DISCUSSION PAPER SERIES

DP15829

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

Shusen Qi, Ralph De Haas, Steven Ongena and Stefan Straetmans

FINANCIAL ECONOMICS

ISSN 0265-8003

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

Shusen Qi, Ralph De Haas, Steven Ongena and Stefan Straetmans

Discussion Paper DP15829
Published 19 February 2021
Submitted 17 February 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

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

Abstract

We study how information sharing between banks influences the geographical clustering of branches. A spatial oligopoly model first explains why branches cluster and how information sharing impacts price competition and equilibrium clustering. With data on 56,555 branches of 614 banks in 19 countries between 1995 and 2012, we test key model hypotheses. We find that information sharing increases branch clustering as banks open branches in localities that are new to them but that are already served by other banks. This branch clustering is associated with less spatial credit rationing as information sharing allows firms to borrow from more distant banks.

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

Keywords: Branch clustering, information sharing, spatial oligopoly model

Shusen Qi - shusenqi@xmu.edu.cn
Xiamen University

Ralph De Haas - dehaasr@ebrd.com
EBRD, Tilburg University and CEPR

Steven Ongena - steven.ongena@bf.uzh.ch
University of Zurich and CEPR

Stefan Straetmans - s.straetmans@maastrichtuniversity.nl
Maastricht University

Acknowledgements
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, participants at the Bocconi University–Sapienza University Conference on “New Frontiers in Banking: from Corporate Governance to Risk Management” (Milan), the American Economic Association ASSA Annual Meeting (Philadelphia), the 11th Swiss Winter Conference on Financial Intermediation (Lenzerheide), the Great China Area Finance Conference (Xiamen), the 7th MoFiR Workshop on Banking (Ancona), the China International Conference in Finance (Hangzhou), the Deutsche Bundesbank–IWH–CEPR Conference on the Future of Financial Intermediation (Eltville am Rhein), the 2nd EBC Network Workshop (Luxembourg), the 32nd 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, 71790601) and the Social Science Foundation of Fujian Province (FJ2019B140). Steven Ongena acknowledges financial support from ERC ADG 2016 - GA 740272 lending.

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

Shusen Qi
Xiamen University

Ralph De Haas
EBRD, Tilburg University, and CEPR

Steven Ongena
University of Zurich, Swiss Finance Institute, KU Leuven, and CEPR

Stefan Straetmans
Maastricht University and University of Antwerp

February 13, 2021

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, participants at the Bocconi University–Sapienza University Conference on "New Frontiers in Banking: from Corporate Governance to Risk Management" (Milan), the American Economic Association ASSA Annual Meeting (Philadelphia), the 11th Swiss Winter Conference on Financial Intermediation (Lenzerheide), the Great China Area Finance Conference (Xiamen), the 7th MoFiR Workshop on Banking (Ancona), the China International Conference in Finance (Hangzhou), the Deutsche Bundesbank–IWH–CEPR Conference on the Future of Financial Intermediation (Eltville am Rhein), the 2nd EBC Network Workshop (Luxembourg), the 32nd 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, 71790601)* and the Social Science Foundation of Fujian Province (FJ2019B140). Steven Ongena acknowledges financial support from *ERC ADG 2016 - GA 740272 lending*.

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

Abstract

We study how information sharing between banks influences the geographical clustering of branches. A spatial oligopoly model first explains why branches cluster and how information sharing impacts price competition and equilibrium clustering. With data on 56,555 branches of 614 banks in 19 countries between 1995 and 2012, we test key model hypotheses. We find that information sharing increases branch clustering as banks open branches in localities that are new to them but that are already served by other banks. This branch clustering is associated with less spatial credit rationing as information sharing allows firms to borrow from more distant banks. (100 words)

Key words: Branch clustering, information sharing, spatial oligopoly model

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

1. Introduction

Over the past two decades, banks across the world have adjusted their branch networks in response to regulatory changes, increased competition, and progress in information and communication technology. Many banks have also pruned their branch networks in the aftermath of the Great Recession. Importantly, these dynamics did not play out in a spatially uniform way within countries. Banks instead increasingly cluster together as they open new branches in economically strong centers while closing branches in sparsely populated areas. The resulting emergence of banking deserts – localities almost entirely devoid of bank branches (Morgan, Pinkovsky and Yang, 2016) – has raised concerns about adverse local effects on firms' funding costs (Bonfim, Noguiera and Ongena, 2021) and concomitant declines in small-business lending and employment growth (Nguyen, 2019).

Despite this increase in spatial clustering, there hardly is any theoretical or empirical research on the drivers of branch location. The scarce existing literature on the spatial clustering of bank branches is either of a rational or behavioral nature (Deller and Sundaram-Stukel, 2012). Clustering is rational when locating near other banks generates external economies of scale or when the demand for banking services is spatially clustered itself (Brown, Guin and Kirschenmann, 2015). In contrast, behavioral explanations regard clustering as the result of "groupthink" or of banks following first movers in an informational cascade model. Due to reputational concerns, bank managers may open a new branch in a neighborhood with pre-existing branches rather than in uncharted territory. In line with such herding, Chang, Chaudhuri and Jayaratne (1997) find that branch openings follow existing branches even if this hurts the profitability of the new branch. Both explanations for geographical branch clustering are hard to test empirically. For instance, it is challenging to evaluate current and expected credit demand across regions (we discuss later how we carefully control for changes in local credit demand). Moreover, bank managers are compensated based on multiple criteria so that branch locational decisions are hard to evaluate separately. It is also problematic to directly measure and compare banks' informational awareness.

The first contribution of our paper is therefore to build a simple spatial oligopoly model that explains branch clustering. While the model is highly stylized, its main building blocks are representational and it yields testable hypotheses about the impact of information sharing between banks on the equilibrium level of clustering (which as we discuss below is a key theme in our analysis). To the best of our knowledge, we are the first to develop a simple model in which bank branch clustering arises in an intuitive way. The main intuition is that while branch clustering increases the likelihood an entrepreneur will visit a locality and obtain a loan, and correspondingly boosts the size of the local banking market (market-size effect), inter-bank proximity also implies more vigorous competition (price-cutting effect).[1] If the first effect dominates, banks earn higher profits by locating closer to each other so that they attract more clients. If the second effect dominates, banks try to decrease competition by dispersing their branches geographically.

Using this theoretical framework, we derive predictions about the impact of a reduction in information asymmetries between banks and borrowers on the equilibrium level of branch clustering. In particular, we assess the impact of the introduction of a formal mechanism (a public credit registry or a private credit bureau) through which banks share hard (that is, codified and transferable) borrower information. 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. Without such information sharing, banks make lending decisions using private information they produce themselves. Information asymmetries then increase with distance, and geographical credit rationing makes it more difficult for entrepreneurs to successfully apply for a loan at a branch that is further away (Petersen and Rajan, 2002; Hauswald and Marquez, 2006). In our model, such distance constraints bind less when banks can credibly share information about loan applicants.

We derive four testable hypotheses. First, our model predicts that sharing hard information increases branch clustering because banks can attract some distant borrowers that were previously too opaque to lend to. Second, the model predicts that information sharing increases the likelihood that banks open branches in new localities where they were not previously active. This is because adding more branches to a locality with pre-existing branches of the *same* bank would not make it a more attractive 'shopping' destination for entrepreneurs. Third, information sharing spurs relationship banks, which rely heavily on soft information in the absence of information sharing, to cluster more when compared with transactional banks.

[1] Our model revolves around an entrepreneur who needs credit to expand her business. She needs to visit a bank branch to know whether she will get a loan or not. Suppose the marginal probability p that a local bank rejects a loan application is 80 percent (we assume for now that each bank only operates one branch). Moreover, assume that screening and the resulting loan decisions are independent across banks (this is a simplified assumption to keep our main story tractable. We relax this assumption later so that loan decisions can be correlated across banks). If the entrepreneur visits a locality with only one bank, the probability of getting a loan is $1\text{-}p$ = 20 percent. Yet, if another locality has six banks, then the probability of getting a loan is more than three times as high, namely 1-$(0.8)^6 \approx 74$ percent (if the probability of loan rejection equals p, the joint probability of getting a loan accepted in a cluster of k banks taking independent loan decisions equals $1 - p^k$). Given this much higher probability, the entrepreneur will be inclined to visit the second locality. At the same time, the higher inter-bank competition in this locality will lower banks' lending rates and profit margins.

Fourth, in countries where information sharing is of higher quality, its impact on branch clustering is more pronounced.

To test our model empirically, we use detailed bank branch data – geographical coordinates and the dates of establishment (and sometimes 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 towns and cities (henceforth: 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 also provides insights into how bank clustering may respond to similar but slower improvements in borrower transparency in more developed banking markets.

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

By way of preview, we find that information sharing has a strong positive effect on bank 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. This branch clustering is more pronounced for relationship banks and in countries where information sharing is more effective.

Lastly, an additional analysis of data on bank-firm relationships shows that, in line with a reduction in geographical credit rationing, information sharing allows firms to borrow from banks that are more distant. In sum, our results show how information sharing makes it more important for banks to move closer to each other than to their borrowers.

This paper contributes to three strands of the literature. First, there is a lack of research that theoretically explains and empirically identifies the fundamental determinants of the physical location of bank branches. In contrast, a rich empirical literature exploits plausibly *exogenous* spatial variation in bank branches – reflecting historical 'quirks' or waves of financial deregulation – to identify the impact of bank-branch clustering on various outcomes.[2] While useful for identification, one ought to bear in mind that outside of these specific settings, branches are unlikely to be spread quasi-randomly across space. The limited literature that does investigate banks' decisions about their branch network mainly focuses on the size of these networks rather than on their geographical distribution.[3] Our contribution is to develop a simple and intuitive framework in which banks rationally trade off the market-size and price-cutting effects of geographical clustering. We then 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.

Second, we add to the literature on the economic impact of information sharing. Theoretical contributions explore how information sharing reduces moral hazard and adverse selection, improves loan quality, and lowers interest rates (Padilla and Pagano, 1997; 2000), and in general may shape competition between banks (Bouckaert and Degryse, 2004; 2006; Bennardo, Pagano, and Piccolo, 2015). On the empirical side, cross-country evidence indicates that information sharing is associated with more private-sector lending, fewer defaults and lower interest rates (Jappelli and Pagano, 2002). Evidence 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

[2] See Jayaratne and Strahan (1996), Beck, Levine and Levkov (2010), Rice and Strahan (2010), Kroszner and Strahan (2014), Favara and Imbs (2015), Célérier and Matray (2019) for the US; Guiso, Sapienza and Zingales (2004), Herrera and Minetti (2007) and Benfratello, Schiantarelli and Sembenelli (2008) for Italy; and Berkowitz, Hoekstra and Schoors (2014) and Bircan and De Haas (2020) for Russia.

[3] Cerasi, Chizzolini and Ivaldi (2002) and Cohen and Mazzeo (2010) investigate the impact of competition on the size of branch networks. Temesvary (2015) shows theoretically and empirically that locational market power allows banks with larger branch networks to charge an interest-rate premium, while Coccorese (2012) incorporates branch decisions in a price competition model.

information leads to a different branch-clustering equilibrium that is associated with less spatial credit rationing.[4]

Third, our findings add to the industrial organization literature on firm location. This literature asserts that customers trade off the utility they derive from products and the geographic distance to the firms where they can buy these products. As a result, firms have greater market power when they are closer to their customers. This literature starts with the Hotelling (1929) model where firms compete and price their products in geographic locations along a line of fixed length. Salop (1979) introduced a circle model on which firms are located and compete. Much sophistication has been built into such competition models over the years. Syverson (2004), for example, extends the Salop model to allow for heterogeneous producer costs and adds asymmetric information among producers about their production costs (see also, e.g., Barros, 1999; Dell'Ariccia, 2001; Kim, Lozano-Vivas and Morales, 2007). Our assumptions are less stringent than those in the original Salop model. In our model, borrowers are uniformly distributed on a two-dimensional plane and banks can cluster in a locality (in contrast to the Salop model where banks are equidistant).

We proceed as follows. Section 2 introduces our theoretical background and hypotheses development 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

We develop a simple spatial oligopoly model to illustrate the trade-off between the market-size and price-cutting effects of bank branch clustering.[5] Specifically, we determine both the number of entrepreneurs who visit a locality to apply for credit (the market size) and the equilibrium loan rate prevailing in that locality (the price).[6] This section sets out the main intuition of the model while Appendix 1 provides a more formal discussion.

[4] Van Cayseele, Bouckaert, and Degryse (1994) analyze theoretically the effect of sharing 'negative' borrower information about past defaults and 'positive' information about indebtedness on the number of branches per bank. Unlike our paper, the authors do not analyze the spatial distribution of branches.

[5] We build on Konishi (2005) who models the spatial concentration of retail stores.

[6] To ensure tractability, we assume that depositors put all their savings in the nearest bank branch and that the introduction of information sharing has no impact on the deposit market, which is much less affected by information asymmetries. Our focus on lending as the key banking activity is consistent with much of the literature (e.g., Stein, 2002; Berger and Udell, 2006; Hauswald and Marquez, 2006, among others). An interesting exception is Park and Pennacchi (2009) who concurrently model credit granting and deposit taking. A number of recent contributions also highlight the importance of deposit taking for value creation in banking (Egan, Lewellen and Sunderam, 2017 and Drechsler, Savov and Schnabl, 2017; 2020). We leave the spatial modelling of the information derived from observing checking account turnover, for example, for future research.

In our model, both entrepreneurs and banking localities (towns and cities) are distributed across a two-dimensional plane, the former uniformly so, the latter in ways we describe below. Each entrepreneur wants to obtain a single loan for which she can apply by travelling to any locality with at least one bank branch. Entrepreneurs face a probability of *not* obtaining a loan when applying. Loan rejection decisions correlate across branches and we assume this correlation is the same for different localities. Loan size is homogeneous across entrepreneurs and normalized to one. Entrepreneurs need to pay a commuting cost to their locality of choice that is proportional to the travel distance. In addition, entrepreneurs pay the equilibrium loan rate prevailing in this locality if they successfully obtain a loan there.

The model consists of three stages. In Stage I, banks open a finite number of branches across localities on the two-dimensional plane. They cluster their branches based on expected profits. In Stage II, entrepreneurs observe the branch locations and consequently receive a signal about the loan rate in each locality. They now decide, based on the project's Net Present Value (NPV) in each locality, which locality to visit. The NPV depends on the distance to the locality (and the associated transportation costs), the probability of successfully applying for a loan there, and the interest rate in case the borrower receives credit. Each entrepreneur visits at most one locality: the one that in expectation gives the highest (positive) net return. If no locality yields a positive NPV, the entrepreneur does not apply for a loan.

Critically, without the sharing of information among banks, information asymmetries between entrepreneurs and banks cause a discrete distance threshold beyond which the probability of an unsuccessful loan application is one.[7] Stated otherwise, due to geographical credit rationing, entrepreneurs know for sure that they will be rejected when applying for a loan at branches beyond this distance threshold.[8] Only below this threshold does the entrepreneur face the usual rejection probability and does she trade off the higher transportation costs of more distant localities (with more branches) against the higher probability of receiving a loan there (at a lower interest rate).[9] Lastly, in Stage III of the model, bank branches in the same

[7] This is similar in spirit to the adverse selection model of contestable local banking markets by Pagano and Jappelli (1993). In that set up, banks can lend to clients in their own town and, at a higher cost, to clients in 'nearby' towns but not to clients in 'distant' towns for whom lending costs are prohibitive.
[8] According to the president of the Italian Bankers' Association, "*the banker's rule of thumb is to never lend to a client located more than three miles from his office*" (quoted in Guiso, Sapienza and Zingales, 2004). The median Belgian small business borrower in Degryse and Ongena (2005) is located 2.5 kilometers (1.6 miles) from the lending branch. In U.S. data analyzed in Petersen and Rajan (2002) and Agarwal and Hauswald (2010) this median distance is 3.7 km (2.3 miles) and 4.2 km (2.6 miles), respectively.
[9] Hence, our model highlights the first-order impact of information sharing (on branching and lending) through the removal of geographical credit rationing ("the extensive margin"). We leave the incorporation of its impact through informational changes in local lending ("the intensive margin") for future research.

locality compete the loan rate down to a local equilibrium level.[10] We assume that branches grant loans at zero marginal cost.

In our model, a larger bank-branch cluster increases an entrepreneur's NPV for two reasons: a higher chance of getting a loan and loans being cheaper. These advantages may be (partially) offset if the locality is distant and transportation costs are high. There also exists a trade-off for the bank. On the one hand, stronger branch clustering expands the local market because entrepreneurs' loan applications are more often accepted in deeper local banking markets (the market-size effect). On the other hand, branch clustering and the associated competition reduce loan rates (the price-cutting effect). This trade-off determines the optimal level of clustering (number of bank branches in the same locality) and makes the relationship between clustering and the expected profit of a branch follow an inverse U-shape. More branches in a locality initially leads to higher profits as the positive market-size effect dominates the negative price-cutting effect. After some optimum, however, opening another branch in a locality drives down profits as the price-cutting effect more than offsets the increase in market size.

The sharing of information among banks impacts the equilibrium level of branch clustering as it eliminates the distance threshold beyond which entrepreneurs cannot successfully apply for loans. Put differently, when borrower information is shared, entrepreneurs can in principle apply in each locality – as long as transportation costs are not prohibitive. This increased competition from distant localities incentivizes banks to make nearby localities more attractive: through branch clustering they aim to attract or retain relatively distant entrepreneurs that are in search of deeper credit markets in which they can apply for a loan from a wider variety of banks. This yields our first testable hypothesis:

HYPOTHESIS 1: *After the introduction of information sharing, different banks increasingly cluster their branches in the same localities.*

Our model predicts that banks exploit the opportunities of sharing borrower information by extending their branch network to localities where adding a branch of their own increases the number of different banks that entrepreneurs can choose from. In contrast, adding more branches of the same bank in a locality where this bank is already present does not make this locality a more attractive 'shopping' destination for distant entrepreneurs because loan

[10] We assume that the equilibrium lending rate is determined by within-locality competition and is unaffected by distant banks. See Ho and Ishii (2011) for empirical evidence on this account.

rejection rates correlate perfectly among branches of the same bank. That is, if one branch of Bank A rejects an applicant then all other branches of Bank A reject this applicant too. This dynamic is also at work after the introduction of information sharing when attracting and retaining borrowers becomes more important. Our second hypothesis is therefore:

HYPOTHESIS 2: *After the introduction of information sharing, banks are more likely to open new branches in localities with no (or fewer) pre-existing own branches (all else equal).*

Our model also speaks to how information sharing differentially affects relationship lenders and transactional lenders, bank types that rely on different lending technologies. Whereas relationship banks depend primarily on long-term lending relationships during which they obtain and exploit proprietary (soft) borrower information, transactional banks instead mostly rely on publicly available (hard) information (Boot, 2000; Mian, 2006; Beck, Ioannidou, and Schäfer, 2018; Beck et al., 2018). Because soft information is more difficult to transport long-distance than hard information, distance thresholds due to informational asymmetries will bind more for relationship banks. We therefore expect the introduction of information sharing, and thus the breakdown of the distance threshold, to impact relationship banks more, leading to an increase in relationship bank clustering in particular. Our third hypothesis is therefore:

HYPOTHESIS 3: *The impact of information sharing on bank clustering is stronger for relationship banks.*

Lastly, the impact of information sharing on branch clustering depends on how effective the information sharing system is. The extent to which information sharing eliminates the distance threshold due to information asymmetries, and thus fosters branch clustering, directly reflects how comprehensive and trustworthy the shared borrower information is. Our fourth and final hypothesis is therefore:

HYPOTHESIS 4: *The impact of information sharing on bank clustering is stronger in countries with higher quality information sharing systems.*

3. 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 localities (towns and cities) across 19 emerging European countries.[11] 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 in these countries at the start and the end of our sample.

[Insert Figure 1 here]

Appendix Table A2 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 2001-07, 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). This means that 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 in each of the 8,536 localities (towns and cities). We do this for every year in the period 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

[11] A team of consultants with extensive banking experience collected these data by contacting banks or downloading data from bank websites. All information was double-checked 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). For more information, see Beck, Degryse, De Haas and Van Horen (2018) and www.ebrd.com/what-we-do/economics/data/banking-environment-and-performance-survey.html.

observations see a new branch opening. Given the small number of branch closures, this percentage is virtually the same for the variable *Net branch opening*. We also count the log of (one plus the) number of pre-existing branches in a locality that are owned by other banks (*No. branches other banks*) and by individual banks (*No. branches own bank*). Variation is substantial, with some localities not being served by any bank whereas some of the largest localities contain many bank branches.

[Insert Table 1 here]

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; Berger et al., 2005). 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.

Second, we merge our data with the bank ownership information of Claessens and Van Horen (2014) 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 (Mian, 2006; Beck, Ioannidou, and Schäfer, 2016). 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, 2002; 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. 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 lender 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, Degryse, De Haas and Van Horen (2018) 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. (2018) and use question Q6 of the 2nd Banking Environment and Performance Survey (BEPS II). As part of this unique survey, the CEOs of banks participated in in-depth, 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 dealing with small businesses: relationship lending; fundamental and cash-flow analysis; business collateral; and personal collateral (personal assets pledged by the entrepreneur). Although, as expected, almost all banks find building a relationship (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.[12] Table 1 also 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 A3 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. There exists substantial variation in the timing of these introductions, 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).[13]

[12] We have this information for slightly over half of all the banks in our sample. Beck et al. (2018) 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.

[13] 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 financial institutions are distributed; more

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 after the introduction of information sharing, we merge our branch data with the Kompass database on firm-bank relationships. Kompass provides information on firms' address, industry and – critically for our purposes – the primary bank relationship.[14] We have these data for the years 2000 and 2005. We collect the geographical coordinates of Kompass firms based on their name and address and identify the name of 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* in kilometers. The median distance between a firm and its primary bank is 1.8 km (Table 1).

The Kompass data also allow us to create proxies for firms' relative opaqueness. We construct the following three dummy variables: whether the firm has a publicly available email address (*Has email address*); whether the firm has a tax number (*Has tax number*); and whether the firm has formal opening/working hours (*Has formal opening hours*). 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, and 74 percent of the firms work based on formal openings hours.

4. Identification

To test our hypotheses, we apply a difference-in-difference-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 over time, and because treatment effects may be heterogeneous across countries, a standard two-way fixed effects framework can yield biased estimates. This will be the case if some already-treated countries (with information sharing in place) incorrectly act as controls for later events (Abraham and Sun, 2018; Goodman-Bacon, 2018; Callaway and Sant'Anna, 2020).

To address this issue, we follow Cengiz, Dube, Lindner, and Zipperer (2019) and create event-specific data sets. Each event includes all observations from the countries in which

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.

[14] Other papers that employ Kompass include Ongena and Şendeniz-Yüncü (2011), Giannetti and Ongena (2012), Kalemli-Özcan, Laeven and Moreno (2018), and Beck, Ongena and Şendeniz-Yüncü (2019).

information sharing (the treatment) is introduced in the same calendar year as well as the observations from all clean control countries for a 6-year panel by event time ($t=-3, \ldots, 2$) with information sharing introduced at $t=0$.[15] Clean control countries are those without any information sharing system in place within the full 6-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 then compare how, within the same country, the introduction of information sharing differentially affects branch openings across localities with different numbers of pre-existing branches from other banks. To test Hypothesis 1, we estimate the linear probability model:

$$\textit{New branch opening}_{ijct} = \beta_1 * \textit{Information sharing}_{ct} * \textit{No. branches other banks}_{ijct} + \beta_2 * \textit{No. branches other banks}_{ijct} + \Phi_{jt} + \Phi_{it} + \Phi_{ij} + \varepsilon_{ijct} \quad (1)$$

Where the dependent variable *New branch opening*$_{ijct}$ is a dummy that equals one if bank i opens a new 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 equals zero otherwise (the level effect of this variable is absorbed by the fixed effects). *No. branches other banks*$_{ijct}$ measures the number of pre-existing branches by banks other than bank i in locality j. Based on our model, we expect β_1 to be positive as the introduction of information sharing induces banks to cluster to attract more borrowers. That is, after the introduction of a credit registry or bureau, banks are more likely to open new branches in localities that already had more branches of other banks to begin with.

The most saturated version of this model includes three types of interactive fixed effects: *locality*year* (Φ_{jt}); *bank*year* (Φ_{it}); and *bank*locality* (Φ_{ij}). We also allow these effects to vary by treatment event. *Locality*year* fixed effects absorb all time-varying and time-invariant historical, social, economic and cultural differences across towns and 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. *Bank*year* fixed effects flexibly account for time variation in individual banks' operational

[15] 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.

strategies and financial health that affect their branch network as a whole. *Bank*locality* fixed effects absorb time-invariant variation across banks in each specific locality. Lastly, ε_{ijct} is the error term and we cluster standard errors at the country*treatment event level.

To test Hypothesis 2, we measure the number of pre-existing branches of the *same* bank in each locality (*No. branches own bank*) and run the following linear probability model:

$$New\ branch\ opening_{ijct} = \beta_1 * Information\ sharing_{ct} * No.\ branches\ own\ bank_{ijct} + \beta_2 * No.\ branches\ own\ bank_{ijct} + \Phi_{jt} + \Phi_{it} + \Phi_{ij} + \varepsilon_{ijct} \quad (2)$$

Our theoretical model predicts that information sharing reduces the probability that banks open a new branch in localities where they themselves already operate one or several branches. We thus expect β_1 to be negative.

To test Hypothesis 3, we also examine whether information sharing differentially affects relationship versus transactional lenders. We do so by further interacting our treatment with *Bank type* and running the following model:

$$New\ branch\ opening_{ijct} = \beta_1 * Information\ sharing_{ct} * No.\ branches\ other\ banks_{ijct} + \beta_2 * Information\ sharing_{ct} * No.\ branches\ other\ banks_{ijct} * Bank\ type_{ic} + \beta_3 * No.\ branches\ other\ banks_{ijct} * Bank\ type_{ic} + \beta_4 * No.\ branches\ other\ banks_{ijct} + \Phi_{jt} + \Phi_{it} + \Phi_{ij} + \varepsilon_{ijct} \quad (3)$$

Bank type is one out 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 theoretical model, we expect information sharing to have a bigger impact on relationship than on transactional lenders so that especially relationship lenders start to open new branches in localities with more pre-existing branches of other banks. That is, we expect both β_1 and β_2 to be positive.

Lastly, we investigate whether the relationship between information sharing and branch clustering is more pronounced when the quality of information sharing is higher (Hypothesis 4). 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 (1) with this variable renders:

$$\textit{New branch opening}_{ijct} = \beta_1 * \textit{Information sharing}_{ct} * \textit{No. branches other banks}_{ijct} + \beta_2 * \textit{Quality information sharing}_{ct} * \textit{No. branches other banks}_{ijct} + \beta_3 * \textit{No. branches other banks}_{ijct} + \Phi_{jt} + \Phi_{it} + \Phi_{ij} + \varepsilon_{ijct} \quad (4)$$

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, we expect β_2 (and β_1) to be positive.

5. Empirical results

5.1. Baseline results

Table 2 presents regression results based on the linear probability models (1) and (2). The dependent variable is *New branch opening*, which indicates whether a bank opens a branch in a particular locality in a particular year. We investigate hypotheses 1 and 2 in columns 1–4 and 5–8, respectively, while increasingly saturating the models with interactive fixed effects. Columns 2 and 6 include *Locality*Year*Treatment event* fixed effects while in columns 3 and 7 we add *Bank*Year*Treatment event* fixed effects. We further saturate the specifications with *Locality*Bank*Treatment event* fixed effects in columns 4 and 8. These granular fixed effects together capture unobserved variation at various levels, including changes in local credit demand, which might otherwise bias our results.

In line with our first hypothesis, columns 1 to 4 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. This effect of establishing a credit registry or credit bureau is also economically significant. Our preferred (most complete) specification in column 4 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 an additional new branch is opened in that locality by 62 percent. The second row of coefficients shows that also in the absence of information sharing, a higher presence of other bank branches increases the chances of additional branches opening. Yet, the introduction of information sharing significantly increases this tendency of banks to cluster together, as predicted by our model.

Next, columns 5 to 8 show that in line with Hypothesis 2 information sharing induces banks to open new branches in localities where they operate fewer existing branches of their own. The effect is again sizable. Column 8 shows that after the introduction of information sharing, a one standard deviation increase in the number of pre-existing own branches in a locality reduces the likelihood that a bank opens another branch in that locality by 6.7 percentage points. In sum, Table 2 shows that the introduction of 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. As a result, the spatial clustering of bank branches intensifies once a country starts operating a credit registry or a credit bureau.

[Insert Table 2 here]

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 (Panel A) and by the bank itself (Panel B). 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 specifications in column 4 (Panel A) and column 8 (Panel B) of Table 2.

Figure 2 reveals sharp changes at the time of the introduction of information sharing in terms of where banks open new branches. Banks become more likely to open new branches in localities with more pre-existing branches from other banks (Panel A) but fewer branches of their own (Panel B). While the magnitude of the estimated effects gets somewhat smaller over time, the effects continue 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. Equally importantly, we only observe slight trends prior to the introduction of information sharing. These leads are also very small relative to the post-treatment effect estimates. 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; and the persistent post-treatment effects all validate our research design.

[Insert Figure 2 here]

5.2. Information sharing, relationship lending, and branch clustering

The introduction of information sharing may not affect all banks equally. In particular, our third hypothesis states that the impact of information sharing will be stronger among relationship banks as compared with transactional banks. In Table 3, we test this hypothesis by further interacting our main 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). From hereon we focus on the two regression specifications that are most saturated with interactive fixed effects.

The first two columns of Table 3 show that while the introduction of information sharing increases the tendency of large banks to cluster 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 third 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 64 and 68 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 3, we assess heterogeneity by bank ownership. As discussed in Section 2, some prior studies have proxied lending technologies by comparing domestic versus foreign banks. The traditional dichotomy is then 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, Degryse, De Haas and Van Horen (2018) 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 and in our sample.

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 3 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 third hypothesis. However, in absolute terms this difference is limited, at about a fourth of the difference between small and large banks.

[Insert Table 3 here]

5.3. The quality of information sharing regimes and branch clustering

Our model posits that the extent to which information sharing successfully eliminates the distance threshold due to information asymmetries, and thus fosters branch clustering, depends directly on how comprehensive and trustworthy the shared borrower information is (Hypothesis 4). 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.

Columns 1 and 2 of Table 4 show that in line with our fourth hypothesis the introduction of 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 18 percent.

In columns 3 and 4, we restrict the data set to only those observations from countries and years in which some form of information sharing was in place. That is, we now focus on the intensive margin of information sharing to see whether, conditional on a credit registry and/or bureau being in place, more effective information sharing is associated with more bank clustering. In line with the first two columns, the results indicate that this is indeed the case.

[Insert Table 4 here]

5.4. Robustness and placebo tests

Instrumental variables regressions

One may worry that the introduction of information sharing in a country is endogenous as it reflects unobservable national circumstances that also bear directly on branch clustering. However, such country and time specific confounds are at least partly controlled for by our *Locality*Year*Treatment event* 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 variables in our most saturated baseline specifications are in fact the interaction terms *Information sharing*Number branches other banks* and *Information sharing*Number branches own bank*. 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 (column 1) or the bank itself (column 2).

Table 5 reports our IV results. The first stages (columns 1 and 2) show 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 are comparable to our baseline results though larger by a factor of three. There are two reasons why the IV estimates may be larger. 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.

[Insert Table 5 here]

Net branch openings and per capita branch openings

Next, we replace our dependent variable, *New branch opening*, by *Net branch opening*. This dummy variable 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 1-4 of Appendix Table A4. They are very similar to our baseline results in Table 2, both statistically and economically.

In columns 5-8 of Table A4, we normalize the number of bank branches by the population of the locality (town or city), using data from the World Cities Database. We construct the variables *No. branches other banks per 1,000 population* and *No. branches own bank per 1,000 population*. Normalizing the presence of bank branches by the local population does not affect our results. This also reflects that our most saturated specifications already include *Locality*Year*Treatment event* 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.

Bank branch clustering in rural versus urban areas

A separate though related issue is that our findings could mostly 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 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 investigate this issue further, 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.[16] We then rerun our baseline (fully saturated) regression specifications on all three samples. Appendix Table A5 shows that our estimates point to a somewhat stronger impact of information sharing in larger localities. Yet,

[16] We use data from the World Bank-EBRD Business Environment and Performance Survey (BEEPS) to divide localities into these three broad size buckets. This allows us to retain more observations as compared with the approach in Table A4 where we collect the exact population size of localities using the World Cities Database.

the impacts in more rural areas are highly significant and economically sizable too. We conclude that our baseline findings do not reflect secular urbanization trends.

Controlling for other country-level reforms

Our sample countries went through a process of significant economic and political transformation after the fall of the Berlin Wall in 1989. Because many structural reforms occurred simultaneously, one may worry that our information-sharing treatment partially picks up other reforms as well. We note though that 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 Table A3 for the timing of these introductions).

To address this concern more formally, we include four additional interactions with the locality-level number of pre-existing bank branches. This allows us to control for key reform dimensions that might confound our estimates of the impact of new information-sharing regimes. The first variable we interact with is a dummy that equals one if a country is a member of the European Union in a particular year and equals zero otherwise. According to the principle of single authorization, a bank authorized to operate in any one EU country can provide its services throughout the whole Single Market. Acceding the EU may therefore expose a country to foreign-bank entry and an associated change in branch clustering dynamics. The countries in our sample were on average an EU member for 17 percent of the sample period (Table 1).

Second, we control for countries' progress with setting up effective competition policies. We take the EBRD Transition Indicator for competition policy, which can range 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"). The average score on this measure across the years and countries in our sample is 2.2 (Table 1). Enhanced competition policy may change the clustering of real economic activity and, via demand effects, eventually 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"). This indicator averages 3.7 across the countries and years in our data set (Table 1). Progress with small-scale privatization makes lending to small businesses more attractive and can therefore have an independent impact on the branching decisions of commercial banks.

Fourth, we employ a measure of how pro-competitive bank regulation is. This measure, taken from the IMF Financial Reforms Database, can range between 0 and 3 and averages 2.8 in our data set (Table 1). 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.

In Appendix Table A6, we first add these additional interaction terms one-by-one (columns 1-4 and 6-9) and then all at the same time (columns 5 and 10). Throughout all specifications, the interaction between information sharing and the pre-existing locality-level number of branches remains statistically significant at the 1 percent level. 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. Interestingly, this differs in columns 4-5 and 9-10, where we (also) control for the state of bank regulation (of which restrictions on bank branching is a key component). When we add an additional interaction term with this variable, our coefficient of interest declines by about two-thirds. This indicates that changes in top-down restrictions on banks' branching decisions had an important impact on local clustering equilibria too. However, even when controlling for this, we find large effects of the introduction of information sharing. A one standard deviation higher number of pre-existing bank branches in a locality increases the probability that a bank opens an additional branch by 17 percent after the introduction of information sharing. Likewise, a one standard deviation higher number of own bank branches in a locality, reduces the likelihood of another branch opening by the same bank by 2.4 percent.

<u>Placebo test</u>

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 (columns 4 and 8 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 Figure 3. The top (bottom) panel shows the estimates related to column 4 (8) of Table 2. The vertical red lines indicate the 95th percentile of this distribution. Reassuringly, we find that in both panels the real coefficient estimate from Table 2 (0.395 for the top figure and -0.125 for the bottom figure) lies outside the corresponding distribution of the placebo treatment coefficients.

[Insert Figure 3 here]

5.5. Information sharing and geographical credit rationing

An important model prediction that we have not yet been able to test with our branch-level data is that the introduction of information sharing reduces spatial credit rationing: firms will be able to borrow from more distant bank branches. To test this prediction empirically, we merge our branch data with information from the Kompass database on firm-bank relationships. We then assume that firms borrow from the nearest branch of their primary bank and use this nearest distance as the *Firm-branch distance* in kilometers.

Of all countries in Kompass, there are four 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 final merged data set contains 9,348 and 4,960 firm records in 2000 and 2005, respectively, across these four countries.

The upper panel of Table 6 shows summary statistics and a two-sample t-test with unequal variances. In the 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 the two comparator countries that did not introduce information sharing during this period (Croatia and Hungary). We analyze this more formally in a difference-in-differences regression framework (lower panel of Table 6). 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 the same period.[17]

If the sharing of hard information reduces geographical credit rationing, allowing firms to borrow from more distant bank branches, then we expect this to be particularly important for firms that are more opaque. For these firms, information asymmetries are more of an issue and the new publicly available information may therefore have more 'bite'. To test whether this is the case, we use the Kompass data to construct three dummy variables that proxy for a firm's opaqueness. We then use these opaqueness proxies to construct triple interaction terms with *Information sharing*. Each model is fully saturated with additional (unreported) interaction terms between the country and year fixed effects and the respective opaqueness proxy.

Columns 2, 3, and 4 of Table 6 present the results. We find that the effect of information sharing on the reduction in spatial credit rationing is about twice as large for relatively opaque firms than for more transparent 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 (here proxied as those without an email address) and only 11.3 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.

[Insert Table 6 here]

6. Concluding remarks

It is well known that branches of different banks tend to cluster spatially. Yet, to date there exists surprisingly little theoretical and empirical research on the drivers of this phenomenon. Our contribution is to use the introduction of information sharing regimes as plausibly exogenous shocks that shift the relative advantages and disadvantages of branch clustering. We

[17] 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.

then observe how these shocks play out at a very disaggregated level (that of individual towns and cities) across a large number of countries.

We start by building a simple spatial oligopoly model of branch clustering. The model focuses on the trade-off between the market-size effect and the price-cutting effect of clustering. It predicts that the sharing of hard borrower information among banks stimulates clustering due to an increase in competition from far-away bank branches. The model also predicts that after the introduction of information sharing, banks are less likely to open additional branches in locations where they already own a branch. Lastly, our model indicates that the impact is more pronounced for relationship banks and in countries with credit registries of higher quality.

In the empirical part of the paper, we test these theoretical predictions by exploiting dynamic information on the geographical locations of bank branches. We find that the establishment of information sharing has a significantly positive impact on bank clustering and that this impact is larger for relationship banks. We also show 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. Moreover, we provide evidence that suggests that due to these changes the average firm is able to borrow from more distant bank branches.

Our results indicate that branch clustering is a function of the public availability of trustworthy hard borrower information. When such information becomes more broadly available, banks can expand their branch network to new localities that they would previously have avoided. At the same time, it becomes more important for banks to cluster together as a higher local variety of banks makes it easier to attract distant customers. In other words, information sharing makes it more important for banks to move closer to each other than to be closer to potential clients.

Taken together, these findings mean that banking markets become more homogenous in terms of composition – as they are served by the same banks that now operate across the country – but less homogenous in terms of size. While the public availability of hard information leads to further clustering of banks in well-served locations, other (smaller) locations may lose out as access to credit deteriorates further. Assessing the real-economic impacts of such spatial variation in access to credit due to information sharing is a promising avenue for further research.

References

Abiad, A., and Mody, A. (2005). Financial reform: What shakes it? What shapes it? *American Economic Review, 95*(1), 66-88.

Abraham, S., and Sun, L. (2018). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Cornell University Working Paper*.

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

Barros, P. P. (1999). Multimarket competition in banking, with an example from the Portuguese market. *International Journal of Industrial Organization, 17*(3), 335-352.

Beck, T., Degryse, H., De Haas, R., and Van Horen, N. (2018). When arm's length is too far: Relationship banking over the credit cycle. *Journal of Financial Economics, 127*(1), 174-196.

Beck, T., Ioannidou, V., and Schäfer, L. (2018). Foreigners vs. natives: Bank lending technologies and loan pricing. *Management Science, 64*(8), 3792-3820.

Beck, T., Levine, R., and Levkov, A. (2010). Big bad banks? The winners and losers from bank deregulation in the United States. *Journal of Finance, 65*(5), 1637-1667.

Beck, T., Ongena, S., and Şendeniz-Yüncü, İ. (2019). Keep walking? Geographical proximity, religion, and relationship banking. *Journal of Corporate Finance, 55*, 49-68.

Benfratello, L., Schiantarelli, F., and Sembenelli, A. (2008). Banks and innovation: Microeconometric evidence on Italian firms. *Journal of Financial Economics, 90*(2), 197-217.

Bennardo, A., Pagano, M., and Piccolo, S. (2015). Multiple bank lending, creditor rights, and information sharing. *Review of Finance, 19*(2), 519-570.

Berger, A.N., L.F. Klapper and G.F. Udell (2001). The ability of banks to lend to informationally opaque small businesses. *Journal of Banking and Finance*, *25*(12), 2127-2167.

Berger, A.N., N.H Miller, M.A. Petersen, R.G. Rajan and J.C. Stein (2005). Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial Economics*, *76*, 237-269.

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

Berger, A.N. and G.F. Udell (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *Economic Journal*, 112(477), F32-F53.

Berger, A.N., and Udell, G.F. (2006). A more complete conceptual framework for SME finance. *Journal of Banking and Finance, 30*(11), 2945-2966.

Berkowitz, D., Hoekstra, M., and Schoors, K. (2014). Bank privatization, finance, and growth. *Journal of Development Economics, 110*, 93-106.

Bircan, Ç., and De Haas, R. (2020). The limits of lending? Banks and technology adoption across Russia. *Review of Financial Studies, 33*(2), 536-609.

Bonfim, D., G. Nogueira, and S. Ongena (2021). "Sorry, we're closed". Bank branch closures, loan pricing, and information asymmetries. *Review of Finance*, forthcoming.

Boot, A.W.A (2000). Relationship banking: what do we know? *Journal of Financial Intermediation* 9, 7-25.

Bouckaert, J., and Degryse, H. (2004). Softening competition by inducing switching in credit markets. *Journal of Industrial Economics, 52*(1), 27-52.

Bouckaert, J., and Degryse, H. (2006). Entry and strategic information display in credit markets. *Economic Journal, 116*(513), 702-720.

Brown, M., Guin, B., and Kirschenmann, K. (2016). Microfinance banks and financial inclusion. *Review of Finance, 20*(3), 907-946.

Brown, M., Jappelli, T., and Pagano, M. (2009). Information sharing and credit: Firm-level evidence from transition countries. *Journal of Financial Intermediation, 18*(2), 151-172.

Callaway, B., and Sant'Anna, P.H.C (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*, forthcoming.

Célérier, C. and A. Matray (2019). Bank-branch supply, financial inclusion, and wealth accumulation. *Review of Financial Studies*, *32*(12), 4767-4809.

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

Cerasi, V., Chizzolini, B., and Ivaldi, M. (2002). Branching and competition in the European banking industry. *Applied Economics, 34*(17), 2213-2225.

Chang, A., Chaudhuri, S., and Jayaratne, J. (1997). Rational herding and the spatial clustering of bank branches: an empirical analysis. *Columbia University Working Paper.*

Claessens, S., and Van Horen, N. (2014). Foreign banks: Trends and impact. *Journal of Money, Credit and Banking, 46*(s1), 295-326.

Coccorese, P. (2012). Banks as 'fat cats': Branching and price decisions in a two-stage model of competition. *Journal of Economics and Business, 64*(5), 338-363.

Cohen, A., and Mazzeo, M. J. (2010). Investment strategies and market structure: An empirical analysis of bank branching decisions. *Journal of Financial Services Research, 38*(1), 1-21.

Cole, R.A., L.G. Goldberg and L.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*, *39*, 227-251.

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

Dell'Ariccia, G. (2001). Asymmetric information and the structure of the banking industry. *European Economic Review*, *45*(10), 1957-1980.

Deller, S., and Sundaram-Stukel, R. (2012). Spatial patterns in the location decisions of US credit unions. *Annals of Regional Science*, *49*(2), 417-445.

Drechsler, I., A. Savov, and Schnabl, P. (2017). The deposits channel of monetary policy. *Quarterly Journal of Economics*, *132*(4), 1819-1876.

Drechsler, I., A. Savov, and Schnabl, P. (2020). Banking on deposits: Maturity transformation without interest rate risk, NBER Working Paper No. 24582, National Bureau of Economic Research.

Egan, M., S. Lewellen, and Sunderam, A. (2017). The cross section of bank value, NBER Working Paper No. 23291, National Bureau of Economic Research.

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

Favara, G., and Imbs, J. (2015). Credit supply and the price of housing. *American Economic Review, 105*(3), 958-92.

Giannetti, M., and Ongena, S. (2012). "Lending by example": Direct and indirect effects of foreign banks in emerging markets. *Journal of International Economics, 86*(1), 167-180.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. *NBER Working Paper.*

Guiso, L., Sapienza, P., and Zingales, L. (2004). Does local financial development matter?. *Quarterly Journal of Economics, 119*(3), 929-969.

Gupta, R. C., and Tao, H. (2010). A generalized correlated binomial distribution with application in multiple testing problems. *Metrika, 71*(1), 59.

Hauswald, R., and Marquez, R. (2006). Competition and strategic information acquisition in credit markets. *Review of Financial Studies, 19*(3), 967-1000.

Herrera, A. M., and Minetti, R. (2007). Informed finance and technological change: Evidence from credit relationships. *Journal of Financial Economics, 83*(1), 223-269.

Ho, K., and Ishii, J. (2011). Location and competition in retail banking. *International Journal of Industrial Organization, 29*(5), 537-546.

Hotelling, H. (1929). Stability in competition. *Economic Journal, 39*(153), 41-57.

Jappelli, T., and Pagano, M. (2002). Information sharing, lending and defaults: Cross-country evidence. *Journal of Banking and Finance, 26*(10), 2017-2045.

Jayaratne, J., and Strahan, P. E. (1996). The finance-growth nexus: Evidence from bank branch deregulation. *Quarterly Journal of Economics, 111*(3), 639-670.

Kalemli-Özcan, S., Laeven, L., and Moreno, D. (2018). Debt overhang, rollover risk, and corporate investment: Evidence from the European crisis. CEPR Discussion Paper No. 12881, Centre for Economic Policy Research, London.

Kim, M., Lozano-Vivas, A., and Morales, A.J. (2007). Multistrategic spatial competition in banking. *International Journal of the Economics of Business, 14*(3), 351-366.

Konishi, H. (2005). Concentration of competing retail stores. *Journal of Urban Economics, 58*(3), 488-512.

Kroszner, R. S., and Strahan, P. E. (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?* (pp. 485-543). University of Chicago Press.

Mian, A. (2006). Distance constraints: The limits of foreign lending in poor economies. *Journal of Finance, 61*(3), 1465-1505.

Morgan, D. P., Pinkovskiy, M. L., and Yang, B. (2016). Banking deserts, branch closings, and soft information. *Federal Reserve Bank of New York Working Paper.*

Nguyen, H. L. Q. (2019). Are credit markets still local? Evidence from bank branch closings. *American Economic Journal: Applied Economics, 11*(1), 1-32.

Ongena, S., and Şendeniz-Yüncü, İ. (2011). Which firms engage small, foreign, or state banks? And who goes Islamic? Evidence from Turkey. *Journal of Banking and Finance, 35*(12), 3213-3224.

Padilla, A. J., and Pagano, M. (1997). Endogenous communication among lenders and entrepreneurial incentives. *Review of Financial Studies, 10*(1), 205-236.

Padilla, A. J., and Pagano, M. (2000). Sharing default information as a borrower discipline device. *European Economic Review, 44*(10), 1951-1980.

Pagano, M., and Jappelli, T. (1993). Information sharing in credit markets. *Journal of Finance, 48*(5), 1693-1718.

Park, K., and Pennacchi, G. (2008). Harming depositors and helping borrowers: The disparate impact of bank consolidation. *Review of Financial Studies, 22*(1), 1-40.

Peria, M. S. M., and Singh, S. (2014). The impact of credit information sharing reforms on firm financing. *World Bank Working Paper.*

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

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

Rice, T., and Strahan, P. E. (2010). Does credit competition affect small-firm finance? *Journal of Finance, 65*(3), 861-889.

Salop, S. C. (1979). Monopolistic competition with outside goods. *Bell Journal of Economics, 10*(1), 141-156.

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

Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy, 112*(6), 1181-1222.

Temesvary, J. (2015). Dynamic branching and interest rate competition of commercial banks: Evidence from Hungary. *International Journal of Industrial Organization, 43*, 98-110.

Van Cayseele, P., Bouckaert, J., and Degryse, H. (1995). Credit market structure and information sharing mechanisms. In: *Market evolution* (pp. 129-143). Springer, Dordrecht.

von Thadden, E.L. (1995), Long-term contracts, short-term investment and monitoring, *Review of Economic Studies*, *62*(4), 557-575.

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

Appendix 1. Theoretical model

In our model, both entrepreneurs and banking localities (towns and cities) are distributed across a two-dimensional plane, the former uniformly so, the latter in ways we describe below. Each entrepreneur has identical project returns r and wants to obtain a single loan for which she can apply by travelling to any locality with at least one bank branch. Entrepreneurs face a probability p of *not* obtaining a loan when applying. Loan rejection decisions correlate across branches with correlation φ. We assume this correlation is the same for different localities.

Loan size is homogeneous across entrepreneurs and normalized to one. Entrepreneurs need to pay the commuting cost to their locality of choice and this cost equals the distance times a positive transportation cost coefficient t. In addition, entrepreneurs pay the equilibrium loan rate prevailing in this locality if they successfully obtain a loan there.

We assume there are two nearby bank localities d and s as well as a more distant bank locality w. Each entrepreneur visits at most one of these three localities to apply for a loan. We focus on the derivation of the equilibrium level of branch clustering (number of bank branches k) in locality d, treating as given the situation in localities s and w. While stylized, this three-locality setting allows us to derive our main testable hypotheses.

The model consists of three stages. In Stage I, banks open a finite number of branches across localities on the two-dimensional plane. They cluster branches based on expected profits. In Stage II, entrepreneurs observe the branch locations and consequently receive a signal about the loan rate in each locality. They now decide, based on the project's Net Present Value (NPV) in each locality, which locality to visit. The NPV depends on the distance to the locality (and the associated transportation costs), the probability of successfully applying for a loan there, and the interest rate in case the borrower receives credit. Each entrepreneur visits at most one locality: the one that in expectation gives the highest (positive) net return. If no locality yields a positive NPV, the entrepreneur does not apply for a loan.

Critically, without the sharing of hard information among banks, information asymmetries between entrepreneurs and banks cause a discrete distance threshold beyond which the probability p of an unsuccessful loan application is 1. Stated otherwise, due to geographical credit rationing, entrepreneurs know for sure that they will be rejected when applying for a loan at branches beyond the distance threshold. Only below this threshold does the entrepreneur face the usual rejection probability $p<1$ and trades off the higher transportation costs of more distant localities against the higher probability of receiving a loan (at a relatively low cost) in distant localities with more branches.

Lastly, in Stage III of the model, bank branches in the same locality compete the loan rate down to a local equilibrium level. We assume that bank branches grant loans at zero marginal cost. We proceed by backward induction and start in Stage III. In locality d with k bank branches the equilibrium loan rate is:

$$i_d = i_0 + i_1/k \tag{1}$$

where i_o stands for the minimum loan rate and i_1 is a markup that banks can extract from nearby borrowers, for a total of oligopoly rent equal to i_d. With more bank branches, the equilibrium loan rate decreases in line with the price-cutting effect of branch presence. We assume that each bank operates a maximum of one branch per locality. This is equivalent to assuming that multiple local branches of the same bank do not compete with each other (making loan rejection rates perfectly correlated among them).

So far, we have focused on the borrowing costs that entrepreneurs face in locality d. In order to derive the entrepreneur's NPV in this locality, we need to know the probability of loan acceptance. This in turn depends on the local number of branches. To start, assume a locality with only two bank branches that exhibit equal loan rejection probabilities p. Moreover, the loan decision is correlated across branches with correlation $\varphi>0$ because both branches possess partially overlapping proprietary borrower information. The joint probability of rejection at both branches then equals (Gupta and Tao, 2010):

$$Prob(2) = p * p + \varphi * \sqrt{p * p * (1-p) * (1-p)} = p^2 + \varphi * p * (1-p) \tag{2}$$

In the case of three branches, we can compare the third branch with the first two branches, while treating those first two as one unit. The joint probability of rejection at all three branches then equals:

$$Prob(3) = p * Prob(2) + \varphi * \sqrt{p * Prob(2) * (1-p) * (1 - Prob(2))} \tag{3}$$

Likewise, if there are k bank branches in locality d, then the joint probability of rejection in locality d is:

$$Prob(k) = p * Prob(k-1) + \varphi * \sqrt{p * Prob(k-1) * (1-p) * (1 - Prob(k-1))} \tag{4}$$

Then in Stage II, given the expected loan rates in each bank locality, an entrepreneur decides which locality to visit by maximizing her project's expected NPV:

$$EP_d = (1 - Prob(k))(r - i_d) - t * R \tag{5}$$

Where R is the geographic distance between the entrepreneur and locality d. If we assume that localities d and s are sufficiently distant such that local entrepreneurs cannot realize a positive NPV in both localities simultaneously, then the marginal entrepreneur wanting to borrow in d should satisfy $EP_d = 0$ and the market area for locality d is determined by a circle around d with radius:

$$R^*_{no\ overlap} = (1 - Prob(k))(r - i_d)/t \tag{6}$$

This also implies that the market areas of localities d and s do not overlap and that there is therefore no competition between d and s.

We can generalize the model to allow for competition among nearby bank localities. We then assume that around locality d there is an infinite number of localities s at a constant distance m. For an entrepreneur situated between locality d and s, where the distance to locality d equals R, the distance to locality s will equal (m-R). The transportation cost for this entrepreneur to visit locality s then equals t(m-R). Assume that each locality s has the same number of branches j. The NPV for the entrepreneur when she borrows in locality s is then:

$$EP_s = \big(1 - Prob(j)\big)(r - i_s) - t(m - R) \tag{7}$$

The equilibrium loan rate at locality s equals:

$$i_s = i_0 + i_1/j \tag{8}$$

Assume that localities d and s are close enough so that the entrepreneur's NPV is positive in both localities. The entrepreneur then opts for the locality that offers her the highest NPV. The marginal entrepreneur is indifferent between borrowing in locality d and s:

$$EP_d = EP_s \tag{9}$$

This gives us the radius *R* of locality *d*:

$$R^*_{overlap} = [(1 - Prob(k))(r - i_d) - (1 - Prob(j))(r - i_s)]/2t + m/2 \tag{10}$$

Therefore, all entrepreneurs for whom the distance to *d* is less than *R* go to locality *d* to apply for a loan. In other words, the market area for locality *d* encompasses a circle around locality *d* with the above radius. Comparing (10) with (6) shows that the market area around locality *d* shrinks because of competition of nearby locality *s* (assuming that *m* is not too large).

Lastly, in Stage I, banks determine the clustering of their branches based on expected profits. Assume, for example, that the market areas of localities *d* and *s* do not overlap.[18] If all bank branches in locality *d* equally share the total market, then the market size of each branch in locality *d* is:

$$S_d = (\pi * R^2)/k = \pi * [(1 - Prob(k))(r - i_d)/t]^2/k \tag{11}$$

The expected profit of each branch in locality *d* is then:

$$E_d = S_d(k) * i_d(k) \tag{12}$$

Banks will not open a branch in locality *d* if the expected profit is below the expected profit of opening a stand-alone branch in a new locality.

In our model, branch clusters increase an entrepreneur's NPV for two reasons: a higher chance of getting a loan and loans being cheaper. These advantages may be (partially) offset if the locality is distant and transportation costs are high. There also exists a trade-off for the bank. On the one hand, branch clustering increases the local market because entrepreneurs' loan applications are accepted more often in deeper banking markets (the market-size effect). On the other hand, branch clustering and the associated competition reduce loan rates (the price-cutting effect). This trade-off determines the optimal level of clustering (number of bank branches *k* in the same locality) and makes the relationship between clustering and the expected profit of a branch follow an inverse U-shape. More branch clustering initially leads to higher profits as the positive market-size effect dominates the negative price-cutting effect. After some

[18] Alternatively, when the market areas of localities *d* and *s* overlap, radius expression (10) needs to be substituted into (11) in order to calculate the market size and resulting bank branch profits.

optimum, however, opening another branch in a locality drives down profits as the price-cutting effect more than offsets the increase in market size.

Crucially, in the absence of the sharing of hard information, entrepreneurs can only apply for a loan in nearby localities d and s. After all, due to geographical credit rationing, the loan-rejection probability in distant locality w is 1. However, when information sharing is introduced the entrepreneur can also choose to apply for a loan in locality w.

Assume there are n branches located in distant locality w and there is a strictly positive additional cost component c. These costs include higher expenses due to long-distance travel as well as agency costs that result from the serious information asymmetries between bank branches and very distant entrepreneurs. The marginal entrepreneur who chooses the far-away locality w should hence satisfy:

$$EP_w = (1 - Prob(n))(r - i_n) - c \geq 0 \tag{13}$$

Assume the market areas for localities d and s do not overlap prior to information sharing and radius (6) applies. With information sharing (and if the transaction cost c is sufficiently small), the fraction of entrepreneurs that still visits bank locality d declines as they are competed away by distant locality w. The marginal entrepreneur who is indifferent between borrowing in locality d or in locality w should satisfy:

$$EP_d = EP_w \tag{14}$$

This gives us the new radius R around locality d, which should be strictly positive. This implies that there are still some borrowers who visit bank locality d to get a loan:

$$R^*_{info\ sharing} = [(1 - Prob(k))(r - i_d) - (1 - Prob(n))(r - i_n) + c]/t \geq 0 \tag{15}$$

The establishment of information sharing introduces competition from localities that are more distant. Entrepreneurs in the periphery of locality d or s may now decide to apply for a loan in the distant locality w. The reduced radius in (15), as compared with the radius under the no information sharing regime (6), reflects this decrease in the market size of locality d. Banks' profit functions (12) also need to be re-optimized. In order to regain the lost market share, banks can cluster their branches even further (increase in k and thus in the probability of loan acceptance at locality *d*) in order to attract (or retain) borrowers who may be tempted to travel to a distant locality otherwise. Yet, information sharing also increases the correlation between loan decisions across branches. This is because different branches now have similar public information about a borrower. This may partly or fully offset the clustering impact of information sharing because the marginal increase in the loan acceptance probability *1-prob(k)* by increasing *k* is dampened for higher values of the loan decision correlation φ.

Figure A1 in the Appendix shows the situation without overlap between the market areas of locality *d* and nearby locality *s*. The larger circle in light grey represents the market area of locality *d* and *s* before information sharing, while the smaller dark circle is the market area afterwards. The market size shrinks as some entrepreneurs – those already at the outer margins of localities *d* and *s* – decide to apply for a loan in locality *w*. Figure A2 depicts the situation with competition among nearby localities. The dashed line around locality *d* represents all the possible nearby localities *s*.

[Insert Figure A1 and A2 here]

We provide a few numerical illustrations to our stylized model. We assume that the probability of loan rejection is 70 percent, both the minimal loan rate and the oligopoly rent is 2 percent, the project return is 10 percent, the transaction cost coefficient equals 1 percent and the commuting cost of applying for a loan in the distant locality *w* is 6. There are 10 bank branches in the distant locality. We first assume that with information sharing, the correlation among bank branches of a loan rejection stays at 0.2. Figure 3 shows the numerical results.

[Insert Figure A3 here]

The comparative statics in the top panel show that before the establishment of information sharing, banks cluster together until there are six branches in locality *d*. The expected profit of each branch is still higher than the expected profit of operating alone. However, adding a seventh branch would push expected profit below the level that could be had when opening this branch in a new locality instead.

The bottom panel of Figure A3 shows that after the establishment of information sharing (which introduces competition from distant bank localities) branch clustering increases significantly to 16 (until the profit of operating alone is higher than with clustering).

Information sharing reduces spatial credit rationing, increases competition, and decreases the market size. Banks in nearby localities now have more incentives to cluster their branches to attract (or retain) borrowers who may be tempted to travel to, and apply in, a distant locality.

Figure A4 shows the numerical results when nearby localities compete with each other. The comparative statics in the top panel show again that our model predicts a certain amount of bank clustering. According to the panel at the bottom, clustering increases from 4 to 14 branches in locality d once information sharing is introduced (we assume that the number of branches in locality s is 20 and that the distance m between locality d and s is 12). That is, increased clustering happens regardless of whether there is overlap in nearby banking markets.

[Insert Figure A4 here]

In short, the sharing of hard information among banks impacts the equilibrium level of branch clustering as it eliminates the distance threshold beyond which entrepreneurs cannot successfully apply for loans. Otherwise stated, when hard borrower information is shared, entrepreneurs can in principle apply in each locality – as long as transportation costs are not prohibitive. This increased competition from distant localities incentivizes banks to make nearby localities more attractive: by further increasing branch clustering they aim to attract relatively distant entrepreneurs that are in search of deeper credit markets in which they can apply for a loan from a wider variety of banks. This yields our first testable hypothesis that after the introduction of information sharing, different banks increasingly cluster their branches in the same localities.

Our model also predicts that banks exploit the opportunities of sharing borrower information by extending their branch network to localities where adding a branch of their own increases the number of different banks that entrepreneurs can choose from. In contrast, adding more branches of the same bank in a locality where this bank is already present does not make this locality a more attractive 'shopping' destination for distant entrepreneurs because loan rejection rates correlate perfectly among branches of the same bank. That is, if a branch of Bank A rejects an applicant, *all* branches of Bank A will reject the applicant. This dynamic is also at work after the introduction of information sharing when attracting and retaining borrowers becomes more important. Our second hypothesis is therefore that after the introduction of information sharing, banks are more likely to open new branches especially in localities with no (or few) pre-existing own branches.

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.**
*Dependent variables (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 (log)	833,916	1.611	1.386	1.561	0.000	7.290
No. branches own bank (log)	833,916	0.428	0.693	0.537	0.000	5.347
No. branches other banks per 1,000 population (log)	200,274	0.023	0.000	0.082	0.000	0.615
No. branches own bank per 1,000 population (log)	200,274	0.001	0.000	0.007	0.000	0.239
*Country characteristics (country*year level)*						
Information sharing	342	0.532	1	0.500	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 level)						
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

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. *Country* * *Treatment Event* clustered 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)	(7)	(8)
Information sharing * No. branches other banks	0.017***	0.334***	0.165**	0.395***				
	(0.000)	(0.003)	(0.014)	(0.002)				
No. branches other banks	0.002	0.030***	0.039***	0.662***				
	(0.221)	(0.002)	(0.002)	(0.000)				
Information sharing * No. branches own bank					-0.073**	-0.080***	-0.070***	-0.125***
					(0.043)	(0.000)	(0.000)	(0.002)
No. branches own bank					-0.012*	-0.002	0.007*	-0.389***
					(0.091)	(0.816)	(0.076)	(0.000)
Information sharing	-0.007				0.085*			
	(0.710)				(0.087)			
Locality * Year * Treatment Event Fixed Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Bank * Year * Treatment Event Fixed Effects	No	No	Yes	Yes	No	No	Yes	Yes
Locality * Bank * Treatment Event Fixed Effects	No	No	No	Yes	No	No	No	Yes
R-squared	0.005	0.433	0.628	0.712	0.009	0.432	0.628	0.727
Observations	833,916	833,916	833,916	833,916	833,916	833,916	833,916	833,916

Table 3

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 as compared with transaction-based 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. *Country * Treatment Event* clustered 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	0.167**	0.407***	0.164**	0.392***	0.128**	0.418***
	(0.012)	(0.001)	(0.014)	(0.002)	(0.013)	(0.000)
Information sharing * No. branches other banks * Bank type	0.012**	0.027***	0.002	0.005	0.005*	0.006**
	(0.015)	(0.002)	(0.607)	(0.262)	(0.096)	(0.048)
No. branches other banks * Bank type	-0.011***	0.005	-0.003***	-0.046***	-0.001	0.063***
	(0.002)	(0.442)	(0.002)	(0.000)	(0.342)	(0.005)
No. branches other banks	0.037***	0.663***	0.040***	0.696***	0.078***	0.697***
	(0.002)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Locality * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Bank * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank * Treatment Event Fixed Effects	No	Yes	No	Yes	No	Yes
R-squared	0.628	0.712	0.628	0.712	0.703	0.767
Observations	833,916	833,916	833,916	833,916	592,383	592,383

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. Columns 1-2 are based on all observations from all countries that introduce information sharing in the same calendar year. These are grouped together and then 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. Columns 3-4 are based on only those countries and years in which information sharing is in place. Table A1 contains the definitions and Table 1 the summary statistics for all variables. *Country * Treatment Event* clustered 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)
Information sharing * No. branches other banks	0.157**	0.383***		
	(0.027)	(0.003)		
Quality information sharing * No. branches other banks	0.122***	0.180***	0.122***	0.273***
	(0.000)	(0.000)	(0.001)	(0.000)
No. branches other banks	0.039***	0.662***	0.196**	1.584***
	(0.002)	(0.000)	(0.022)	(0.000)
Locality * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes
Bank * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes
Locality * Bank * Treatment Event Fixed Effects	No	Yes	No	Yes
R-squared	0.628	0.712	0.652	0.791
Observations	833,916	833,916	81,240	81,240

Table 5

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 stages (columns 1-2) 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 (column 1) or the bank itself (column 2). The instruments in these first stages are 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 (column 1) or the bank itself (column 2). The dependent variable in the second stage (columns 2 and 3) 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. *Country * Treatment Event* clustered 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	Information sharing * No. branches own bank	New branch opening	
	(1)	(2)	(3)	(4)
% neighboring countries introduced information sharing	0.174***			
* No. branches other banks	(0.000)			
% neighboring countries introduced information sharing		0.202***		
* No. branches own bank		(0.000)		
Information sharing * No. branches other banks			1.174***	
			(0.000)	
No. branches other banks	0.008***		0.622***	
	(0.000)		(0.000)	
Information sharing * No. branches own bank				-0.422***
				(0.000)
No. branches own bank		0.102***		-0.343***
		(0.000)		(0.000)
F-Statistic	4,005	8,615		
Locality * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes
Bank * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes
Locality * Bank * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes
Observations	833,916	833,916	833,916	833,916

Table 6

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. P-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

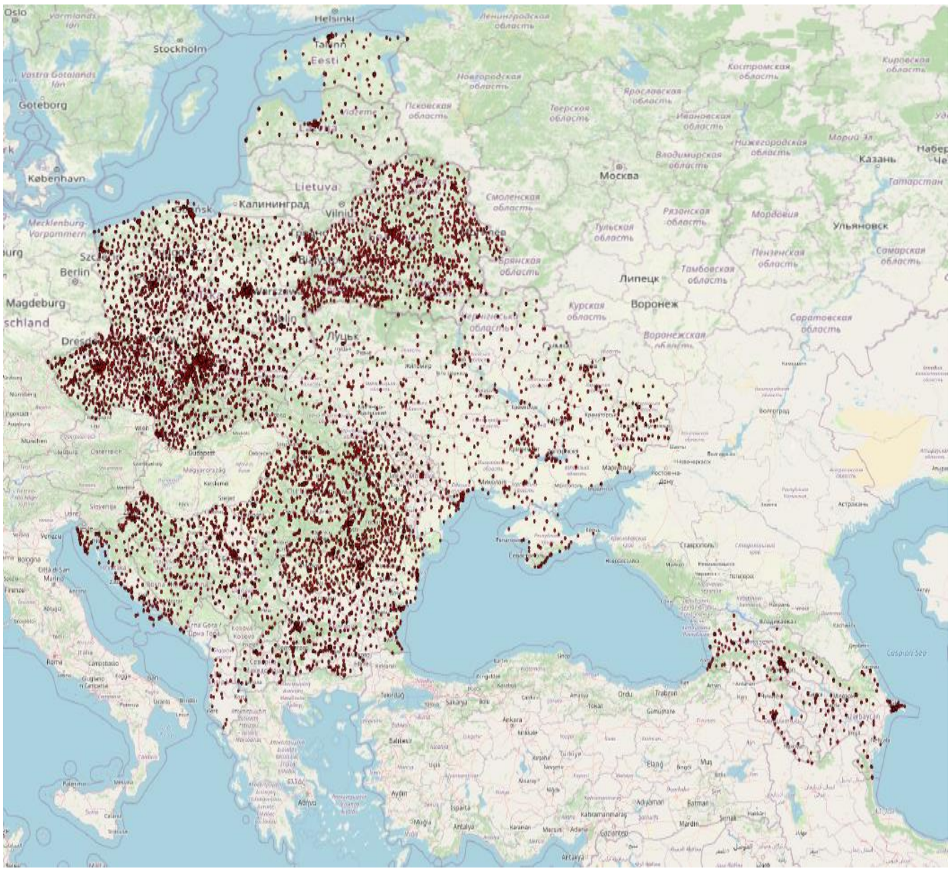
Dependent variable → Firm-branch distance (in km)											
Czech Republic (Introduced information sharing in 2002)						**Poland (Introduced information sharing in 2001)**					
	Obs.	Mean	St. Dev.	5%	95%		Obs.	Mean	St. Dev.	5%	95%
2000	1,650	3.01	5.16	2.76	3.26	2000	5,286	19.13	56.57	17.60	20.65
2005	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)
Information sharing	15.14***	19.15***	21.02***	19.48***
	(0.000)	(0.000)	(0.000)	(0.000)
Information sharing*Has email address		-7.89***		
		(0.001)		
Information sharing*Has tax number			-15.77***	
			(0.003)	
Information sharing*Has formal opening hours				-11.63***
				(0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
R-squared	0.027	0.030	0.030	0.029
Observations	14,308	14,308	14,308	14,308

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.

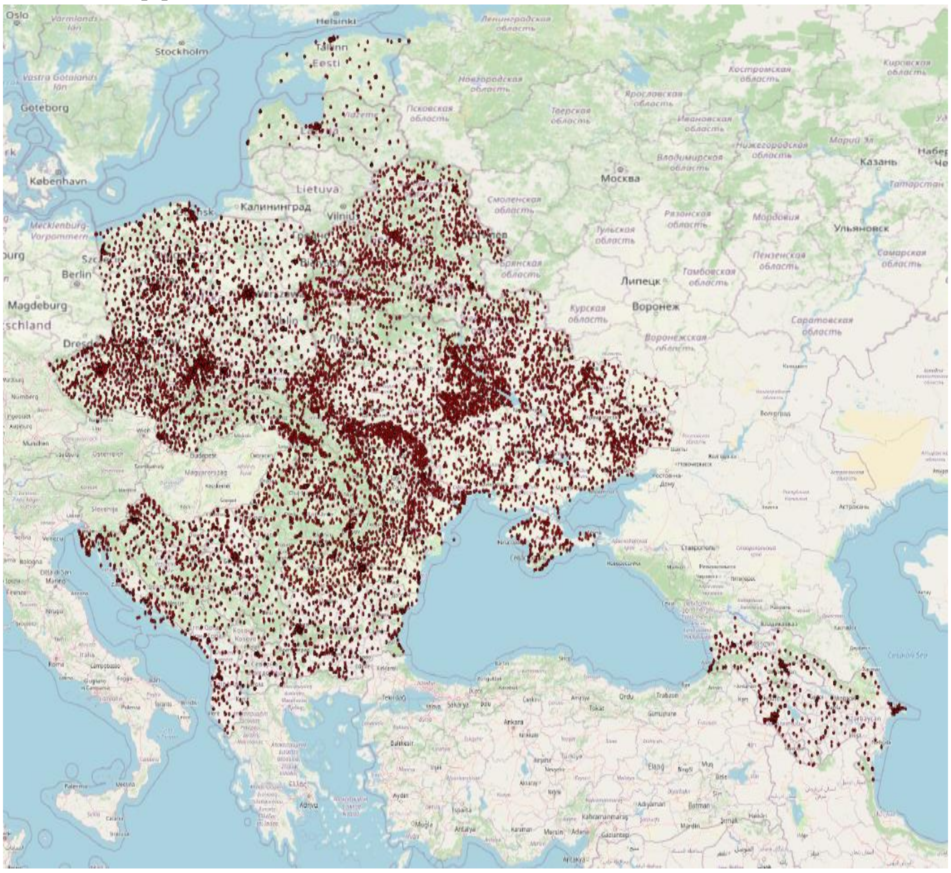
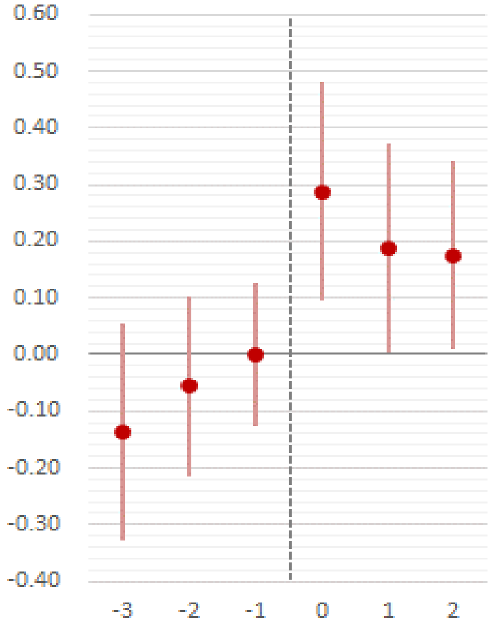


Figure 2

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 either the number of pre-existing branches of other banks in a locality (Panel A) or with the number of pre-existing branches of the bank itself (Panel B). All coefficients are based on specifications with the same interactive fixed effects and covariates as in column 4 (Panel A) and column 8 (Panel B) of Table 2.

Panel A



Panel B

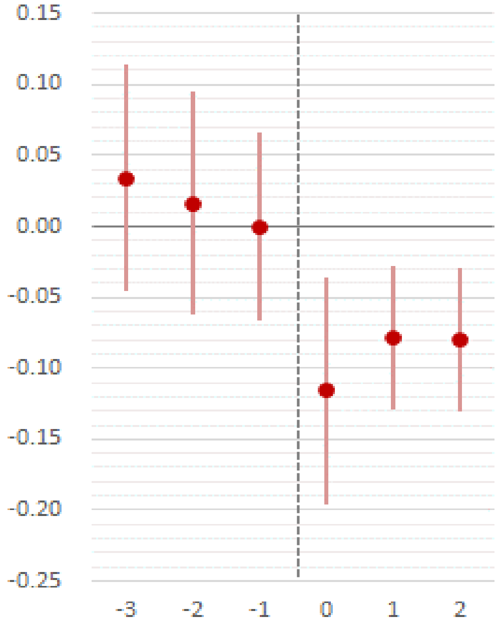
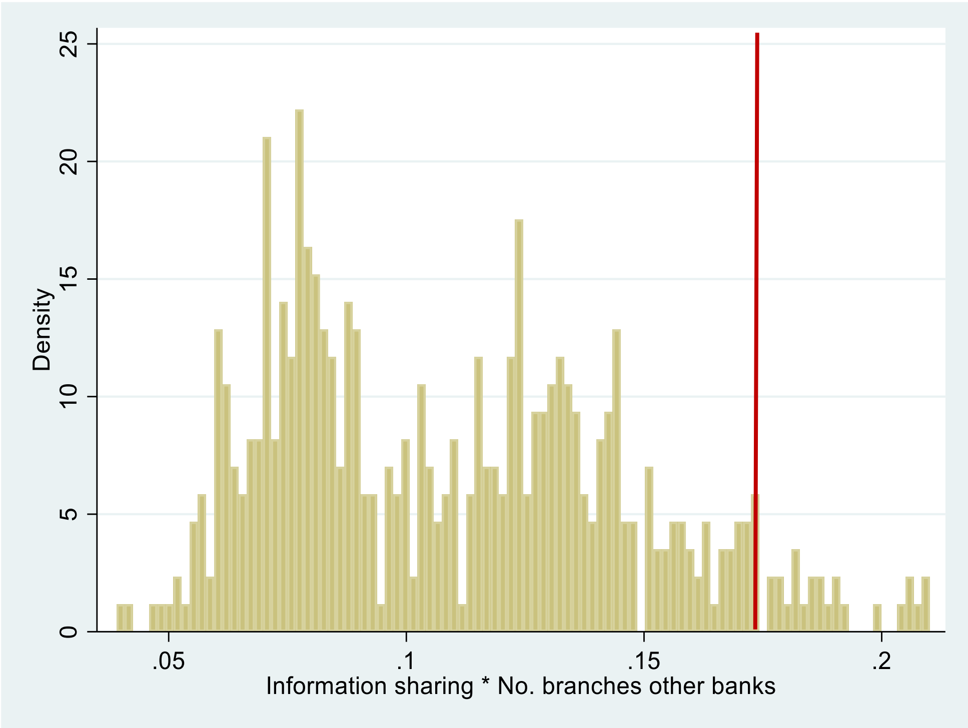


Figure 3

Information Sharing and the Clustering of Bank Branches: Placebo Treatments

These figures present 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 regressions (columns 4 and 8 of Table 2) are 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 these placebo treatments. The top (bottom) panel shows the estimates related to column 4 (8) of Table 2. The vertical red lines indicate the 95th percentile of this distribution. In both panels the real coefficient estimate from Table 2 (0.395 for the top figure and -0.125 for the bottom figure) is outside the corresponding distribution of the placebo treatment coefficients.



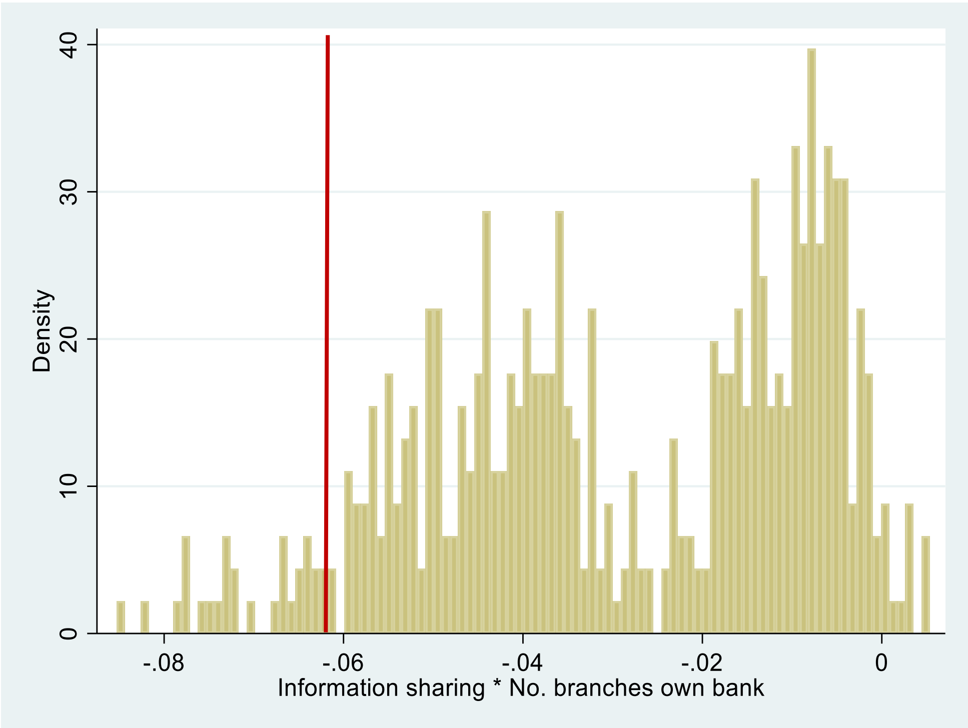


Table A1

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) which was conducted among 611 banks across 32 countries. "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. CvH: Claessens and Van Horen (2014).

	Definition	***Data Sources***
Dependent variables		
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 (log)	Log number of existing branches of other banks within a locality in a year	BEPS II
No. branches own bank (log)	Log number of existing branches of the same bank within a locality in a year	BEPS II
No. branches other banks per 1,000 population (log)	Log number of existing branches of other banks within a locality per 1,000 inhabitants in a year	BEPS II, WCD
No. branches own bank per 1,000 population (log)	Log number of existing branches of the same bank within a locality per 1,000 inhabitants in a year	BEPS II, WCD
Independent variables		
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
Quality information sharing	= 0 to 6, higher values indicate a higher quality of information sharing in the country in that year	Doing Business
Domestic bank	= 1 if more than 50% of a bank's shares are foreign-owned, = 0 otherwise	CvH
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-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
EU membership	= 1 if a country is part of the European Union in a particular year, = 0 otherwise	European Commission
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

Table A2

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 A3

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 A4

Information Sharing and the Clustering of Bank Branches: 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)-(4) 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 (5)-(8), the number of existing bank branches are normalized by the local population in 1,000 persons and the dependent variable measures whether a bank opens a new branch in a locality in a year. Table A1 contains all definitions and Table 1 the summary statistics for each variable. *Country * Treatment Event* clustered 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			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Information sharing * No. branches other banks	0.165**	0.395***						
	(0.014)	(0.002)						
No. branches other banks	0.039***	0.663***						
	(0.002)	(0.000)						
Information sharing * No. branches own bank			-0.070***	-0.125***				
			(0.000)	(0.002)				
No. branches own bank			0.007*	-0.389***				
			(0.090)	(0.000)				
Information sharing * No. branches other banks per 1,000 population					1.477**	3.241**		
					(0.023)	(0.042)		
No. branches other banks per 1,000 population					-0.471**	10.814***		
					(0.015)	(0.000)		
Information sharing * No. branches own bank per 1,000 population							-1.190**	-2.089**
							(0.017)	(0.036)
No. branches own bank per 1,000 population							0.402**	-9.698***
							(0.015)	(0.000)
Locality * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank * Treatment Event Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.628	0.712	0.628	0.727	0.501	0.624	0.501	0.624
Observations	833,916	833,916	833,916	833,916	200,274	200,274	200,274	200,274

Table A5

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. *Country * Treatment Event* clustered 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	0.377***		0.575***		0.625***	
	(0.010)		(0.001)		(0.002)	
No. branches other banks	0.820***		0.915***		0.500***	
	(0.000)		(0.000)		(0.000)	
Information sharing * No. branches own bank		-0.129***		-0.119***		-0.150***
		(0.000)		(0.002)		(0.000)
No. branches own bank		-0.575***		-0.513***		-0.211***
		(0.000)		(0.000)		(0.000)
Locality * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Bank * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.702	0.729	0.663	0.700	0.684	0.697
Observations	234,790	234,790	98,766	98,766	95,777	95,777

Table A6

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. *Country* * *Treatment Event* clustered robust *p*-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of statistical significance, respectively.

Dependent variable →	New branch opening									
X →	No. branches other banks					No. branches own banks				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Information sharing * X	0.396***	0.386***	0.383***	0.110***	0.100***	-0.125***	-0.122***	-0.121***	-0.045***	-0.034***
	(0.002)	(0.003)	(0.002)	(0.000)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
EU membership * X	-0.158***				0.040	0.032				-0.021
	(0.324)				(0.793)	(0.401)				(0.547)
Competition policy * X		0.101***			-0.020		-0.025			-0.018
		(0.009)			(0.818)		(0.100)			(0.345)
Small-scale privatisation * X			0.146***		0.137**			-0.047***		-0.043**
			(0.001)		(0.015)			(0.007)		(0.027)
Pro-competition bank regulation * X				-0.044***	-0.043*				-0.029*	-0.029**
				(0.005)	(0.058)				(0.054)	(0.035)
X	0.662***	0.444***	0.133	0.777***	0.319	-0.389***	-0.333***	-0.214***	-0.287***	-0.091
	(0.000)	(0.000)	(0.443)	(0.000)	(0.150)	(0.000)	(0.000)	(0.001)	(0.000)	(0.290)
Locality * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank * Year * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Locality * Bank * Treatment Event Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.712	0.712	0.713	0.721	0.721	0.727	0.727	0.727	0.731	0.732
Observations	833,916	833,916	833,916	650,601	650,601	833,916	833,916	833,916	650,601	650,601

Figure A1

Impact of Information Sharing on Branch Clustering without Overlap of Bank Localities

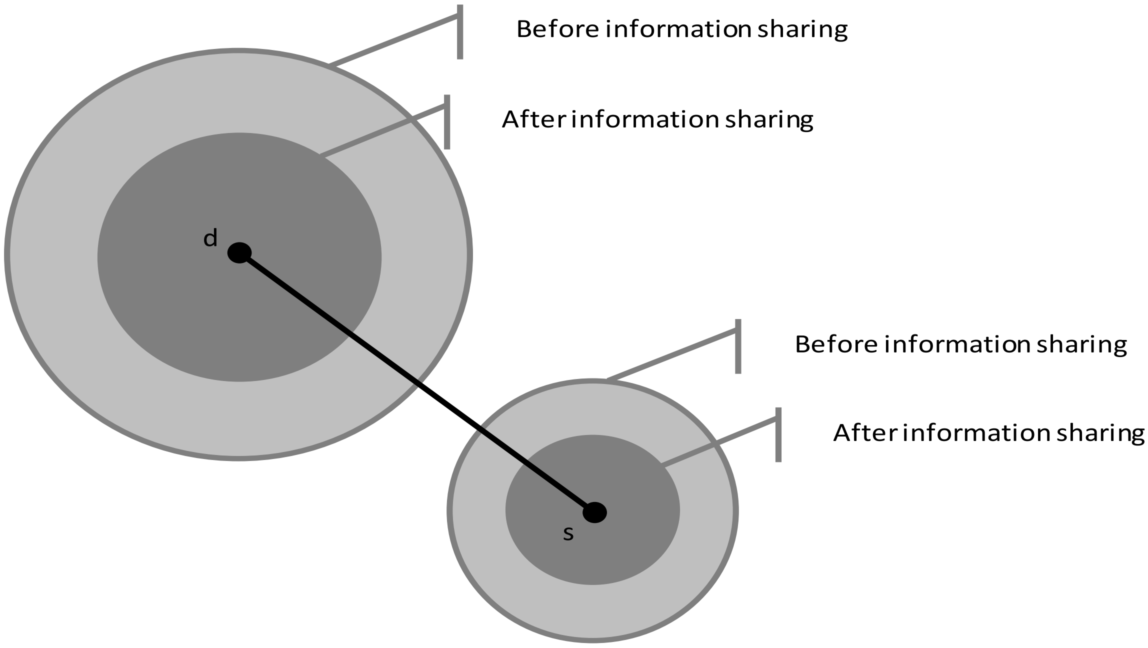


Figure A2

Impact of Information Sharing on Branch Clustering with Overlap of Bank Localities

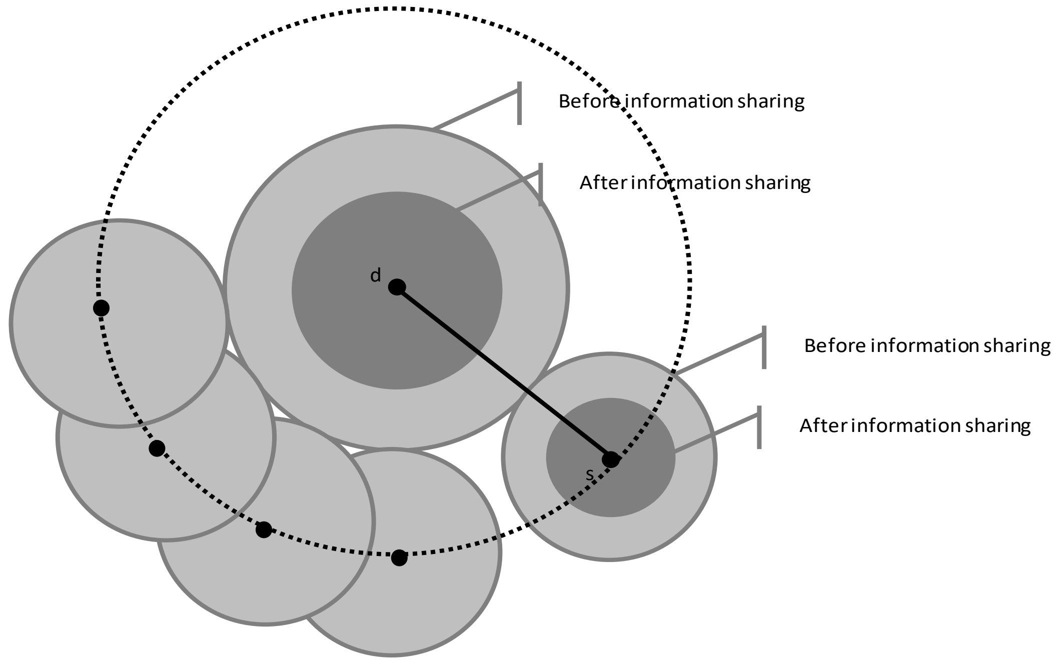
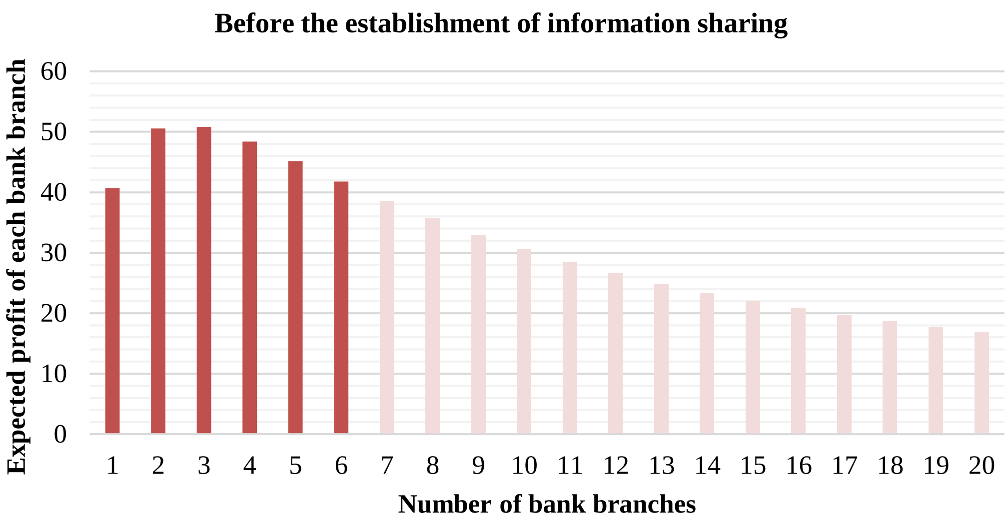


Figure A3

Branch Clustering after the Establishment of Information Sharing

No Overlap Among Bank Localities

This figure presents comparative statics based on a calibration of our theoretical model that assumes no overlap among nearby bank localities. We assume that the probability of loan rejection is 70 per cent; the minimal loan rate is 2 percent; the oligopoly rent is 2 percent; the project return is 10 per cent; the commuting cost coefficient is 1 per cent; the correlation among bank branches of the loan-rejection probability is 0.2; and the commuting cost of applying for a loan in a distant locality is 6. There are 10 bank branches in the distant locality w. The vertical axis shows the expected profit of each bank branch and the horizontal axis shows the number of bank branches. Darker (lighter) shades indicate that the expected profit of opening a new branch in locality d is larger (smaller) than the expected profit (shown by the first column at the very left) of opening a new branch in a new locality without pre-existing branches. Before the establishment of information sharing, banks cluster together until there are 6 branches in locality d. The expected profit of each of these 6 branches is still higher than the expected profit of operating alone (which is just above 40). Adding a 7th branch would, however, push expected profit below the profit that could be had when opening that additional branch in a new locality instead. After the introduction of information sharing (which introduces competition from distant bank localities) branch clustering increases significantly to 16 (until the profit of operating alone is higher than of clustering).



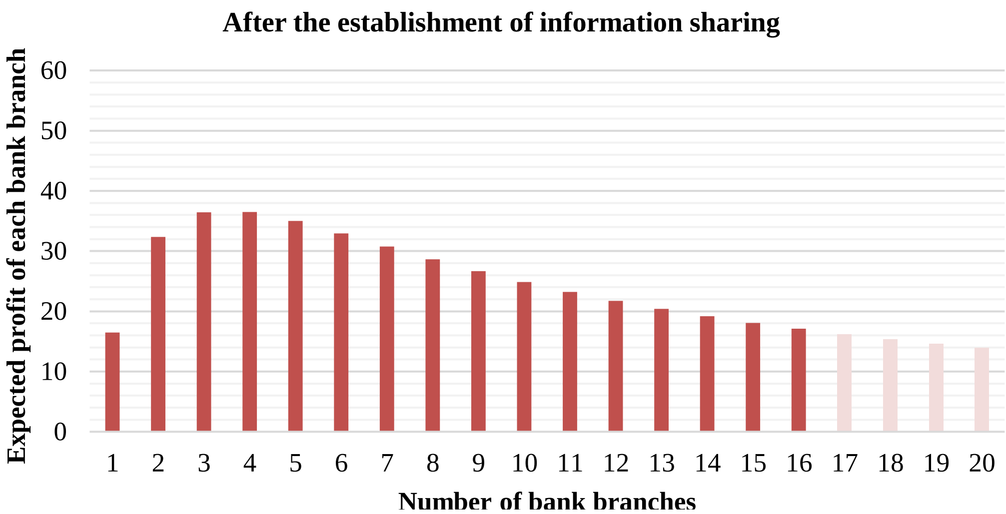


Figure A4

Branch Clustering after the Establishment of Information Sharing

Overlap Among Bank Localities

This figure presents comparative statics based on a calibration of our theoretical model with overlap among nearby bank localities. We assume that the probability of loan rejection is 70 per cent; the minimal loan rate is 2 percent; the oligopoly rent is 2 percent; the project return is 10 per cent; the commuting cost coefficient is 1 per cent; the correlation among bank branches of the loan rejection probability is 0.2; and the commuting cost to a distant locality is 6. There are 10 bank branches in distant locality w. The number of bank branches in locality s is 20 and the distance m between locality d and s is 12. The vertical axis shows the expected profit of each bank branch and the horizontal axis the number of branches. Darker (lighter) shades indicate that the expected profit of opening a new branch in locality d (shown by the first column on the very left) is larger (smaller) than the expected profit of opening such a branch in a locality without pre-existing branches. Before the establishment of information sharing, banks cluster together until there are 4 branches in locality d. The expected profit of each of these branches is still higher than the expected profit of operating alone (which is just above 40). Adding a 5th branch would push expected profit below the profit that could be had when opening the additional branch in a new locality instead. With information sharing (which introduces competition from distant localities) branch clustering increases significantly to 14 (until the profit of operating alone is higher than of clustering).