Move a Little Closer? Information Sharing and the Spatial Clustering of Bank Branches

Shusen Qi[*] Ralph De Haas[†] Steven Ongena[‡] Stefan Straetmans[§] and Tamas Vadasz[¶‖]

April 10, 2024

Abstract

We present a model of credit market competition to derive key hypotheses about how information sharing between banks influences the spatial clustering of their branches. We then test these hypotheses using data on 56,555 branches owned by 614 banks across 19 countries. We find that information sharing incentivizes banks to establish branches in localities that are new to them but that are already served by other banks. The resultant branch clustering is associated with reduced spatial credit rationing, as information sharing enables firms to access credit from more distant banks. These findings underscore how information sharing makes it more important for banks to move closer to each other rather than closer to their borrowers.

JEL classification: D43, G21, G28, L13, R51

Keywords: Information sharing, branch clustering, spatial bank competition

[*]Xiamen University and Peking University (`shushenqi@xmu.edu.cn`).

[†]European Bank for Reconstruction and Development, KU Leuven, and CEPR (`dehaasr@ebrd.com`).

[‡]University of Zurich, Swiss Finance Institute, KU Leuven, NTNU Business School, and CEPR (`steven.ongena@bf.uzh.ch`).

[§]Maastricht University and University of Antwerp (`s.straetmans@maastrichtuniversity.nl`).

[¶]KU Leuven (`tamas.vadasz@kuleuven.be`).

[‖]The authors thank Martin Brown, Stephanie Chan, Francesca Cornelli, Hans Degryse, Thomas Mathä, Marco Pagano, Francesc Rodriguez Tous, Vahid Saadi, Koen Schoors, Nathaniel Young, Yunqi Zhang, Kuncheng Zheng, and participants at the Bocconi University–Sapienza University Conference on "New Frontiers in Banking: from Corporate Governance to Risk Management", the American Economic Association ASSA Annual Meeting, the Swiss Winter Conference on Financial Intermediation (Lenzerheide), the Great China Area Finance Conference (Xiamen), the MoFiR Workshop on Banking, the China International Conference in Finance (Hangzhou), the Deutsche Bundesbank–IWH–CEPR Conference on the Future of Financial Intermediation (Eltville am Rhein), the EBC Network Workshop (Luxembourg), the European Economic Association Annual Meeting (Lisbon), and seminar participants at the University of Antwerp, Xiamen University, Shandong University, and Huazhong University of Science and Technology (Wuhan) for useful comments. The views expressed are those of the authors and not necessarily those of the EBRD. Shusen Qi acknowledges financial support from the National Natural Science Foundation of China (71903164) and the Social Science Foundation of Fujian Province (FJ2019B140).

1 Introduction

In many emerging markets and developing countries, credit markets remain segmented geographically (Canales and Nanda, 2012; De Haas, Lu and Ongena, 2023). Due to a lack of reliable and publicly available borrower information, banks typically limit their lending to local firms. Such spatial credit rationing disproportionately impacts small and medium-sized enterprises (SMEs), given their opaqueness and reliance on bank funding. Credit market segmentation can create stark geographical variation in firms' borrowing capabilities and may therefore perpetuate spatial disparities in economic activity (Jayaratne and Strahan, 1996)

To attenuate the negative consequences of segmented credit markets, various countries have introduced systems for the voluntary or mandatory exchange of borrower information among banks. The aim of these systems is to reduce information asymmetries between banks and borrowers, allow banks to lend over a greater distance, and integrate local credit markets and increase their competitiveness. Against this background, we study both theoretically and empirically, how the introduction of information sharing shapes the spatial configuration of banks' branch networks. Analyzing changes in banks' branch footprint is important as recent evidence shows how policy-induced shifts in the geographical location of bank branches can have substantial real-economic consequences, especially in low-income settings.[1]

Our first contribution is to construct a stylised model to guide intuition about how information sharing between banks affects branch clustering. Building on an extensive literature on relationship lending (Rajan, 1992; Von Thadden, 2004; Hauswald and Marquez, 2006) banks in our model have private information about their customers. Credit market competition is therefore subject to adverse selection, which raises interest rates and the incumbent banks' profits above the competitive benchmark. Adverse selection also acts as a barrier to entry—as in Dell'Ariccia, Friedman and Marquez (1999)—because uninformed entrants likely obtain a low (or zero) market share of creditworthy customers.

In this setting, we consider the introduction of a formal mechanism through which banks

[1]See, for example, Barboni, Field and Pande (2022) on India; Fonseca and Matray (2022) on Brazil; and Ji, Teng and Townsend (2023) on Thailand.

share hard (i.e., codified and transferable) borrower data: an information sharing system such as a public credit registry or a private credit bureau. Such data can include 'negative' information about prior defaults and late payments of loan applicants as well as 'positive' information about whether they have outstanding debt elsewhere. Information sharing decreases banks' information advantage and increases competition in the credit market. This is reflected in a higher fraction of valuable customers an uninformed rival bank—either an existing competitor in a locality or a new entrant—can poach from informed incumbents.

Our model differentiates between soft and hard information, thereby providing insights into the disparities between lending practices in rural and urban areas. In rural areas, where the social distance between borrowers and lenders is smaller (Uzzi, 1999), lending relies mostly on soft information that cannot be hardened as part of an information sharing system. In urban areas, personal connections play a lesser role and banks rely more on hard information. This conceptual framework yields four testable hypotheses about the impact of information sharing on branch concentration and clustering.

First, information sharing increases the likelihood that banks open branches in localities with more pre-existing branches of other banks. *Within locality*, this means that small banks competing with relatively many rival branches increase their presence by opening new branches. *Across localities*, banks tend to open branches where many other banks are already present. In such areas, the impact of information sharing is the largest, as entrant branches can attract many new clients from incumbent rivals. Second, this result holds both on the intensive and the extensive margin. That is, as a specific case, banks are more likely to enter a new locality and open a *first* branch in localities where many other banks already operate. This effect in particular increases the spatial clustering of branches. Third, the impact of information sharing on branch clustering is more pronounced when information sharing is of higher quality. Fourth, information sharing especially influences the branch clustering of relationship banks. These banks interact frequently with borrowers to obtain proprietary information and rely more on such soft information than transactional banks. Because soft information is largely private, it is inaccessible to relationship banks that consider entering a new local market. In contrast, transactional banks rely more on public hard information and are less affected by a lack of access to soft information. When an information sharing regime effectively hardens a subset of soft information and makes it public, this disproportionately benefits relationship banks as it mitigates their key entry barrier.

To test these predictions, we use detailed bank branch data—geographical coordinates and the dates of establishment (and/or closure) of each branch—from 19 Eastern European countries. Our sample covers 56,555 branches from 614 banks that were active during the period 1995-2012 across 8,536 villages, towns or cities (henceforth indicated as 'localities'). The data set further contains information on the ownership of these branches so that we can distinguish between branches of different types of banks. Eastern Europe constitutes a natural testing ground for our model because information asymmetries are pervasive while creditor rights remain relatively weak (Brown, Jappelli and Pagano, 2009).

Importantly, during our sample period many eastern European countries institutionalized information sharing among lenders—either through a public credit registry or through a private credit bureau. We use the introduction of these information sharing regimes as country-level shocks that push banks towards a new clustering equilibrium. This setting can also provide insights into how bank clustering can respond to similar but slower improvements in borrower transparency in more developed banking markets.

In terms of empirical methodology, we implement a difference-in-differences-in-differences framework with the treatment (presence of information sharing) varying across countries and years. We follow Cengiz, Dube, Lindner and Zipperer (2019) to handle the staggered treatment timing where different countries introduced information sharing at different points in time. We then compare how, within the same country, information sharing differentially affects branch openings across localities with different numbers of pre-existing bank branches. This strategy mitigates selection biases and, through the use of granular fixed effects, alleviates concerns about omitted variables. In particular, we saturate our specifications with *locality*year*treatment event* fixed effects; *bank*year*treatment event* fixed effects; and *locality*bank*treatment event* fixed effects. This removes the possibility that branch clustering merely reflects local firms' demand for credit or other potential confounders, such as the depth of local labor markets.

By way of preview, we find that information sharing has a strong positive effect on branch clustering. Banks are more likely to open new branches in localities where they did not yet operate but where other banks were already present, and this effect is stronger in large localities. This clustering is also more pronounced for relationship banks and in countries where information sharing is more effective. Moreover, due to information sharing, banks become more likely to open a branch in a new locality while simultaneously closing a branch elsewhere in the country.

An analysis of auxiliary data provides compelling evidence that, following the implementation of information sharing, banks exhibit a growing tendency to cluster their branches within regions as defined at various levels in the *Nomenclature des Unités Territoriales Statistiques* (Nomenclature of Territorial Units for Statistics or NUTS). Importantly, in rural areas with a low population density, we do not see an absolute decline in branch density, while in more densely populated urban areas there is a significant increase in local branch density because of the introduction of information sharing. In other words, due to information sharing, smaller localities lose out in relative but not in absolute terms.

These results are confirmed by an analysis in which we compare the distance between a bank's headquarters and all its individual domestic branches, before and after the introduction of information sharing. We observe, in line with distance constraints, that in general banks are less likely to operate branches in localities that are further from their headquarters. Importantly, however, this 'gravitational pull' weakens with the introduction of information sharing: banks become willing to open branches in more distant locations.

Lastly, we also provide suggestive evidence showing that, in line with a reduction in geographical credit rationing, information sharing allows firms to borrow from banks that are more distant. Taken together, these results show how information sharing makes it more important for banks to move closer to each other rather than closer to their borrowers.

This paper contributes to three strands of the literature. First, we provide new insights into the causes and consequences of spatial agglomeration in the supply of financial services.

In line with Robinson's (1952) adage that "where enterprise leads, finance follows", banks tend to locate branches in areas with fast population and economic growth, and where the demand for their services is high (Carbo-Valverde, Hannan and Rodriguez-Fernandez, 2011; Brown, Guin and Kirschenmann, 2016). At the same time, an influential 'finance-and-growth' literature shows that the supply of credit not just passively follows economic activity but also has a causal impact on subsequent growth (see Levine (2005) for an overview). Relatedly, recent work exploits exogenous spatial variation in bank branches—reflecting historical 'quirks' or waves of financial deregulation. This literature shows that the local presence of bank branches remains a first-order determinant of economic dynamism and continues to have strong effects on socioeconomic outcomes including employment, inequality, and firm innovation.[2] Our contribution is to develop an intuitive framework to study how information sharing among banks reshapes the spatial footprint of branch networks. We test our model predictions in a rich international context, using the introduction of information sharing as country-level shocks that push banks towards a new clustering equilibrium (while controlling for initial real agglomeration).

Second, we add to the literature on the economic impact of information sharing. Cross-country evidence indicates that information sharing is linked to more private-sector lending, fewer defaults, and lower interest rates (Jappelli and Pagano, 2002). Evidence also suggests that (voluntary) private credit bureaus are more effective than (mandatory) public registries in this regard (Martinez Peria and Singh, 2014). Yet, it remains unclear exactly how information sharing affects bank behavior. We uncover an important mechanism: the central availability of hard borrower information leads to a different branch-clustering equilibrium, in which local credit markets get "de-segmented", and that is associated with less spatial credit rationing.

Third, our model adds to the theoretical literature on reducing informational asymmetries in banking. Earlier contributions explore how information sharing reduces moral hazard and adverse selection, improves loan quality, lowers interest rates (Padilla and Pagano, 1997, 2000), and shapes inter-bank competition (Bouckaert and Degryse, 2004, 2006; Bennardo, Pagano and

[2]See Jayaratne and Strahan (1996) and Kroszner and Strahan (2014) for the US; Guiso, Sapienza and Zingales (2004) for Italy; and Bircan and De Haas (2020) for Russia.

Piccolo, 2015).[3] Our approach is similar in spirit to Dell'Ariccia et al. (1999) who endogenize entry in a market subject to adverse selection. In contrast to their work, and consistent with the tradition of modeling adverse selection (Von Thadden, 2004), our model features simultaneous bidding and endogenous information acquisition (Hauswald and Marquez, 2006). In contrast to the latter paper, we focus on branch entry incentives rather than the strategic use of information acquisition in obtaining market share. That is, we do not allow for location-based price discrimination and our banks compete in a Bertrand-way with asymmetric information being the only friction among banks. We model information sharing formally, which is reminiscent of the public signal in Hauswald and Marquez (2003). Our model makes explicit that the shared information is a subset of the insider's information set, and allows us to study the link between information sharing and information production.

We proceed as follows. Section 2 introduces our theoretical model and develops hypotheses after which Section 3 describes our data. Section 4 then sets out our methodology and Section 5 reports the empirical findings. Section 6 concludes.

2 Theory and hypotheses

2.1 General setup

To derive testable hypotheses about the impact of information sharing on bank clustering and concentration, we develop a stylized model of credit market competition subject to adverse selection. We consider a locality populated by a continuum of mass M penniless borrowers with access to a project. The project requires 1 unit of investment in period $t = 1$, which borrowers can obtain from branches of competing banks. The project is either high-quality (type H) with probability q, or low-quality (type L) with complementary probability $(1 - q)$.

[3] Van Cayseele, Bouckaert and Degryse (1994) analyze theoretically the effect of sharing 'negative' borrower information (past defaults) and 'positive' information (current indebtedness) on the number of branches per bank. Unlike our paper, the authors do not analyze the spatial distribution of branches. Other key contributions include Gehrig and Stenbacka (2007)—on the dynamic patterns of loan rates paid by various groups of customers—and Karapetyan and Stacescu (2014) who also study endogenous information acquisition as a response to information sharing.

An L-type project fails with a probability of 1 (returning 0), while an H-type project yields an expected return of $R > 1$. We identify the borrower-type with the type of their project.

Both the number of banks competing for borrowers, their branch network, and the bank-client matching are determined exogenously prior to the borrowing stage. Borrowers have established a relationship with exactly one legacy bank, which receives an informative signal about the borrower's type in the subsequent credit phase. This prior relationship can be thought of either as having a payment account with the bank, or a past lending relationship.

Before the credit market opens, an information sharing regime may be introduced. Following this, further bank entry is possible. A new entrant bank has immediate access to the information included in the information sharing system, but has no legacy customers in the given locality. Banks adjust their number of branches in any given locality as a function of the number of customers they can obtain in the credit market. A single branch can efficiently serve up to $\overline{m}$ clients for a per-customer cost of k, which we normalize to zero. If a bank obtains more than $\overline{m}$ customers, it must open a new branch. On the other hand, if a branch handles too few clients, say less than $\underline{m}$, then the fixed cost of operating a branch dominates, resulting in a unit cost of $K > k$. We assume $\overline{m} > 2\underline{m}$, so that a bank with at least $\underline{m}$ customers can always operate a branch in a locality in a cost-efficient way.

Information structure. Insider banks use a costly screening technology that generates a signal $s \in \{\ell, h\}$ about the borrower. The signal has the following conditional distribution:

$$Prob[\ell|L] = \phi \leqslant 1 \qquad \text{and} \qquad Prob[h|H] = 1 \tag{1}$$

This means that while an ℓ-signal is evidence of an L-type, an h-signal is an informative but imperfect signal of an H-type.[4] A completely uninformative technology ($\phi = 0$) means that the bank identifies all borrowers, and q of them erroneously, as H-type. The signal quality, ϕ, is endogenously determined by the bank's costly screening- and information acquisition effort.

[4]Risky borrowers (successfully or unsuccessfully) attempt to hide this information from the bank. A similar bad news structure is used in He, Huang and Zhou (2023) and Thanassoulis and Vadasz (2023).

Specifically, to generate a signal with quality ϕ, costs the bank $\frac{1}{2}c\phi^2$ per borrower screened. We interpret ϕ as 'soft' information accumulated during the bank-borrower relationship, which—absent an institutional arrangement—is the inside bank's proprietary information.[5]

When an information sharing regime is in place, a subset of such accumulated soft information is 'hardened'—encoded in an inter-operable digital form—and distributed to competitors. For example, banks may condense various pieces of text-based information (meeting notes, financial projections, market commentary, etc.) into one numerical borrower rating that becomes meaningful to other lenders and gets shared via the new information sharing regime. However, not all soft information can be hardened and subsequently shared. Intuitively, the informed bank retains some private information about the riskiness of the borrower which either cannot be codified without significant loss of information (for example, impressions through face-to-face interactions, local rumors, etc.) or may not reach the legal requirement to post them in the credit registry (delinquencies not reaching formal default conditions, erratic payment flows, etc.). To capture this formally, let the signal of the less informed bank be $\tilde{s} \in \{\tilde{h}, \tilde{\ell}\}$, where

$$Prob[\tilde{\ell}|L] = \nu < \phi \qquad Prob[\tilde{h}|H] = 1 \qquad \text{and} \qquad Prob[\ell|\tilde{\ell}] = 1 \tag{2}$$

Thus, the quality of information shared among banks is determined by the free parameter $\nu \in [0, \phi]$. For the sake of generality, this can be thought of as an arbitrary weakly increasing function of the total information production ϕ. The first inequality implies that only a subset of information can be shared. The last equation is the sharing condition: the outsider bank can only be informed about the riskiness of the client if the insider bank is also informed.[6]

Both types of banks can price discriminate based on their respective signals. Therefore, an insider bank offers $r(h)$ upon observing an h-signal and $r(\ell)$ upon observing an ℓ-signal, while an uninformed bank offers $r(\tilde{h})$ and $r(\tilde{\ell})$, respectively. Notice that our model captures

[5]The classification of information into soft and hard components is in line with a large prior literature, see for example Agarwal, Ambrose, Chomsisengphet and Liu (2011), Agarwal and Ben-David (2018), Liberti (2018), and Liberti and Petersen (2019).

[6]This feature, among others, distinguishes our model from Hauswald and Marquez (2003) and others where uninformed banks observe an *independent* public signal.

no information sharing as a limiting case: as $\nu \to 0$, the uninformed bank never receives an $\tilde{\ell}$-signal and their $\tilde{h}$-signal becomes completely uninformative.

The timing of the model is designed to concisely but realistically capture the market's reaction to the introduction of information sharing. Banks' decisions to enter new markets is slow-moving and information sharing comes as a surprise to the incumbents. However, banks can relatively quickly adapt their screening- and information producing strategies as a response to the new information environment. Potential further entrants—which is again a slow-moving decision—will have to compete on the basis of the adjusted information structure and endogenize the impact of information sharing on their entry and pricing decisions.

2.2 Results and hypotheses

First, we solve for equilibrium with arbitrary fixed private information ϕ and information sharing $\nu \leqslant \phi$, and analyze the impact of the introduction of information sharing on credit prices and profits.[7] For brevity, we refer to the more informed bank as "informed" (index i) and the relatively less informed banks as "uninformed" (index u). Both types can price discriminate based on their respective signals. However, the informed bank's signal is more precise so that the uninformed competitor necessarily obtains a larger share of lower quality loans.[8] In line with the literature on pricing under adverse selection (Von Thadden, 2004; Hauswald and Marquez, 2006), there is no pure-strategy Nash equilibrium of the pricing game. The structure of equilibrium credit prices and profits is derived in Proposition 1.

Proposition 1 *Let $\underline{r}(\nu)$ be the break-even rate for an uninformed bank (defined formally in the Appendix). The credit market equilibrium for a given ϕ and ν is as follows:*

1. *All banks reject credit for borrowers with a low (ℓ or $\tilde{\ell}$) signal.*

2. *If $\underline{r}(\nu) < R$, the informed (resp. uninformed) banks randomize credit prices for their h-signal type customers over the interval $[\underline{r}(\nu), R]$ according to continuous distributions*

[7] At the time of posting loan prices, the equilibrium values of these parameters have already been realized.

[8] Bofondi and Gobbi (2006) show that entrant banks experience higher default rates than incumbents due to information asymmetries.

$F_i(r)$ and $F_u(r)$. The informed bank places a positive probability mass on R, while the uninformed banks deny credit with some positive probability.

3. *If $\underline{r}(\nu) > R$ the uninformed banks reject the applicant with probability 1 and the informed bank acts as a local monopolist over its captured clients.*

4. *The uninformed banks obtain zero profit. The informed bank obtains a positive expected profit π_i per unit insider customer of:*

$$\pi_i = \begin{cases} \pi_{comp} := (1-q)(\phi-\nu) & \text{if } \ \underline{r}(\nu) < R; \\ \pi_{mon} := (1-q)\phi + (qR-1) & \text{otherwise} \end{cases} \tag{3}$$

5. *There exist no price equilibrium in which a new entrant enters but operates cost-inefficiently.*

Proof. *See Appendix A.1* ■

First, notice that Proposition 1 characterizes the static competitive equilibrium absent information sharing with the substitution $\nu = 0$. This is what banks expect a priori as an equilibrium outcome. It in turn determines the initial market structure and the quality of private information. Second, Proposition 1 describes the credit market for both the incumbents and a potential new entrant following the introduction of information sharing of quality ν.

The break-point in the profit function arises due to a feasibility constraint: the maximum interest rate a bank can charge is R. If ν is low or zero, the break-even interest rate $\underline{r}(\nu)$ is high. The uninformed bank then cannot profitably bid for customers and the informed bank effectively becomes a local monopolist over its informationally captured borrowers.[9]

The next Proposition quantifies the probability that a borrower switches banks with information sharing:

[9] Empirically, this would correspond to extremely low levels of switching, such as in a local market where competition is severely hindered by strong adverse selection.

Proposition 2 *For any fixed ϕ and ν, an informed incumbent retains the following share of its creditworthy (h-signal) customers:*

$$\lambda_i = \frac{1}{2}\left(1 + \left[\frac{\pi_i}{\pi_{mon}}\right]^2\right) \tag{4}$$

The retained market share ceteris paribus increases in private information ϕ and decreases in information sharing ν.

Proof. *See See Appendix A.2* ■

Proposition 2 states that the probability that the informed bank can retain a client decreases following the introduction of information sharing. Intuitively, information sharing enables uninformed rivals to identify some high-risk borrowers who try to switch from their informed insider. This mitigates adverse selection and intensifies price competition, leading to a decrease in interest rates and a reduction in banks' profits. As rival banks can now compete more aggressively, they can poach valuable customers more easily from the insider bank, despite their informational disadvantage. This results in more frequent switching.

The ability to poach clients from rivals is particularly attractive for banks who compete with a larger number of branches of other banks. To see this, consider a natural interpretation of our setup where banks are heterogeneous in their number of branches in a locality at the beginning of $t = 1$. Such heterogeneity can result from previous bank mergers or from other historical reasons. Branches of the same bank do not compete with each other, so the enhanced ability to win over clients who approach the bank is particularly valuable for smaller challenger banks or new entrants that can now successfully attract business away from their large legacy competitors with many pre-existing branches. As their customer base grows and exceeds (a multiple of) $\overline{m}$, such smaller banks with many rival branches in a locality are more likely to open new branches themselves. This decreases bank concentration within a locality and yields our first testable hypothesis:

Hypothesis 1a: (Concentration effect). Information sharing increases the likelihood that banks facing a large number of competitor branches within a locality open new branches. Thus, the concentration of bank branches within a locality decreases.

The analysis so far has taken the quality of information as given and uniform. Yet, the information environment may differ across localities. In particular, rural areas, where social ties are stronger (Uzzi, 1999) tend to be characterized by a relative abundance of 'traditional' soft information, which cannot be hardened as part of an information sharing system. As DeYoung, Glennon, Nigro and Spong (2019) argue: "*If one accepts the conventional wisdom that rural communities are closer knit than urban communities—i.e., they are places where "everyone knows each other's business"—then this information grapevine will include the business community. This will provide rural banks with a relatively low-cost endowment of soft information about local businesses, over and above what is available to urban banks*".[10]

In contrast, in large urban areas it is typically relatively easier to produce information that can be hardened, for example, due to more active information production by large accounting firms. Allee and Yohn (2009) show that a firm's size positively predicts whether its financial statements are compiled, reviewed or audited by a professional accounting firm. Given the typically smaller size of rural firms, this would mean that loan applications from such firms are less likely to be supported by audited financial statements.

Furthermore, information production is endogenous, and incumbents may strategically respond to the introduction of an information sharing regime by altering their information acquisition strategies. We characterize this in the next Proposition:

[10] DeYoung et al. (2019) show that even when holding the geographic distance between borrowers and lenders constant, there are differences in how rural and urban banks interact with their clients. Kittiakarasakun (2010) shows that in urban areas, banks rely more on verifiable hard information while rural branches tend to lend to nearby customers about whom they have personal knowledge. Further evidence to support the assertion that small banks, which are more active in rural areas, rely more on soft information is given by Stein (2002) and Berger, Miller, Petersen, Rajan and Stein (2005).

Proposition 3 *The optimal signal quality chosen by banks conditional on the information sharing regime characterized by $\nu(\phi)$ is:*

$$\phi^{\star}(\nu) = \begin{cases} \min\{\frac{1-q}{c}\left(1 - \frac{\partial \nu}{\partial \phi}\right); 1\} & \textit{if} \quad \underline{r}(\nu) < R \\ \min\{\frac{1-q}{c}; 1\} & \textit{otherwise} \end{cases} \tag{5}$$

Proof. *See See Appendix A.3* ■

Proposition 3 shows that the banks' strategic response to the introduction of information sharing crucially depends on the nature of the local information environment. In rural areas, banks rely more on non-transferable soft information, and retain a sizable information advantage even after the introduction of information sharing. In contrast, in large localities, banks tend to rely more on information that can be hardened and shared. Our model would capture this by $\partial\nu/partial\phi$ being larger in urban areas. Banks in such locations therefore reduce their information acquisition effort more after an information sharing system is introduced. This further damages their information advantage and in turn, according to Proposition 2, increases the market share a rival can capture. We conclude that the introduction of information sharing has a more pronounced effect in large localities, where more branches tend to be opened as a result. This yields our next hypothesis:

Hypothesis 1b *(Clustering effect). Information sharing increases the likelihood that banks open new branches in localities with more pre-existing branches of other banks (all else equal). This increases the spatial clustering of bank branches.*

So far, our model has established that information sharing facilitates new branch openings. It does so especially in larger localities, where it makes smaller rivals more competitive, relative to larger incumbents, on the intensive margin. This argument extends in a straightforward way to the extensive margin, where banks enter new localities for the first time. Without information sharing, a potential entrant bank would attain little (or no) market share, and would be unable to operate even one single branch in the efficient part of its cost curve. Part (v) of Proposition 1 shows that entry is prevented if the potential entrant would end up in the inefficient part of its cost curve. Such a market size effect is more pronounced when adverse selection concerns are more serious. In particular, if $\underline{r}(\nu = 0) \geqslant R$, which would be the case with large information asymmetries, the entrant cannot gain any market share. This would completely prevent further entry even without additional entry costs considerations. Information sharing reduces the informational barriers to enter and allows the entrant to gain market share. Consistent with the previous discussion, this is especially the case in localities where relatively many other branches are already present. This leads to the following variant of Hypothesis 1:

Hypothesis 2: Information sharing increases the likelihood that banks enter a new locality by opening a first branch in localities with more pre-existing branches of other banks (all else equal). This increases the spatial clustering of bank branches on the extensive margin.

Taken together, hypotheses (1) and (2) tell us that following the introduction of an information sharing regime: (i) banks compete more aggressively and gain new clients more easily especially when they face incumbent rivals with a large number of branches, and (ii) banks face less information barriers to enter new localities, and especially do so in larger ones. The combined result is that banks tend to open branches and/or increase their presence in localities where already many other bank branches are present.

Next, information sharing has two effects in the model. First, from Equation 4, it invariably increases the switching probability, and therefore increases clustering on the intensive and extensive margins. Second, Equation (3) suggests that it decreases banks' profits as information rents decrease and competition intensifies.[11] However, it is likely that private information remains an important component in lending decisions, as information sharing is restricted to a subset of hard information. We therefore expect that the first effect dominates in all empirically interesting cases:

[11] By construction, banks obtain zero economic profits in the limit of full information sharing. When information production is endogenous and costly, this naturally implies zero information production. While this prediction is theoretically sound—and in the spirit of the Grossman and Stiglitz (1980) paradox—it is empirically irrelevant.

Hypothesis 3: *The impact of information sharing on bank clustering is stronger in countries with higher-quality information sharing systems.*

Finally, our model speaks to how information sharing differentially affects relationship and transactional lenders. Suppose banks differ in their lending technologies: relationship banks rely more on the private (soft) signals they collect, whereas transaction banks primarily make their lending decisions based on publicly available (hard) information. Information sharing can then especially affect relationship banks for two reasons. First, without information sharing, the lack of access to soft information is an entry barrier that mostly affects relationship banks. These banks in turn benefit disproportionally from the introduction of information sharing. Second, while both types of banks can access the newly 'hardened' information, it is likely that transaction banks, with their focus on public information, may be less experienced in decoding and interpreting such information. It follows that especially in localities with many competing legacy branches, relationship banks benefit more from information sharing. They become more willing to enter new markets, leveraging the advantages gained from the shared information. This leads to our fourth hypothesis:

Hypothesis 4: *The impact of information sharing on bank clustering is stronger for relationship banks.*

3 Empirical setting and data

3.1 Empirical setting

After the fall of the Berlin Wall in 1989, the countries of Central and Eastern Europe started a process of profound economic transition from socialist, centrally planned economies to capitalist market economies. The first decade of this transition process was highly disruptive. Large-scale privatization and widespread liberalization brought about a massive reallocation of labor and capital across and within economic sectors, rectifying the distortions inherited from central planning (Roland, 2000).

An important element of the transition process was the development of well-functioning banking sectors. After having recapitalized state banks during the first half of the 1990s, countries started to privatize these banks. This often involved selling majority stakes to foreign investors, mostly multinational banking groups from Western Europe. These foreign-owned banks currently regard Emerging Europe as a second home market where they compete with the remaining domestic banks. In many countries, foreign banks presently own between 20 and 90 percent of all banking assets (De Haas, Korniyenko, Pivovarsky and Tsankova, 2015).

After the disruptive reforms of the 1990s, policy makers realized that liberalization and privatization needed to be complemented with new institutions to support the functioning of markets and private enterprise. It was during this later wave of institutional reforms that countries introduced credit registries and bureaus. Appendix Table A2 shows that the average credit registry (credit bureau) was only introduced in 2003 (2005).

The above background to our empirical setting helps to clarify to what extent our findings carry over to other countries that introduced, or will introduce, information sharing. It can be argued that the introduction of information sharing relatively soon after a period of economic upheaval, as was the case in Emerging Europe, was not a one-off anomaly. In fact, historically, credit registries and bureaus have often been introduced (and continue to be introduced) in countries that recently experienced structural transformation. For example, Germany established its public credit registry in 1934, in between the two World Wars, while Ukraine only introduced a registry in 2019 after an economic crisis in 2014–2016 (it had been operating a credit bureau since 2007). While this suggests that our setting is representative, it also underlines the importance of absorbing local time-varying developments in population and economic activity and of robustness checks that control for concomitant policy changes and reforms. We discuss both issues in the sections to come.

3.2 Data

To test our hypotheses, we use the introduction of information sharing regimes as country-level shocks that push banks towards a new clustering equilibrium. This approach requires time-varying data on branch locations for countries that introduce information sharing—either through a public credit registry or through a private credit bureau—at different points in time. We have access to information on the geographical coordinates of 56,555 branches owned by 614 banks in 8,536 villages, towns or cities (i.e., 'localities') across 19 emerging European countries. These branches represent over 95 percent of all bank assets in our sample countries. A team of consultants with extensive banking experience collected these data by contacting banks or downloading data from bank websites. We double-checked all information with the banks as well as with the SNL Financial database. This data collection exercise was part of the second Banking Environment and Performance Survey (BEPS II).[12] The data paint a precise, complete and gradually changing picture—reflecting branch openings and closures—of the banking landscape during the years 1995 to 2012. Figure 1 depicts the spatial branch distribution at the start and the end of our sample.

Appendix Table A3 summarizes the number of branches that opened or closed by year and country: 31,927 (1,065) branches opened (closed) during our sample period. Many branches were established during 2001–2007, a period of rapid credit growth. The expansion of branch networks slowed down after the global financial crisis when fewer branches opened while branch closures (rare before the crisis) accelerated. Approximately half of all branch openings took place when a country had a credit registry or bureau in place.

The unit of observation in our main analysis is the bank-locality-year (see Section 4). For each bank in our data set, we track the number of existing branches (if any), the number of newly opened branches, and the number of closed branches. For each branch, we know its geo-coordinates and its address (street; name of village, town or city (the 'locality') and postal code). We use this information to aggregate all branches up to the locality level. We have

[12] For more details, see Beck, Degryse, De Haas and Van Horen (2018a) and `www.ebrd.com/what-we-do/economics/data/banking-environment-and-performance-survey.html`.

information on 8,536 such localities for every year during 1995–2012. The resulting dependent variables capture the opening of new bank branches across localities and over time. Table 1 contains summary statistics while Appendix Table A1 provides all definitions.

New branch opening is a dummy variable that captures whether a particular bank opens a new branch in a locality in a given year. *Net branch opening* is a dummy that also takes branch closures into account: it equals one if in a particular year and locality a bank adds at least one branch in net terms (that is, the number of branch openings minus closures is strictly positive), and equals zero otherwise. Table 1 shows that on average 4 percent of all bank-locality-year observations see a new branch opening. Given the small number of branch closures, this percentage is virtually the same for *Net branch opening*.

We also count the number of pre-existing branches in a locality that are owned by other banks (*No. branches other banks*). To handle zeros, we use an inverse hyperbolic sine transformation.[13] Variation is substantial, with some localities not being served by any bank whereas some of the largest localities contain many bank branches. We consistently measure this time-invariant variable in the year in which information sharing is introduced.[14]

To contrast the impact of information sharing on relationship lenders versus transactional lenders, we use three empirical proxies of banks' lending techniques: size, ownership, and a direct measure of a bank's main lending technology. We first classify a bank as small if the number of branches it operates is strictly below the country median. The existing literature suggests that small banks are more likely to apply relationship-lending techniques and hence have a comparative advantage in lending to small and informationally opaque firms. In contrast, large banks tend to be better at lending to larger and more transparent firms (Cole, Goldberg and White, 2004). We therefore expect that the introduction of information sharing affects smaller banks more. Table 1 shows that 32 percent of all banks in our data set are small and that these banks own 10 percent of all bank branches in our sample.

[13]Alternatively, we could have taken the log of 1 plus the number of pre-existing branches. This is less ideal as it involves an arbitrary manipulation of the original variable. All results are nevertheless very similar when using that alternative approach.

[14]We provide a robustness test in Appendix Table A12 to show that our results hold when we instead measure *No. branches other banks* in the year before information sharing is introduced.

Second, we merge our data with hand-collected bank ownership information to distinguish between branches of foreign and domestic banks. A bank is classified as foreign if at least half of its equity is in foreign hands. Domestic banks can possess a comparative advantage in reducing information asymmetries vis-à-vis local firms (Beck, Ioannidou and Schäfer, 2018b). In this view, domestic banks tend to have a deeper understanding of local businesses and typically base their lending decisions on 'soft' qualitative information on these firms (Berger and Udell, 1995; Petersen and Rajan, 2002). In contrast, foreign banks may have difficulties in processing soft information and therefore tend to grant loans on a transaction-by-transaction basis using standardized decision methodologies (Berger, Klapper and Udell, 2001). Table 1 shows that only 43 percent of the banks in our country sample are still in domestic hands, reflecting the high levels of foreign direct investment in these banking systems (see also Section 3.1). Domestic banks tend to be relatively large and on average account for 51 percent of all bank branches.

Third, we determine more directly whether a bank is a relationship or a transactional lender when providing credit to small businesses. Recent contributions argue that foreign banks, just like their domestic competitors, can successfully lend to small businesses (Berger and Udell, 2006). Indeed, Beck et al. (2018a) show that among both domestic *and* foreign banks in emerging Europe, large proportions operate as relationship lenders. Banks' ownership and their lending techniques may thus be more orthogonal than previously thought.

To characterize banks' lending technologies, we follow Beck et al. (2018a) and use question Q6 of the 2$^{\text{nd}}$ Banking Environment and Performance Survey (BEPS II). As part of this unique survey, bank CEOs participated in face-to-face interviews in 2012. Question Q6 asked CEOs to rate on a five-point scale the importance (frequency of use) of the following techniques when lending to small businesses: relationship lending; fundamental and cash-flow analysis; business collateral; and personal collateral (personal assets pledged by the entrepreneur). Although almost all banks find knowledge of the client of some importance, 59 percent of the banks find building relationships "very important", while the rest considers it only "important" or "neither important nor unimportant". We categorize banks that find client relationships to be "very important" as relationship lenders and all other banks as transactional lenders.[15] Table 1 shows that while relationship banks make up 59 percent of all banks, they own only 45 percent of all branches (among the banks for which we have data on lending technologies). This confirms that relationship lenders are typically somewhat smaller than transactional lenders.

Next, we collect data on the introduction of information sharing regimes from the World Bank Doing Business database. Appendix Table A2 shows that during 1995–2012, 13 out of the 19 countries in our data set introduced a public credit registry and 15 a private credit bureau. The timing of these introductions varies substantially, which is crucial for our empirical identification. We also measure the quality of these information sharing regimes through the World Bank Doing Business credit information index. This index ranges from zero to six and reflects the rules and practices that affect the coverage, scope, and accessibility of credit information (higher values indicate information sharing that is more effective).[16] Unconditional (conditional) on either a credit registry or a credit bureau being in place, the average quality score across countries and years is 1.3 (2.4).

Lastly, to test whether firms can borrow from more distant bank branches in the presence of information sharing, we merge our branch data with the Kompass database. Kompass provides firms' address, industry and—critically for our purposes—the primary bank relationship. We have these data for the years 2000 and 2005. We collect the geographical coordinates of Kompass firms and identify their primary bank. We then match each Kompass firm to all the branches of their primary lender (using BEPS II information) and calculate the distance from the firm to each of these branches. We assume that firms borrow from the nearest branch of their primary bank and use this nearest distance as the *Firm-branch distance*. The median distance between a firm and its primary bank is 1.8 km (Table 1).

[15] We have this information for just over half of the banks in our sample. Beck et al. (2018a) use credit registry data to show that when CEOs consider relationship lending to be very important, according to BEPS II, this is indeed reflected in the lending practices of their bank.

[16] A score of one is assigned for each of six features: both positive credit information (outstanding loans and on-time repayments) and negative information (late payments and defaults) are distributed; data on both firms and individuals are distributed; data from retailers, utility companies, and banks are distributed; more than two years of historical data are distributed; data on loan amounts below one percent of income per capita are distributed; and by law borrowers have the right to access their data in the largest credit bureau or registry.

Kompass also allows us to create four dummy variables as proxies for firms' opaqueness: whether the firm has a publicly available email address (*Has email address*); whether the firm has a tax number (*Has tax number*); whether the firm has formal opening/working hours (*Has formal opening hours*); and whether the firm is at least ten years old (*Established for more than 10 years*). Table 1 shows that 60 percent of all firms have a publicly available email address, almost 74 percent of them have an official tax number, 74 percent of the firms operate on the basis of formal opening hours, and 74 percent is at least ten years old.

4 Identification

To test our hypotheses, we apply a difference-in-differences-in-differences (DDD) framework in which (i) the treatment (the presence of information sharing) varies across countries and years and (ii) localities within countries are affected differentially depending on the pre-existing bank branch structure. Because our treatment is introduced in a staggered fashion, and because treatment effects may be heterogeneous across countries, a standard two-way fixed effects framework can yield biased estimates (Goodman-Bacon, 2018).

To address this issue, we follow Cengiz et al. (2019) and create event-specific data sets. Each event includes all observations from the countries in which information sharing (the treatment) is introduced in the same calendar year as well as the observations from all clean control countries for a six-year panel by event time ($t=-3, \ldots, 2$) with information sharing introduced at $t=0$.[17] Clean control countries are those without an information sharing system during the full six-year event window. We stack these event-specific data sets to estimate a single average DDD result. Aligning events by event time instead of calendar time is equivalent to a setting where all events happen simultaneously. This approach avoids biases due to the negative weighting of some events (which can occur in a staggered design) or due to heterogeneous treatment effects (Goodman-Bacon, 2018). We can now compare how, within the same country,

[17] In addition to the nineteen countries in our data set, we also have bank branch data for Hungary, Lithuania, and Slovenia. However, because these countries already introduced information sharing in 1994 or 1995, we cannot include them in our analysis, as no pre-treatment data are available.

the introduction of information sharing differentially affects branch openings across localities with different numbers of pre-existing branches from other banks. To test Hypotheses 1 and 2, we estimate the linear probability model:

$$\begin{aligned} \text{Branch opening}_{ijct} = {} & \beta_1 \cdot \text{Information sharing}_{ct} \cdot \text{No. branches other banks}_{ijc} \\ & + \beta_2 \cdot \text{No. branches other banks}_{ijc} + \Phi_{jt} + \Phi_{it} + \Phi_{ij} + \epsilon_{ijct} \end{aligned} \tag{6}$$

The dependent variable *Branch opening*$_{ijct}$ is either *New branch opening*$_{ijct}$ (Hypothesis 1) or *First branch opening*$_{ijct}$ (Hypothesis 2). These are dummies that equal one if bank i opens a new (first) branch in locality j of country c in year t, and equals zero otherwise. *Information sharing*$_{ct}$ is a dummy that equals one if banks in country c share borrower information in year t, and that equal zero otherwise (the level effect of this variable is absorbed by the fixed effects). *No. branches other banks*$_{ijc}$ measures the number of pre-existing branches by banks other than bank i in locality j, measured at $t = 0$.[18] Based on our model, we expect β_1 to be positive as the introduction of information sharing decreases concentration and induces banks to cluster to attract more borrowers.

We gradually saturate the model with three types of high-dimensional fixed effects: *locality*year* (Φ_{jt}); *bank*year* (Φ_{it}); and *bank*locality* (Φ_{ij}).[19] *Locality*year* fixed effects absorb all time-varying and time-invariant historical, social, economic and cultural differences across villages, towns or cities. Importantly, this includes local trends in credit demand that may affect the location choice of banks. These fixed effects also wipe out local variation in labor markets or in the available IT infrastructure. Moreover, they remove the between-locality variation in the variable *No. branches other banks*$_{ijc}$, leaving the estimation of β_1 and β_2 to rely on *within-locality* differences in the number of competing branches. Thus, the identification is based on comparing the differential response to information sharing of 'small' banks in a locality—those facing a large number of rival branches—with 'large' banks that have relatively

[18] We fix the number of branches at $t = 0$ so we do not condition on a variable affected by the treatment.

[19] We also interact these fixed effects with cohort indicators, which is more conservative than including the fixed effects on their own (Gormley and Matsa, 2011).

few competitors. A positive β_1 can then be interpreted as evidence in support of hypothesis H1a, the concentration effect.

*Bank*year* fixed effects account for time variation in banks' strategies and financial health that affect their branch network as a whole. Identification of β_1 and β_2 then exploits *cross-locality variation* in rival branches for the same bank, thus focusing on *where* a bank opens a branch after controlling for its average response to information sharing. Therefore, including *Bank*year* fixed effects allows us to analyze the presence of clustering effects (H1b).[20]

Finally, *Bank*locality* fixed effects absorb time-invariant variation across banks in each locality. The variation we capture when including all fixed effects is thus at the bank*locality*year level and picks up whether smaller banks tend to open new branches in larger markets with more existing other bank branches after the introduction of information sharing. ϵ_{ijct} is the error term and we cluster standard errors at the *country*treatment event* level.[21]

We next investigate whether the relationship between information sharing and branch clustering is more pronounced when the quality of information sharing is higher (Hypothesis 3). The time-varying variable *Quality information sharing*$_{ct}$ measures the rules and practices affecting the accessibility, coverage, scope, and quality of the borrower information that is publicly available. Augmenting the base regression with this variable renders:

$$\begin{aligned}\text{New branch opening}_{ijct} = {} & \beta_1 \cdot \text{Information sharing}_{ct} \cdot \text{No. branches other banks}_{ijc} \\ & + \beta_2 \cdot \text{Quality information sharing}_{ct} \cdot \text{No. branches other banks}_{ijc} \\ & + \beta_3 \cdot \text{No. branches other banks}_{ijc} + \Phi_{jt} + \Phi_{it} + \Phi_{ij} + \epsilon_{ijct}\end{aligned} \tag{7}$$

Quality information sharing$_{ct}$ is by definition only available for country-years in which banks exchange borrower information (that is, when *Information sharing*$_{ct}$ equals one). It equals zero if there is no information sharing in a specific country and year. Based on our theoretical model,

[20]By including both types of fixed effects, we estimate the combined effect of information sharing on clustering and concentration.

[21]We use linear models so we can fully saturate them with the aforementioned fixed effects. Including these fixed effects in a non-linear (e.g., probit) model would introduce an incidental parameters problem: parameter estimates would not converge to their true value as the number of parameters grows with the number of observations.

we expect β_2 (and β_1) to be positive.

Lastly, to test Hypothesis 3, we examine whether information sharing differentially affects relationship versus transaction lenders. We do so by further interacting our treatment with *Bank type* in the following model:

$$\begin{aligned}\text{New branch opening}_{ijct} = {} & \beta_1 \cdot \text{Information sharing}_{ct} \cdot \text{No. branches other banks}_{ijc} \\ & + \beta_2 \cdot \text{Information sharing}_{ct} \cdot \text{No. branches other banks}_{ijc} \cdot \text{Bank type}_{ic} \\ & + \beta_3 \cdot \text{No. branches other banks}_{ijc} \cdot \text{Bank type}_{ic} \\ & + \beta_4 \cdot \text{No. branches other banks}_{ijc} + \Phi_{jt} + \Phi_{it} + \Phi_{ij} + \epsilon_{ijct}\end{aligned} \tag{8}$$

Bank type is one of three time-invariant proxies for whether a bank is a relationship lender: a small bank dummy, a domestic bank dummy, or a dummy for whether relationship lending is the main technique when lending to small businesses. Based on our model, we expect information sharing to have a bigger impact on relationship than on transactional lenders so that especially relationship lenders start to open branches in localities with more pre-existing branches of other banks. That is, we expect both β_1 and β_2 to be positive.

5 Empirical results

5.1 Baseline results

Table 2 presents regression results based on linear probability model (6) to investigate our first hypothesis. The dependent variable is *New branch opening*, indicating whether a bank opens a branch in a particular locality in a particular year. We start in column 1 with our basic difference-in-differences-in-differences setting without any fixed effects. We then gradually saturate the model with granular fixed effects that capture unobserved variation at various levels, including changes in local credit demand, which might otherwise bias results.

Overall, and in line with our first hypothesis, the estimates show that when a country introduces information sharing, banks become more likely to open new branches in localities with more pre-existing branches of competitor banks. Column 2, which (only) includes *Bank*Year* fixed effects, focuses on within-bank variation across localities in the same year. This allows us to zoom in on where a bank opens a new branch in a given year, after controlling for its average response to the introduction of information sharing. We find evidence for a clustering effect due to information sharing.

In column 3, we (only) include *Locality*Year* fixed effects. These absorb all time-varying historical, social, economic, and cultural differences across localities, including local trends in credit demand that may affect the location choice of banks, and wipe out local variation in labor markets or in the available IT infrastructure. These fixed effects allow us to compare how different banks—with different numbers of pre-existing branches in the same locality in the same year—react differentially to the introduction of information sharing. We find clear evidence for a concentration effect due to information sharing.

Next, column 4 combines both sets of fixed effects and hence estimates a combined clustering and concentration effect. In column 5, we saturate this specification further with *Locality*Bank* fixed effects. Doing so absorbs all time-invariant variation across banks in each locality, and the possibility that our results are driven by any unobservable differences across bank-locality pairs. These most saturated specifications reveal that the overall impact of establishing a credit registry or credit bureau is economically significant. The estimated coefficient in column 5 indicates that with information sharing in place, a one standard deviation higher number of pre-existing bank branches in a locality increases the probability that a bank opens an additional new branch in that locality by 33 percent.[22]

The last column of Table 2 (column 6), includes three additional interaction terms between *No. branches other banks* and other country-level reforms. Our sample countries went through a process of economic and political transformation after 1989. One may hence worry that our information sharing treatment partially picks up other reforms as well. We note though that

[22]When we assess the separate impact of the introduction of credit registries versus credit bureaus, we find that both types of information sharing influence bank branch clustering similarly. We present these results in Appendix Table A4.

much of the structural reform agenda was heavily concentrated in the first decade of transition (EBRD, 2013)—that is, *before* countries introduced information sharing (see Section 3.1 and Table A2).

To address this concern formally, we include three additional interactions with the local number of pre-existing bank branches. The first variable we interact with is a dummy that is one if a country is a member of the European Union in a particular year and is zero otherwise. According to the principle of single authorization, a bank authorized to operate in one EU country can provide its services throughout the Single Market. Acceding the EU may therefore expose a country to foreign-bank entry and an associated change in branch clustering dynamics.

Second, we control for progress with setting up effective competition policies. We take the EBRD Transition Indicator for competition policy, which ranges between 1 ("no competition legislation and institutions") and 4+ ("Standards and performance typical of advanced industrial economies: effective enforcement of competition policy; unrestricted entry to most markets"). Enhanced competition policy may change the clustering of real economic activity and, via demand effects, influence the clustering of the supply of financial services.

Third, we interact with the EBRD Transition Indicator for small-scale privatization. This indicator also ranges between 1 ("Little progress") and 4+ ("Standards and performance typical of advanced industrial economies: no state ownership of small enterprises; effective tradability of land"). Progress with small-scale privatization makes SME lending more attractive and can therefore have an independent impact on banks' branching decisions.

The interaction between information sharing and the pre-existing locality-level number of branches remains statistically significant at the 1 percent level in column 6. Moreover, controlling for the impact of EU membership, competition policy, and progress with small-scale privatization hardly makes a dent in the magnitude of the coefficient. Throughout all remaining tables, we will systematically control for these additional interaction terms (as indicated by the line "Controls Other Reforms" at the bottom of these tables).

Next, to investigate our second hypothesis, on extensive margin effects, Table 3 uses *First branch opening* as the dependent variable: a dummy that is one when a bank opens its first ever branch in a particular locality and year. We reshape our data in such a way that after a bank has opened a first branch in a locality, we no longer observe that bank*locality pair in the data. In this setup, we can no longer include locality*year fixed effects because all banks that do not have a branch face the same number of branches by competing banks. This specification therefore exploits the general within-bank variation across localities. We find that information sharing induces banks to open branches in localities where they did not yet operate themselves but where relatively many other banks were already present. The estimated coefficient in column 2 indicates that with information sharing in place, a one standard deviation higher number of pre-existing bank branches in a locality increases the probability that a bank opens its first ever branch in that locality by 1.6 percent.

To gain more insights into the dynamics at play, we conduct an event-study analysis where we define an event as the year in which a country introduces an information sharing regime. We present results for a six-year window around these events (the year of introduction is $t=0$). Figure 2 shows the pre- and post-trends for the probability of opening a new branch in localities with more pre-existing branches owned by other banks. All estimates are expressed as changes relative to event date $t=-1$ (the estimates for which we normalize to zero) and based on the most saturated specification in column 6 of Table 2.

Figure 2 reveals sharp changes in where banks open new branches at the time of the introduction of information sharing. Banks become more likely to open new branches in localities with more pre-existing branches from other banks. While the magnitude of the estimated effects gets smaller over time, the impact continues to be substantial and statistically significant three years out. This suggests that the introduction of information sharing pushes banks towards a durable new clustering equilibrium.

We observe some slight trends prior to the introduction of information sharing, although these leads are statistically insignificant and small relative to the post-treatment effect estimates. To investigate formally how robust our results are to deviations from the parallel trends assumption, we follow Rambachan and Roth (2023). We assume that the observed small pre-trend differences persist and can be extrapolated into the treatment period. Extrapolation can either be linear or we can allow the slope of the differential trend to change between consecutive periods by a parameter Z. Appendix Figure A1 reports the results. It shows the confidence set for the treatment parameter of interest: the number of pre-existing branches of other banks in a locality. The baseline DiD estimate is shown in red and the estimates that allow for extrapolated pre-trends are shown in blue. The first blue estimate corresponds to $Z=0$, which indicates a linear extrapolation of the pre-existing trend. Our results are robust to this extrapolation. Moving further to the right, we let Z increase to allow for deviations from the pre-trend during the treatment period. The estimates remain very stable.

Taken together, the sharp changes that we observe at $t=0$ in terms of where banks open new branches; the lack of substantial pre-treatment trends; the robustness of our results to extrapolating the limited pre-treatment effects into the treatment period; and the persistent post-treatment effects all support our research design.

Finally, as banks respond to the introduction of information sharing by opening branches in areas where relatively many other banks are already present, we expect that, at the same time, they become less inclined to open new branches in localities where they themselves already operate branches. Appendix Table A5 shows that this is indeed the case.

5.2 The quality of information sharing and branch clustering

We expect that the extent to which information sharing successfully eliminates distance thresholds, and thus fosters branch clustering and concentration, depends on how comprehensive and trustworthy the shared borrower information is (Hypothesis 3).

In Table 4, we now test whether in countries that introduce a particularly effective information sharing system, subsequent bank branch clustering is stronger. We can only measure the variable *Quality information sharing* in countries with information sharing in place; in countries without information sharing we set this variable to zero. From hereon we focus on the two regression specifications that are most saturated with interactive fixed effects.

We find, in line with our third hypothesis, that information sharing boosts branch clustering particularly in countries where the system is more effective. The results in column 2 indicate that an improvement of the registry quality by one point (out of six) increases branch clustering due to information sharing by 7.7 percent.

5.3 Information sharing, relationship lending, and clustering

Information sharing may not affect all banks equally. Our fourth hypothesis states that the impact of information sharing will be stronger among relationship banks as compared with transactional banks. In Table 5, we test this hypothesis by further interacting our interaction term—*Information sharing*No. branches other banks*—with the variable *Bank type*. *Bank type* is one of three proxies for a bank's reliance on relationship lending: whether the bank is relatively small (columns 1-2); whether it is domestically owned (columns 3-4); and whether its CEO finds relationship lending a very important technique to provide credit to small businesses (columns 5-6).

The first two columns of Table 5 show that while the introduction of information sharing increases the tendency of large banks to cluster and concentrate their branches, this impact is somewhat larger for small banks. To the extent that smaller banks rely more on relationship lending, this finding is therefore in line with our fourth hypothesis. Economically, when information sharing is introduced, a one standard deviation higher number of pre-existing branches of competitor banks in a locality increases the probability of a new branch opening by 34 and 38 percent for large and small banks, respectively. This difference does not simply reflect that small banks are more likely to open new branches. Instead, it shows that conditional on a new branch being opened, small banks are particularly likely to do so in a locality with more pre-existing branches once information sharing is introduced.

Next, in columns 3 and 4 of Table 5, we assess heterogeneity by bank ownership. As discussed in Section 2, some prior studies have proxied lending technology by bank ownership. The traditional dichotomy is that domestic banks are mostly relationship lenders while foreign banks rely more on transactional lending. Yet, we find no evidence for heterogeneous effects of information sharing by bank ownership. The triple interaction terms in columns 3 and 4 are small and imprecisely estimated. This null result is in line with Beck et al. (2018a) who find that the often (implicitly) assumed 'bank stereotype' that domestic banks are relationship lenders while foreign banks are transactional lenders, does not necessarily hold in reality—at least not in the emerging markets in their sample and in ours.

We then proceed by using our most direct measure of a bank's main lending technique when dealing with small businesses. The results in columns 5 and 6 of Table 5, show that while information sharing leads to more branch clustering among transactional lenders (as shown by the coefficient for the interaction term in the first line) the impact on relationship banks is even larger—again in line with our fourth hypothesis. However, in absolute terms this difference is limited, at about a half of the difference between small and large banks.

5.4 Robustness, placebo tests, and extensions

Instrumental variables regressions. One may worry that the introduction of information sharing is endogenous as it reflects unobservable national circumstances that also bear directly on branch clustering. However, country and time specific confounds are at least partly controlled for by our *Locality*Year* fixed effects. A related issue concerns reverse causality whereby the structure of a country's banking sector influences the (timing of) the introduction of information sharing. To alleviate this concern, we instrument the introduction of information sharing in a country-year with the percentage of all neighboring countries that introduced information sharing in the past five years (Martinez Peria and Singh, 2014). This instrumentation strategy builds on the notion that financial reforms tend to converge regionally (Abiad and Mody, 2005). The exclusion restriction is that the introduction of information sharing in nearby countries only has an impact on domestic bank clustering via an increase in the probability that information sharing is introduced domestically as well.

Because the *country*year*-level variable *Information sharing* gets absorbed by our interactive fixed effects, the endogenous variable in our most saturated baseline specification is the interaction term *Information sharing*Number branches other banks*. We make use of the fact that interactions of instruments with exogenous variables are valid instruments for endogenous variables interacted with exogenous variables (Wooldridge, 2002, p. 122). As first-stage instruments we therefore use interaction terms between the percentage of neighboring countries that introduced information sharing in the previous five years and a locality-level measure of the number of pre-existing branches of other banks.

Table 6 reports our IV results. The first stage (columns 1 and 2) shows a strong and positive correlation between the introduction of information sharing in neighboring countries in the recent past (interacted with the local pre-existing branch structure) and the introduction of a credit registry or bureau in the country of observation (similarly interacted). The second-stage estimates (columns 3 and 4) are comparable to our baseline results though slightly larger (especially in column 3, which excludes *Locality*bank* fixed effects). There are two main reasons why this may be the case. First, as discussed above, information sharing may have emerged later in countries with relatively strong branch clustering to begin with. Correcting for this endogenous treatment timing then increases the (IV) estimate. A second explanation is the Local Average Treatment Effect (LATE) when the impact of information sharing on branch clustering differs across countries. If information sharing has a larger impact on branch clustering in complier countries (that is, those countries where the introduction of a credit registry or a credit bureau had been delayed by a lack of 'example' information sharing systems in neighboring countries) than in non-complier countries, then the IV estimates will be larger than their OLS counterparts.

Information sharing and within-bank distance constraints. We expect that with the introduction of information sharing, it becomes easier for banks to operate branches in localities that are further away.[23] To empirically test whether this is indeed the case, we use a granular geographic approach and identify for each bank its core (i.e. headquarter) location. We then

[23]Oberfield, Rossi-Hansberg, Trachter and Wenning (2024) study the deregulation of large US banks in the 1990s and show that distance to headquarters is an important determinant of the evolution of bank branch locations during this wave of deregulation.

compute the number of branches of each bank with core location i in any other destination location j (*No. branches own bank*) in a given year. We also calculate the distance between core location i and each destination location j (*Distance to bank HQ*).

With these data in hand, we run gravity-style regressions where *No. branches own bank* is the dependent variable and *Distance to bank HQ* is the independent variable. We also include the interaction between *Distance to bank HQ* and *Information Sharing*, to see how the gravity coefficient changes with the introduction of information sharing. We follow the trade literature (Silva and Tenreyro, 2006) and use the Pseudo-Poisson Maximum Likelihood (PPML) estimator to estimate these gravity equations while handling the presence of many zeros. More specially, we use the estimator by Correia, Guimarães and Zylkin (2020), which allows for the inclusion of multiple high-dimensional fixed effects.

We present the results in Table 7. The first three columns present the baseline results, using increasingly saturated specifications in terms of fixed effects. As expected, we observe that, in general, banks are less likely to operate branches in localities that are further away from their central headquarters. Importantly, this "gravitational pull" weakens with the introduction of mandatory information sharing: banks become willing to open branches in more distant locations. In column 4, we also add two triple interaction terms: one with the size of the destination locality and one with the size of the core locality (we also fully saturate this model with all possible additional bilateral interaction terms but do not show the related coefficients). We find that the erosion of gravitational forces is smaller when destination localities are larger. This is intuitive, as it indicates that information sharing is especially useful in allowing banks to overcome distance constraints to localities that are further away and that are smaller. Lastly, we find that the reduction in gravitational pull is larger for banks headquartered in larger credit markets. Since banks are typically headquartered in the capital city, this is essentially a country-specific effect.

Credit bureaus versus credit registries. In Appendix Table A4, we distinguish between introducing a public credit registry versus a private credit bureau. We find similar effects for the opening of private bureaus and public registers, although the effect of private bureaus is somewhat larger. This is in line with work by Martinez Peria and Singh (2014), showing that (voluntary) private credit bureaus are more effective than (mandatory) public registries.

Net branch openings and per capita branch openings. Next, we replace our dependent variable, *New branch opening*, by *Net branch opening*. This dummy also takes the closure of local bank branches into account by measuring whether there is a net increase in the number of branches of a bank in a specific locality and year. We present the results in columns 1-2 of Appendix Table A6. They are very similar to our baseline results in Table 2, both statistically and economically.

In columns 3 to 6 of Table A6, we normalize the number of branches by the population of the locality, using data from the World Cities Database. We construct the variable *No. branches other banks per 1,000 population*. Our results go through. This similarity reflects the presence in our most saturated specifications of *Locality*Year* fixed effects, so that we effectively compare how different banks—with different numbers of pre-existing branches in the *same* locality (of a given population size) in the same year—react differentially to the introduction of information sharing at the country level. Note also that we were able to collect precise population data for about a quarter of all observations. It is reassuring that our results hold up well in this (non-random) sub-sample.

In Appendix Table A7, we replace the black box of our fixed effects with locality-level measures of the number of bank branches per 1,000 inhabitants, annual population growth, and local economic growth as proxied by the change in nighttime light intensity. When we explicitly control for these locality characteristics (instead of absorbing this variation through fixed effects) our results hold continue to go through. In the main tables, we keep our fully saturated baseline regressions in which fixed effects absorb all observable and unobservable variation. This also allows us to work with the largest possible sample.

Next, Table A8 explores simultaneous branch closures and openings. This analysis yields some interesting results. First, the last two coefficient rows in this table show consistently that *before* the introduction of information sharing, banks are less likely to open a new branch in a locality when they close another branch elsewhere in the same year. This indicates that in the absence of information sharing, banks use both margins (closures and openings) in parallel to expand (or shrink) their network. Interestingly, this pattern changes after the introduction of information sharing. The first two columns show that with information sharing in place, banks are more likely to open a branch in a new locality when they close a branch elsewhere in the same year (although this effect is imprecisely estimated in column 2). This holds true especially when banks close a branch in a relatively underbanked locality (that is, a locality in the bottom quartile in terms of bank density)—cf. columns 3 and 4.[24]

In columns 5 to 8, we test whether after the introduction of information sharing, banks are especially more likely to open a branch in an already heavily banked locality when they close a branch elsewhere. We find strong evidence for this, both when we look at branch closures in general (columns 5-6) and when we focus on branch closures in sparse banking markets (columns 7-8). This indicates that the increased clustering due to information sharing works as a double-edged sword: closures in less densely populated areas go hand-in-hand with new bank openings in densely banked areas.

Lastly, in Appendix Table A9, we assess whether there is a more general and direct impact of the introduction of information sharing on branch closings. This turns out not to be the case. Together with the results in Table A8, this indicates that while information sharing does not have a direct effect on the likelihood that a branch in more densely banked localities gets closed down, we do find that in case of such a closure, banks become more likely to replace such a branch with a new one in a more densely banked area.

Branch clustering in rural versus urban areas. A separate issue is that our findings could reflect branch clustering in specific parts of countries. A secular urbanization process can induce a disproportionate increase in the opening of new bank branches in urban areas. We

[24] We also test whether information sharing is associated with an increased likelihood that singleton branches close down (branches that are the only branch in a locality). We do not find that such solo branches are more likely to close down in general nor are they more likely to close down after the introduction of information sharing.

may then pick up clustering forces in urban areas that are largely unrelated to (but coincide with) the introduction of information sharing regimes. Figure 2, which shows *sharp* changes in clustering behavior right after the introduction of information sharing regimes, should already partly dispel concerns about gradual trends driving our estimates.

To look into this in more detail, we superimpose a grid consisting of 30x30 km cells on the 19 countries in our dataset. We use gridded population data from NASA (CIESIN, 2021) and calculate for each cell the local population density (number of inhabitants to the size of the cell) and the local bank branch density (number of bank branches to the size of the cell). Figure 3 provides a binned scatterplot that groups all grid cells into 20 equal-sized bins, based on their population density. We plot the mean population density (horizontal axis) and branch density (vertical axis) for each bin. We do this for the average in the (country-specific) years before (blue) and after (red) the introduction of information sharing. The lines reflect a linear OLS fit with 95 percent confidence intervals.[25]

Figure 3 shows that the introduction of information sharing is associated with an increase in bank branch density across the whole population density distribution, including in rural (less densely populated) areas. Yet, branch density increases most in densely populated areas, suggesting that in relative terms banks open more branches in larger agglomerations than in sparsely populated areas. Information sharing has enabled this relative shift as banks no longer need to be present "everywhere" but can now instead count on (potential) borrowers to travel to larger agglomerations with a deeper credit market.

Relatedly, Figure 4 explores the geo-spatial dimension of our data in more descriptive detail. We use Eurostat's Nomenclature of Territorial Units for Statistics (NUTS) hierarchy to divide each country into NUTS 1 'major socio-economic regions' and split these further into smaller NUTS 3 regions. We then take geographic boundary shape files for all NUTS 1 and the related NUTS 3 sub-regions and use the geo-coordinates of individual branches to match them to these (sub-) regions. We do this for every year, so that we get an annual spatial picture of where

[25]This binned scatterplot looks very similar when we partial out country fixed effects or use more granular grid cells at the 5x5 km level.

bank branches are located within the NUTS hierarchy.

For each NUTS 1 region, we then calculate a Herfindahl-Hirschman Index (HHI) as a measure of spatial concentration of bank branches across the NUTS 3 sub-regions in that NUTS 1 region. We calculate this HHI three years before and three years after the introduction of information sharing in the respective country. We also calculate, again at the NUTS 1 level and over the same six-year window, the percentage change in the share of branches located in the most densely banked NUTS 3 sub-region (that is, the NUTS 3 region with the largest share of all bank branches in that NUTS 1 region).

The dashed kernel density plot in Figure 4 (Panel A) shows the distribution of the percentage change in the NUTS 1 level HHI indices. Likewise, the solid plot shows the distribution of the percentage change in the share of all branches clustered in the most densely banked sub-region. It is striking that both distributions indicate how in the six-year window around the introduction of information sharing, bank branches tend to agglomerate more within individual NUTS 1 regions. The strengthening of financial agglomeration occurs soon after the introduction of information sharing (see also Figure 2) and is hence unlikely to reflect concomitant changes in the clustering of people or economic activity. Indeed, Panel B of Figure 4 shows clearly that there is no equivalent increase in the concentration of local population over the same period. This further underlines that the sharp increase in bank branch clustering around the time of the introduction of credit registries and bureaus indeed reflects sudden information-regime changes rather than gradual demographic shifts.

To investigate this issue more formally, we split our sample into localities with less than 50,000 inhabitants; localities with between 50,000 and 250,000 inhabitants; and localities with over 250,000 inhabitants.[26] We then rerun our baseline regression specifications on all three samples. Appendix Table A10 confirms that our estimates point to a somewhat stronger impact of information sharing in the largest localities. Yet, the impacts in more rural areas are highly significant and economically sizable too. We therefore conclude that our baseline findings do

[26] We use the World Bank-EBRD Business Environment and Performance Survey (BEEPS) to divide localities into these size buckets. This allows us to retain more observations as compared with the approach in Table A6 where we collect the exact population size of localities using the World Cities Database.

not mainly reflect secular urbanization trends.

Lastly, another way to directly investigate the impact of information sharing on branch closures and openings in localities of different size, is to divide all localities into quintiles based on the number of pre-existing branches. We do so and then run a regression like the one in column 6 of Table 2 while replacing the continuous variable *No. branches other banks* with four dummies that classify each locality according to its quintile in the cross-locality distribution of the number of bank branches. We take the middle quintile as a benchmark. Appendix Figure A2 shows that the impact of information sharing on branch closings is close to zero as estimated in each quintile (relative to the middle quintile). We therefore do not find that smaller localities lose out in terms of branch closings due to the introduction of information sharing.

We provide a similar exercise for branch openings. Interestingly, here we find that the impact of information sharing on branch openings increases monotonically across quintiles. The impact of the introduction of information sharing is much smaller in the lower quintiles (localities with fewer pre-existing bank branches). We visualize these results in Appendix Figure A3. This figure shows that because of the introduction of information sharing, banks are more likely to open new branches in areas with more pre-existing branches and that this positive relationship is driven by larger localities. In fact, in smaller localities, the introduction of information sharing actually somewhat reduces the likelihood that small banks expand their local presence when they get access to publicly available borrower information.[27] The fact the coefficient on lower quintiles is negative, suggests that, at least in terms of new branch openings, there is some zero-sum impact.[28]

This analysis of quintile effects on branch closings and openings lines up nicely with the visual evidence we presented in Figure 3. In that figure, we did not see an absolute decline in branch density at any point of the population density distribution (the red line never gets

[27] We conduct F-tests to examine the statistical significance of the differences between the quintile-specific estimates. We find that the difference between the top (that is, the 5th) quintile and the bottom (1st) quintile, as well as the difference between the top (5th) quintile and the 2nd quintile are both statistically significant at the 1 percent level (p-values are 0.01 in both cases).

[28] The results also indicate that the impact is mostly linear. Therefore, we use a continuous variable as the baseline approach throughout this paper.

below the blue one) while only towards the right of the population density distribution do we observe a significant increase in local branch density because of the introduction of information sharing. In other words, due to information sharing, smaller localities lose out in relative but not in absolute terms.

Controlling for additional reforms. In column 6 of Table 2, we showed that adding additional interactions between *No. branches other banks* and other types of reforms, leaves the main coefficient largely unchanged. Appendix Table A11 shows our results are also robust to adding these variables one-by-one in the first three columns. Moreover, in column 4, we employ a measure of how pro-competitive bank regulation is. It measures whether the government allows the entry of new domestic banks; whether there are restrictions on bank branching; and whether the government allows banks to engage in a wide range of activities. This variable therefore provides a direct gauge of whether governments constrain banks' branching decisions in a top-down manner. Where and when such constraints bind less, it is easier for banks to optimize their branching decisions, including in response to the introduction of new information sharing regimes.

When we add an interaction term with this variable in columns 4 and 5, our coefficient of interest declines by about two-thirds. Yet, this merely reflects that we do not observe this variable for all countries: moving from column 3 to columns 4 and 5, we experience a drop in the number of observations of 22 percent. If we re-estimate this specification on the same smaller sample but without the additional interaction term, the coefficient remains the same. This indicates that the decline in coefficient when moving from column 3 to 4 is due to the reduction in sample size resulting from certain countries dropping out of the sample.

Placebo test. We conduct a placebo test in which, within each treatment event, we randomize the countries that introduced information sharing. For example, the 2001 treatment event consists of all countries that introduced information sharing in 2001 (the real treated) as well as all countries that did not have or introduce information sharing in the six-year window around 2001 (the clean controls). Suppose the number of real treated countries in 2001 is three. We then randomly pick three placebo treatment countries from the set of all real treated and clean controls in the 2001 event sample. We do this for each event, stack the resulting randomized event samples, and rerun our baseline regressions (column 6 of Table 2) to estimate the coefficient of our interaction terms of interest. We repeat this process 500 times and plot the distribution of the point estimates for these placebo treatments in Appendix Figure A4. The vertical red line indicates the 95$^{\text{th}}$ percentile of this distribution. Reassuringly, we find that the real coefficient estimate from Table 2 (0.242) lies outside the corresponding distribution of the placebo treatment coefficients.

Information sharing and geographical credit rationing. One model implication that we have not yet tested is that information sharing reduces spatial credit rationing as firms will be able to borrow from more distant branches. We present some suggestive empirical evidence in this direction by merging our branch data with information from the Kompass database on firm-bank relationships. We assume that firms borrow from the nearest branch of their primary bank and use this nearest distance as the *Firm-branch distance* in kilometers.

There are four Kompass countries that introduced information sharing between 2000 and 2005 and that are also included in our BEPS data: the Czech Republic, Estonia, Latvia, and Poland. Because the bank information in Kompass and in BEPS can only be matched poorly for Estonia and Latvia, we focus on the Czech Republic and Poland. These countries introduced information sharing in 2002 and 2001, respectively. We also include two countries that did not introduce information sharing between 2000 and 2005. There are four such BEPS countries (Croatia, Hungary, Slovak Republic, and Ukraine) but because the matching of bank information is very poor for the Slovak Republic and Ukraine, we focus on the first two. We thus compare the change in firm-branch distance between 2000 and 2005 in two countries that introduced information sharing during this period (Czech Republic and Poland) with the change in firm-branch distance in two similar countries that did not (Croatia and Hungary).

The top panel of Appendix Table A13 shows summary statistics and a two-sample t-test with unequal variances. In countries that introduced information sharing between 2000 and 2005, firms on average borrow from more distant bank branches in 2005 than in 2000 (2 km and 8 km further for the Czech Republic and Poland, respectively). In contrast, firms do not borrow from more distant branches in comparator countries that did not introduce information sharing during this period (Croatia and Hungary). In the lower panel of Table A13, we analyze this in a difference-in-differences framework. We cluster standard errors by country using the wild cluster bootstrap-t to account for the small number of clusters (Cameron, Gelbach and Miller, 2008). Column 1 shows that after the introduction of information sharing, firms borrow from branches that are around 15 km further away as compared with firms in countries that did not introduce information sharing during these years.[29]

If sharing hard information reduces spatial credit rationing, thus allowing firms to borrow from more distant bank branches, then we expect this to be less important for firms that are relatively transparent. For these firms, agency problems are less of an issue and the new publicly available information may have less 'bite'. To test for this, we use the Kompass data to construct four dummy variables that proxy for a firm's overall transparency: whether the firm has a publicly available email address; whether it has an official tax number; whether it advertises formal opening hours; and whether it is at least ten years old.

Along all these dimensions, we expect firms to be more transparent to potential creditors and agency problems to be less severe. For example, the financial statements of firms with a formal tax number will convey more credible information to banks and may therefore facilitate access to credit (Gatti and Honorati, 2008). Likewise, firm age has been shown to correlate positively with access to credit. Theoretical work going back to Diamond (1989) argues that older firms suffer less from informational asymmetries as they have accumulated reputation over time. We then use these proxies to construct triple interaction terms with *Information sharing*. Each model is fully saturated with additional (unreported) interaction terms between

[29]We compare the average distance between firms and their lender for a cross-section of firms in 2000 with that of a cross-section of firms in 2005. This average distance can increase faster in countries that introduce information sharing because existing borrowers switch to a new, more distant lender or because previously credit rationed borrowers now have access to a larger variety of (more remote) lenders.

the country and year fixed effects and the respective opaqueness proxy.

Columns 2-5 of Appendix Table A13 present the results. We find suggestive evidence that the effect of information sharing on the reduction in spatial credit rationing is about half the size for transparent as compared to more opaque firms. For instance, while the average effect of information sharing is an increase in the firm-bank distance of 15.1 km (column 1), column 2 shows that this effect is 19.2 km for more opaque firms (those without an email address) and only 11.3 km for less opaque firms (with an email address). Because of these differential impacts, opaque and less opaque firms partially converge in terms of the geographical radius within which they can successfully seek out attractive borrowing opportunities.

6 Concluding remarks

This paper provides a stylized model of credit market competition that helps to clarify how information sharing between banks influences branch clustering. Using this theoretical framework, we derive key predictions and test these by exploiting dynamic information on the geographical locations of bank branches. We use the introduction of information sharing regimes across a large number of countries, at different points in time, as plausibly exogenous shocks that shift the relative advantages and disadvantages of branch clustering.

In line with our theoretical priors, we show that information sharing has a positive impact on bank clustering and concentration and that this impact is stronger for relationship banks. We also find that after the introduction of information sharing, banks are more likely to establish new branches in localities where they themselves did not yet have a branch presence. Closures in sparsely banked localities go hand-in-hand with new bank openings in already densely banked areas. Due to these spatial shifts, we observe that bank branches start to agglomerate more within individual NUTS 1 regions. Lastly, we provide suggestive evidence indicating that due to these changes the average firm is able to borrow from more distant bank branches. In summary, information sharing makes it more important for banks to move closer to each other than to be closer to potential clients. We expect these findings to apply also in other emerging markets experiencing structural change, especially those with a Communist legacy such as China.

An important implication of our results is that banking markets become more homogeneous in terms of composition—as they are served by the same banks that now operate across the country—but less homogeneous in terms of size. While the public availability of hard information leads to further clustering of bank branches in well-served locations, we do not (yet) find strong evidence that (smaller and more rural) locations are losing out in absolute terms, at least not in the period and context we study. In relative terms, rural areas are loosing out, as they see little new bank branches open.

Assessing the real-economic impacts of such spatial variation in access to credit due to information sharing is a promising avenue for further research. It would be particularly interesting to study how the agglomeration of bank branches may widen regional disparities. For example, while older rural firms with an established operational and borrowing track record may benefit from information sharing (as they can now be screened by more distant bank branches), younger firms (without a track record) may lose out.

Moreover, information sharing may also have important indirect and second-round effects in the longer run that go beyond the immediate impact on branch clustering only. As banks adjust their branch footprint, firms in locations that see an increase in branch density will benefit from better access to credit at more advantageous terms. This can in turn stimulate local economic growth: financial agglomeration can shape real agglomeration. In the longer term, the associated higher demand for credit, including from retail customers, may feed back through the establishment of further branches. Understanding the longer-term economic impacts of information sharing across space and over time will be a fruitful area for future research.

References

Abiad, Abdul and Ashoka Mody (2005) "Financial reform: What shakes it? What shapes it?," *American Economic Review*, Vol. 95, No. 1, pp. 66–88.

Agarwal, Sumit and Itzhak Ben-David (2018) "Loan prospecting and the loss of soft information," *Journal of Financial Economics*, Vol. 129, No. 3, pp. 608–628.

Agarwal, Sumit, Brent W. Ambrose, Souphala Chomsisengphet, and Chunlin Liu (2011) "The role of soft information in a dynamic contract setting: Evidence from the home equity credit market," *Journal of Money, Credit and Banking*, Vol. 43, No. 4, pp. 633–655.

Allee, Kristian D and Teri Lombardi Yohn (2009) "The demand for financial statements in an unregulated environment: An examination of the production and use of financial statements by orivately held small businesses," *Accounting review*, Vol. 84, No. 1, pp. 1–25.

Barboni, Giorgia, Erica Field, and Rohini Pande (2022) "Rural banks can reduce poverty: Experimental evidence from 870 indian villages,"Technical report, Working paper.

Beck, Thorsten, Hans Degryse, Ralph De Haas, and Neeltje Van Horen (2018a) "When arm's length is too far: Relationship banking over the credit cycle," *Journal of Financial Economics*, Vol. 127, No. 1, pp. 174–196.

Beck, Thorsten, Vasso Ioannidou, and Larissa Schäfer (2018b) "Foreigners vs. natives: Bank lending technologies and loan pricing," *Management Science*, Vol. 64, No. 8, pp. 3792–3820.

Bennardo, Alberto, Marco Pagano, and Salvatore Piccolo (2015) "Multiple bank lending, creditor rights, and information sharing," *Review of Finance*, Vol. 19, No. 2, pp. 519–570.

Berger, Allen N., Leora Klapper, and Gregory F. Udell (2001) "The ability of banks to lend to informationally opaque small businesses," *Journal of Banking and Finance*, Vol. 25, No. 12, pp. 2127–2167.

Berger, Allen N., Nathan H. Miller, Mitchell A. Petersen, Raghuram G. Rajan, and Jeremy C. Stein (2005) "Does function follow organizational form? Evidence from the lending practices of large and small banks," *Journal of Financial Economics*, Vol. 76, pp. 237–269.

Berger, Allen N. and Gregory F. Udell (1995) "Relationship lending and lines of credit in small firm finance," *Journal of Business*, Vol. 68, No. 3, pp. 351–381.

——— (2006) "A more complete conceptual framework for SME finance," *Journal of Banking and Finance*, Vol. 30, No. 11, pp. 2945–2966.

Bircan, Cagatay and Ralph De Haas (2020) "The limits of lending? Banks and technology adoption across Russia," *Review of Financial Studies*, Vol. 33, No. 2, pp. 536–609.

Bofondi, Marcello and Giorgio Gobbi (2006) "Informational barriers to entry into credit markets," *Review of Finance*, Vol. 10, No. 1, pp. 39–67.

Bouckaert, Jan and Hans Degryse (2004) "Softening competition by inducing switching in credit markets," *Journal of Industrial Economics*, Vol. 52, No. 1, pp. 27–52.

——— (2006) "Entry and strategic information display in credit markets," *Economic Journal*, Vol. 116, No. 513, pp. 702–720.

Brown, Martin, Tullio Jappelli, and Marco Pagano (2009) "Information sharing and credit: Firm-level evidence from transition countries," *Journal of Financial Intermediation*, Vol. 18, No. 2, pp. 151–172.

Brown, Martin, Benjamin Guin, and Karolin Kirschenmann (2016) "Microfinance banks and financial inclusion," *Review of Finance*, Vol. 20, No. 3, pp. 907–946.

Cameron, A. Colin, Jonah Bt. Gelbach, and Douglas L. Miller (2008) "Bootstrap-based improvements for inference with clustered errors," *Review of Economics and Statistics*, Vol. 90, No. 3, pp. 414–427.

Canales, Rodrigo and Ramana Nanda (2012) "A darker side to decentralized banks: Market power and credit rationing in SME lending," *Journal of Financial Economics*, Vol. 105, No. 2, pp. 353–366.

Carbo-Valverde, Santiago, Timothy H. Hannan, and Francisco Rodriguez-Fernandez (2011) "Exploiting old customers and attracting new ones: The case of bank deposit pricing," *European Economic Review*, Vol. 55, No. 7, pp. 903–915.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019) "The effect of minimum wages on low-wage jobs," *Quarterly Journal of Economics*, Vol. 134, No. 3, pp. 1405–1454.

Cole, Rebel A., Lawrence G. Goldberg, and Lawrence J. White (2004) "Cookie cutter vs. character: The micro structure of small business lending by large and small banks," *Journal* *of Financial and Quantitative Analysis*, Vol. 39, pp. 227–251.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin (2020) "Fast Poisson estimation with high-dimensional fixed effects," *Stata Journal*, Vol. 20, No. 1, pp. 95–115.

De Haas, Ralph, Yevgeniya Korniyenko, Alexander Pivovarsky, and Teodora Tsankova (2015) "Taming the herd? Foreign banks, the Vienna Initiative, and crisis transmission," *Journal of Financial Intermediation*, Vol. 24, No. 3, pp. 325–355.

De Haas, Ralph, Liping Lu, and Steven Ongena (2023) "Close competitors? Bilateral bank competition and spatial variation in firms' access to credit," *Journal of Economic Geography*, Vol. 23, No. 6, pp. 1237–1271.

Dell'Ariccia, Giovanni, Ezra Friedman, and Robert Marquez (1999) "Adverse selection as a barrier to entry in the banking industry," *The RAND Journal of Economics*, pp. 515–534.

DeYoung, Robert, Dennis Glennon, Peter J Nigro, and Kenneth Spong (2019) "Small business lending and social capital: Are rural relationships different?," *Journal of Entrepreneurial Finance*, Vol. 21, No. 2, p. 4.

Diamond, Douglas W (1989) "Reputation acquisition in debt markets," *Journal of Political Economy*, Vol. 97, No. 4, pp. 828–862.

EBRD (2013) *Transition Report 2013: Stuck in Transition?*, London: European Bank for Reconstruction and Development.

Fonseca, Julia and Adrien Matray (2022) "The real effects of banking the poor: Evidence from Brazil,"Technical report, National Bureau of Economic Research.

Gatti, Roberta and Maddalena Honorati (2008) "Informality among formal firms: Firm-level, cross-country evidence on tax compliance and access to credit," *World Bank Policy Research Working Paper*, No. 4476.

Gehrig, Thomas and Rune Stenbacka (2007) "Information sharing and lending market competition with switching costs and poaching," *European Economic Review*, Vol. 51, No. 1, pp. 77–99.

Goodman-Bacon, Andrew (2018) "Difference-in-differences with variation in treatment timing," Working Paper 25018, National Bureau of Economic Research.

Gormley, Todd A and David A Matsa (2011) "Growing out of trouble? Corporate responses to liability risk," *Review of Financial Studies*, Vol. 24, No. 8, pp. 2781–2821.

Grossman, Sanford J. and Joseph E. Stiglitz (1980) "On the impossibility of informationally efficient markets," *American Economic Review*, Vol. 70, No. 3, pp. 393–408.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales (2004) "Does local financial development matter?" *Quarterly Journal of Economics*, Vol. 119, No. 3, pp. 929–969.

Hauswald, Robert and Robert Marquez (2003) "Information technology and financial services competition," *Review of Financial Studies*, Vol. 16, No. 3, pp. 921–948.

——— (2006) "Competition and strategic information acquisition in credit markets," *Review of Financial Studies*, Vol. 19, No. 3, pp. 967–1000.

He, Zhiguo, Jing Huang, and Jidong Zhou (2023) "Open banking: Credit market competition when borrowers own the data," *Journal of Financial Economics*, Vol. 147, No. 2, pp. 449–474.

Jappelli, Tullio and Marco Pagano (2002) "Information sharing, lending and defaults: Cross-country evidence," *Journal of Banking and Finance*, Vol. 26, No. 10, pp. 2017–2045.

Jayaratne, Jith and Philip E. Strahan (1996) "The finance-growth nexus: Evidence from bank branch deregulation," *Quarterly Journal of Economics*, Vol. 111, No. 3, pp. 639–670.

Ji, Yan, Songyuan Teng, and Robert M Townsend (2023) "Dynamic bank expansion: Spatial growth, financial access, and inequality," *Journal of Political Economy*, Vol. 131, No. 8, pp. 2209–2275.

Karapetyan, Artashes and Bogdan Stacescu (2014) "Information sharing and information acquisition in credit markets," *Review of Finance*, Vol. 18, No. 4, pp. 1583–1615.

Kittiakarasakun, Jullavut (2010) "Does location impact how banks make their lending decisions," in *Paper presented at the Southwestern Finance Association 2010 Annual Meeting*.

Kroszner, Randall S. and Philip E. Strahan (2014) "Regulation and deregulation of the US banking industry: Causes, consequences, and implications for the future," in *Economic regulation and its reform: what have we learned?*: University of Chicago Press, pp. 485–543.

Levine, Ross (2005) "Finance and growth: Theory and evidence," in Philippe Aghion and Stephen Durlauf eds. *Handbook of Economic Growth*, Vol. 1, Chap. 12, pp. 865–934.

Liberti, José María (2018) "Initiative, incentives, and soft information," *Management Science*, Vol. 64, No. 8, pp. 3714–3734.

Liberti, José María and Mitchell A. Petersen (2019) "Information: Hard and soft," *Review of Corporate Finance Studies*, Vol. 8, No. 1, pp. 1–41.

Martinez Peria, Maria Soledad and Sandeep Singh (2014) "The impact of credit information sharing reforms on firm financing?" *World Bank Policy Research Working Paper*, No. 7013.

Oberfield, Ezra, Esteban Rossi-Hansberg, Nicholas Trachter, and Derek T Wenning (2024) "Banks in space," *National Bureau of Economic Research Working Paper No. 32256*.

Padilla, A. Jorge and Marco Pagano (1997) "Endogenous communication among lenders and entrepreneurial incentives," *Review of Financial Studies*, Vol. 10, No. 1, pp. 205–236.

——— (2000) "Sharing default information as a borrower discipline device," *European Economic Review*, Vol. 44, No. 10, pp. 1951–1980.

Petersen, Mitchell A. and Raghuram G. Rajan (2002) "Does distance still matter? The information revolution in small business lending," *Journal of Finance*, Vol. 57, No. 6, pp. 2533–2570.

Rajan, Raghuram G. (1992) "Insiders and outsiders: The choice between informed and arm's-length debt," *Journal of Finance*, Vol. 47, No. 4, pp. 1367–1400.

Rambachan, Ashesh and Jonathan Roth (2023) "A more credible approach to parallel trends," *Review of Economic Studies*, Vol. forthcoming.

Robinson, Joan (1952) *The Rate of interest and Other essays*, London: MacMillan.

Roland, Gerard (2000) *Transition and Economics: Politics, markets and firms*: MIT Press.

Silva, JMC Santos and Silvana Tenreyro (2006) "The log of gravity," *Review of Economics and Statistics*, Vol. 88, No. 4, pp. 641–658.

Stein, Jeremy C. (2002) "Information production and capital allocation: Decentralized versus hierarchical firms," *Journal of Finance*, Vol. 57, No. 5, pp. 1891–1921.

Thanassoulis, John E. and Tamas Vadasz (2023) “The cost of banking with naivety and adverse selection,” *CEPR Press Discussion Paper No. 18171. https://cepr.org/publications/dp18171.*

Uzzi, Brian (1999) “Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing,” *American Sociological Review*, Vol. 64, No. 4, pp. 481–505.

Van Cayseele, Patrick, Jan Bouckaert, and Hans Degryse (1994) “Credit market structure and information sharing mechanisms,” in *Studies in Industrial Organization*, Dordrecht: Kluwer Academic Publishers, pp. 129–143.

Von Thadden, Ernst-Ludwig (2004) “Asymmetric information, bank lending and implicit contracts: The winner’s curse,” *Finance Research Letters*, Vol. 1, No. 1, pp. 11–23.

Wooldridge, Jeffrey M. (2002) *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Table 1
Summary Statistics

This table provides the number of observations, mean, median, standard deviation, minimum and maximum for all variables used in the analysis.

Variable	**Obs.**	**Mean**	**Median**	**St. Dev.**	**Min.**	**Max.**
*Bank and locality characteristics (bank*locality*year level)*						
New branch opening	833,916	0.040	0	0.195	0	1
Net branch opening	833,916	0.039	0	0.195	0	1
No. branches other banks	833,916	1.962	1.818	1.816	0.000	7.457
No. branches other banks per 1,000 population	200,274	0.149	0.128	0.108	0.000	0.587
No. branches own bank	376,884	1.246	1	4.154	0	209
Distance to HQ	376,884	0.249	0.210	0.179	0.000	1.120
First branch opening	833,916	0.031	0	0.172	0	1
New branch opening in top quartile	833,916	0.025	0	0.157	0	1
Closed branch	833,916	0.005	0	0.072	0	1
Closed branch in bottom quartile	833,916	0.003	0	0.056	0	1
*Locality characteristics (locality*year level)*						
No. bank branches per 1,000 population	200,274	0.142	0.121	0.100	0.000	0.625
Population growth rate	165,948	-0.232	0.000	1.306	-33.853	15.158
Night-time light growth rate	830,636	0.060	-0.006	0.639	-1.000	108.500
Size of destination locality	376,884	2.784	3	1.516	1	5
Size of core locality	376,884	1.590	2	0.571	1	5
Population density	14,308	5.596	5.440	1.151	1.714	7.430
*Country characteristics (country*year level)*						
Information sharing	213	0.254	0	0.436	0	1
Public credit registry	213	0.169	0	0.376	0	1
Private credit bureau	213	0.122	0	0.328	0	1
Quality information sharing	342	1.289	0	2.132	0	6
EU membership	342	0.167	0	0.373	0	1
Competition policy	342	2.226	2.33	0.699	1	3.67
Small-scale privatisation	342	3.701	4	0.683	1	4.33
Pro-competition bank regulation	121	2.802	3	0.641	0	3
Bank characteristics (bank level)						
Small bank	614	0.318	0	0.466	0	1
Domestic bank	614	0.430	0	0.495	0	1
Relationship bank	316	0.592	1	0.492	0	1
Bank characteristics (branch level)						
Small bank	56,555	0.104	0	0.262	0	1
Domestic bank	56,555	0.505	1	0.500	0	1
Relationship bank	38,439	0.446	0	0.497	0	1
Firm characteristics						
Firm-branch distance	14,308	15.447	1.809	45.266	0.010	443.515
Has email address	14,308	0.602	1	0.489	0	1
Has tax number	14,308	0.736	1	0.441	0	1
Has formal opening hours	14,308	0.743	1	0.437	0	1
Established for more than 10 years	14,308	0.735	1	0.441	0	1

Table 2

Information Sharing and the Clustering of Bank Branches

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variable measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p*-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening					
	(1)	(2)	(3)	(4)	(5)	(6)
Information sharing * No. branches other banks	0.017***	0.005**	0.236***	0.133***	0.244***	0.242***
	(0.000)	(0.047)	(0.000)	(0.000)	(0.000)	(0.000)
No. branches other banks	0.001	0.010***	-0.026*	-0.022***		
	(0.414)	(0.000)	(0.055)	(0.001)		
Information sharing	-0.012					
	(0.468)					
EU membership * No. branches other banks						-0.076
						(0.352)
Competition policy * No. branches other banks						-0.009
						(0.851)
Small-scale privatisation * No. branches other banks						0.075*
						(0.063)
Bank * Year Fixed Effects	No	Yes	No	Yes	Yes	Yes
Locality * Year Fixed Effects	No	No	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	No	No	No	No	Yes	Yes
R-squared	0.005	0.413	0.432	0.628	0.697	0.697
Observations	833,916	833,916	833,916	833,916	833,916	833,916

Table 3

Information Sharing and the Clustering of Bank Branches: First Branch Openings

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variable measures whether a bank opens its first branch ever in a locality in a year and the sample is cut off afterwards. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p* -values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	First branch opening	
	(1)	(2)
Information sharing * No. branches other banks	0.009***	0.012***
	(0.008)	(0.002)
No. branches other banks	0.013**	
	(0.016)	
Controls Other Reforms	Yes	Yes
Bank * Year Fixed Effects	Yes	Yes
Locality * Bank Fixed Effects	No	Yes
R-squared	0.558	0.710
Observations	424,795	424,795

Table 4

Quality of Information Sharing and the Clustering of Bank Branches

This table reports linear probability regressions to estimate the relationship between the quality of a country's information-sharing regime and bank branch clustering. The dependent variable measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p*-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening	
	(1)	(2)
Information sharing * No. branches other banks	0.124***	0.236***
	(0.004)	(0.000)
Quality information sharing * No. branches other banks	0.051***	0.077***
	(0.000)	(0.000)
No. branches other banks	-0.078**	
	(0.029)	
Controls Other Reforms	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes
Bank * Year Fixed Effects	Yes	Yes
Locality * Bank Fixed Effects	No	Yes
R-squared	0.628	0.697
Observations	833,916	833,916

Table 5

Information Sharing, Relationship Lending, and the Clustering of Bank Branches

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering by relationship lenders versus transaction lenders. The dependent variable measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p*-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening					
	(1)	(2)	(3)	(4)	(5)	(6)
Bank type →	Small banks		Domestic banks		Relationship banks	
Information sharing * No. branches other banks	0.131***	0.251***	0.128***	0.241***	0.116***	0.256***
	(0.001)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
Information sharing * No. branches other banks * Bank type	0.014***	0.026***	-0.000	0.001	0.007***	0.012**
	(0.002)	(0.000)	(0.991)	(0.704)	(0.006)	(0.012)
No. branches other banks * Bank type	-0.014***		-0.002***		-0.002*	
	(0.000)		(0.002)		(0.091)	
No. branches other banks	-0.081**		-0.078**		-0.099**	
	(0.026)		(0.030)		(0.015)	
Controls Other Reforms	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Bank * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	No	Yes	No	Yes	No	Yes
R-squared	0.628	0.697	0.628	0.697	0.702	0.754
Observations	833,916	833,916	833,916	833,916	592,383	592,383

Table 6

Information Sharing and the Clustering of Bank Branches: IV Results

This table reports IV regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variable in the first stage is an interaction term between a dummy variable indicating whether in a given year and country an information-sharing system is in place and a locality-level measure of the number of pre-existing branches of other banks. The instrument is an interaction term between the percentage of neighboring countries that introduced information sharing in the previous five years and a locality-level measure of the number of pre-existing branches of other banks. The dependent variable in the second stage measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

	First stage		**Second stage**	
Dependent variable →	Information sharing * No. branches other banks		New branch opening	
	(1)	(2)	(3)	(4)
% neighboring countries introduced information sharing	0.014***	0.229***		
* No. branches other banks	(0.000)	(0.000)		
Information sharing * No. branches other banks			0.229***	0.250***
			(0.000)	(0.000)
No. branches other banks	-0.216***		-0.028***	
	(0.000)		(0.000)	
F-stat	6351.57	3812.62		
Controls Other Reforms	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes	Yes	Yes
Bank * Year Fixed Effects	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	No	Yes	No	Yes
Observations	833,916	833,916	833,916	833,916

Table 7

Information Sharing and Within-Bank Distance Constraints

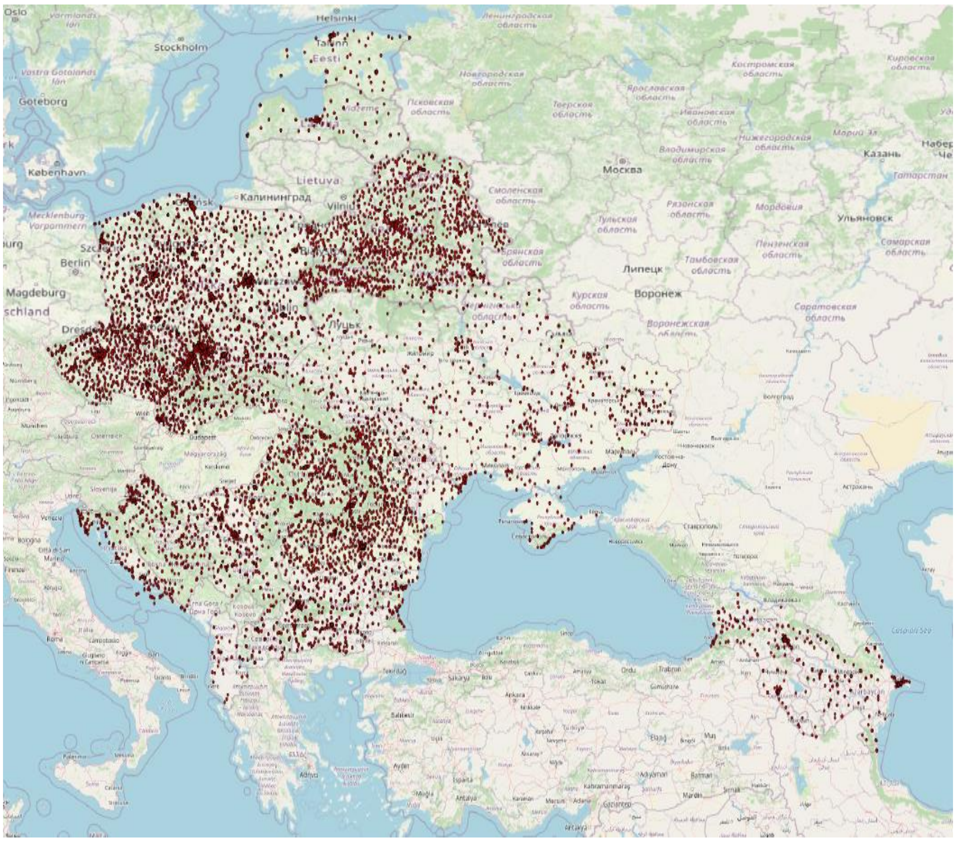
This table reports Pseudo-Poisson Maximum Likelihood (PPML) regressions to estimate how the number of a bank's own branches in a locality depends on (i) the distance between the bank's HQ and the locality, (ii) whether the country has an information sharing regime or not, and (iii) the interaction of these two variables. In column (4), we further interact with the size of the destination/core locality (all other bilateral interactions are included but not reported). Table A1 contains the definitions and Table 1 the summary statistics for all variables. Bank * Year-level cluster-robust *p*-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	No. branches own bank			
	(1)	(2)	(3)	(4)
Distance to bank HQ	-0.975***	-1.291***	-1.437***	-2.558***
	(0.000)	(0.000)	(0.000)	(0.000)
Information sharing * Distance to bank HQ	1.049***	0.607***	0.604***	1.142***
	(0.000)	(0.001)	(0.001)	(0.002)
Information sharing * Distance to bank HQ * Size destination locality				-0.411***
				(0.000)
Information sharing * Distance to bank HQ * Size core locality				0.576**
				(0.044)
Information sharing	0.463***	-0.015	-0.014	-0.253
	(0.000)	(0.866)	(0.824)	(0.129)
Year Fixed Effects	No	Yes	Yes	Yes
Country Fixed Effects	No	Yes	Yes	Yes
Bank Fixed Effects	No	No	Yes	Yes
R-squared	0.038	0.082	0.136	0.350
Observations	376,884	376,884	376,884	376,884

Figure 1

Distribution of Localities with Bank Branches in 1995 and in 2012

Panel A. This map plots all localities in our data set with at least one bank branch in 1995



Panel B. This map plots all localities in our data set with at least one bank branch in 2012

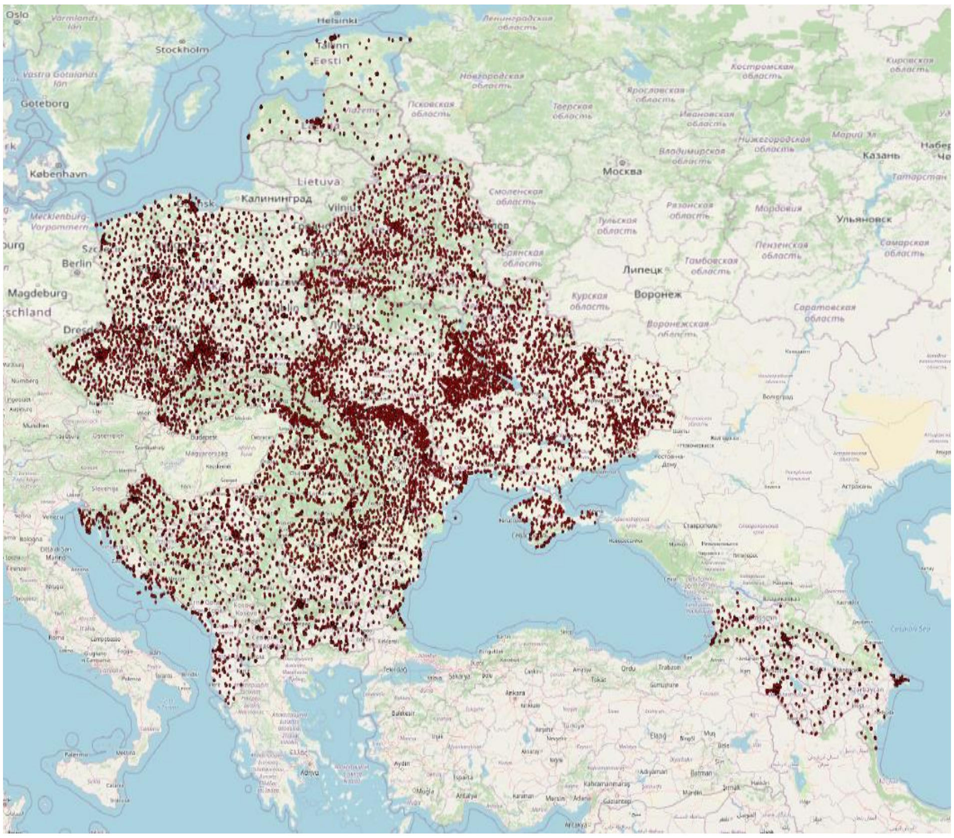


Figure 2

Information Sharing and the Clustering of Bank Branches: Event Study

This figure summarizes an event-study analysis in which a binary variable indicating whether a bank opens a new branch in a locality in a particular year is regressed on a set of year dummies around the introduction of information sharing in a country at $t=0$, each interacted with the number of pre-existing branches of other banks in the locality. One year before the introduction is omitted as the baseline. All coefficients are based on specifications with the same interactive fixed effects and covariates as in column 6 of Table 2.

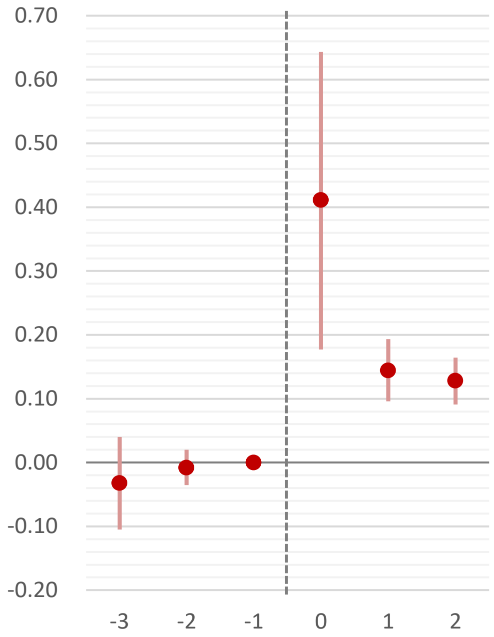


Figure 3

Population Density and Bank Branch Density Before and After the Introduction of Information Sharing

This binned scatter plot reflects spatial data at the 30x30 km grid level across 19 countries in Emerging Europe. For each grid cell, the local population density (number of inhabitants to the size of the cell) is calculated as well as the local bank branch density (number of bank branches to the size of the cell). The scatterplot groups all grid cells into 20 equal-sized bins, based on their population density, after which the mean population density (horizontal axis) and branch density (vertical axis) is plotted for each bin. This is done for the average in the years before (blue) and after (red) the introduction of information sharing. The lines reflect a linear OLS fit with 95% confidence intervals.

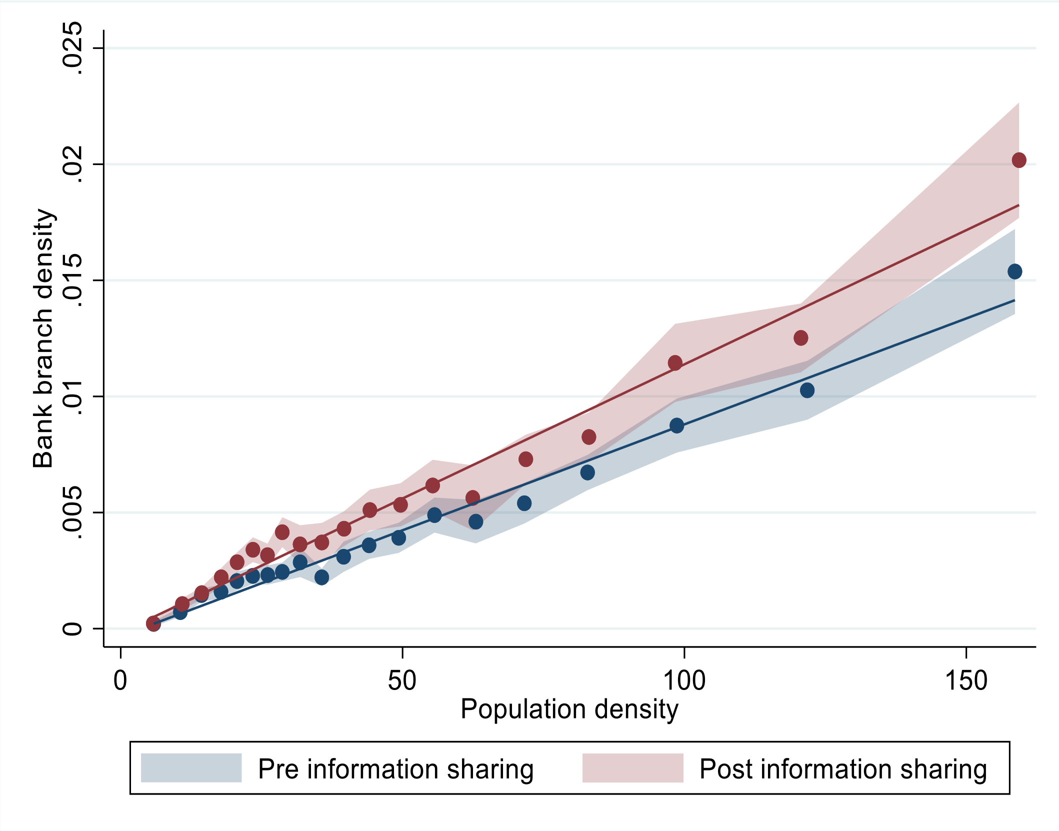
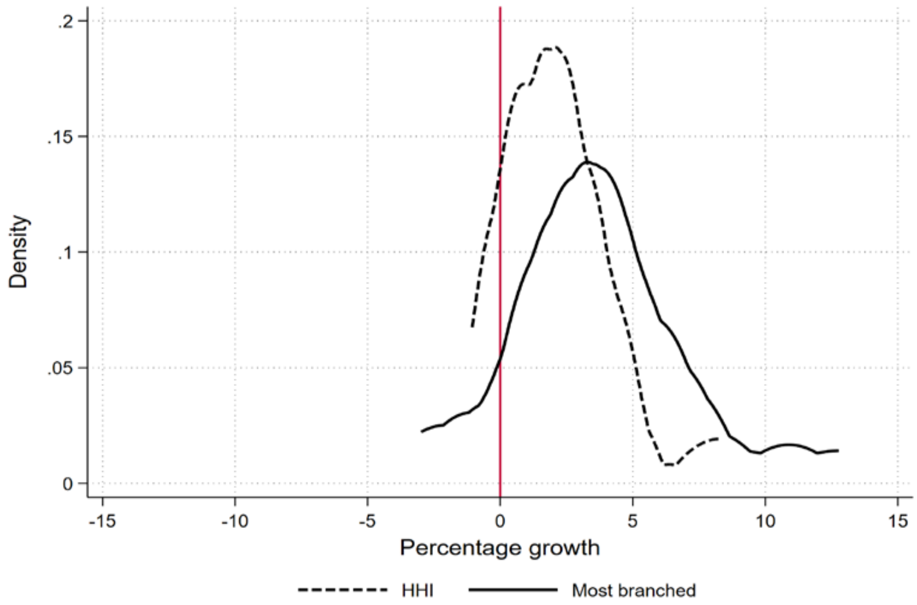


Figure 4

Information Sharing and the Clustering of Bank Branches at the NUTS 1 Level

Panel A summarizes the change in geo-spatial concentration of bank branches at the NUTS 1 level in the three years after as compared with the three years before the introduction of information sharing. Each country is divided into NUTS 1 'major socio-economic regions', each of which is further split into smaller NUTS 3 regions, using the official Eurostat Nomenclature of Territorial Units for Statistics (NUTS) hierarchy. For each NUTS 1 region, we calculate a Herfindahl-Hirschman Index (HHI) as a measure of spatial concentration of bank branches across the NUTS 3 sub-regions in that NUTS 1 region. We calculate this HHI three years before and three years after the introduction of information sharing in the respective country. The dashed kernel density plot shows the distribution of the percentage change in these NUTS 1 level HHI indices over this period. In a similar vein, the solid kernel density plot shows the distribution of the percentage change at the NUTS 1 level of the share of all branches clustered in the most densely banked NUTS 3 region. Panel B shows similarly the distribution of changes in the geo-spatial concentration of populations within NUTS 1 regions (data source: Eurostat).

Panel A



Panel B

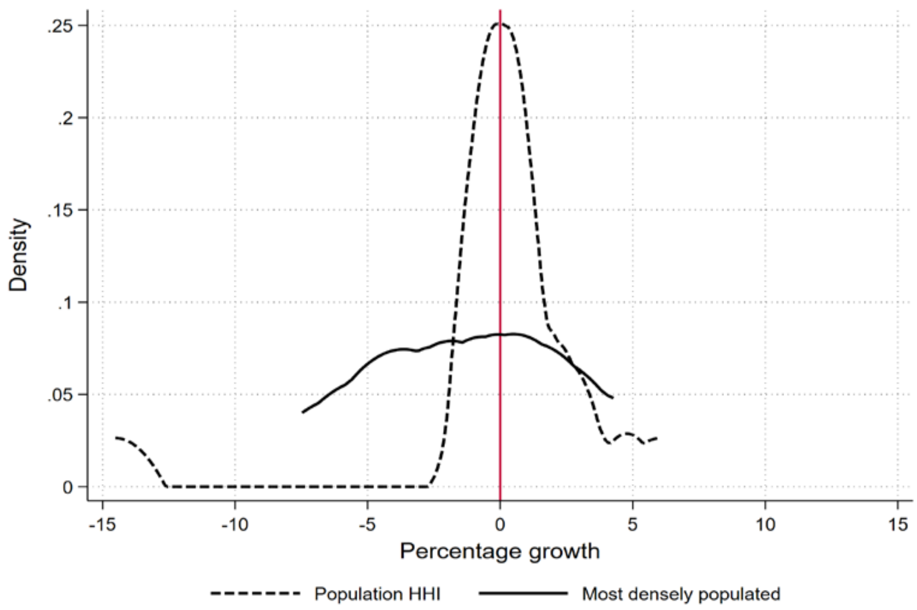


Table A1

Variable Definitions and Sources

This table provides the definition and data sources for all variables used in the analysis. BEPS II is the second round of the EBRD Banking Environment and Performance Survey (BEPS). "Doing Business" is the Doing Business Database by the World Bank. "Kompass" refers to the Kompass business directory. EBRD TI refers to the EBRD transition indicators. IMF FRD is the IMF Financial Reform Database. WCD is the World Cities Database. NOAA is the National Oceanic and Atmospheric Administration. NASA is the National Aeronautics and Space Administration.

	Definition	***Data Sources***
Bank and locality characteristics		
New branch opening	= 1 if there is any bank branch opening in a locality in a year, = 0 otherwise	BEPS II
Net branch opening	= 1 if the number of bank branch openings is larger than the number of bank branch closures in a locality in a year, = 0 otherwise	BEPS II
No. branches other banks	IHS transformed no. existing branches of other banks within a locality in the year when information sharing is introduced	BEPS II
No. branches other banks per 1,000 populati	IHS transformed no. existing branches of other banks within a locality per 1,000 inhabitants in year introduction information sharing	BEPS II, WCD
No. branches own bank	Number of existing branches of own bank within a locality in a year	BEPS II
Distance to HQ	Distance between a locality and the HQ locality in 1,000 km	BEPS II
First branch opening	= 1 if there is the first branch of a bank is opened in a locality in a year, = 0 otherwise	BEPS II
New branch opening in top quartile	= 1 if any branch opens in a locality belonging to the top 25% most densely banked localities in a country and year, = 0 otherwise	BEPS II
Closed branch	= 1 if a bank closed any branch in a country and year, = 0 otherwise	BEPS II
Closed branch in bottom quartile	= 1 if bank closed a branch in a locality belonging to the bottom 25% most densely banked localities in a country and year, = 0 otherwise	BEPS II
Locality characteristics		
No. bank branches per 1,000 population	IHS transformed no. existing branches of all banks within a locality per 1,000 inhabitants in year introduction information sharing	BEPS II, WCD
Population growth rate	Annual percentage change in the size of the population in a locality	WCD
Night-time light growth rate	Annual percentage change in the locally emitted night-time light as measured by satellite imagery	NOAA/NASA
Size of destination locality	= 1 to 5, indicating the quintiles of destination locality within a country on the number of bank branches	BEPS II, WCD
Size of core locality	= 1 to 5, indicating the quintiles of core locality with bank HQ within a country on the number of bank branches	BEPS II, WCD
Population density	Log population density within a 30x30 km grid	WCD
Country characteristics		
Information sharing	= 1 if there is information sharing (credit registry and/or credit bureau) in the country in that year, = 0 otherwise	World Bank/EBRD
Public credit registry	= 1 if there is public credit registry in the country in that year, = 0 otherwise	World Bank/EBRD
Private credit bureau	= 1 if there is private credit bureau in the country in that year, = 0 otherwise	World Bank/EBRD
Quality information sharing	= 0 to 6, higher values indicate a higher quality of information sharing in the country in that year	Doing Business
EU membership	= 1 if a country is part of the European Union in a particular year, = 0 otherwise	European Commissior
Competition policy	= 1 to 4+, higher values indicate that a country has created more market-based competition policies and institutions	EBRD TI
Small-scale privatisation	= 1 to 4+, higher values indicate more progress of a country in terms of the privatisation of small- and medium-sized enterprises	EBRD TI
Pro-competition bank regulation	= 0 to 3, higher values indicate fewer entry barriers in the banking sector of a country in a given year	IMF FRD
Bank characteristics		
Domestic bank	= 1 if more than 50% of a bank's shares are foreign-owned, = 0 otherwise	Bank websites
Relationship bank	= 1 if according to the bank CEO relationship lending is "very important" when providing credit to SMEs, = 0 otherwise	BEPS II
Small bank	= 1 if the no. branches of a bank is below the median no. branches operated by banks in a country and year, = 0 otherwise	BEPS II
Firm characteristics		
Firm-branch distance	Distance to the nearest branch of a firm's primary bank in km	Kompass
Has email address	= 1 if the firm has an email address, = 0 otherwise	Kompass
Has tax number	= 1 if the firm has a tax number, = 0 otherwise	Kompass
Has formal opening hours	= 1 if the firm has listed formal opening hours in Kompass, = 0 otherwise	Kompass
Established for more than 10 years	= 1 if the firm has been established for more than 10 years, = 0 otherwise	Kompass

Table A2

Introduction of Information Sharing

This table provides an overview of the introduction years of public credit registries and private credit bureaus in our 19 sample countries. N.a.: No credit bureau or registry has as yet been introduced in this country. Source: World Bank Doing Business Database and EBRD.

Country	Public Credit Registry	Private Credit Bureau
Albania	2008	2009
Armenia	2003	2004
Azerbaijan	2005	n.a.
Belarus	2007	n.a.
Bosnia & Herzegovina	2006	2001
Bulgaria	1999	2005
Croatia	n.a.	2007
Czech Republic	2002	2002
Estonia	n.a.	2001
Georgia	n.a.	2005
Latvia	2003	n.a.
Moldova	n.a.	2011
Montenegro	2008	n.a.
North Macedonia	1998	2010
Poland	n.a.	2001
Romania	2000	2004
Serbia	2002	2004
Slovak Republic	1997	2004
Ukraine	n.a.	2007

Table A3

Overview of Branch Openings and Closures

This table provides an overview of the opening and closure of branches in our dataset by year (left) and by country (right).

Year	Opened branches	Closed branches	Country	Opened branches	Closed branches
1995	2,388	0	Albania	443	11
1996	489	0	Armenia	448	19
1997	546	0	Azerbaijan	335	13
1998	525	0	Belarus	2,481	9
1999	543	0	Bosnia & Herzegovina	617	10
2000	974	6	Bulgaria	1,405	100
2001	1,361	3	Croatia	608	48
2002	1,389	7	Czech Republic	382	19
2003	2,571	9	Estonia	60	56
2004	4,307	34	Georgia	703	108
2005	2,122	20	Latvia	195	9
2006	2,535	19	Moldova	1,300	180
2007	7,833	61	Montenegro	206	12
2008	1,753	92	North Macedonia	189	16
2009	548	199	Poland	3,192	51
2010	709	223	Romania	2,053	177
2011	1,060	201	Serbia	1,080	227
2012	274	191	Slovak Republic	153	0
			Ukraine	16,077	0
Total	*31,927*	*1,065*	*Total*	*31,927*	*1,065*

Table A4

Information Sharing and the Clustering of Bank Branches: Public Credit Registries and Private Credit Bureaus

This table reports linear probability regressions to estimate the impact of the introduction of either a public credit registry or a private credit bureau on bank branch clustering. The dependent variable measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p*-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening				First branch opening			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Public credit registry * No. branches other banks	0.050*	0.140**			0.014*	0.022***		
	(0.082)	(0.018)			(0.087)	(0.006)		
Private credit bureau * No. branches other banks			0.145***	0.261***			0.006***	0.007**
			(0.000)	(0.000)			(0.009)	(0.014)
No. branches other banks	-0.110***		-0.068*		0.012**		0.013**	
	(0.000)		(0.070)		(0.028)		(0.018)	
Controls Other Reforms	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes	Yes	Yes	No	No	No	No
Bank * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.627	0.696	0.628	0.697	0.558	0.710	0.558	0.710
Observations	833,916	833,916	833,916	833,916	424,795	424,795	424,795	424,795

Table A5

Information Sharing and the Clustering of Bank Branches: Own Branches

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variable measures whether a bank opens a branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening	
	(1)	(2)
Information sharing * No. branches own bank	-0.094***	-0.105**
	(0.007)	(0.037)
No. branches own bank	0.035***	
	(0.000)	
Controls Other Reforms	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes
Bank * Year Fixed Effects	Yes	Yes
Locality * Bank Fixed Effects	No	Yes
R-squared	0.632	0.698
Observations	833,916	833,916

Table A6

Information Sharing, Net Branch Openings, and Branch Openings per 1,000 Inhabitants

This table reports regressions to estimate the impact of the introduction of information sharing on bank branch clustering using the Cengiz et al. (2019) methodology to address the potential concern in staggered treatment timing. The dependent variable in columns (1)-(2) measures whether on a net basis, a bank increases its number of branches in a locality in a year (the number of newly opened branches exceeds the number of closed branches). In columns (3)-(6) the number of existing bank branches is normalized by the local population in 1,000 persons and the dependent variable measures whether a bank opens a new branch (columns 3-4) or a first branch (columns 5-6) in a locality in a year. Table A1 contains all definitions and Table 1 the summary statistics for each variable. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p* -values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable →	Net branch opening		New branch opening		First branch opening	
	(1)	(2)	(3)	(4)	(5)	(6)
Information sharing * No. branches other banks	0.128***	0.242***				
	(0.002)	(0.000)				
No. branches other banks	-0.080**					
	(0.028)					
Information sharing * No. branches other banks per 1,000 population			1.730***	3.130*	0.099***	0.083*
			(0.007)	(0.070)	(0.000)	(0.071)
No. branches other banks per 1,000 population			-0.906***		-0.085***	
			(0.002)		(0.000)	
Controls Other Reforms	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes	Yes	Yes	No	No
Bank * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	No	Yes	No	Yes	No	Yes
R-squared	0.628	0.697	0.503	0.614	0.613	0.748
Observations	833,916	833,916	200,274	200,274	108,816	108,816

Table A7

Information Sharing and the Clustering of Bank Branches

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variable measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p*-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening				First branch opening			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Information sharing * No. branches other banks	0.023***	0.022***	0.018***	0.023***	0.024*	0.030**	0.001	0.027**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.059)	(0.021)	(0.964)	(0.041)
No. branches other banks	0.009	0.017**	0.007	0.018**	0.033	0.076***	0.019	0.076***
	(0.263)	(0.050)	(0.507)	(0.032)	(0.123)	(0.000)	(0.615)	(0.000)
Information sharing	-0.042***	-0.044***	-0.016	-0.045***	0.106	0.096	0.203	0.090
	(0.000)	(0.000)	(0.295)	(0.000)	(0.333)	(0.389)	(0.198)	(0.419)
No. bank branches per 1,000 population	-0.047			-0.054	0.184***			0.195***
	(0.384)			(0.371)	(0.000)			(0.037)
Population growth rate		-0.001		-0.001		-0.001		-0.001
		(0.404)		(0.386)		(0.494)		(0.649)
Night light growth rate			-0.004***	0.011**			-0.006**	0.021***
			(0.004)	(0.038)			(0.025)	(0.003)
Controls Other Reforms	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	No	No	No	No	No	No	No	No
Bank * Year Fixed Effects	No	No	No	No	No	No	No	No
Locality * Bank Fixed Effects	No	No	No	No	No	No	No	No
R-squared	0.021	0.022	0.010	0.023	0.050	0.054	0.036	0.058
Observations	200,274	165,948	830,636	165,528	108,816	87,876	423,244	87,668

Table A8

Information Sharing, Branch Closures, and the Clustering of Bank Branches

This table reports linear probability regressions to estimate the impact of the introduction of information sharing and branch closures on bank branch clustering. The dependent variable measures whether a bank opens a new branch in a locality (columns 1-4) or whether a bank opens a new branch in a locality ranked in the top quartile based on the total number of other bank branches in a year (columns 5-8). Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p*-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening				New branch opening in top quartile (no. branches)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Information sharing * Closed branch	0.078***	0.052			0.074***	0.046*		
	(0.005)	(0.116)			(0.001)	(0.066)		
Information sharing * Closed branch in bottom quartile (no. branches)			0.063***	0.046***			0.056***	0.038***
			(0.000)	(0.005)			(0.000)	(0.003)
Closed branch	-0.037***	-0.037***			-0.053***	-0.035***		
	(0.000)	(0.004)			(0.000)	(0.000)		
Closed branch in bottom quantile (no. branches)			-0.025***	-0.025***			-0.032***	-0.022***
			(0.001)	(0.001)			(0.000)	(0.001)
Controls Other Reforms	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank * Year Fixed Effects	No	No	No	No	No	No	No	No
Locality * Bank Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.428	0.537	0.523	0.642	0.428	0.537	0.523	0.642
Observations	833,916	833,916	833,916	833,916	833,916	833,916	833,916	833,916

Table A9

Information Sharing and the Closure of Bank Branches

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variable measures whether a bank closes a branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p* -values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	Branch closure	
	(1)	(2)
Information sharing * No. branches other banks	-0.001	-0.002
	(0.564)	(0.280)
No. branches other banks	-0.000	
	(0.991)	
Controls Other Reforms	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes
Bank * Year Fixed Effects	Yes	Yes
Locality * Bank Fixed Effects	No	Yes
R-squared	0.361	0.485
Observations	833,916	833,916

Table A10

Information Sharing and the Clustering of Bank Branches in Different Sized Towns and Cities

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering in localities with different population sizes. The dependent variable measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p* -values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

In cities with a population of:	*Less than 50,000*		*50,000 to 250,000*		*More than 250,000*	
Dependent variable →	New branch opening					
	(1)	(2)	(3)	(4)	(5)	(6)
Information sharing * No. branches other banks	0.136***	0.235***	0.135***	0.228***	0.213***	0.321***
	(0.002)	(0.002)	(0.002)	(0.000)	(0.003)	(0.000)
No. branches other banks	-0.111***		-0.060		0.156*	
	(0.002)		(0.337)		(0.073)	
Interacted Country-Level Reforms	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Bank * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	No	Yes	No	Yes	No	Yes
R-squared	0.625	0.677	0.591	0.654	0.572	0.681
Observations	234,790	234,790	98,766	98,766	95,777	95,777
Dependent variable →	First branch opening					
	(7)	(8)	(9)	(10)	(11)	(12)
Information sharing * No. branches other banks	-0.001	-0.000	0.024**	0.030***	0.022***	0.038***
	(0.881)	(0.980)	(0.030)	(0.006)	(0.010)	(0.001)
No. branches other banks	0.017		0.002		0.008	
	(0.149)		(0.801)		(0.184)	
Controls Other Reforms	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	No	No	No	No	No	No
Bank * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	No	Yes	No	Yes	No	Yes
R-squared	0.554	0.712	0.691	0.799	0.678	0.792
Observations	115,143	115,143	45,335	45,335	52,258	52,258

Table A11

Information Sharing and the Geographical Clustering of Bank Branches: Controlling for Other Country-Level Reforms

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variable measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening				
	(1)	(2)	(3)	(4)	(5)
Information sharing * No. branches other banks	0.245***	0.238***	0.241***	0.071***	0.074***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)
EU membership * No. branches other banks	-0.080				0.041
	(0.319)				(0.628)
Competition policy * No. branches other banks		0.072***			-0.051
		(0.009)			(0.449)
Small-scale privatisation * No. branches other banks			0.073**		0.080*
			(0.043)		(0.080)
Pro-competition bank regulation * No. branches other banks				-0.007	-0.000
				(0.580)	(0.993)
Locality * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Bank * Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes
R-squared	0.697	0.697	0.697	0.708	0.708
Observations	833,916	833,916	833,916	650,601	650,601

Table A12

Information Sharing and the Clustering of Bank Branches: Lagged Branch Structure

This table reports linear probability regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variable measures whether a bank opens a new branch in a locality in a year. Observations from all countries that introduce information sharing in the same calendar year are grouped together and combined with the observations from not (yet) treated (control) countries within a six-year window around the introduction year. Following Cengiz et al. (2019) these event-specific data sets are then stacked to estimate a single coefficient. Table A1 contains the definitions and Table 1 the summary statistics for all variables. All fixed effects are interacted with the event cohort dummy. Country-level cluster-robust *p* -values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening		First branch opening	
	(1)	(2)	(3)	(4)
Information sharing * No. branches other banks at *t=-1*	0.092***	0.166***	0.009***	0.012***
	(0.003)	(0.000)	(0.007)	(0.002)
No. branches other banks at *t=-1*	-0.071**		0.014**	
	(0.014)		(0.015)	
Controls Other Reforms	Yes	Yes	Yes	Yes
Locality * Year Fixed Effects	Yes	Yes	No	No
Bank * Year Fixed Effects	Yes	Yes	Yes	Yes
Locality * Bank Fixed Effects	No	Yes	No	Yes
R-squared	0.627	0.697	0.558	0.710
Observations	833,916	833,916	424,795	424,795

Table A13

Information Sharing and Spatial Credit Rationing

This table reports, by country, summary statistics for the variable *Firm-branch distance* and regressions to estimate the impact of the introduction of information sharing on spatial credit rationing. All diff-in-diff-in-diff regressions in the lower panel are fully saturated with additional (unreported) interaction effects between the year and country dummies and the firm characteristics. Standard errors are clustered by country and *p*-values based on the wild cluster bootstrap-t, which accounts for the small number of clusters (Cameron, Gelbach and Miller, 2008) are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable → Firm-branch distance (in km)

Czech Republic (Introduced information sharing in 2002)						**Poland (Introduced information sharing in 2001)**					
	Obs.	Mean	St. Dev.	5%	95%		Obs.	Mean	St. Dev.	5%	95%
2000	1,650	3.01	5.16	2.76	3.26	2000	5,286	19.13	56.57	17.60	20.65
2005	1,892	5.01	14.02	4.38	5.64	2005	1,242	27.22	68.88	23.38	31.05
2005-2000	2.00***					2005-2000	8.09***				
Croatia (Introduced information sharing in 2007)						**Hungary (Introduced information sharing in 1995)**					
	Obs.	Mean	St. Dev.	5%	95%		Obs.	Mean	St. Dev.	5%	95%
2000	953	16.65	48.97	13.54	19.77	2000	1,459	24.08	34.51	22.31	25.85
2005	409	20.92	47.43	16.31	25.53	2005	1,417	8.54	13.65	7.83	9.25
2005-2000	4.26					2005-2000	-15.54***				

Difference-in-Difference (-in-Difference) regression

	(1)	(2)	(3)	(4)	(5)
Information sharing	15.14***	19.15***	21.02***	19.48***	18.136***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Information sharing * Has email address		-7.89***			
		(0.000)			
Information sharing * Has tax number			-15.77***		
			(0.000)		
Information sharing * Has formal opening hours				-11.63***	
				(0.000)	
Information sharing * Established for more than 10 years					-4.663*
					(0.051)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes
R-squared	0.027	0.030	0.030	0.029	0.028
Observations	14,308	14,308	14,308	14,308	14,308

Population Density

	(6)	(7)
Population density	-2.288***	-1.429***
	(0.000)	(0.001)
Population density * Information sharing		-2.131***
		(0.001)
Information sharing		27.542***
		(0.000)
Year Fixed Effects	Yes	Yes
Country Fixed Effects	Yes	Yes
R-squared	0.025	0.031
Observations	14,308	14,308

Figure A1

Sensitivity to Linear and Non-Linear Deviations from Parallel Trends

This figure shows confidence sets for the treatment parameter of interest – the number of pre-existing branches of other banks in a locality – following Rambachan and Roth (2023). The original estimate is shown in red and the estimates allowing for deviations from pre-trends – fixed length confidence intervals – in blue.

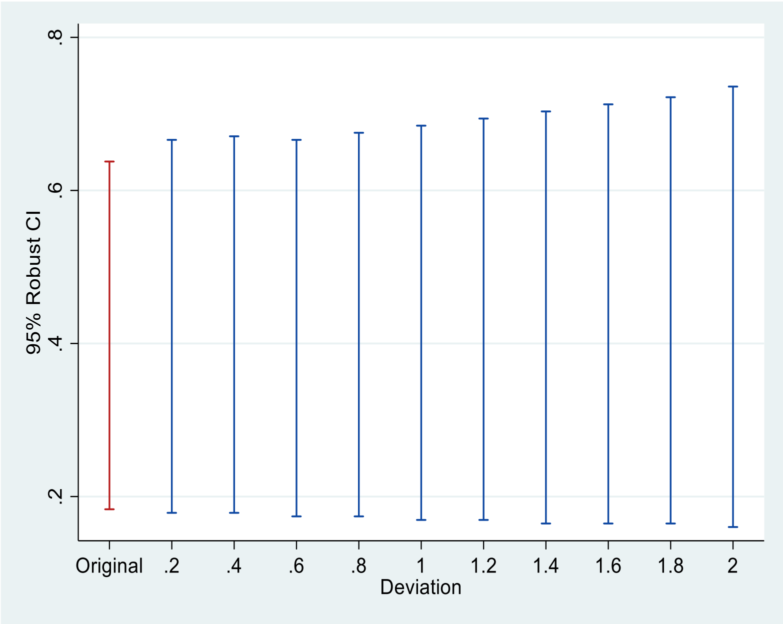


Figure A2

Information Sharing and the Closing of Bank Branches: Quintile Effects

This figure depicts the estimated coefficients for the interaction terms of a regression specification in which the dependent variable is a binary variable indicating whether a bank closes a branch in a locality in a particular year. This variable is regressed on four interaction terms between the information sharing dummy and four quintile dummies. Localities within a country are classified into quintiles based on *No. branches other banks*. The interaction with the middle quintile is omitted as the baseline. The regression also includes the same covariates as in column 6 of Table 2.

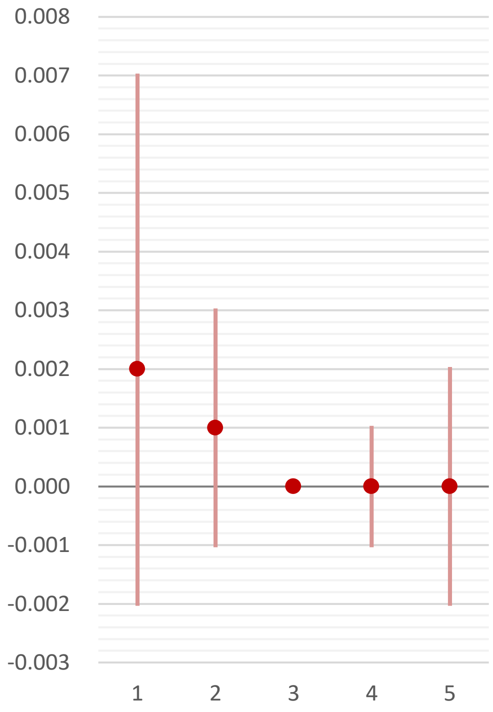


Figure A3

Information Sharing and the Clustering of Bank Branches: Quintile Effects

This figure depicts the estimated coefficients for the interaction terms of a regression specification in which the dependent variable is a binary variable indicating whether a bank opens a new branch in a locality in a particular year. This variable is regressed on four interaction terms between the information sharing dummy and four quintile dummies. Localities within a country are classified into quintiles based on *No. branches other banks*. The interaction with the middle quintile is omitted as the baseline. The regression also includes the same covariates as in column 6 of Table 2.

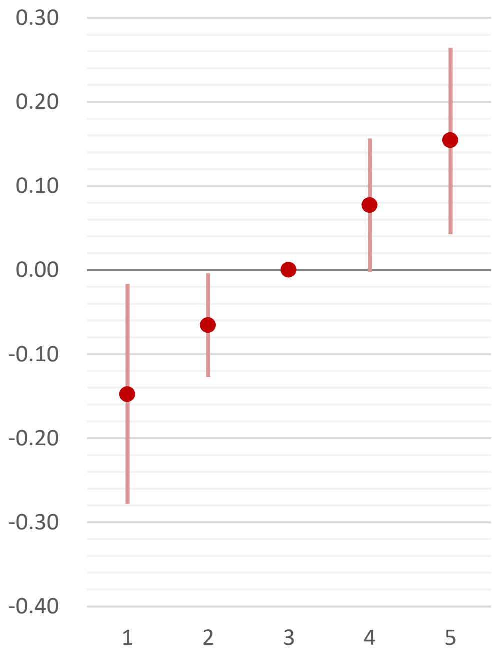
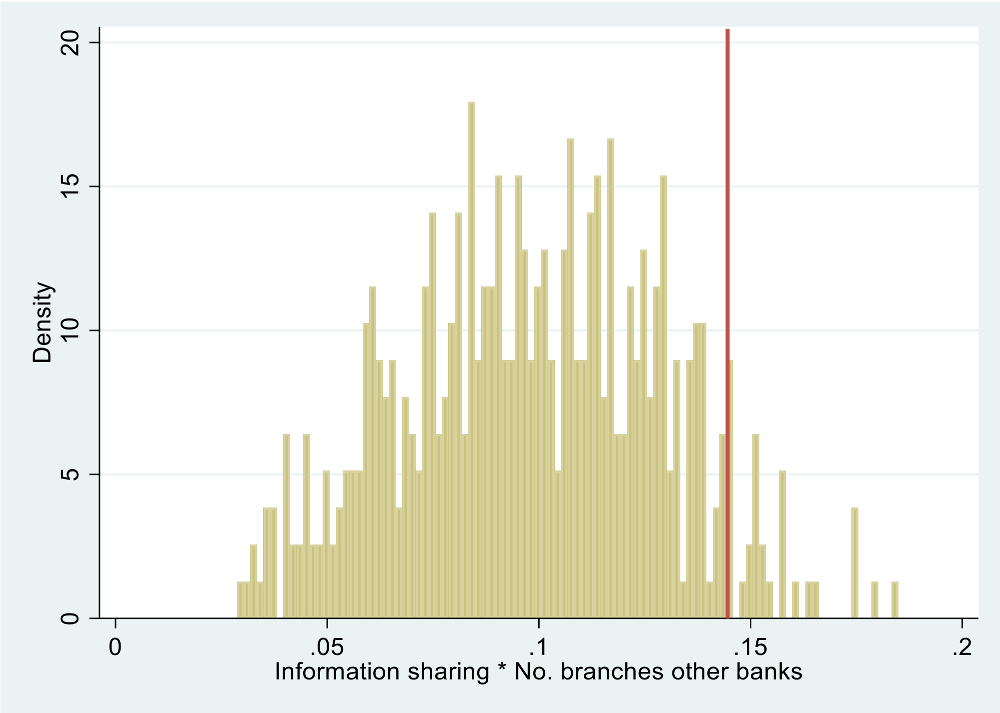


Figure A4

Information Sharing and the Clustering of Bank Branches: Placebo Treatment

This figure presents the results of placebo tests in which, within each treatment event, the countries that introduced information sharing are randomized. Using that randomized sample, the baseline regression (column 6 of Table 2) is rerun to estimate the coefficient estimate of the interaction term of interest. We repeat this process 500 times and plot the distribution of the point estimates for this placebo treatment. The vertical red line indicates the 95th percentile of this distribution. The real coefficient estimate from Table 2 (0.242) is outside the corresponding distribution of the placebo treatment coefficients.



Appendix A Online Appendix: Proofs

Appendix A.1 Proof of Proposition 1

Let ρ and μ denote the probability that a h-signal (resp. $\tilde{h}$-signal) is correct:

$$\rho := Pr[H|h] = \frac{q}{1-\phi(1-q)} \qquad \text{and} \qquad \mu := Pr[H|\tilde{h}] = \frac{q}{1-\nu(1-q)} \tag{A.1}$$

Without information sharing ($\nu = 0$) we obtain $Pr[H|\tilde{h}] = q$, the unconditional probability. Note that in this case $Pr[\tilde{\ell}] = 0$.

Consider first the competition between banks where all branches operate efficiently at cost level k=0. Let i and u index the relatively more 'informed' bank and the 'uninformed' banks, respectively. The banks' actions consist of a set of prices conditioned on the respective signals, $\{r(h), r(\ell), r(\tilde{h}), r(\tilde{\ell})\}$. We concentrate on symmetric equilibria, that is, all uninformed banks play the same strategy. By construction of the signal structure, whenever the informed bank receives an h-signal, the uninformed banks receive an $\tilde{h}$-signal. But whenever the informed bank receives an ℓ-signal, the uninformed banks may observe $\tilde{\ell}$ or $\tilde{h}$. Trivially, when the signal is ℓ or $\tilde{\ell}$, banks reject the customer and therefore make zero profit on the client.

The informed bank's strategy. It is standard that there is no pure-strategy Nash-equilibrium and that the pricing distributions for high-type borrowers are continuous with a common upper and lower boundary (e.g., Von Thadden (2004)). Let $r(h) \sim F_i(r)$ and $r(\tilde{h}) \sim F_u(r)$. When playing the lower boundary of the distributions, an uninformed bank wins with probability 1 and must break even. This defines the lower boundary of the price dispersion,[30] $\underline{r} := \frac{1}{\mu}$. When playing $\underline{r}$, the insider obtains the following per-customer expected profit on each h-type:

$$\pi_i(h) = \frac{\rho - \mu}{\mu} \tag{A.2}$$

The insider bank must be indifferent between any action which is played in equilibrium, so $\pi_i(h)$

[30] The uninformed banks' per-customer profit when winning at r with certainty is $\mu \cdot r - 1$.

is the insider bank's profit throughout her mixing distribution. Suppose that the number of banks (including the insider) the borrower approaches is $n \geqslant 2$. Then the insider's indifference condition is:

$$[1 - F_u(r)]^{n-1} \cdot (\rho r - 1) = \frac{(\rho - \mu)}{\mu}$$

where F_u is the CDF of the uninformed bank. Thich implies:

$$F_u(r) = 1 - \left[\frac{(\rho - \mu)}{\mu(\rho r - 1)}\right]^{1/(n-1)} = 1 - \left[\frac{(1-q)(\phi - \nu)}{\phi(1-q) - 1 + rq}\right]^{1/(n-1)} \tag{A.3}$$

The maximum repayment is R. Therefore, the uninformed banks deny credit with probability $1 - F_u(R)$. The total profit of the informed bank is the unit profit multiplied by the probability of receiving an h-signal. From (A.1) we have $Pr[h] = \frac{q}{\rho}$, so:

$$\pi_i = \frac{q}{\rho} \cdot \frac{\rho - \mu}{\mu}$$

Using (A.1) again we obtain the total profit function:

$$\pi_i = (1-q)(\phi - \nu) \tag{A.4}$$

When ν is low enough so that outsider does not play (that is, $\underline{r}(\nu) > R$), the insider serves h-types borrowers as a monopolist at R and obtains $\pi_i = \rho R - 1$ per each h-customer, and a profit of

$$\pi_i = \frac{q}{\rho}(\rho R - 1) = \phi(1-q) + (qR - 1). \tag{A.5}$$

Continuity is trivial at the critical $\hat{\nu}$ defined as $\underline{r}(\hat{\nu}) = R$.

The uninformed bank's strategy. First, note that when an uninformed bank observes $\tilde{\ell}$, it is playing against $F_i(r)$ with some probability strictly less than one. Otherwise the informed bank rejects and the uninformed bank suffers from adverse selection. Using the definition of the signal we obtain:[31]

$$Pr[\ell|\tilde{h}] = \frac{\rho - \mu}{\rho} \qquad \text{and} \qquad Pr[h|\tilde{h}] = \frac{\mu}{\rho}$$

Notice that conditional on playing against F_h, the borrower is H-type with probability ρ rather than μ. With these in mind, the outsider is indifferent between playing any r if and only if such r satisfies:

$$Pr[h|\tilde{h}] \cdot (1 - F_i)(1 - F_o)^{N-2}(\rho \cdot r - 1) + Pr[\ell|\tilde{h}] \cdot (1 - F_o)^{N-2} \cdot (-1) = 0 \tag{A.6}$$

Therefore

$$F_i(r) = 1 - \frac{Pr[\ell|\tilde{h}]}{Pr[h|\tilde{h}](\rho r - 1)} = 1 - \frac{\rho - \mu}{\mu(\rho r - 1)} = 1 - \frac{(1-q)(\phi - \nu)}{\phi(1-q) + qr - 1}$$

The distribution $F_i(r)$ is increasing in r but is less than 1 for all r whenever $\mu < \rho$. This implies that the informed bank must place a positive mass on the maximum interest rate R.

Note that with $\mu \to \rho$ (full information sharing) the informed bank's profit disappears and the distributions are degenerate at $F_i(r) = F_u(r) = 1$, which means all banks charge the lower boundary $\underline{r} = \frac{1}{\rho} = \frac{1}{\mu}$, the competitive interest rate.

No equilibrium with inefficient branch network. We prove that if a bank would have to operate its branch network on the inefficient part of its cost curve, there exists no price equilibrium in which that bank participates. This situation is particularly relevant for a new entrant, which does not have legacy customers in a locality and must obtain market share from informed incumbents. Let F_K denote the strategy of such bank and allow for the possibility of $F_K \neq F_u$. We establish the following claims:

Claim (1): the support of F_K cannot overlap with the support of F_u.

Proof: suppose there is an interval $r \in [a, b]$ which is a common support for all three distribu-

[31] $Pr[h] = Pr[h|\tilde{h}]Pr[\tilde{h}] + Pr[h|\tilde{\ell}]Pr[\tilde{\ell}]$ therefore $Pr[h|\tilde{h}] = \frac{\mu}{\rho}$

tions. Then the inefficient bank is indifferent if and only if

$$Pr[h|\tilde{h}] \cdot (1 - F_i)(\rho r - 1 - c) + Pr[\ell|\tilde{h}] \cdot (-1) = 0$$

This indifference condition would imply an F_i mixing distribution which differs from that of implied by (A.6). This is a contradiction.

Claim (2): The support of F_K cannot be non-overlapping with supp(F_K) > supp(F_u).

Proof: in this case the uninformed efficient bank must participate with probability less than one, otherwise the entrant would never win an $\tilde{h}$-signal client. This implies that the uninformed-efficient bank makes zero profit in equilibrium. Let $\hat{r}$ be the supremum of the support of F_u and the (common) infimum of F_K. Playing in an ϵ-neighborhood of $\hat{r}$, the inefficient bank wins with the same (ϵ-close) probability as the efficient bank, or strictly less, if the informed bank puts a positive probability mass on $\hat{r}$. In any case, the inefficient bank would achieve a profit which is strictly less than that of the uninformed efficient bank. Because the uninformed efficient bank makes zero profit, the inefficient bank would achieve negative profit. This contradicts the individual rationality constraint of the bank which decides to not compete for customers.

Appendix A.2 Proof of Proposition 2

We want to calculate $Prob[r_i < r_u]$, that is, the probability that the informed bank's offer is less than all other uninformed banks' offer. We can write

$$Pr[r_i < \{r_u\}_j] = \int_{\underline{r}}^{R} f_i(r)[1 - F_u(r)]^{n-1} dr + (1 - F_i(R)) \cdot (1 - F_u(R))^{(n-1)}$$

After substitutions and simplifications we obtain:

$$\lambda_i = \frac{1}{2} + \frac{(1-q)^2(\phi - \nu)^2}{2((\phi(1-q) + (1 - qR))^2}$$

Using Equations (A.4) and (A.5) the result follows. Derivatives follow using straightforward algebra.

Appendix A.3 Proof of Proposition 3

Let m denote the market share of the bank as an insider (i.e., the total mass of its informationally captured market). Its aggregate net profit is:

$$\Pi(m, I) = m\left(\pi_i(\phi) - \frac{1}{2}c\phi^2\right) \tag{A.7}$$

The optimal investment is pinned down by the condition $\partial\Pi/\partial\phi = 0$. With information sharing and for $\underline{r} < R$, the optimality condition (noting that $\phi \in [0, 1]$) gives

$$(1-q)(1-\frac{\partial v}{\partial \phi}) - c\phi = 0$$

The case of $\underline{r} \geqslant R$ follows trivially and we obtain:[32]

$$\phi^\star = \begin{cases} \min\{\frac{(1-q)}{c}\left(1-\frac{\partial v}{\partial\phi}\right); 1\} & \text{if } \underline{r} < R \\ \min\{\frac{(1-q)}{c}; 1\} & \text{otherwise} \end{cases} \tag{A.8}$$

[32]We assume that the derivative $\frac{\partial \nu}{\partial \phi}$ is such that uniqueness is guaranteed.