessed Income*) (4)
GT3	−0.343**	1.118***	1.169***	−0.884**
	(−2.06)	(3.75)	(3.78)	(−2.49)
log Driving Distance	−0.119***	0.191***	0.191***	−0.205***
	(−20.44)	(10.36)	(10.28)	(−11.84)
GT3 × log Driving Distance	0.057***	−0.110***	−0.114***	0.150***
	(5.60)	(−3.82)	(−3.81)	(7.62)
Constant	7.551***	5.923***	1.091***	16.704***
	(84.84)	(30.62)	(5.61)	(73.60)
City × Year-Month FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Control FE	Yes	Yes	Yes	Yes
Observations	93,993	93,993	93,993	86,026
R^2	0.533	0.683	0.668	0.430

Notes: The table replaces baseline geographic distance with routed driving distance in the interaction specifications. The dependent variables are the log of Approved Credit Line, Interest Rate, Interest Spread, and the log of Bank Assessed Income. All columns include City × Year-Month, Bank, Industry, and Control fixed effects. Column (4) omits Income Group fixed effects. Standard errors are clustered by province × year-month.

C.3 Pointwise Effects Along the Distance Distribution

Table C.2: Pointwise GT3 Effects Along the Distance Distribution

Panel A. Common loan sample

Point	x	ln(*Approved Credit Line*)	Interest Rate	Interest Spread
x_0	0.0000	0.0831	0.3090**	0.3315**
p10	1.3433	0.1528	0.1708	0.1881
p25	2.4245	0.2089	0.0596	0.0727
p50	3.1507	0.2466*	-0.0151	-0.0048
p75	3.5910	0.2694**	-0.0604	-0.0518
p90	4.0897	0.2953**	-0.1117	-0.1051
Pre GT3 max	5.4649	0.3666***	−0.2532*	−0.2519*
Post GT3 max	5.8912	0.3888***	−0.2970**	−0.2974**

Panel B. Bank Assessed Income sample

Point	x	ln(*Bank Assessed Income*)
x_0	0.0000	0.1722
p10	1.6346	0.4256
p25	2.4245	0.5481*
p50	3.1507	0.6606**
p75	3.5777	0.7269**
p90	3.9705	0.7877***
Pre GT3 max	5.4649	1.0194***
Post GT3 max	5.8912	1.0855***

Notes: The table reports pointwise GT3 effects, $\widehat{\beta}_{GT3} + x\widehat{\beta}_{GT3\times\ln(Distance)}$, from the interaction specifications underlying Figure 6. Panel A uses the common loan sample for credit quantity and pricing outcomes. Panel B uses the Bank Assessed Income sample. Rows p10 through p90 use percentiles of the corresponding pre-GT3 distance distribution, followed by the pre-GT3 and post-GT3 maxima. Stars are based on pointwise linear-combination tests using the variance matrix from Table 5.

D Information Scarcity and Bank Specialization

D.1 One-Digit Industry Estimates

Table D.1 reports separate estimates for one-digit industries with at least 3,000 loan observations.

Table D.1: Approved Credit Line Effects by One-Digit Industry

Panel A.

Dep. Var.: ln(*Approved Credit Line*)	A	C	E	F	G
GT3	0.120	0.116***	0.066	0.238***	0.096
	(0.71)	(3.18)	(1.06)	(3.93)	(1.18)
Constant	5.898***	6.690***	6.557***	6.562***	6.534***
	(65.09)	(288.44)	(225.11)	(180.88)	(150.01)
Year-Month FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes
Control FE	Yes	Yes	Yes	Yes	Yes
Observations	3,484	34,370	11,453	63,386	3,681
R^2	0.603	0.647	0.537	0.525	0.702

Panel B.

Dep. Var.: ln(*Approved Credit Line*)	H	I	L	O
GT3	0.064	0.821***	0.160*	0.164
	(0.69)	(6.26)	(1.68)	(1.38)
Constant	6.425***	6.471***	6.750***	6.376***
	(122.85)	(55.63)	(114.02)	(97.23)
Year-Month FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
Control FE	Yes	Yes	Yes	Yes
Observations	4,761	7,077	4,514	3,574
R^2	0.580	0.475	0.636	0.541

Notes: Each column reports a separate loan-level regression for a one-digit industry subsample. The dependent variable is the log of Approved Credit Line. All columns include Year-Month, City, Bank, and Control fixed effects. Industry fixed effects are omitted by construction. Industry letters follow GB/T 4754 (2017). Standard errors are clustered by province × year-month.

D.2 Bank Specialization Measure

Following Blickle et al. (2026), Bank Specialization measures the distance between a bank's local industry portfolio and the corresponding province-quarter industry portfolio. Let ω_{bpqj} denote bank b's approved-credit-line share in industry j, province p, and quarter q. Let ω_{pqj} denote the corresponding industry share in the province-quarter portfolio:

$$\omega_{bpqj} = \frac{\sum_{\ell \in \mathcal{L}_{bpqj}} CL_\ell}{\sum_{\ell \in \mathcal{L}_{bpq}} CL_\ell}, \qquad \omega_{pqj} = \frac{\sum_{\ell \in \mathcal{L}_{pqj}} CL_\ell}{\sum_{\ell \in \mathcal{L}_{pq}} CL_\ell}.$$

The bank-province-quarter specialization measure is

$$Spec_{bpq} = \frac{1}{2} \sum_j \left| \omega_{bpqj} - \omega_{pqj} \right|.$$

Here, CL_ℓ is the Approved Credit Line for loan ℓ. The sets $\mathcal{L}_{bpqj}$ and $\mathcal{L}_{bpq}$ contain the corresponding bank-province-quarter loans, while $\mathcal{L}_{pqj}$ and $\mathcal{L}_{pq}$ contain the corresponding province-quarter loans. Larger values indicate that a bank's local lending portfolio is more concentrated in industries whose shares differ from the province-quarter benchmark.

The pre-GT3 measure used in Table 9, column (1), is constructed only from pre-GT3 loans. The loan-level assignment is leave-one-out, so the focal loan does not mechanically determine its assigned specialization measure. The resulting pre-GT3 measure is standardized to mean zero and unit standard deviation using its pre-GT3 distribution. Local Excess Specialization in Table 9, column (2), applies the same absolute-deviation formula to the contemporaneous bank × province × year-quarter portfolio.