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

Qi, Shusen; de Haas, Ralph; Ongena, S.R.G.; Straetmans, Stefan

Publication date:
2018

Document Version
Early version, also known as pre-print

CentER

Discussion Paper

No. 2018-038

MOVE A LITTLE CLOSER? INFORMATION SHARING AND THE SPATIAL CLUSTERING OF BANK BRANCHES

By

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

17 September 2018

ISSN 0924-7815
ISSN 2213-9532

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

September 14, 2018

Shusen Qi

Xiamen University

Ralph De Haas

EBRD and Tilburg University

Steven Ongena

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

Stefan Straetmans

Maastricht University and University of Antwerp

The authors thank Jaap Bos, Martin Brown, Stephanie Chan, Hans Degryse, Thomas Mathä, Vahid Saadi, Koen Schoors, Francesc Rodriguez Tous, Yunqi Zhang, Kuncheng Zheng, participants at the 2018 Great China Area Finance Conference (Xiamen), the 2018 Nankai University Young Scholars in Finance Conference (Tianjin), the 7th MoFiR Workshop on Banking (Ancona), the 11th Swiss Winter Conference on Financial Intermediation (Lenzerheide), the 2017 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), the 2018 AEA/ASSA Annual Meeting (Philadelphia) and seminar participants at the School of Economics at 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. Qi acknowledges financial support from *National Natural Science Foundation of China (71790601)* and *Fundamental Research Funds for the Central Universities (20720181028)*. 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. We construct a spatial oligopoly model with price competition that explains why bank branches cluster and how the introduction of information sharing impacts clustering. Dynamic data on 59,333 branches operated by 676 banks in 22 countries between 1995 and 2012 allow us to test the hypotheses derived from our model. We find that information sharing spurs banks to open branches in localities that are new to them, but that are already well served by other banks. Information sharing also 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 technologies. Figure 1 illustrates the resulting time variation in the number of bank branches across a variety of countries. The figure also shows vividly how many American and European banks have pruned their branch networks in the aftermath of the Great Recession.

[Insert Figure 1 here]

What cannot be seen in Figure 1 is that these dynamics did not play out in a geographically uniform way *within* countries. Banks increasingly cluster together as they close branches in sparsely populated areas while opening new ones in economically stronger centers. For example, of the 600 UK branch closures between April 2015 and April 2016, over 90 percent were in areas with a below-median household income. In contrast, about two-thirds of all branch openings occurred in wealthier neighborhoods (Reuters, 2016). A similar trend can be observed in the United States, where branch clustering is mirrored by an increase in "banking deserts": localities entirely devoid of bank branches (Morgan, Pinkovsky and Yang, 2016).[1]

Yet, despite this increase in geographic (spatial) clustering, hardly any theoretical or empirical research exists on the drivers of branch location. The scarce existing literature on the determinants of 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 virgin 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.

[1] Nguyen (2017) finds that (merger-related) branch closures in the US cause prolonged declines in small-business lending and employment growth. These impacts are highly local and dissipate within six to eight miles.

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 main contribution of our paper is therefore to build a spatial oligopoly model that explains branch clustering behavior and that yields testable hypotheses about the impact of information sharing on the equilibrium level of clustering. To the best of our knowledge, we are the first to develop such a stylized model in which bank branch clustering arises in an intuitive way.

Our model revolves around an entrepreneur who needs credit to expand her business. She needs to visit a bank branch and talk to a loan officer to know whether she will get a loan or not. Suppose the probability of getting a loan from any local branch is 20 percent. Moreover, assume that the success probabilities across different bank branches are unrelated.[2] If the entrepreneur visits a locality with only one bank branch, the probability of getting a loan is 20 percent. Yet, if another locality has six bank branches, then the probability of getting a loan is more than three times as high, i.e., $1-(1-0.2)^6 \approx 74$ percent.[3] Given this much higher probability, the entrepreneur will be tempted 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. An important intuition of our model is therefore that while branch clustering increases the size of the local banking market (market-size effect), inter-bank proximity also implies more vigorous competition (price-cutting effect). If the first effect dominates, bank branches earn higher profits by locating closer to each other so that they can attract more clients. If the second effect dominates, banks will 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 branch clustering. In particular, we assess the impact of the introduction of a formal mechanism (such as a public credit registry or private credit bureau) through which banks share hard (that is, codified and transferable) borrower information. Such data include both 'negative' information about prior defaults and late payments of loan applicants and 'positive' information about whether they have outstanding debt elsewhere. Without such information sharing in place, banks need to make their lending decisions mainly on the basis of proprietary (soft) information that they collect themselves. In that case, information asymmetries increase with distance and

[2] We relax this assumption later so that success probabilities can be correlated across banks.
[3] If the probability of getting a loan equals q, for two branches locally present the probability equals $q + (1-q)q = q + q - q^2 + 1 - 1 = 1 - (1-q)^2$, etc.

geographical credit rationing makes it 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 hard information about loan applicants. Information sharing then allows banks to lend to more distant entrepreneurs. We derive four testable hypotheses.

First, our model predicts that sharing hard information increases bank branch clustering and competition because banks can attract distant borrowers that were previously too opaque to lend to. Second, the model predicts that the likelihood that banks open a new branch in a locality where they already operate branches declines, because adding more branches to the same locality will not make it a more attractive ‘shopping’ destination. Third, information sharing spurs domestic banks (that rely more on soft information) to cluster, while it helps foreign banks (that rely more on hard information) to expand into new localities. Finally, information sharing synchronizes different banks’ loan-approval decisions and hence dampens the market-size effect of clustering. Our model accordingly predicts that when information sharing becomes very effective, branch clustering gradually becomes less pronounced.

To empirically test our model, we use unique and detailed bank branch data – geographical coordinates and the dates of establishment (and sometimes closure) of each branch – from 22 Eastern European countries. Our sample covers 59,333 branches that were active within the period 1995-2012. The data set further contains information on the banks that own these branches, which enables us to distinguish between branches of domestic versus foreign 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). Moreover, many Eastern European countries institutionalized information sharing among lenders – either through a public credit registry or a private credit bureau – during our sample period. We use the introduction of such information-sharing regimes as country-level exogenous shocks that push banks towards a new clustering equilibrium. This setting can also provide insights into how bank clustering may respond in more developed banking markets to similar but slower improvements in borrower transparency.

In terms of methodology, we implement a differences-in-differences approach to evaluate the impact of the introduction of information sharing in some countries compared with other sample countries where information sharing had not yet been introduced. This strategy enables us to mitigate selection biases and, by allowing for country, bank and time fixed effects, alleviates concerns about omitted variables.

By way of preview, we find that information sharing has a strong positive effect on bank branch clustering. Information sharing makes it more likely that banks open new branches in localities where they did not yet operate but where other banks were already present. The analysis of additional Kompass data on bank-firm relationships shows that, in line with a reduction in geographical credit rationing, information sharing allows firms to borrow from more distant banks. Lastly, we find that clustering becomes less pronounced over time in countries with more effective information sharing systems. All these findings are in line with our theoretical predictions.

This paper contributes to three strands of the literature. First, there is a clear 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 density on various outcomes.[4] 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. Moreover, the limited literature that does investigate banks' decisions concerning their branch network mainly focuses on the size of these networks rather than on their geographical distribution.[5] 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 impacts 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). On the empirical side, cross-country evidence indicates that information sharing is associated

[4] See Jayaratne and Strahan (1996), Beck, Levine and Levkov (2010), Rice and Strahan (2010), Kroszner and Strahan (2014), and Favara and Imbs (2015) 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) for Russia.

[5] 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.

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 information leads to a different branch-clustering equilibrium which is in turn associated with less spatial credit rationing.[6]

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 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. Our assumptions are less stringent than those in the Salop (1979) model that is used extensively in the literature on bank competition (see Barros, 1999; Dell'Ariccia, 2001 and Kim, Lozano-Vivas and Morales, 2007). 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 a simple spatial oligopoly model of branch clustering 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. Theoretical model

2.1. Model intuition and construction

We develop a spatial oligopoly model to formalize the trade-off between the market-size and price-cutting effects of bank clustering.[7] Specifically, we determine both the number of

[6] 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.

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

entrepreneurs who visit a locality to apply for credit (the market size) and the equilibrium loan rate prevailing in that locality (the price).[8]

In our model, both entrepreneurs and banking localities (towns or cities) are uniformly distributed across a two-dimensional plane. 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. This probability is correlated 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 branch clustering 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 locations of the branches and consequently receive a signal about the loan rate in each locality.[9] They now decide, based on the expected return of borrowing in each locality, which locality to visit. The expected return 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 expected return, 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

---

[8] 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, a market much less affected by information asymmetries. Our focus on credit granting as a key banking activity is consistent with much of the literature (e.g., Stein, 2002; Berger and Udell, 2006; Hauswald and Marquez, 2006, among many others). An interesting exception is Park and Pennacchi (2009) who concurrently model credit granting and deposit taking. We leave the spatial modelling of the information derived from observing checking account turnover, for example, for future research.

[9] The local loan rate depends on the number of branches and hence the intensity of bank competition.

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.[10] 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.

Finally, in Stage III of the model, bank branches in the same locality compete the loan rate down to a local equilibrium level.[11] 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 the oligopoly rent that banks can extract from nearby borrowers. With more bank branches, the equilibrium loan rate decreases in line with the price-cutting effect of branch presence.

To determine the probability that a loan application is rejected, we start from the case of two bank branches. The probability of rejection at the first and the second branch both equal p and the probability is correlated across branches with correlation φ. Because of this interdependence (different banks possess partially overlapping proprietary (soft) information about the same borrower), when a borrower gets rejected by one bank branch, the rejection probability is also higher in another branch. Therefore, the joint probability of rejection at both branches 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:

[10] 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 SME 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.

[11] 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.

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

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

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

Then in Stage II, given the expected loan rates in each bank locality, an entrepreneur decides which locality to visit by maximizing the expected profit:

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

Where *R* is a threshold distance or radius between the borrower location and locality *d*. If we assume that there is no overlap between localities *d* and *s*, then the marginal entrepreneur should satisfy $EP_d = 0$ and we have:

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

We can also generalize the model to allow for competition among nearby bank localities: that is, when the market areas of locality *d* and *s* overlap. We assume that around locality *d* there is an infinite number of localities *s*. The distance between all these localities *s* and locality *d* is *m*. For an entrepreneur located between locality *d* and *s*, and if the distance to locality *d* equals *R*, the distance to locality *s* will equal (*m*-*R*), so that the transportation for this entrepreneur to visit locality *s* will equal *t*(*m*-R). Each locality *s* has the same number of bank branches *j*. The expected profit for the entrepreneur to be had in locality *s* is then:

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

The equilibrium loan rate at locality *s* equals:

$$i_s = i_0 + i_1/j \quad (8)$$

Assume that locality *d* and *s* are close enough so that the expected profit of visiting each locality is positive. The borrowers then compare the expected profit of both options and the marginal borrower is indifferent between 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* choose to 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. If all bank branches in locality *d* equally share the total market, then the market size of each branch in locality *d* is:

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

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

$$E_d = S_d * i_d \tag{12}$$

Finally, in Stage I, banks determine the clustering of their branches based on expected profits. They 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 expected return 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 prefer denser banking markets (the market-size effect). On the other hand, branch density and the associated competition reduce loan rates (the price-cutting effect). This trade-off determines the optimal level of clustering and makes the relationship between clustering and the expected profit of a branch follow an inverse U-shape. Denser branching 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.

In the absence of the sharing of hard information, entrepreneurs can only apply for a loan in nearby localities d and s. 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.[12] The establishment of information sharing thus increases competition in each banking locality and decreases the market size (as entrepreneurs shop for loans elsewhere). Banks in nearby localities now have more incentives to cluster their branches in order to attract (or retain) borrowers who may be tempted to travel to a distant locality and apply there.

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:

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

Note that with information sharing, the inter-branch correlation of the probability that an entrepreneur cannot get a loan at any branch may also increase. This is because different branches now have similar public information about a borrower. This reduces the benefit of branch clustering and decreases the market-size effect.

If the transaction cost *c* is sufficiently small, then the fraction of entrepreneurs that still visits bank locality *d* declines. The marginal entrepreneur who is indifferent between going to locality *d* and locality *w* should satisfy:

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

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

[12] Hence our model highlights the first-order impact of the introduction of information sharing (on branching and lending) through the removal of geographical credit rationing (i.e., "the extensive margin"). We leave the multifaceted incorporation of its impact through the informational changes in local lending (i.e., "the intensive margin") for future research.

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

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*.

2.2. Hypothesis development

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

[Insert Figure 2 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. Adding a seventh branch would, however, push expected profit below the level that could be had when opening this branch in a new locality instead.

The bottom panel of Figure 2 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 a distant locality and apply there.

Figure 3 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 3 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. When hard borrower information is shared, entrepreneurs can in principle apply in each locality – as long as transportation costs are not prohibitive. Realizing this, banks start to cluster in order to attract more 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 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 are perfectly correlated among branches of the same bank. That is, if an applicant gets rejected by a branch of Bank A it will get rejected by *all* branches of Bank A in the same locality. This impact is more important after the introduction of information sharing when attracting and retaining borrowers becomes more vital. 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 few) pre-existing own branches.*

Third, our model implies that information sharing can affect domestic and foreign banks differently if these banks rely on other lending technologies. Domestic banks often rely on long-term lending relationships during which they exploit proprietary (soft) borrower information whereas foreign banks focus on transactional lending based on publicly available (hard) information (Mian, 2006 and Beck, Ioannidou, and Schäfer, 2017). This means that in the absence of information sharing, distance thresholds due to informational asymmetries can bind more for domestic banks. In our model, this amounts to borrowers facing a higher agency cost c when applying at a domestic bank branch as compared with an equidistant foreign branch (recall that c comprises both agency and travel costs). The introduction of information sharing then affects domestic banks more because the overall reduction in c is larger, leading to an increase in domestic bank clustering in particular.

Figure 4 illustrates this prediction with numerical results. We assume that prior to information sharing the cost of long-distance lending was higher for domestic (6.5) than for foreign banks (6.2). This reflects that relationship lending by domestic banks involves higher distance-related agency costs. With information sharing, the cost of screening distant clients is equalized at 6 as both bank types can now use the credit registry or bureau. This change is larger for domestic banks (-0.5) than for foreign banks (-0.2). The model shows accordingly that the clustering response is stronger for domestic banks (both with and without overlap among nearby localities). Our third hypothesis is therefore:

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

[Insert Figure 4 here]

Finally, the impact of information sharing on branch clustering depends on how effective the information-sharing system works. We already know that information sharing increases the correlation between different banks' loan decisions as all banks can now use the same public hard information in addition to their own proprietary soft information. As a matter of fact, the effectiveness of information sharing and related loan decision correlations can be so high that this induces a negative market-size effect and a decrease in bank branch clustering.

To see this in our calibration exercise, we gradually increase the correlation from 0.2 to 0.3 and compare equilibrium clustering (Figure 5). The horizontal axis shows the correlation among banks in loan rejection decisions (a higher correlation indicates a more effective information-sharing system) and the vertical axis shows bank branch clustering in equilibrium. We indeed observe that when lending decisions across banks become increasingly correlated, there is a decline in the market-size effect and therefore in branch clustering. Our fourth and final hypothesis is therefore:

HYPOTHESIS 4: *Information sharing increases branch clustering but this relationship turns negative once the effectiveness of information sharing – as measured by the interbank correlation in lending decisions – becomes sufficiently high. The relationship between the effectiveness of information sharing and branch clustering thus displays an inverse U-shape.*

[Insert Figure 5 here]

3. Data

Table 1 provides summary statistics of our dependent variables (branch data) and independent variables. Appendix Table A1 provides all definitions.[13]

3.1. Branch data

To test our hypotheses, we require time-varying data on branch locations for countries in which information sharing – either through a public credit registry or through a private credit bureau – is introduced at different points in time. We therefore collected information on the geographical coordinates of 59,333 branches operated by 676 banks across 22 emerging European countries.[14] These data paint a precise and gradually changing picture – reflecting branch openings and closures – of the banking landscape during the years 1995 to 2012. Figure 6 shows the geographical branch distribution in these countries at the start and the end of our sample. During our sampling period, bank started to gradually cluster more as indicated by a 17.9 percent increase in the cross-locality Herfindahl-Hirschman index.

[Insert Figure 6 here]

[13] Appendix Table A2 contains a correlation matrix of all variables.

[14] A team of consultants with extensive banking experience collected the data by contacting banks or downloading data from bank websites. 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 http://www.ebrd.com/what-we-do/economics/data/banking-environment-and-performance-survey.html.

Appendix Table A3 summarizes the number of branches that opened or closed by year and country: 33,716 (1,365) branches opened (closed) during our sample period. Many new 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.

For each branch, we know the identity of the parent bank. By merging our data with the bank ownership data in Claessens and Van Horen (2014) we 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. We further distinguish between greenfield foreign banks (de novo banks established from scratch) and take-over banks that were formed when a foreign bank acquired a domestic one.

We take the 33,716 branch openings during our sample period as our main unit of observation.[15] This allows us to test whether the introduction of information sharing encouraged banks to open branches in different types of localities. Table 1 shows that approximately half of all branch openings took place when a country had a credit registry or bureau in place (*Information sharing*). 44 per cent of all branch openings were by a foreign bank and about a third of these were by greenfield foreign banks.

[Insert Table 1 here]

Our main dependent variables capture the local clustering of banks across countries and over time. *No. branches all banks* measures the number of pre-existing bank branches in the proximity of a newly opened branch. For this and all other dependent variables, we use two methods to match a new branch with all nearby existing branches. First, we draw circles with a 2 or 5 km radius around the geo-coordinates of the new branch and count the number of existing bank branches within that circle. Second, we count the number of existing branches within the same locality (town or city) as where the new branch is located. Table 1 shows that the median new branch is surrounded by 17 pre-existing branches within a 2-km radius, 23 branches within a 5-km radius, and 25 branches within the same locality.

Branch same bank is a dummy that indicates whether the bank that opens a new branch already operated one or more branches in the same area (circle or locality). The

[15] Our data contain 59,333 branches (owned by 676 banks) of which 33,716 (owned by 532 banks) opened during 1995-2012. The remaining 144 banks did not open branches during this period. Because relatively few branches were closed during our sample period, we do not separately investigate the impact of such closures on clustering.

probability of pre-existing branches of the same bank being present is 33 (37) per cent when the surrounding area is measured as a circle with a 2 km (5 km) radius and 38 per cent when branches are matched by locality. While the median number of pre-existing branches is zero, there is wide variation and the average number is three.

Lastly, we measure local credit market concentration by constructing a Herfindahl Hirschman Index (HHI) for each circle or locality around a new branch:

$$HHI_i = \sum_{i=1}^{N} s_i^2 \tag{16}$$

Where s_i stands for the market share, measured in branches, of bank *i* in the market and *N* is the number of banks.

3.2. Data on information sharing

Data on the introduction of information-sharing regimes are from the World Bank Doing Business database, the EBRD, and on-line sources. Appendix Table A4 shows that during 1995-2012, fifteen countries introduced a public credit registry and eighteen a private credit bureau. There exists substantial variation in the timing of the introduction of information sharing which is helpful for our empirical identification.

The upper chart in Figure 7 shows the average number of pre-existing branches *of the same bank* around newly opened branches in the four years before and the four years after the establishment of information sharing at *t=0*. In line with our second hypothesis, after the introduction of information sharing, new branches open in localities where banks have fewer pre-existing branches *of their own* (that is, banks spread out their own branch network). The second and third chart show that, in line with our first hypothesis, these new branches tend to cluster in more competitive markets (with more existing branches from other banks and hence a lower HHI).

[Insert Figure 7 here]

Lastly, we measure the quality of the information-sharing regime in a country by means of the World Bank Doing Business credit information index. The index ranges from 0 to 6 and reflects rules and practices affecting the coverage, scope and accessibility of credit information from either a public credit registry or a private credit bureau (higher values indicate more effective information sharing).[16] Among our 22 countries the average value of *Quality information sharing* is 1.45 but there is wide variation with a standard deviation of 2.16 (Table 1).

4. Methodology

To test our hypotheses, we apply a differences-in-differences framework with multiple groups (countries) and time periods (Wooldridge, 2007). To test Hypothesis 1, we estimate the following benchmark model:

$$N_{ijct} = \alpha_c + \alpha_t + \beta_1 T_{ct} + \varepsilon_{ijct} \quad (17)$$

Where *i* indicates a new bank branch; *j* the locality in which the new branch opens (circle or town/city); *c* indicates the country and *t* refers to the year. N_{ijct} is the number of *pre-existing* branches when branch *i* opens in locality *j* (*No. of branches all banks*). T_{ct} (*Information sharing*) is a dummy equal to 1 if banks in country *c* share borrower information in year *t*. α_c and α_t are country and year fixed effects.[17] According to our model, β_1 is expected to be positive as the introduction of information sharing induces bank branches to cluster so as to attract more borrowers. We also run this model with the locality-level HHI index on the left-hand side (HHI_{ijct}). We then expect β_1 to be negative as smaller information asymmetries due to information sharing mean that new branches tend to open in more competitive local markets with a lower *pre-existing* HHI.

To test Hypothesis 2, we measure the number of existing branches of the same bank in the locality where a branch opens (*No. branches same bank*): N_{ijct}^{own}. We also construct a dummy version of this variable (D_{ijct}^{own}) that is equal to 1 if $N_{ijct}^{own} > 0$ and 0 otherwise (*Branch same bank*). We run the following regression model:

$$N_{ijct}^{own} = \alpha_c + \alpha_t + \beta_1 T_{ct} + \varepsilon_{ijct} \quad (18)$$

[16] A score of 1 is assigned for each of six features: (1) Both positive credit information (outstanding loan amounts and on-time repayments) and negative information (late payments and defaults) are distributed; (2) Data on both firms and individuals are distributed; (3) Data from retailers, utility companies, and financial institutions are distributed; (4) More than two years of historical data are distributed; (5) Data on loan amounts below 1 percent of income per capita are distributed; and (6) By law, borrowers have the right to access their data in the largest credit bureau or registry.

[17] The inclusion of bank fixed effects is partially compromised by a number of banks that enter, exit and merge in each country during the sample period. We return to this issue in the robustness section.

According to our model, information sharing makes banks less likely to open new branches in localities where they already operate branches themselves. We thus expect β_1 to be negative.

To examine whether information sharing differentially impacts domestic and foreign banks, we run interaction regressions. Assuming that domestic (foreign) banks are more oriented towards relationship (transaction) lending, we expect domestic banks to be more affected by the introduction of information sharing and therefore have more incentives to cluster. Let F_i be a dummy equal to '1' if a branch belongs to a foreign bank. The sum of β_1 and β_3 is now the treatment effect for foreign banks:

$$N_{ijct} = \alpha_c + \alpha_t + \beta_1 T_{ct} + \beta_2 F_i + \beta_3 T_{ct} * F_i + \varepsilon_{ijct} \tag{19}$$

Lastly, we investigate to what extent the effectiveness of information sharing matters for branch clustering (Hypothesis 4). The time-varying variable $Quality_{ct}$ measures the rules and practices affecting the accessibility, coverage, scope and quality of credit information available through information sharing (*Quality information sharing*). Augmenting the base regression (1) with this variable renders:

$$N_{ijct} = \alpha_c + \alpha_t + \beta_1 T_{ct} + \beta_2 * Quality_{ct} + \varepsilon_{ijct} \tag{20}$$

Note that $Quality_{ct}$ is only available for country-years in which banks exchange borrower information (i.e., $T_{ct} = 1$). The value equals zero if there is no information sharing in a specific year and country. Based on our model we expect β_1 (β_2) to be positive (negative). Very effective information sharing ensures that all banks have access to the same comprehensive information about loan applicants. This increases the inter-bank correlation in loan decisions and dampens the market-size effect of clustering. After all, it becomes less attractive for borrowers to travel to (distant) localities with many branches (as the probability that a loan application gets rejected at each branch becomes very similar). In the extreme case, when all branches share all information and process these data in the same way, there is no difference between having just one branch or having many branches in a locality. Increasingly effective information sharing therefore starts to dampen clustering at one point (all else equal).

5. Empirical results

5.1. Baseline results

Table 2 presents our baseline results (cf. equations (17) and (18)). For each dependent variable – *No. branches all banks*, *Branch same bank*, *No. branches same bank,* and *HHI* – we show three regression outcomes for different proxies of the dependent variables. In the first two regressions, we match newly-opened branches with all surrounding pre-existing branches within a 2-km or a 5-km radius. In the third specification, we match new branches with all existing branches in the same locality. All regressions include country and year fixed effects.

In line with our first hypothesis, columns 1-3 show that once a country introduces information sharing, banks start to open new branches in localities with more pre-existing branches compared to countries where information sharing has not (yet) been introduced. Our results are qualitatively similar when matching new branches with existing branches in a 2 (5) km radius (columns 1 and 2) or within the same locality (column 3). The impact of a credit registry is economically significant: column 3 shows that once information sharing is introduced, banks choose to locate new branches in towns and cities that have 55 more pre-existing bank branches. This is a large effect given that the average locality in our data set contains 129 branches.[18]

In contrast, and in line with hypothesis 2, columns 4 to 9 show that information sharing induces banks to open new branches in localities where they have fewer existing branches of their own. This effect is again sizable: after the establishment of information sharing, banks are 11 percentage points less likely to locate a new branch in cities where they already own one or more branches of their own (column 6). The number of pre-existing own branches is reduced by almost 3 (the average is 4). Estimates in columns 10 to 12 provide further evidence on this fact to show that after the introduction of information sharing, banks start to open new branches in less concentrated markets – as indicated by a lower HHI index. Or in other words, more different banks are going to enter into each locality. This aligns with our first result.

[Insert Table 2 here]

[18] The dependent variables in columns 1-3 and 7-9 are non-negative integers. When we use a Poisson estimator all our results continue to hold at the 1 per cent level.

5.2. Heterogeneous impacts by bank type

Not all banks may be equally affected by information sharing. To analyze such heterogeneity, we interact in Table 3 the information sharing variable with dummies that identify foreign banks (in line with eq. (19)). The first three columns of Table 3 indicate that while information sharing affects the clustering of all banks, the impact is significantly stronger for domestic banks. For instance, after the establishment of information sharing, domestic banks tend to open new branches in localities with 74 more existing branches (compared with the situation before information sharing). This number is only 34 (74.39 - 40.88) for foreign banks.

The first three columns of Table 3 also show that *before* the introduction of information sharing, foreign banks opened branches in localities with more pre-existing branches. That is, compared to domestic banks, foreign banks typically added branches to well-established local banking markets rather than venturing into new territory. A possible explanation is that compared to domestic banks, foreign banks have built up less information about domestic clients and therefore mimic the locational choices of domestic competitors. This effect is economically significant: foreign banks locate new branches in cities with 84 more pre-existing branches. The introduction of information sharing therefore partially levels the playing field between foreign and domestic banks: it allows foreign banks to open branches in underserved markets that they previously avoided.[19]

Lastly, columns 4-9 in Table 3 show that with information sharing foreign banks tend to open branches in new localities where they had no or few branches before. Prior to information sharing, foreign banks were more likely than domestic banks to cluster their own branches together but, again, the registry appears to allow foreign banks to geographically spread out their branch networks. After the establishment of information sharing, foreign banks are 4 percentage points less likely than domestic banks to open new branches in a locality with prior branches of their own.

[Insert Table 3 here]

[19] In Appendix Table A5 we take this analysis one step further by distinguishing between greenfield and take-over foreign banks. As expected, the introduction of information sharing is especially beneficial to greenfield foreign banks: these relatively young and inexperienced banks cluster less after the introduction of information sharing and can spread out their branch networks more widely.

5.3. Heterogeneous impacts by effectiveness of the information-sharing regime

Table 4 assesses our fourth hypothesis that in countries with particularly effective information sharing, branch clustering will gradually level off.[20] More specifically, our model predicts that the market-size effect of clustering declines when the inter-bank correlation in loan-rejection probabilities increases. The intuition is that banks receive increasingly similar hard information about applicants and are therefore more likely to make the same lending decisions (as the role of their proprietary, soft information becomes relatively less important). Note that only in countries with information sharing in place, we can calculate the variable *Quality information sharing* (in countries without information sharing, this variable is zero).

The first three columns in Table 4 show that, in line with our fourth hypothesis, the presence of information sharing leads to more branch clustering but that this increase becomes smaller for more effective registries. The results in column 3 indicate that an improvement of the registry quality by 2 points (out of 6, about one standard deviation) reduces branch clustering due to information sharing from 51 to 38 pre-existing branches per locality. We also directly include *Quality information sharing* and its square term to examine the shape of the relationship. The results in the last three columns of Table 4 confirm the existence of a reverse U-shaped relationship between the quality of information sharing and bank branch clustering as predicted by our theoretical model.

[Insert Table 4 here]

5.4. Robustness and placebo tests

We subject our results to various robustness and placebo tests. First, 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. We absorb any time-invariant unobservable variation through country fixed effects, and region-wide time variation through year fixed effects. However, we cannot control for unobservable country-specific and time-varying variables. We therefore instrument the introduction of information sharing in a country and year with the percentage of all neighboring countries that introduced information sharing in the past five years (Martinez Peria and Singh, 2014). This instrument 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

[20] We only show results for the dependent variable *No. branches all banks* as this is the only variable for which our theoretical model yields a clear prediction for the sign of the coefficient.

countries only has an impact on domestic bank clustering via an increase in the probability that information sharing is introduced domestically as well.

Table 5 reports our IV results. The first stage shows a strong and positive correlation between the introduction of information sharing in neighboring countries in the recent past and the introduction of a credit registry or bureau in the country of observation. Both the Cragg-Donald Wald F-statistic and the Kleibergen-Paap F-statistic indicate that our instrument is strong. Moreover, the second-stage estimates are qualitatively very similar to our OLS baseline results, suggesting that endogeneity in the introduction of information sharing is not driving our results.

[Insert Table 5 here]

As we noted before, the inclusion of bank fixed effects is partially compromised by several banks that enter, exit and merge before and after the introduction of information sharing. Despite this complicating issue, we include bank fixed effects for all banks present during part of or the entire sample period. We rerun all regressions in our main Table 2 and report the estimates in Appendix Table A6. Overall, the results are very similar as the coefficient estimates are close in sign, size and significance levels.

Table 6 reports a battery of robustness tests related to the clustering of our standard errors. Our baseline approach throughout the paper is to report robust standard errors without clustering. This is because our regressions are based on the full population of branch openings rather than a (clustered) sample of openings (Abadie et al., 2017). Yet, because our information-sharing treatment is at the country level, it may nevertheless be advisable to cluster at this level. Since we include country (and year) fixed effects, clustering by country should only matter in case treatment effects are heterogeneous (Abadie et al., 2017). This does not appear to be the case. Table 6 shows that our results hold when clustering standard errors by bank (columns 1-3), country (columns 4-6), or year (columns 7-9). Moreover, our results go through when clustering by bank*locality (each bank can have several observations in the same locality if it opens branches in that locality at different points in time, columns 10-12), by country*year (columns 13-15), or by bank*year (columns 16-18).

[Insert Table 6 here]

Finally, Table 7 reports a placebo test. For each year, we keep the same number of information sharing introductions but instead of using the actual event countries, we assume that information sharing was introduced in another country group of the same size. This placebo country group is chosen randomly out of the total set of countries that at that point in time had not (yet) introduced information sharing. This approach thus preserves the cross-country trend in information sharing introductions, but randomly reallocates these events from the actual to placebo countries. We repeat this random reallocation 500 times and report the average estimation results. As expected, the results disappear in this placebo test, suggesting that it is unlikely that our results reflect unobservable characteristics or linear trends.

[Insert Table 7 here]

5.5. Extension: 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 empirically test this prediction, we merge our branch data with information from the Kompass database on firm-bank relationships. Kompass provides firm-level data including address, industry, and – critically for our purposes – the primary bank relationship(s) (see also Giannetti and Ongena, 2012 and Ongena, Peydró and Van Horen, 2015). 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 from their primary lender (using BEPS II information) and calculate the distance from the firm to each of these branches. 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, Slovakia, and Ukraine) but because the matching of bank information is very poor for Slovakia 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 (the 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 8 shows summary statistics and a two-sample t-test with unequal variances. In the countries that introduced information sharing between 2000 and 2005 (the Czech Republic and Poland), 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 also test this more formally in a differences-in-differences regression framework (lower panel of Table 9). Column (1) shows that after the introduction of information sharing, firms borrow from branches that are around 15 km further away as compared to firms in countries that did not introduce information sharing during the same period.

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 relatively opaque firms. For these firms, information asymmetries were initially more of an issue and the new publicly available information will therefore have more ‘bite’. To test whether this is indeed the case, we use the Kompass data to construct three dummy variables that proxy for a firm’s opaqueness. These are 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*). 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 8 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 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 8 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 then observe how these shocks play out at a very disaggregated level (that of individual villages, towns, and cities) across a variety 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 have a presence. Finally, our model indicates that more effective information sharing systems gradually dampens branch clustering.

In the empirical part of the paper, we then 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 domestic banks. We also show that after the establishment of information sharing banks are more likely to locate new branches in localities where they themselves did not have a branch presence yet. Importantly, we show that as a result of these changes the average firm is able to borrow from more distant bank branches.

Taken together, our results indicate that branch clustering is a function of the public availability of trustworthy, hard borrower information. When such information is more broadly available, banks – especially new players such as foreign-owned 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 relatively more important for banks to move closer to each other than to be closer to their potential clients.

Together, these effects 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

Abadie, A., S. Athey, G. Imbens and J. Wooldridge, (2017). When should you adjust standard errors for clustering? NBER Working Paper No. 24003.

Abiad, A. and A. Mody (2015). Financial reform: What shakes it? What shapes it? *American Economic Review, 95*(1), 66-88.

Agarwal, S., and Hauswald, R. (2010). Distance and private information in lending. *Review of Financial Studies, 23*(7), 2757-88.

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.H.L., Degryse, H., De Haas, R.T.A. and N. van Horen (2018), When arms' length is too far. Relationship lending over the credit cycle, *Journal of Financial Economics*, *127*, 174-196.

Beck, T.H.L., V. Ioannidou, and L. Schäfer (2017), Foreigners vs. natives: Bank lending technologies and loan pricing, *Management Science*, forthcoming.

Beck, T., Levine, R. and A. Levkov (2010). Big bad banks? The winners and losers from bank deregulation in the United States. *Journal of Finance, 65*(5), 1712-30.

Benfratello, L., F. Schiantarelli and A. Sembenelli (2008), "Banks and Innovation: Microeconometric Evidence on Italian Firms", *Journal of Financial Economics, 90*(2), 197-217.

Berger, A. N., & Mester, L. J. (2003). Explaining the dramatic changes in performance of US banks: Technological change, deregulation, and dynamic changes in competition. *Journal of Financial Intermediation, 12*(1), 57-95.

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

Berkowitz, D., M. Hoekstra, and K. Schoors (2014), "Bank Privatization, Finance, and Growth", *Journal of Development Economics, 114*, 93-106.

Bharath, S. T., Dahiya, S., Saunders, A., and Srinivasan, A. (2007). So what do I get: A bank's view of lending relationships. *Journal of Financial Economics, 85*(2), 368-419.

Brown, M., Guin, B., and Kirschenmann, K. (2015). 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.

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, mimeo.

Claessens, S., & 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.

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. *The Annals of Regional Science, 49*(2), 417-445.

Favara, G. and J. Imbs (2015). Credit supply and the price of housing. *American Economic Review, 105*(3), 958-992.

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

Guiso, L., P. Sapienza and L. Zingales (2004). Does local financial development matter?, *Quarterly Journal of Economics, 119,* 929-69.

Gupta, R. C., and Tao, H. (2010). A generalized correlated binomial distribution with application in multiple testing problems. *Metrika, 71*(1), 59-77.

Hauswald, R., and R. Marquez. 2006. Competition and Strategic Information Acquisition in Credit Markets. *Review of Financial Studies, 19,* 967–1000.

Herrera, A.M. and R. Minetti (2007), Informed finance and technological change: Evidence from credit relationships. *Journal of Financial Economics, 83*, 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. *The Economic Journal, 39*(153), 41-57.

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

Jayaratne, J. and P.E. Strahan (1996). The finance-growth nexus: Evidence from bank branch deregulation, *Quarterly Journal of Economics, 111*(3), 639-70.

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.

Martinez-Peria, M. S., and Singh, S. (2014). The impact of credit information sharing reforms on firm financing? *World Bank Policy Research Working Paper*, (7013).

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

Morgan, D., Pinkovsky, M. and B. Yang (2010), Banking Deserts, Branch Closings, and Soft Information, *Liberty Street Economics,* March 7, Federal Reserve Bank of New York.

Nguyen, H-L. Q. (2017). Are credit markets still local? Evidence from bank branch closings, *American Economic Journal: Applied Economics,* forthcoming.

Ongena, S., Peydró, J. L. and N. Van Horen (2015). Shocks abroad, pain at home? Bank-firm-level evidence on the international transmission of financial shocks. *IMF Economic Review, 63*(4), 698-750.

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.

Park, K., and Pennacchi, G. (2009). Harming depositors and helping borrowers: The disparate impact of bank consolidation. *Review of Financial Studies, 22*(1), 1-40.

Petersen, M.A., and R.G. Rajan (2002). Does Distance Still Matter? The Information Revolution in Small Business Lending. *Journal of Finance, 57*, 2533–70.

Poghosyan, T., and Poghosyan, A. (2010). Foreign bank entry, bank efficiency and market power in Central and Eastern European Countries. *Economics of Transition, 18*(3), 571-598.

Reuters (2016), Britain's Poorer Communities, Hit Hardest by Bank Closures, May Face Shut Out, available at http://uk.reuters.com/article/uk-britain-banks-branches-idUKKCN0Z60BP.

Rice, T. and P.E. Strahan (2010), Does credit competition affect small-firm finance? *Journal of Finance, 65*(3), 861-89.

Salop, S. C. (1979). Monopolistic competition with outside goods. *Bell Journal of Economics*, 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., "Credit market structure and information sharing mechanisms," in A. van Witteloostuijn, eds, *Studies in Industrial Organization*. Dordrecht: Kluwer Academic Publishers, 1994, pp. 129-143.

Wooldridge, J. (2007). What's new in econometrics?, mimeo. NBER Summer Institute.

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.
Year of branch opening	33,716	2004	2005	4.06	1995	2012
Year of bank establishment	33,716	1992	1992	14.26	1873	2011
Dependent variables						
No. branches all banks w/i 2 km	33,716	90	17	202	0	1,575
No. branches all banks w/i 5 km	33,716	110	23	209	0	1,575
No. branches all banks w/i same locality	33,716	129	25	232	0	1,574
Branch same bank w/i 2 km	33,716	0.33	0	0.47	0	1
Branch same bank w/i 5 km	33,716	0.37	0	0.48	0	1
Branch same bank w/i same locality	33,716	0.38	0	0.49	0	1
No. branches same bank w/i 2 km	33,716	3	0	9.35	0	202
No. branches same bank w/i 5 km	33,716	3	0	10.12	0	202
No. branches same bank w/i same locality	33,716	4	0	11.95	0	202
HHI w/i 2 km	33,716	0.21	0.13	0.24	0	1
HHI w/i 5 km	33,716	0.21	0.13	0.24	0	1
HHI w/i same locality	33,716	0.22	0.14	0.24	0	1
Independent variables						
Information sharing	33,716	0.55	1	0.50	0	1
Quality information sharing	17,807	1.45	0	2.16	0	6
Branch by foreign bank	33,716	0.44	0	0.50	0	1
Branch by greenfield foreign bank	33,716	0.14	0	0.35	0	1
Has email address	14,308	0.60	1	0.49	0	1
Has tax number	14,308	0.74	1	0.44	0	1
Has formal opening hours	14,308	0.74	1	0.44	0	1

Table 2
Information Sharing and the Geographical Clustering of Bank Branches

This table reports OLS and probit (columns 4-6) regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variables measure the pre-existing branch structure in each locality where a new bank branch is opened. Table A1 contains all definitions and Table 1 the summary statistics for each variable. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable →	No. branches all banks w/i			Branch same bank w/i			No. branches same bank w/i			HHI w/i		
	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Information sharing	84.32***	69.91***	55.45***	-0.08***	-0.10***	-0.11***	-0.67***	-1.51***	-2.68***	-0.01***	-0.01**	-0.02***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.023)	(0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.212	0.177	0.147	0.111	0.115	0.118	0.053	0.054	0.076	0.065	0.059	0.057
Observations	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716

Table 3
Information Sharing and the Geographical Clustering of Foreign Bank Branches

This table reports OLS and probit (columns 4-6) regressions to estimate the impact of the introduction of information sharing on subsequent bank branch clustering by foreign banks versus domestic banks. The dependent variables measure the pre-existing branch structure in each locality where a new bank branch is opened. Table A1 contains all definitions and Table 1 the summary statistics for each variable. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable →	No. branches all banks w/i			Branch same bank w/i			No. branches same bank w/i			HHI w/i		
	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Information sharing	112.51***	95.18***	74.39***	-0.05***	-0.07***	-0.09***	0.31	-0.77***	-2.30***	0.02***	0.01*	-0.00
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.167)	(0.002)	(0.000)	(0.008)	(0.090)	(0.962)
Branch by foreign bank	79.38***	83.68***	84.30***	0.06***	0.05***	0.03***	0.24	0.10	-0.22	0.04***	0.01***	0.01
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.124)	(0.550)	(0.243)	(0.000)	(0.001)	(0.107)
Information sharing * Branch by foreign bank	-60.29***	-54.19***	-40.88**	-0.06***	-0.06***	-0.04***	-2.05***	-1.56***	-0.80***	-0.06***	-0.04***	-0.04***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.223	0.190	0.160	0.112	0.116	0.118	0.056	0.056	0.076	0.068	0.061	0.059
Observations	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716

Table 4
Quality of Information Sharing and Bank Branch Clustering

This table reports OLS regressions to estimate the impact of the introduction of information sharing and of the quality of this information sharing system on subsequent bank branch clustering. The dependent variables measure the pre-existing branch structure in each locality where a new bank branch is opened. Table A1 contains all definitions and Table 1 the summary statistics for each variable. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable →	No. branches all banks w/i					
	2 km	5 km	Locality	2 km	5 km	Locality
	(1)	(2)	(3)	(4)	(5)	(6)
Information sharing	75.30***	64.90***	51.48***			
	(0.000)	(0.000)	(0.000)			
Information sharing * Quality info. sharing	-5.17***	-5.72***	-6.55***			
	(0.000)	(0.000)	(0.000)			
Quality information sharing				58.31***	52.85***	45.63***
				(0.000)	(0.000)	(0.000)
Quality information sharing^2				-14.92***	-13.66***	-12.02***
				(0.000)	(0.000)	(0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.195	0.151	0.109	0.199	0.154	0.112
Observations	17,807	17,807	17,807	17,807	17,807	17,807

Table 5
Robustness: IV Framework

This table reports IV regressions to estimate the impact of the introduction of information sharing on subsequent bank branch clustering. The instrument in the first stage is the percentage of neighboring countries that introduced information sharing in the previous five years. The dependent variables in the second stage measure the pre-existing branch structure in each locality where a new bank branch is opened. Table A1 contains all definitions and Table 1 the summary statistics for each variable. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

	First stage	Second stage											
Dependent variable →	Information sharing	No. branches all banks w/i			Branch same bank w/i			No. branches same bank w/i			HHI w/i		
		2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
% neighboring countries that introduced information sharing in the past five years	0.28***	-	-	-	-	-	-	-	-	-	-	-	-
	(0.000)	-	-	-	-	-	-	-	-	-	-	-	-
Information sharing	-	139.84***	67.66***	9.79	-0.37***	-0.21**	-0.31***	-5.86***	-9.30***	-12.27***	-0.10**	-0.11***	-0.14***
	-	(0.000)	(0.000)	(0.745)	(0.000)	(0.012)	(0.000)	(0.000)	(0.000)	(0.000)	(0.014)	(0.008)	(0.001)
Cragg-Donald Wald F statistic	511.52	-	-	-	-	-	-	-	-	-	-	-	-
Kleibergen-Paap rk Wald F statistic	317.05	-	-	-	-	-	-	-	-	-	-	-	-
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716

Table 6
Robustness: Clustering of Standard Errors

This table reports OLS regressions to estimate the impact of the introduction of information sharing on subsequent bank branch clustering. The dependent variables measure the pre-existing branch structure in each locality where a new bank branch is opened. Standard errors are clustered by bank (columns 1-3), country (columns 4-6), year (columns 7-9), bank*locality (columns 10-12), country and year (columns 13-15) or bank and year (columns 16-18). Table A1 contains all definitions and Table 1 the summary statistics for each variable. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable →	No. branches all banks w/i								
	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Information sharing	84.32***	69.91***	55.45***	84.32***	69.91**	55.45*	84.32**	69.91**	55.45*
	(0.000)	(0.000)	(0.007)	(0.010)	(0.029)	(0.070)	(0.017)	(0.037)	(0.073)
Clustering Standard Errors	Bank	Bank	Bank	Country	Country	Country	Year	Year	Year
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.212	0.177	0.147	0.212	0.177	0.147	0.212	0.177	0.147
Observations	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716

Dependent variable →	No. branches all banks w/i								
	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Information sharing	84.32***	69.91***	55.45***	84.32***	69.91***	55.45***	84.32***	69.91***	55.45***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.003)
Clustering Standard Errors	Bank*Locality	Bank*Locality	Bank*Locality	Country*Year	Country*Year	Country*Year	Bank*Year	Bank*Year	Bank*Year
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.212	0.177	0.147	0.212	0.177	0.147	0.212	0.177	0.147
Observations	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716

Table 7
Placebo Test: Random Introduction of Information Sharing Across Countries

This table reports placebo OLS regressions to estimate the impact of the introduction of information sharing on subsequent bank branch clustering. The dependent variables measure the pre-existing branch structure in each locality where a new bank branch is opened. For each year we keep the actual number of credit registry introductions but instead of using the actual countries that introduced information sharing in that year, we use the same number of placebo countries (randomly picked from the total set of countries that in that year had not (yet) introduced information sharing). We repeat this randomization and estimation 500 times and present average results. Table A1 contains all definitions and Table 1 the summary statistics for each variable. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable →	No. branches all banks w/i			Branch same bank w/i			No. branches same bank w/i			HHI w/i		
	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Information sharing	-4.40	-3.57	-2.71	0.00	0.01	0.00	-0.04	0.01	0.08	-0.08	-0.00	-0.00
	(0.177)	(0.357)	(0.592)	(0.779)	(0.738)	(0.712)	(0.732)	(0.969)	(0.752)	(0.741)	(0.713)	(0.765)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716

Table 8
Information Sharing and Spatial Credit Rationing

This table reports, by country, summary statistics for the variable *Firm-branch distance* and OLS 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. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

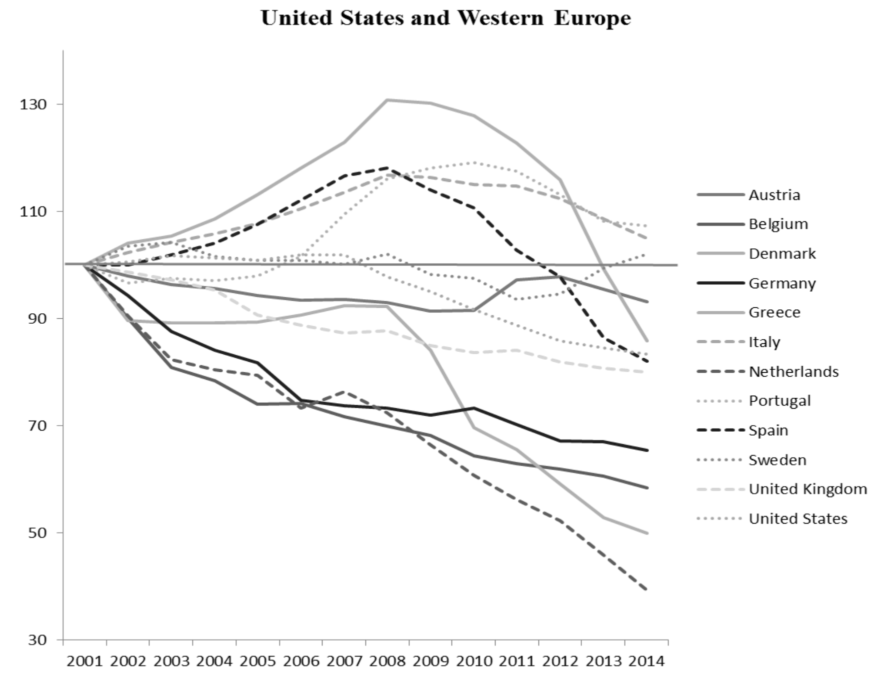
Dependent variable → Firm-branch distance (in km)

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
Bank Branch Dynamics in the U.S. and Europe: 2001-2014

These charts show the development of the total number of bank branches by country. If a foreign bank has several branches in a host country then these are counted as a single branch in the case of European countries (in line with ECB methodology). Source: FDIC and ECB.



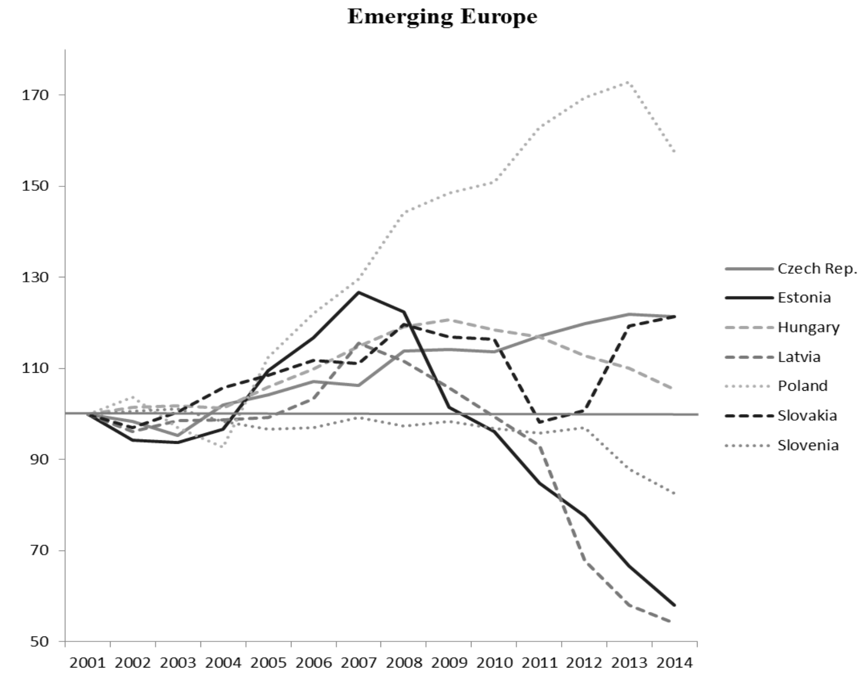
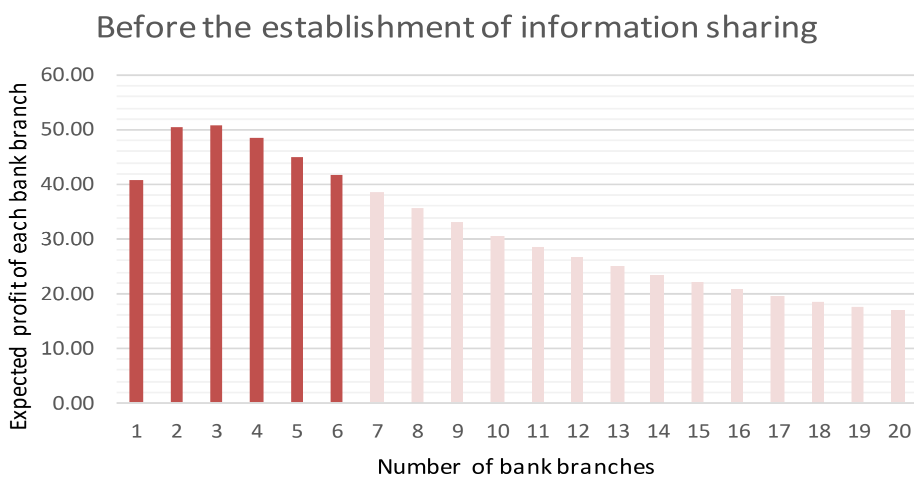


Figure 2
Branch Clustering after the Establishment of Information Sharing
No Overlap Among Bank Localities

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 presents 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 clustering).



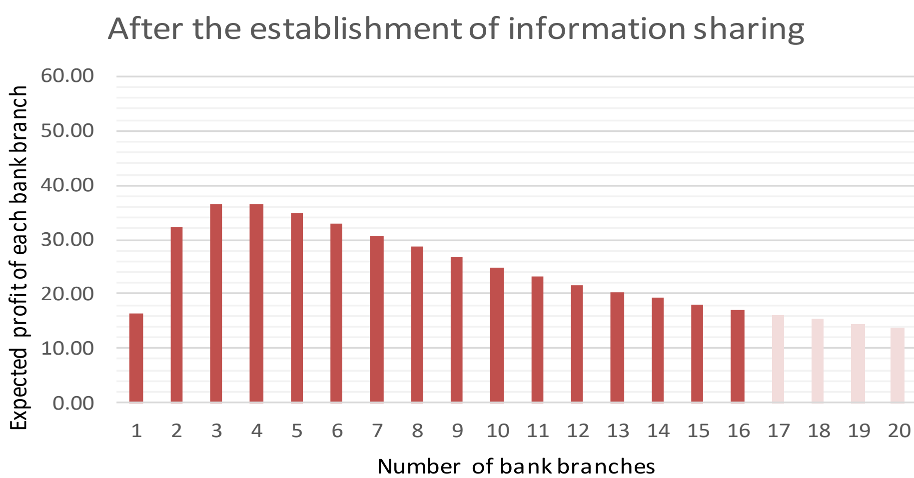
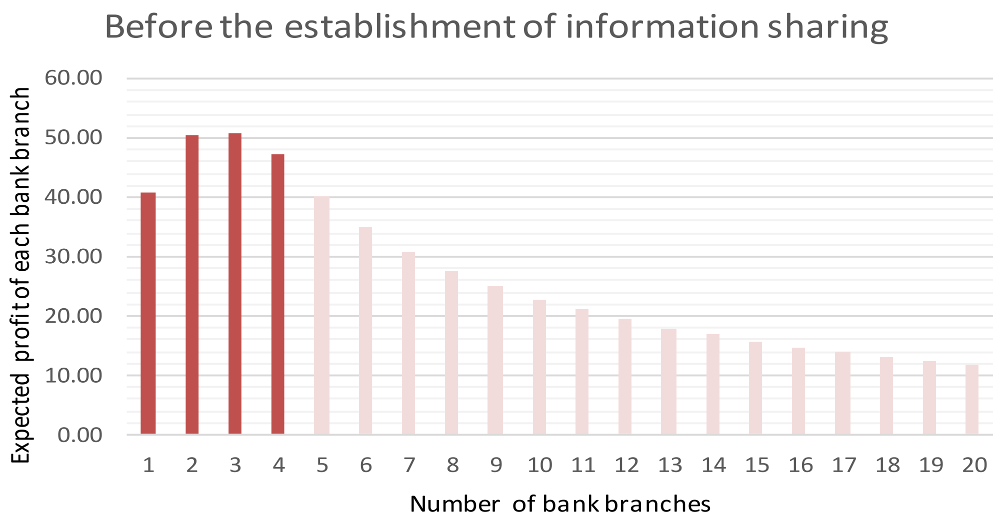


Figure 3
Branch Clustering after the Establishment of Information Sharing
Overlap Among Bank Localities

This figure presents comparative statics based on a calibration of our theoretical model that assumes 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 number of bank branches in locality s is 20 and the distance m between locality d and s is 12. The vertical axis presents 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 (shown by the first column at the very left) is larger (smaller) than the expected profit of opening such a branch in a new 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 4 branches is still higher than the expected profit of operating alone (which is just above 40). Adding a 5th 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 14 (until the profit of operating alone is higher than clustering).



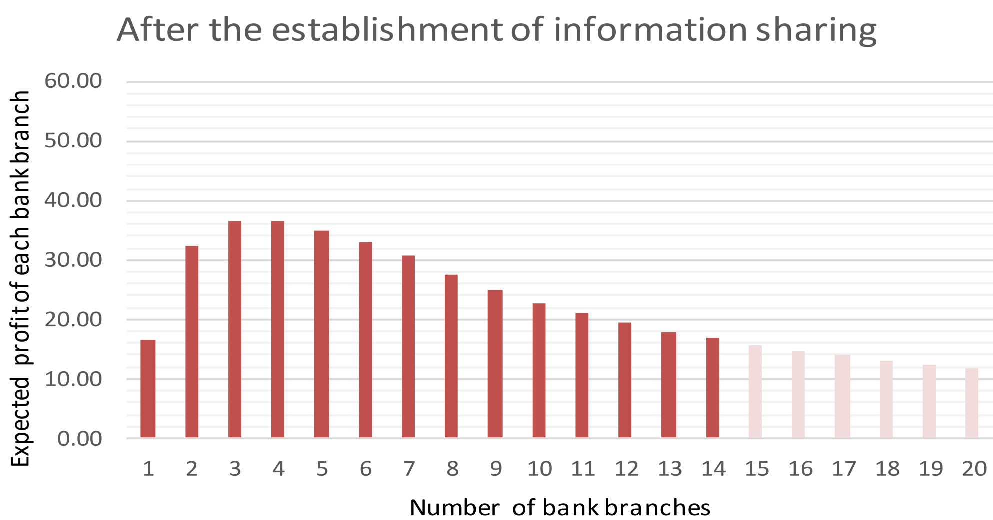


Figure 4
The Impact of Information Sharing on Domestic and Foreign Banks

This figure analyzes the difference between the impact of information sharing on domestic and foreign banks. Information sharing reduces the cost of screening distant borrowers from 6.5 to 6 for domestic banks and from 6.2 to 6 for foreign banks. The introduction of information sharing therefore induces more branch clustering by domestic banks than by foreign banks. Dark (light) bars show results without (with) overlap among nearby localities.

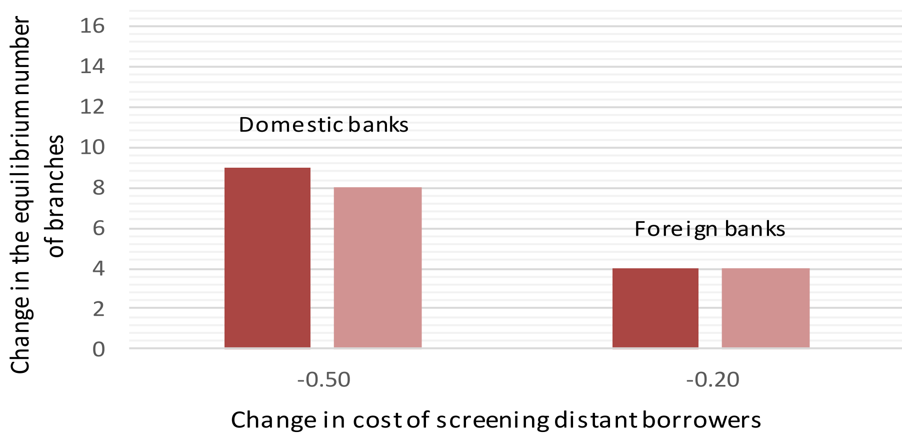


Figure 5
Quality of Information Sharing

This figure shows the declining equilibrium level of branch clustering when the inter-branch correlation of loan-rejection rates increases as information sharing systems become more effective (from 0.2 to 0.3). Bank branch clustering decreases when the quality of information sharing increases (with higher correlation among bnaks in loca rejection decisions).

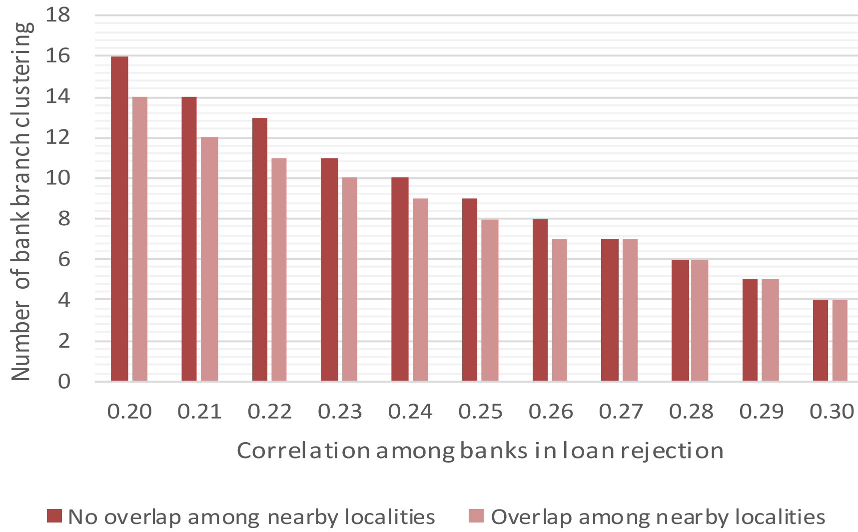


Figure 6
Distribution of Localities with Bank Branches in 1995 and in 2012

Panel A. This map plots all localities in our dataset with at least one bank branch in 1995. Source: BEPS II survey.

Panel B. This map plots all localities in our dataset with at least one bank branch in 2012. Source: BEPS II survey.

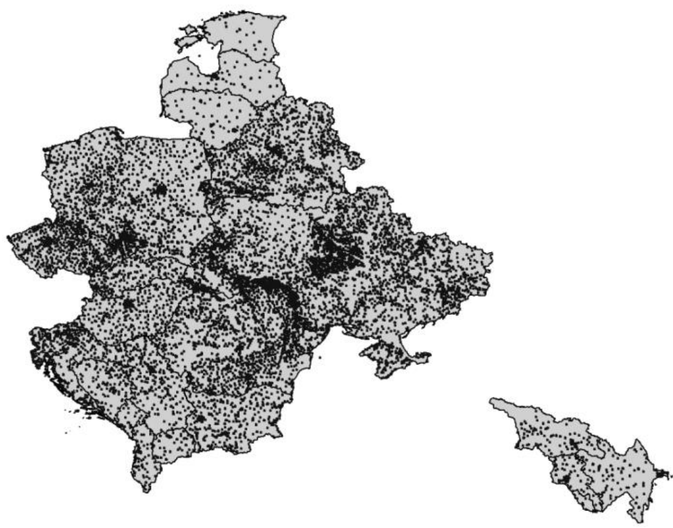


Figure 7
Branch Clustering Before and During Information Sharing

The first two charts show the trend in the number of pre-existing branches of the same bank and all banks, respectively, within a 2 km radius around newly opened branches in the four years before and the four years after the establishment of information sharing at t=0. The third chart shows the same for the HHI index in the 2 km around newly opened branched. Table A1 contains all variable definitions.

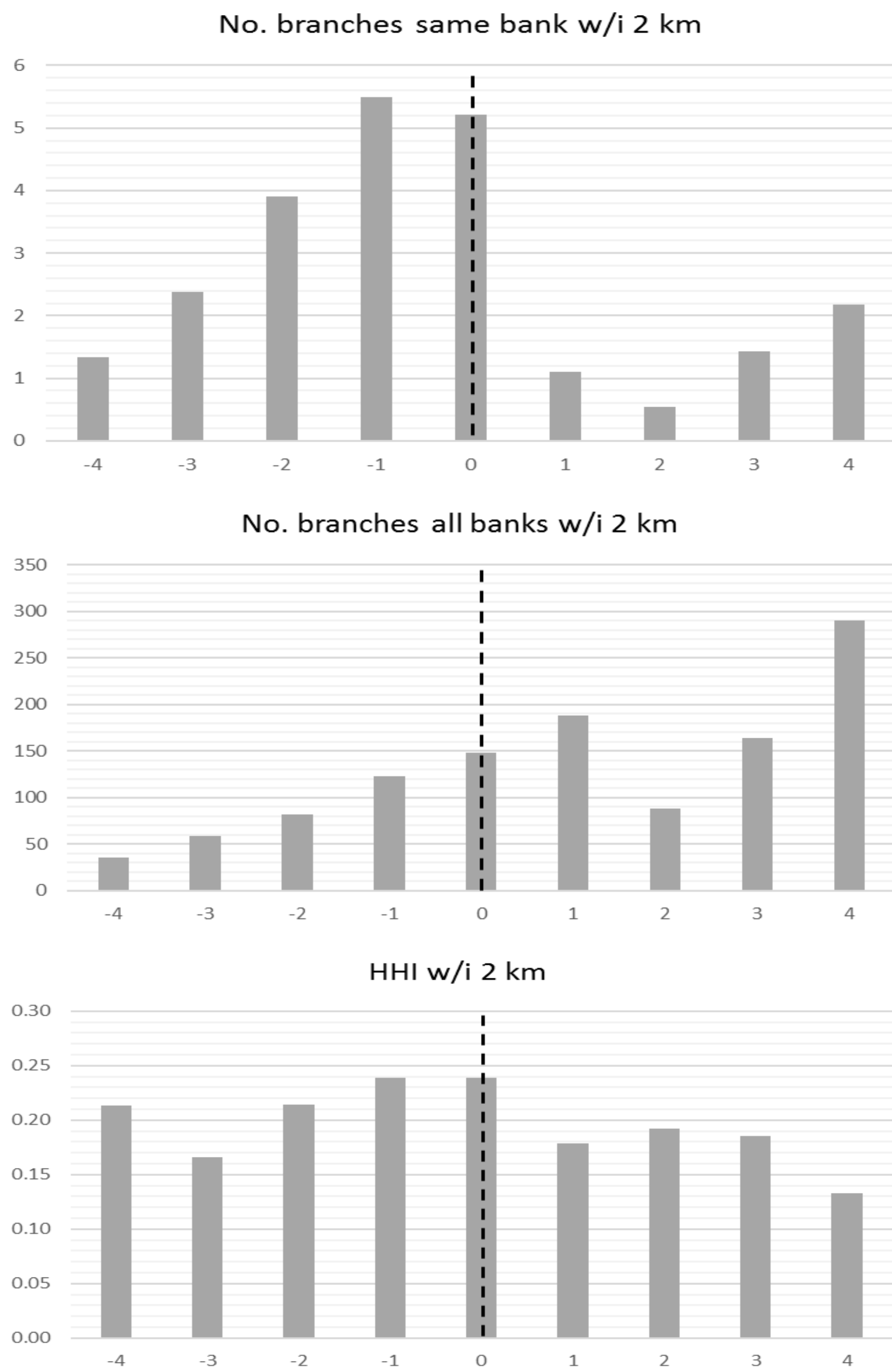


Table A1
Variable Definitions and Sources

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 in 32 countries among 611 banks. "Own calculations" indicates authors' own calculations based on BEPS II; "Doing Business" is the Doing Business Database by the World Bank.

Variable Name	***Definition***	***Data Sources***
Year of branch opening	The year a bank branch was opened	BEPS II
Year of bank establishment	The year a bank was established or entered a country (either as a greenfield or by taking over an existing bank)	Various sources
Dependent variables		
No. branches all banks w/i x km	Number of existing branches of all banks within a radius x kilometres around the new branch	BEPS II
No. branches all banks w/i same locality	Number of existing branches of all banks within the same locality of the new branch	BEPS II
Branch same bank w/i x km	= 1 if there is an existing branch of the same bank within a radius of x kilometers around the new branch, = 0 otherwise	BEPS II
Branch same bank w/i same locality	= 1 if there is an existing branch of the same bank within the same locality of the new branch, = 0 otherwise	BEPS II
No. branches same bank w/i x km	Number of existing branches of the same bank within a radius of x kilometres around the new branch	BEPS II
No. branches same bank w/i same locality	Number of existing branches of the same bank within the same locality of the new branch	BEPS II
HHI w/i x km	Herfindahl-Hirschman Index of bank concentration within a radius of x kilometers around the new branch	BEPS II
HHI w/i same locality	Herfindahl-Hirschman Index of bank concentration within the same locality of the new branch	BEPS II
Independent variables		
Information sharing	= 1 if there is information sharing in the country in that year, = 0 otherwise	Various sources
Quality information sharing	= 0 to 6, measures the quality of information sharing in the country	Doing Business
Branch by foreign bank	= 1 if the new branch is opened by a foreign bank, = 0 otherwise	Claessens and Van Horen (2014)
Branch by greenfield foreign bank	= 1 if the new branch is opened by a foreign greenfield bank, = 0 otherwise	Claessens and Van Horen (2014)
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

Table A2
Correlation Matrix

This table reports a correlation matrix for our main variables. Tables A1 and 1 contain all definitions and summary statistics, respectively. Correlation coefficients (p-values) are listed in the first (second) row.

	Year of branch opening	Year of bank establishment	No. branches all banks w/i same locality	Branch same bank w/i same locality	No. branches same bank w/i same locality	HHI w/i same city	Information sharing	Quality information sharing	Branch by foreign bank	Branch by greenfield foreign bank
Year of branch opening	1.0000									
Year of bank establishment	0.0569	1.0000								
	(0.000)									
No. branches all banks w/i same locality	0.2416	0.1453	1.0000							
	(0.000)	(0.000)								
Branch same bank w/i same locality	0.2755	-0.1306	0.3205	1.0000						
	(0.000)	(0.000)	(0.000)							
No. branches same bank w/i same locality	0.1822	-0.1146	0.3605	0.4413	1.0000					
	(0.000)	(0.000)	(0.000)	(0.000)						
HHI w/i same city	-0.0410	-0.0351	-0.2423	-0.1073	-0.1158	1.0000				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
Information sharing	0.5974	0.0862	0.2077	0.0860	0.0965	-0.0575	1.0000			
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
Quality information sharing	0.2427	-0.0383	-0.0280	0.0143	-0.0040	-0.0709	0.1281	1.0000		
	(0.000)	(0.000)	(0.000)	(0.009)	(0.462)	(0.000)	(0.000)			
Branch by foreign bank	-0.0275	0.0236	0.0743	-0.0119	0.0009	-0.0302	0.1996	0.0961	1.0000	
	(0.000)	(0.000)	(0.000)	(0.029)	(0.872)	(0.000)	(0.000)	(0.000)		
Branch by greenfield foreign bank	-0.0576	0.2605	0.0637	-0.1306	-0.0881	-0.0666	0.1282	0.0556	0.4656	1.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Table A3
Overview of Branch Openings and Closures

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

Year	Opened branches	Closed branches	Country	Opened branches	Closed branches
1995	2,391	0	Albania	443	11
1996	490	0	Armenia	448	19
1997	603	0	Azerbaijan	335	13
1998	552	0	Belarus	2,481	9
1999	555	0	Bosnia	617	10
2000	987	6	Bulgaria	1,405	100
2001	1,440	3	Croatia	608	48
2002	1,440	11	Czech Rep.	382	19
2003	2,735	10	Estonia	60	56
2004	4,675	36	Georgia	703	108
2005	2,391	24	Hungary	1,538	287
2006	2,700	25	Latvia	195	9
2007	7,999	66	Lithuania	94	0
2008	1,928	111	Macedonia	189	16
2009	665	269	Moldova	1,300	180
2010	789	272	Montenegro	206	12
2011	1,095	287	Poland	3,192	51
2012	281	248	Romania	2,053	177
			Serbia	1,080	227
			Slovakia	153	0
			Slovenia	157	16
			Ukraine	16,077	0
Total	*33,716*	*1,368*	*Total*	*33,716*	*1,368*

Table A4
Introduction of Information Sharing

This table provides an overview of the introduction years of public credit registries and private credit bureaus in our 22 sample countries. N.a.: No credit bureau or registry has as yet been introduced in this country. Source: World Bank Doing Business Database, EBRD and various publications and websites.

Country	Public Credit Registry	Private Credit Bureau
Albania	2008	2009
Armenia	2003	2004
Azerbaijan	2005	n.a.
Belarus	2007	n.a.
Bosnia	2006	2001
Bulgaria	1999	2005
Croatia	n.a.	2007
Czech Republic	2002	2002
Estonia	n.a.	2001
Georgia	n.a.	2005
Hungary	n.a.	1995
Latvia	2003	n.a.
Lithuania	1995	2003
Macedonia	1998	2010
Moldova	n.a.	2011
Montenegro	2008	n.a.
Poland	n.a.	2001
Romania	2000	2004
Serbia	2002	2004
Slovak Republic	1997	2004
Slovenia	1994	2008
Ukraine	n.a.	2007

Table A5
Information Sharing and the Geographical Clustering of Greenfield Foreign Bank Branches

This table reports OLS and probit (columns 4-6) regressions to estimate the impact of the introduction of information sharing on subsequent bank branch clustering by greenfield foreign banks. The dependent variables measure the pre-existing branch structure in each locality where a new bank branch is opened. Table A1 contains all definitions and Table 1 the summary statistics for each variable. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable →	No. branches all banks w/i			Branch same bank w/i			No. branches same bank w/i			HHI w/i		
	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Information sharing	105.85***	87.99***	68.01***	-0.03***	-0.05***	-0.06***	0.12	-0.73***	-2.02***	0.00	-0.00	-0.01**
	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.541)	(0.001)	(0.000)	(0.814)	(0.978)	(0.036)
Branch by greenfield foreign bank	73.81***	84.76***	94.00***	-0.02**	-0.03***	-0.03***	-1.17***	-1.27***	-1.53***	-0.02***	-0.03***	-0.03***
	(0.000)	(0.000)	(0.000)	(0.017)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Branch by M&A foreign bank	68.93***	69.51***	67.00***	0.11***	0.11***	0.09***	0.55***	0.58***	0.40**	0.03***	0.01	-0.00
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.045)	(0.000)	(0.109)	(0.415)
Information sharing * Branch by greenfield foreign bank	-50.92***	-47.72***	-40.17***	-0.10***	-0.11***	-0.10***	-0.80***	-1.43***	-2.10***	-0.04***	-0.03***	-0.03***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Information sharing * Branch by M&A foreign bank	-54.51***	-45.30***	-30.72***	-0.08***	-0.07***	-0.06***	-2.30***	-1.50***	-0.13	-0.02**	-0.00	0.00
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.677)	(0.015)	(0.477)	(0.508)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.224	0.191	0.161	0.121	0.127	0.126	0.057	0.059	0.083	0.071	0.063	0.061
Observations	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716

Table A6
Robustness: Information Sharing and the Geographical Clustering of Bank Branches, Including Bank Fixed Effects

This table reports OLS and probit (columns 4-6) regressions to estimate the impact of the introduction of information sharing on bank branch clustering. The dependent variables measure the pre-existing branch structure in each locality where a new bank branch is opened. Table A1 contains all definitions and Table 1 the summary statistics for each variable. Robust p-values are reported in parentheses. ***, **, * correspond to the 1%, 5%, and 10% level of significance, respectively.

Dependent variable →	No. branches all banks w/i			Branch same bank w/i			No. branches same bank w/i			HHI w/i		
	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality	2 km	5 km	Locality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Information sharing	18.82***	10.54**	-2.30	-0.06***	-0.07***	-0.04***	-1.01***	-2.24***	-3.92***	-0.05***	-0.04***	-0.02***
	(0.000)	(0.011)	(0.667)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.431	0.385	0.330	0.264	0.285	0.280	0.218	0.225	0.255	0.149	0.135	0.127
Observations	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716	33,716

Figure A1
Impact of Information Sharing on Branch Clustering without Overlap of Bank Localities

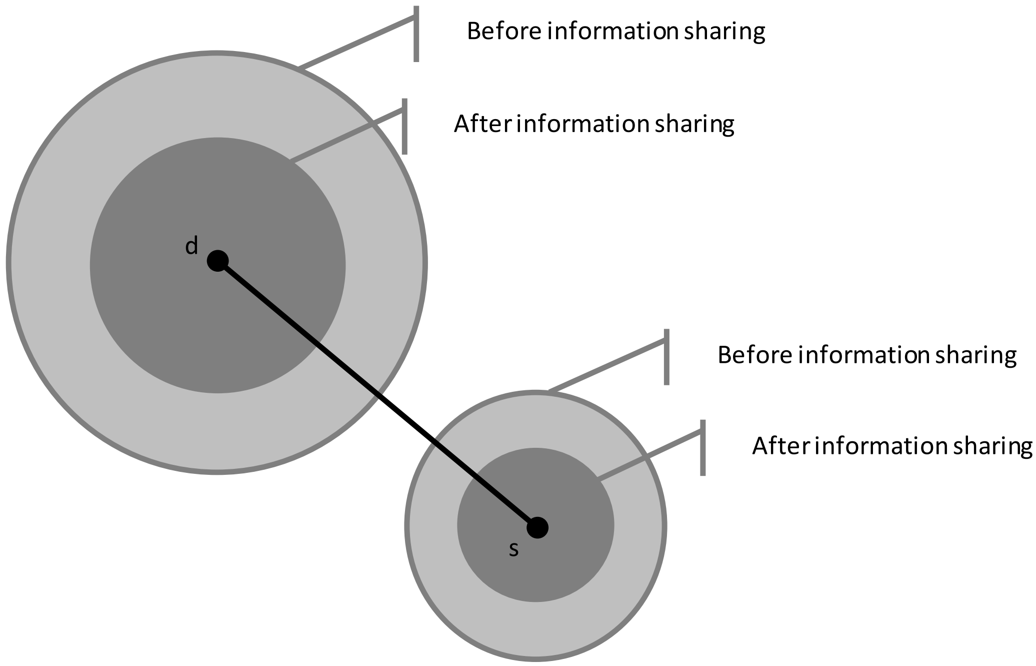


Figure A2
Impact of Information Sharing on Branch Clustering with Overlap of Bank Localities