Matching in the Corporate Loan Market

Fabian Gunzinger*

Abstract

We understand very little about what drives match sorting patterns in the corporate loan market – about why banks and firms lend to and borrow from the counterparties they do. In this paper, I use a large firm-level dataset from Japan and a many-to-many matching framework to make two contributions to such an understanding: I show that a set of information-based variables, like geographical distance and a previous lending relationship, and the objective of banks to focus or diversify their loan portfolio both drive match sorting patterns. These results are largely consistent with what theory predicts, and they help us better understand the formation of networks of interdependencies between firms and banks, which play an important role in determining the robustness of the financial system.

*Department of Economics and Lincoln College, University of Oxford. fa.gunzinger@gmail.com.
I am grateful to my supervisors Felix-Reed Tsochas and Peyton Young for their guidance and encouragement. And I thank Daniel Fricke, Barbara and Pierre-Alain Schlaffer-Bruchez, Konstantin Büchel, Alex Teytelboym, Marina Gertsberg, and Patrick Dendorfer for helpful comments and suggestions. I am grateful to the Berrow Foundation for generously supporting my graduate studies at Oxford. Finally, without the incredible technical support of Andy Gittings at the University of Oxford's Advanced Research Computing facility, the code to produce my results might still be running as I am writing this – many thanks to him for his time and effort, and to ARC for access to its facilities (http://dx.doi.org/10.5281/zenodo.22558).

Contents

1 Introduction

In 2013, Tekken (a construction company) was one of 396 firms that took out a new loan from Mizuho Bank. Not among these firms was Nishimatsu (also a construction company), which was instead one of 273 firms that borrowed from Resona Bank in that year. In the language of the matching literature, Tekken was matched to Mizuho and Nishimatsu was matched to Resona. We know very little about what drives these match sorting patterns in the corporate loan market – about why banks and firms match up with the counterparties that they do.

But these matching patterns have important implications for financial stability.[1] They shape the structure of networks of interdependencies among banks and between banks and firms, and the structure of these networks, in turn, determines to an important degree the robustness of the financial system and thus the frequency and the extent of financial crises.[2] Therefore, understanding what drives match sorting patterns in the corporate loan market is an important part of understanding the role of networks in causing and amplifying financial crises.

In this paper, I use a large firm-level dataset from Japan that covers 5073 firms and 179 banks between 1980 and 2014, and a many-to-many matching framework introduced by Fox (2016), to make two contributions to such an understanding. First, I show that a set of information-based characteristics drives match sorting patterns; banks and firms are more likely to agree to new loans if they are geographically close and when a bank has previously lent either to a firm directly or to other firms from the same sector. Theory suggests this is so because in a world of asymmetric information,

[1]The corporate loan market is central to economic performance, so that an understanding of its workings might also indirectly benefit a number of other areas of economic research such as long-run economic growth, because of the corporate loan market's role in capital allocation (Goldsmith 1969); business cycle research, because search frictions in the credit market amplify the business cycle (Wasmer and Weil 2004) and because frictions in the loan market are a key link between financial instability and macroeconomic downturns (Bernanke et al. 1999, Hall 2010); and work on monetary policy implementation, because the loan market is an important transmission channel of monetary policy (Beck et al. 2014, Hachem 2011).

[2]Financial crises are costly: they can lead to cumulated output losses of 15 to 20 percent of annual GDP (Hoggarth et al. 2002), amplify the adverse effects of economic downturns (Dell'Ariccia et al. 2008), and weaken government's fiscal position by increasing public debt levels (Reinhart and Rogoff 2013).

where firms are better informed about their credit-worthiness than banks are, and where banks monitor firms to close that informational gap, these characteristics reduce information asymmetries and monitoring costs, and thus make matches more likely. I also show that – contrary to earlier findings and to what theory suggests – similar relative size of banks and firms does not increase the likelihood of a match between a bank and a firm. Second, I show that banks' objective to focus or diversify their loan portfolio drives match sorting patterns; banks have complementarities between their borrowers so that their decision of whether to lend to any particular firm depends on what other firms they can lend to.

Relationship to prior literature: My work lies at the intersection of three strands of literature: relationship banking, matching, and financial networks. Relationship banking studies the nature of bank-firm relationships and the optimal number of lending relationships of firms. A large number of theoretical and empirical papers explores the determinants of switching costs and switching probabilities, relationship duration and scope, and the number and intensity of banking relationships. The main insights of the theoretical literature are that such relationships can facilitate Pareto-improving transactions, but can also lead to less discipline in the loan market and to higher borrowing rates for firms. The empirical literature shows that whether the positive or negative effects dominate varies for different types of banks and firms and across countries.[3]

The matching literature has two strands. One is concerned with the existence of stable, optimal, and unique allocations, the other with identifying the preferences of agents in matching markets. The seminal contribution to the former is Gale and Shapley (1962), who introduce the "deferred acceptance algorithm" and show that it leads to stable, optimal, and unique allocations in marriage markets and college admissions. The paper has inspired a large literature exploring stability, optimality, and uniqueness in different settings, and variants of the deferred acceptance algorithm have

[3]For a survey of the overall literature, see Boot (2000). Theoretical contributions are synthesised in Freixas and Rochet (2008), empirical ones in Degryse et al. (2009).

since been widely and fruitfully applied to solve allocation problems such as assigning pupils to schools, aspiring doctors to hospitals, and kidney donors to recipients. Recent advances in that literature by Hatfield et al. (2013) and in particular Azevedo and Hatfield (2015), who are the first to prove equilibrium existence if agents have complementarities in preferences, underpin my empirical analysis.[4]

The study of agents' preferences in matching markets goes back to Becker (1973), who first tackled the challenge of inferring the valuation functions of agents in a setting where there is data on actual matches but not transaction prices. Like Gale and Shapley, Becker's market of interest was that of marriages, and a large number of studies has since shed further light on how people choose their spouses.[5] In addition, the approach inspired by Becker has also been applied to study CEO pay, the allocation of microcredit, firm mergers, the allocation of faculty offices, the allocation of medical residents and many other areas.[6] The empirical framework of Fox (2016) uses as its theoretical foundation the result of Azevedo and Hatfield (2015), while its objective of identifying elements of match utility functions is in the tradition of Becker (1973).

Matching studies relationship formation, and a set of relationships gives rise to a network. Thus, matching is the process by which networks – including financial networks – form.[7] The literature on financial networks is concerned with the formation process and the stability properties of networks of interdependencies between actors in the financial system. Most of this work is focused on the network of interbank lending obligations and the main message from the literature is that network structure – together with other factors such as leverage and the size of institutions – can be a critical determinant of the likelihood and of the extent of contagion. Furthermore,

[4]For a discussion on the practical applications of matching, see Roth (2002) or Roth (2015). Key theoretical contributions are discussed in Che et al. (2013).

[5]See, for instance: Choo and Siow (2006), Dupuy and Galichon (2014), Hitsch et al. (2010), Chiappori et al. (2012), and Echenique et al. (2013).

[6]See Gabaix and Landier (2008) and Terviö (2008) for studies on CEO pay, Ahlin (2016) on microcredit allocation, Akkus et al. (2016) on firm mergers, Baccara et al. (2012) on faculty office allocations, and Agarwal (2015) on the allocation of medical residents and for a discussion of yet more applications.

[7]The literature on networks, and networks in economics, is too large to discuss here so that I focus on financial networks. For surveys on the broader literature, see Jackson (2008), Jackson (2014), and Jackson et al. (forthcoming).

contagion is often non-monotonic: the system is robust to shocks up to a certain size-threshold or to a particular set of institutions, but fragile to shocks above that threshold and to another set of institutions.[8]

Contribution of this paper: This paper is motivated by what I think is a blindspot in the relationship banking and financial network literature: an answer to the question of what characteristics help explain how banks and firms select their counterparties in the corporate loan market. Insights from relationship banking provide guidance on how to address this question, but most of the literature does not, to my knowledge, address it directly. Similarly, while some recent contributions in the financial network literature explore network formation, these models are highly stylised and do not help us answer that question.[9]

One paper that does address the question is Chen and Song (2013), who use data on bank-firm lending in the United States between 2000 and 2003 and a similar estimation approach based on Fox (2010). They find that distance, relative size and a previous lending relationship are predictors of a match between a bank and a firm. My first set of results corroborates, refines, and repudiates different parts of their findings: corroborates because I also find that geographical proximity and a previous lending relationship make a match between a bank and a firm more likely; refines because in addition to taking into account prior lending between a bank and a firm, I also account for prior lending between a bank and other firms from the same sector and find that, while both make a new match more likely, prior lending directly to the firm increases that likelihood by more; repudiates because contrary to their results, I find that similar relative size of a bank and a firm does not make a match between these two parties more likely.

My work differs from Chen and Song (2013) in that I use a many-to-many matching framework rather than a many-to-one matching framework. This seems more appropri-

[8]Seminal contributions are Allen and Gale (2000), Freixas et al. (2000), and Eisenberg and Noe (2001). For recent reviews of the literature, see Glasserman and Young (2016) and Hüser (2015).

[9]Among the papers on financial network formation are Acemoglu et al. (2015), Farboodi (2015), and Navarro and Castiglionesi (2016).

ate given that in my dataset, firms tend to have multiple banking relationships at any given time. Except for Fox (2016), who uses a many-to-many matching framework to study the market for car parts, I am not aware of any other empirical implementation of this framework. More importantly, the framework of Fox (2016) allows agents to have complementarities in their preferences, which allows me to estimate the relative importance of banks' objective to focus or diversify their portfolio as a driver of match sorting patterns. These results are, to the best of my knowledge, novel

2 Institutional context

Japan has a banking-oriented financial system, in which bank loans are the main source of funding for firms. The banking system is segregated into different types of banks, which differ in their geographical focus and in hte range of services they provide. The system is a legacy from World War II policy, whose aim it was to restrict competition among banks and to stabilise the banking system. While deregulation in the 1980s and 1990s has blurred the lines between different bank types, they still exist today.

The five main types of banks, and the ones for which I have data, are *city banks*, large universal banks with nationwide branches that dominate all segments of the Japanese banking sector, including corporate lending; *regional banks*, medium sized banks whose operations are – unsurprisingly – regionally focused; *second-tier regional banks*, which are also regionally focused but tend to focus on, and be an important source of funding of, small and medium sized enterprises; *long-term credit banks*, which no longer exist in their original form but whose purpose was to provide long-term funding in the post war period; and *trust banks*, which manage funds on behalf of their clients and usually invest long-term.[10]

Banks have strong ties to the firms they lend to. As Peek and Rosengren (2005) point out, corporate lending by Japanese banks is not only driven by their profit motive and careful credit analysis, but also by a government mandated responsibility and a

[10]A more comprehensive discussions on different bank types and other aspects of the Japanese financial system can be found in Hoshi and Kashyap (2001) and in Uchida and Udell (2014).

perceived "national duty" to support troubled firms. This attitude is reflected in the response to the 2007-09 financial crises. Harada et al. (2015) note that banks were encouraged by their main regulatory agency, the Financial Services Agency, to grant and to roll over loans to struggling firms and – under certain conditions – were allowed to exclude these loans when reporting their non-performing loans.

The responsibility to support a struggling firm is particularly strong for that firm's *main bank*. According to Aoki and Patrick (1995), almost every Japanese firm has a main bank, and almost every bank serves as main bank for some firms.[11] The relationship between a firm and its main bank involves five aspects: lending, public debt issuance, equity cross holdings, the provision of business settlement accounts, as well as the supply of information services and managerial resources.[12] The main bank often acts as a firm's primary lender and as such assumes two additional sets of responsibilities. First, when the firm is in financial distress, the main bank is the main source of funding, oversees the financial rescue, and – if necessary – coordinates restructuring and dismantling of the firm. It is thus likely that main banks take on most of the aforementioned mandated social responsibility for supporting struggling firms. Second, the main banks tend to act as the primary monitor of the firm; while firms borrow substantial sums from banks other than their main bank, those banks often delegate the entire monitoring – ex ante screening of potential projects, interim information gathering on ongoing business, and ex post verification of investment projects – to the main bank.

The main bank system has important implications for my analysis. Main bank relationships grew out of policy aimed to support economic recovery after the second

[11] There is an overlap between the main bank system and another uniquely Japanese institution: the *keiretsu* group. A keiretsu is a large industrial group of firms from different sectors centred around a large bank, often a city bank. The relationships within the group are characterised by a set of formal and informal ties including extensive cross-holding of shares and interlocking boards. For firms that are affiliated with a keiretsu, the keiretsu bank acts as the main bank. Because for my analysis it is the nature of the main bank relationship that matters, I do not distinguish between keiretsu and non-keiretsu main banks.

[12] A considerable literature, including Hoshi and Kashyap (2001), discusses the advantages and disadvantages of the main bank system. Among the benefits are intensive monitoring, secure access to credit, a reduced likelihood of financial distress, and reduced duplication of monitoring tasks. Potential downsides are the ability of main banks to extract rents, the difficulty for firms to access outside funding if the main bank is distressed, and "evergreening" (Peek and Rosengren 2005, Caballero et al. 2008).

world war, and the matching of firms to their main banks might well have been driven by factors other than profit maximisation. In addition, the main bank's strong responsibility to support struggling firms means that some of the loans between a main bank and a firm – those granted in times of firm distress – are driven not by profit maximisation but by that social responsibility motive. Because of that, I exclude loans that were granted by main banks from my analysis. This has the advantage that loans granted based on factors beyond what my model can capture are excluded. But it also comes at a cost: the information-based characteristics in my model are potential drivers of match sorting patterns because they make monitoring cheaper. To the extent that non-main banks delegate monitoring to the main bank, the importance of these information-based variables will be lesser.

3 Theoretical framework and estimation

The credit market is a matching market. Just as a bank that is willing to lend needs to find a firm that is willing to borrow, a firm that is willing to borrow needs to find a bank that is willing to lend. One implication of this need for mutual consent is that there is scarcity; banks and firms can and want to agree to a limited number of new loan contracts only, so that whenever two parties match up, it becomes a little less likely for everyone else in the market to match up with either of them. Another implication is that inferring agents' preferences is difficult. In a commodity market, knowing an agent's choice set is relatively simple because she could have bought everything on offer as long as she was willing and able to pay the quoted price. With the help of some revealed preference argument it is then relatively straightforward to infer something about the agent's preferences. In a matching market, we generally observe actual matches only and do not know whom our agent could also have matched up with. This is also the case in my dataset, where I only observe the loan contracts banks and firms actually agreed to, but I do not see what set of banks each firm could alternatively have borrowed form and what set of firms each bank could also have lent to. The challenge for empirical matching models is thus to infer preferences from observed matching patterns only.

When we think of banks and firms as utility maximisers, understanding why they match up with the counterparties they do comes down to understanding what characteristics of those counterparties drives their utility from a match: the higher their utility from matching to a counterparty with certain characteristics, the higher the likelihood that banks and firms will actually (try to) match up with that counterparty. This section introduces the conceptual framework of Azevedo and Hatfield (2015) that gives rise to match utility functions of banks and firms, and the empirical approach of Fox (2016) that provides a tool to identify the relative importance of different characteristics in determining match utilities.

3.1 Framework

This subsection introduces a special case of the more general model by Azevedo and Hatfield (2015) that establishes the existence of a unique, stable and efficient equilibrium in a two-sided matching market with transferable utility, heterogeneous preferences, and complementarities in preferences. Their model is more general than mine in that they consider two-sided trading networks in which agents can engage in multiple trades on both sides of a contract. In my setting, this would mean having only one agent type that can act both as a bank and as a firm.

The environment: The market for corporate loans has two sides, a finite and exogenously given set B of lending *banks*, and a finite and exogenously given set F of borrowing *firms*. B and F are disjoint so that each *agent* $i \in I \equiv B \cup F$ is either a bank or a firm. In the language of networks, the market for corporate loans is a bipartite network. Agents can engage in loan contracts ω, and each loan contract is associated with a bank $b(\omega) \in B$ and a firm $f(\omega) \in F$. The set of all possible loan contracts is Ω. There is one possible loan contract between each bank and each firm, to which the parties can either agree to or not, so that Ω is finite and $|\Omega| = |B| \times |F|$.[13] If agent i

[13]There is one possible contract between each bank-firm pair because the variable of interest in the empirical analysis is binary: it is whether or not a bank and a firm have agreed to a new loan contract

is a bank, $\Psi_i \subseteq \Omega$ is the set of all loan contracts in which it is the lender, while if i is a firm, $\Phi_i \subseteq \Omega$ is the set of all loan contracts in which it is the borrower. We can think of ω as representing the notional amount of a loan, while the terms and conditions of the loan (such as interest rate, maturity data, amount and quality of collateral) are represented by a *transfer* p_ω. The vector of transfers for all trades in the market is $p \in \mathbb{R}^{|\Omega|}$.

Agents are endowed with utility functions. If bank i is the lender in all loan contracts in Ψ_i, its utility is

$$v^i(\Psi_i) + \sum_{\omega \in \Psi_i} p_\omega, \tag{1}$$

where $v^i(\Psi_i)$ is the *valuation function*, which takes values in $(-\infty, \infty)$. A bank has *complementarities* in preferences if its valuation from matching with a certain firm depends on which other firms it is matched to. To illustrate, let ω_1 and ω_2 be two loan contracts with identical transfers between bank i and two different firms and suppose that i's preferences are given by

$$v^i(\omega_1, \omega_2) \succ v^i(\varnothing) \succ v^i(\omega_1) \sim v^i(\omega_2),$$

so that the bank prefers not lending at all to lending to only one of the firms, but prefers lending to both firms to not lending at all. Thus, the bank's utility from lending to one firm – and hence its willingness to do so – depends on whether it can also lend to the other firm. Banks that aim to focus or diversify their loan portfolio in a certain way have preferences of that kind. To capture this aim I make no assumption on the functional form of $v^i(\Psi_i)$ and will parametrise it below such as to allow for complementarities.

In contrast, I assume that firms do not have complementarities in preferences. If

in a given year of not, while the notional amount of that loan is irrelevant.

firm i is the borrower in all loan contracts in Φ_i, its utility is given by

$$\sum_{\omega \in \Phi_i} v^i(\omega) - p_\omega, \tag{2}$$

where $v^i(\omega)$ is again the valuation function. Without complementarities in preferences, the firm's total valuation of borrowing from a given set of banks is the sum of its valuation of each contract. For banks that do not lend and firms that do not borrow, $v^i(\varnothing)$ can be normalised to zero, and to ensure that banks only act as lenders and firms only act as borrowers, we can define $v^i(\Phi_i, \Psi_i) = -\infty$.

An *allocation* A is a map $A : I \to \Delta(\mathcal{P}(\Omega))$, which specifies for each agent i a probability distribution A^i over the space of all possible sets of loan contracts; $A^i(\Psi_i)$ is the probability that bank i is the lender in each contract $\omega \in \Psi_i$, and $A^i(\Phi_i)$ is the probability that firm i is the borrower in each contract $\omega \in \Phi_i$.[14]

Equilibrium definition: An *arrangement* (A, p) consists of an allocation A and a vector of transfers p. (A, p) is a competitive equilibrium if A is incentive compatible and feasible.[15] A is incentive compatible if for all banks, $A^i(\Psi_i) > 0$ only if

$$\Psi_i \in \underset{\tilde{\Psi}_i \subseteq \mathcal{P}(\Omega)}{\arg\max} \left(v^i(\tilde{\Psi}_i) + \sum_{\omega \in \tilde{\Psi}_i} p_\omega \right),$$

and for all firms, $A^i(\Phi_i) > 0$ only if

$$\Phi_i \in \underset{\tilde{\Phi}_i \subseteq \mathcal{P}(\Omega)}{\arg\max} \left(\sum_{\omega \in \tilde{\Phi}_i} v^i(\omega) - p_\omega \right).$$

Thus, incentive compatibility requires that banks and firms engage with positive probability in a set of contracts only if doing so maximises their utility.

[14]For concreteness, let $\Omega = \{\omega_1, \omega_2\}$. Then $\mathcal{P}(\Omega) = \{\{\varnothing\}, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}\}$. If agent i (be it a bank or a firm) is equally likely to choose each of those four possible sets of contracts, then $A^i = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, and $A^i(\{\omega_1\}) = \frac{1}{4}$.

[15]In Azevedo and Hatfield (2015), there are two additional technical conditions for the existence of an equilibrium. First, the economy needs to be regular in the sense that the integral of absolute values of the agents' utilities must be finite. Second, agents need to be able to supply any small net demand for contracts to rule out cases where there is positive demand but no supply for a certain finitely priced contract.

To define feasibility, it is useful to define the excess demand for loan ω by agent i under allocation A as

$$Z^i_\omega(A) \equiv \sum_{\{\omega\}\subseteq\Phi_i\subseteq\Omega} A^i(\Phi_i) - \sum_{\{\omega\}\subseteq\Psi_i\subseteq\Omega} A^i(\Psi_i).$$

The first term sums over all possible sets of borrowing contracts of which ω is an element, and adds up the probabilities that agent i is engaged in each of them. This gives the probability that agent i borrows under contract ω. Similarly, the right term sums over all possible sets of lending contracts of which ω is an element to give the probability that i lends under contract ω. The difference between the two gives a probabilistic version of excess demand for agent i for loan contract ω. Because each agent is either a bank or a firm, Z^i_ω will always be non-positive for banks and non-negative for firms.

The set of agents is I, and the market is given by a Lebesgue measurable distribution η over I, defined over a σ-algebra, with $\eta(I) < \infty$. Economy-wide excess demand for ω under allocation A can then be found by integrating over the excess demands of all agents, which yields

$$Z_\omega(A) \equiv \int_I Z^i_\omega(A) d\eta. \tag{3}$$

Allocation A is feasible if $Z_\omega(A) = 0$ for each loan contract $\omega \in \Omega$. Feasibility thus requires the decisions of banks and firms to be consistent – banks cannot lend more than firms borrow.

Equilibrium existence and assumptions: Azevedo and Hatfield (2015) show (in Section 6 of their paper) that, under certain assumptions, an equilibrium of the type described above exists, is unique, and is efficient in the sense that it maximises a social welfare function of the form

$$\int_I \left(\sum_{\Phi_i} v^i(\Phi_i) A^i(\Phi_i) + \sum_{\Psi_i} v^i(\Psi_i) A^i(\Psi_i) \right) d\eta. \tag{4}$$

Their paper is the first to prove equilibrium existence in a setting that allows for both heterogeneous preferences and complementarities. Previous work on stable matchings assumes that all agents have substitutable preferences. The first assumption required for the result to hold is that agents all have quasilinear utility functions. In the utility functions of banks and firms defined above quasilinearity results from the fact that transfers enter utility additively.[16] The second assumption is that agents derive utility only from loan contracts in which they are personally involved so that there are no externalities. Finally, and most crucially, the results hold for economies with a continuum of agents of each agent type i. In my loan market setting, the number of agents of each type corresponds to the number of banks and firms in my sample, both of which are finite. But given the large number of banks and firms, one can think of the loan market as an approximation to the infinite number of agents world of Azevedo and Hatfield (2015).

3.2 Parametrisation

Mapping the model of Azevedo and Hatfield (2015) to the data requires two basic steps: a decision on how to parametrise the value functions, and a procedure to identify the structural parameters. For the remainder of this section, I use the framework of Fox (2016) to achieve both of these steps.

In the model of Azevedo and Hatfield (2015) agents have complete information as they know all match-relevant characteristics and the utility functions of all other agents. I maintain this assumption. But because it is unlikely that my dataset captures all match-relevant characteristics, the first step in specifying the empirical counterparts of the valuation functions (1) and (2) is to make an assumption about how observable and unobservable characteristics drive those valuations.

Separability of valuation functions: In Azevedo and Hatfield (2015), an agent type i is distinct from other agents by a unique set of characteristics. In the empirical

[16]Fleiner et al. (2015) point out that quasilinear utility can be a strong assumption because it is violated whenever firms face financing constraints, or when the behaviour of their management is influenced by "wealth effects" or risk aversion.

setting, these characteristics are of two types: observable characteristics j from a finite set J, and unobservable characteristics k from a possibly infinite set K. An agent type i is a unique combination of observable and unobservable characteristics, so that $i = (j, k)$. For convenience, I will sometimes refer to distinct sets of characteristics as types: i is a *full agent type*, j is an *observable agent type*, and k is an *unobservable agent type*.

Each contract ω encodes information about the observable characteristics of the bank $b(\omega)$ and firm $f(\omega)$ associated with the contract, so that $b(\omega), f(\omega) \in J$. This assumption about contracts has an important implication: if agents have preferences over contracts and contracts only encode information about the observable types of the two parties involved in the contract, then agents only have preferences over observable types of their counterparties. This observation gives rise to the separability assumption: I assume that the valuation of firm i from being the borrower in all contracts in Φ_i is separable in the valuations from its own observable and unobservable types. The valuation of i can then be written as

$$v^i(\Phi_i) = \pi^j(\Phi_i) + \epsilon^k_{\Phi_i},$$

where $\pi^j(\Phi_i)$ is the valuation of i's observable type j, and $\epsilon^k_{\Phi_i}$ is the valuation of i's unobservable type k.[17] When i is a bank, the expression is symmetric for a given loan portfolio Ψ_i. To reiterate: observability is about what can be observed in the data, not about what agents can observe when choosing whom to match up with – I assume that they can observe everything that is relevant. Thus, $\pi^j(\Phi_i)$ is the valuation firm i gets from having certain characteristics that are observable in the data (it might be located in Tokyo) and from borrowing from banks with a particular set of characteristics that are also observable in the data (they might all be located in Osaka), while $\epsilon^k_{\Phi_i}$ is

[17] Agents have preferences over observable types j of their counterparties, not over contracts per se. So, when I write $v^i(\Phi_i)$ for the valuation of firm i from being the borrower in all trades $\omega \in \Phi_i$, I use that notation as shorthand to refer to the valuation of i from being a borrower of all banks associated with the contracts in Φ_i. Each of these banks has a particular set of observable characteristic j, which is what actually drives the match value of the firm. Defining ω as encoding all the observable characteristics of both parties is what makes this shorthand notationally correct.

the valuation firm i gets from having certain characteristics that are not observable in the data (maybe it prefers borrowing from city banks) and from being matched to counterparties with characteristics that are observable in the data (some of the borrowers might indeed be city banks).

Observable valuation: I assume that $\pi^j(\Phi_i)$ is known up to a finite vector of parameters β, and that these parameters are homogenous among banks and among firms but not necessarily between banks and firms, so that

$$\pi^j(\Phi_i) = \pi_{\beta^f}(\Phi_i).$$

I also assume that the valuation function is linear in these parameters, so that

$$\pi_{\beta^f}(\Phi_i) = X(\Phi_i)'\beta^f, \tag{5}$$

where X is a vector of observable characteristics of the parties associated with all loan contracts in Φ_i, and β^f is the vector of structural parameters of a firm's valuation function. The expression for when agent i is a bank is again symmetric for a given loan portfolio Ψ_i and a vector of structural parameters β^f. The structural parameters in $\beta = (\beta^f, \beta^b)$ are the parameters that I seek to identify.

Unobservable valuation: The distribution over full agent types, $\eta(I)$, and the definition of an agent i as a unique combination of observable and unobservable characteristics so that $i = (j, k)$, give rise to a conditional distribution of unobservable valuations. Associated with each unobservable type k is a realisation of the vector $\epsilon^k = (\epsilon^k_{\Phi_i})_{\Phi_i \subseteq \Omega}$, in which typical element $\epsilon^k_{\Phi_i}$ is the valuation of agent i's unobservable type k from being engaged in the set of contracts in Φ_i. The vector ϵ^k contains as many elements as there are possible sets of contracts. The vector is of finite length because the set of contracts Ω is finite so that the number of possible sets of contracts, $\mathcal{P}(\Omega)$, is also finite. Given $\eta(I)$, we can thus form a joint probability distribution over the valuations of unobservable types, $F(\epsilon^k|j)$, for every observable type j. In essence, given that we have a probability distribution over full agent types i, and given that $i = (j, k)$, observing j probabilistically pins down k and the vector of its valuations.

In contrast to the observable valuation functions, I do not assume that the distribution of unobservable valuations is known up to a finite vector of parameters. One implication of this is that the maximum score estimator allows for heteroskedasticity: it allows the distribution $F(\epsilon^k|j)$ to vary across observable agent types j, and it allows the distribution of a given agent type j to be different in each market m (a year, say), so that we have $F(\epsilon^k|j, m)$. This flexibility is a virtue of the estimator, as semiparametric discrete choice estimators commonly impose homoskedasticity.

3.3 Identification

To identify the structural parameters I use the semiparametric matching maximum score estimator introduced by Fox (2016). The estimator is semiparametric because I assume above that the valuation of observable type j is known up to a finite vector of parameters (the parametric part), while I make no such assumption about the distribution of the valuations of unobservable types k (the nonparametric part).

The maximum score estimator has two main advantages over more widely used single-agent discrete choice models. First, it accounts for scarcity, the fact that a match between two agents reduces the likelihood for all other agents to match up with either of those two agents. Second, it is computationally more efficient because in contrast to logit or probit models, there is no need to integrate over all possible realisations of the vector of unobservable characteristics.[18]

Identification of the matching maximum score estimator is based on three elements: a matching maximum score inequality, a rank order property, and an objective function. This subsection discusses these three elements in turn.

Matching maximum score inequality: A matching maximum score inequality, g, contains on its left-hand side the observable valuations of two distinct banks and two

[18] For a comparison of the matching maximum score estimator with a binary logit model, see Akkus et al. (2016).

distinct firms that are associated with two matches, ω_1 and ω_2, that we observe in the data. On its right-hand side, g contains the observable valuations of those same four parties from two counterfactual matches, ω_3 and ω_4, that we do not observe in the data and that we create by swapping the matching partners from the two actual matches ω_1 and ω_2. All remaining matches in the economy remain the same. The inequality states that the valuation of the four parties from the actual matches is higher than their valuations from the counterfactual matches. Transfers do not enter the story because they sum to zero between the lender and the borrower of each contract. Using the definition of observable valuations in (5), we can write g as

$$\begin{aligned} & X(\Psi_{b(\omega_1)})'\beta^b + X(\Phi_{f(\omega_1)})'\beta^f + X(\Psi_{b(\omega_2)})'\beta^b + X(\Phi_{f(\omega_2)})'\beta^f \\ > & X(\Psi'_{b(\omega_1)})'\beta^b + X(\Phi'_{f(\omega_1)})'\beta^f + X(\Psi'_{b(\omega_2)})'\beta^b + X(\Phi'_{f(\omega_2)})'\beta^f, \end{aligned}$$

where Φ and Ψ are the set of all counterparties of firms and banks under the actual matching, and Φ' and Ψ' are their sets of counterparties under the counterfactual matching. For instance, for bank $b(\omega_1)$, the lender in the actually observed contract ω_1, it will be the case that $\omega_1 \in \Psi_{b(\omega_1)}$. If $b(\omega_1)$ is also the lender in the counterfactual contract ω_3, and given that all other matches in the economy do not change, it is then also the case that $\Psi'_{b(\omega_1)} = (\Psi_{b(\omega_1)} \setminus \omega_1) \cup \{\omega_3\}$. To economise on notation, we can define the *joint valuation function* of the bank and the firm associated with contract ω as

$$\Pi_\beta(\omega) = X(\Psi_{b(\omega)})'\beta^b + X(\Phi_{b(\omega)})'\beta^f \tag{6}$$

and then rewrite the inequality as

$$\Pi_\beta(\omega_1) + \Pi_\beta(\omega_2) > \Pi_\beta(\omega_3) + \Pi_\beta(\omega_4). \tag{7}$$

Why does the inequality hold? In Azevedo and Hatfield (2015), the equilibrium allocation is unique and maximises the social welfare function (4). Thus, if we assume that the allocation we observe in the data is an equilibrium (as we do), then any alternative allocation – such as the one resulting from swapping the counterparties of the matches ω_1 and ω_2 – will lower social welfare. And because we also assume that there are no externalities, the reduction in welfare is necessarily driven by a reduction in the valuations of the parties directly involved in the swap. If there are no unobservable match valuations then this reasoning implies that (7) holds. If there are unobservable match valuations – and in practice there almost certainly are – we need an assumption about the relationship between observable valuations, unobservable valuations, and the choice probabilities that drive match sorting patterns. This assumption is the rank order property.[19]

Rank order property: The rank order property is a revealed preference argument with regard to match sorting patterns: it says that agents are more likely to engage in trades that yield them a higher observable valuation. The rank order property is the key identification assumption and the only requirement we make on the conditional distribution of the unobservable match valuations, $F(\epsilon^k|j)$. Fox (2016) shows that the rank order property holds and that, given that it does, the matching maximum score estimator consistently identifies β, the parameter vector of interest.[20]

Let the probability that the firm of observable type j chooses the set of contracts

[19]There is a reason why we consider two actual contracts and then compare their match value to two counterfactual contracts formed by cross-matching partners, rather than just comparing one actual contract with one counterfactual contract. For the matching maximum score inequality to have meaning, the observable types of the agents on both sides of the inequality need to be the same. In the fully general model of Azevedo and Hatfield (2015), two agents can engage in a variety of different contracts and so considering one of those contracts on each side of the inequality would be sufficient. Because – in their language – agents are allowed to act as both sellers and buyers of certain trade ω, the counterfactual to an actually observed contract could simply be a contract where the same two parties swap their roles as buyer and seller. In my two-sided market framework this is not possible, as it would require banks to borrow and firms to lend. Hence, to have legitimate and meaningful comparisons with the same set of agents, two actual contracts are required, chosen such that a cross-matching of partners produces counterfactual, rather than actually existing, loan contracts.

[20]The proof of the rank order property and thus the consistency of the matching maximum score estimator partially rely on properties of competitive equilibrium in Azevedo and Hatfield (2015). But the two key assumptions are on the vector of unobservable match values ϵ^k. First, ϵ^k needs to have full support in $\mathbb{R}^{dim(\epsilon^k)}$. Second, ϵ^k needs to have an exchangeable distribution for a given j. The distribution is exchangeable if for a permutation ρ, we have that $F(\epsilon^k|j) = F(\rho(\epsilon^k)|j)$ for all permutations ρ.

Φ conditional on it choosing either Φ_1 or Φ_2 be given by

$$A^j(\Phi|\Phi_1,\Phi_2) = \frac{A^j(\Phi)}{A^j(\Phi_1) + A^j(\Phi_2)}$$

The rank order property then states that the matching maximum score inequality (7) holds if and only if

$$\begin{aligned} &A^{f(\omega_1)}(\Phi_{f(\omega_1)}|\Phi_{f(\omega_1)},\Phi'_{f(\omega_1)}) \times A^{f(\omega_2)}(\Phi_{f(\omega_2)}|\Phi_{f(\omega_2)},\Phi'_{f(\omega_2)}) \\ \geq\ &A^{f(\omega_1)}(\Phi'_{f(\omega_1)}|\Phi_{f(\omega_1)},\Phi'_{f(\omega_1)}) \times A^{f(\omega_2)}(\Phi'_{f(\omega_2)}|\Phi_{f(\omega_2)},\Phi'_{f(\omega_2)}). \end{aligned} \tag{8}$$

So, conditional on engaging in the actual contracts ω_1 and ω_2 or the counterfactual contracts ω_3 and ω_4, agents are more likely to engage in the actual contracts whenever their combined observable valuations are higher from doing so than from engaging in the counterfactual contracts. In other words: the choice probabilities of the agents are rank ordered by their observable valuations. The condition involves firm probabilities only because bank and firm probabilities are linked through the feasibility condition (3): if firms are more likely to engage in one set of contracts than in another set of contracts, then so are banks – if they were not, the market for these contracts would not clear in expectation, which would violate feasibility.

How plausible is the assumption of the rank order property in my context? Quite plausible, I think. One driver of unobservable match valuation that could have violated the assumption is the role of main banks. This is because, as discussed above, a new loan contract between a main bank and a firm might not be driven by profit maximisation motives but the bank's obligation to support the firm in times of financial distress. To the extent that my definition of a main bank incorrectly classifies main-bank lending – which in some cases it surely will – I cannot fully eliminate that issue by dropping all loans from main banks, but these cases should not drive my overall results. Another reason to be confident in the assumption is that by including past loans as an explanatory variable, I indirectly capture the effect of unobservable valuations on matching decisions. One of my explanatory variables is whether there was a loan between a bank and a firm during the previous three years, which is an outcome that was also driven by unobservable match valuations of both parties. Thus, only drivers of unobservable match valuations that were not to the same extent driving decisions in the previous three years could violate the rank order property.

Objective function: The matching maximum score inequality is the basis for identification and the rank order property ensures that identification based on this inequality is consistent. The objective function is the third element of the apparatus: it is the tool that implements the identification procedure computationally. The function is

$$Q(\beta) = \sum_{m \in M} \sum_{g \in G_m} 1[\Pi_\beta(\omega_1) + \Pi_\beta(\omega_2) > \Pi_\beta(\omega_3) + \Pi_\beta(\omega_4)], \tag{9}$$

where each market m corresponds to one of the 31 years in my dataset, and where G_m is the set of all possible matching maximum score inequalities, g, that can be formed based on the number of active banks and firms in market m. $1[\cdot]$ is the indicator function, which evaluates to one if a given matching maximum score inequality is satisfied and to zero if it is not. Thus, the objective function value (or the score), $Q(\beta)$, is the number of satisfied inequalities for a given value of the parameter vector. Estimation of the matching maximum score estimator proceeds as follows: for a given guess of the parameter vector β, the estimator evaluates each inequality, raises the score by one for each satisfied inequality, and then chooses β such that the score is maximised. The estimator thus exploits the implication of Azevedo and Hatfield (2015) that the valuations of non-observed matchings should be lower than that of observed matchings, and chooses β such that this claim is true most often in the data.

While the second summation of the objective function is over all possible inequalities of a given market, the estimator is consistent when estimated based on a subsample of those inequalities. This is a key advantage of the estimator because it makes identification in large markets computationally feasible. In my dataset, banks and firms agree to 12,500 new loans in an average year. Estimating on all possible inequalities that can be formed from those loans would be computationally very demanding if not impossible. Instead, I randomly sample 20,000 inequalities from each market. To generate one inequality, I first randomly select two banks from the set of all active banks and then for each of these banks randomly select a firm that is not a counterparty of the other bank.

The maximum score objective function is a step function (it increases by 1 every time an inequality is satisfied) and is thus not continuous. Hence, optimisation needs to be based on a method that does not rely on the gradient. Differential evolution is such a method. I use the tool kit for maximum score estimation provided by Fox and Santiago (2015), which uses the differential evolution optimisation routine implemented in Mathematica. To generate the reported results I run the optimisation routine a total of 20 times, each time with 250 initial points and a maximum of 15,000 iterations. I get the same results each time. More detail on the estimation procedure is provided in Appendix A.

Finally, because I do not have data on transfers, I need to normalise β by setting one of its elements to either -100 or 100 (which is akin to the scale normalisation in logit and probit models). I normalise the *relationship to firm* variable, a dummy that indicates whether a bank and a firm have agreed to a new loan in the previous three years. I choose whether to normalise it to -100 or 100 by optimising (9) for both cases and choosing the normalisation for which the score is higher. The magnitudes of all other parameters are then interpreted as the importance of the respective characteristic relative to the importance of *relationship to firm* in driving match valuation.[21]

3.4 Inference

Given that I have data on multiple markets (the 31 years covered in the dataset), there are two possible asymptotic arguments to conduct inference for the matching maximum score estimator. The first is to hold fixed the number of observed participants in a given market and to let the number of observed markets go to infinity. The second is to hold the number of observed markets fixed, and let the number of observed participants in

[21] I normalise to 100 rather than the more commonly used 1 because my estimates are such that the latter would lead to very small estimates for some of my characteristics.

each market go to infinity. I follow the latter approach because it has an appealing link to the model of Azevedo and Hatfield (2015): the limiting case of increasing the number of observed participants is their world of a continuum of agents, for which we know that a unique, stable and efficient equilibrium exists. The argument that links the empirical market and the continuum market is then the following: we know that for a market with a continuum of agents of each type there exists a unique and stable competitive equilibrium. If we assume that this equilibrium is played by all agents in the continuum market, then we can use a subset of those agents to identify the structural parameters of that equilibrium as long as estimation is consistent when based on only a subset of all matches. As discussed above, the matching maximum score estimator is consistent in this case, so that we can consistently identify the structural parameters of the equilibrium game. The key assumption is that all agents in the continuum market play the same equilibrium.

To produce confidence intervals based on this asymptotic argument, I follow Fox (2016) in using the subsampling procedure of Romano and Shaikh (2008). The procedure is implemented in Fox and Santiago (2015). I use subsampling because, as discussed in Fox and Santiago (2015), the limiting distribution of the maximum score estimator is too complex for inference and the conventional bootstrap method of sampling with replacement at the original sample size is inconsistent.

4 Data

4.1 Dataset description

I use data from the Nikkei Economic Electronic Databank System (NEEDS), which provides firm-level financial data for all companies listed on a Japanese stock exchange.[22] My dataset covers 179 banks and 5073 firms between the years 1980 and 2013. After cleaning the data, I am left with 14,310,318 unique bank-firm-year observations.

[22] List of exchanges: Tokyo, Osaka, Nagoya, Kyoto, Hiroshima, Fukuoka, Niigata, Sapporo, and JASDAQ. For firms that have merged during the sample period, NEEDS provides data only for the surviving institutions. Pre-merger data is not available for firms deceased through mergers.

I focus on Japan because firm-level financial data can be accessed from Nikkei for a fee and is thus readily (though not cheaply) available. In contrast, the same data for other major economies is either collected by regulators and available only under certain restrictions (such as in Germany) or is not available at all (such as in the US). But focusing on Japan has benefits that go beyond data availability: it is the world's third largest economy, the banking system plays an important role for firm financing, and, as Hoshi and Kashyap (2010) demonstrate, the Japanese economic experience over the past decades can hold important lessons for other economies.

I use data from three different NEEDS databases.[23] The Corporate Attributes Database (Nikkei 2008) provides basic company information such as company name, listing information, location, and industry. The Corporate Financial Database (Nikkei 2013) provides detailed financial information on industrial and financial companies. I use annualised data, which is based on quarterly financial statements (Yuho reports) that listed companies are mandated to submit to the Financial Services Agency. The fiscal year of most firms is from April to March, so that annualised data is based on the four Yuho reports submitted by the end of June, September, December and March.[24] The third source I use is the Corporate Borrowing from Financial Institutions Database (Nikkei 2012), which provides information on (total, long, and short-term) corporate borrowing vis-a-vis individual financial institutions. Data is based on original Nikkei research, and provided for all non-financial listed companies, which means all companies that are not classified as banks, insurers, or the Tokyo Exchange Foreign Section.

4.2 Variable definitions

New loan: A firm and a bank (that is not the firm's main bank) have agreed to a new loan contract in a given year if, at the end of said year, the outstanding loan volume between the firm and the bank has increased compared to the previous year. This corresponds to a year-on-year increase in the variable *DBT* in Nikkei (2012). If

[23]For an overview of all available databases, visit https://www.nikkeieu.com/needs/.

[24]The fiscal year does not end in March for all firms. Fricke and Roukny (2017) use the same dataset as I do and find that dropping those firms from the sample does not affect the results.

a bank and a firm have agreed to a new loan contract, I say that they are matched to each other. Given that my dataset starts in 1980, I cannot calculate the variable for that year and drop all observations for that year from the data.

To identify a firm's main bank, I define it as the largest short-term creditor over the entire sample period for firms that borrow short-term, and as the largest long-term creditor for firms that only borrow long-term. More commonly, the main bank is defined simply as a firm's largest overall creditor. I choose the definition based on short-term lending because Aoki and Patrick (1995) highlight that some firms are known to borrow more from non-main banks than from their main bank, which makes the definition based on overall lending misleading. Furthermore, the same authors note that while main banks supply both long and short-term loans, being the main short-term creditor is a hallmark of main banks because short-term loans are the channel through which they exercise control over firm's ongoing financing and operational decisions. Out of all 195,936 new loans in my dataset, 61,196 are between a firm and its main bank. I drop all of those observations from my data.

My definition of a new loan has two caveats. First, given that I have data on end-of-year loan volumes rather than on individual loans, an increase in the volume could be the result of multiple new loans or the net effect of new loans and repayments of old loans. For my analysis this is unimportant, and I use the terminology "new loan" simply as a shorthand. Second, the new loans dummy is zero in cases where a firm has repaid old loans, which is the case in 313,307 bank-firm-year observations, or about 2 percent of all observations. This means that by optimising the objective function (9), I could draw an inequality where the counterfactual matching is between a firm and a bank that have reduced their loan exposure, so that the reasoning behind the counterfactual comparison might not be fully appropriate.

Rank difference: Rank difference is the absolute difference between a bank's size relative to all other banks and a firm's size relative to all other firms in a given year. To calculate relative sizes of bank and firms, I calculate their respective percentile rank among their peers based on asset size. Data on asset size is given by variables

B11098 (for banks) and *B01110* (for firms) in Nikkei (2013). I have no data on asset size for 1338 bank-year observations and 56,691 firm-year observations, mostly because a firm or a bank was not active in a given year (it might not have existed yet or stopped operating). This leads to 13,216,085 missing observations for the rank difference variable. I drop all of those observations from the data.

Distance: Distance is the number of kilometres between the headquarters of a bank and a firm. To generate the variable I use zip codes for bank and firm headquarters from Nikkei (2008), use the batch geocoding tool from *doogal*[25] to obtain longitudes and latitudes, and then use the Python toolbox *GeoPy* to calculate the distance between each bank and each firm based on Vincenty's formulae. Geocoding locates the headquarters of 21 banks in the United States (2 in Manhattan, 19 on Japan Street in Port Charlotte). I correct for this by manually looking up the actual location of those headquarters, all of which are in Japan. Geocoding also locates the headquarters of 187 firms outside of Japan. Because these are less than 5 percent of all firms in my dataset, I drop these firms from the sample. This leads to a loss of 1,037,663 observations (31 years × 179 banks × 187 firms).

The variable has two limitations. Whenever a firm interacts with a bank's local branch, a firm subsidiary interacts with a bank's headquarters, or a subsidiary interacts with a local branch, then distance between headquarters is not the relevant metric. This will often be the case for loans from city banks, which are headquartered in Tokyo or Osaka and operate branches nationwide. The second limitation is that I only have postcodes for 2013, so that the distance between headquarters is correctly calculated for all bank-firm pairs where neither party has relocated during the sample period.

Relationship to firm: A bank has a relationship to a firm in a given year if the two parties have agreed to a new loan contract in at least one of the previous three years. I construct the variable using the *new loan* variable defined above and set the relationship dummy to one whenever the new loan dummy was one in at least one of

[25] https://www.doogal.co.uk/BatchGeocoding.php.

the three previous years.

Relationship to sector: A bank has a relationship to a sector in a given year if it has agreed to a new loan contract with at least one firm of that sector in at least one of the previous three years. I classify firms into sectors based on the Nikkei Medium Classification Industry Code, which is available as variable *NKILM* in Nikkei (2012). The code classifies each firm uniquely into one of 36 different sectors. The construction of the variable is analogous to that of the *relationship to firm* variable above, with the difference that I define a new loan based on an increase in bank-sector loan volumes, rather than an increase in bank-firm loan volumes. I cannot construct the two relationship variables for the years 1980, 1981, and 1982 and drop all data for those years from my sample.

Bank HHI: To capture a bank's objective to focus or diversify its loan portfolio, I construct a Herfindahl-Hirschman Index (HHI) based on the bank's sector exposures.[26] If x_s is a bank's loan exposure to sector $s \in S$, and if $x = \sum_{s \in S} x_s$ is the bank's total exposure, then

$$\text{HHI} = \sum_{s \in S} \left(\frac{x_s}{x} \right)^2 .$$

The HHI is thus the sum of squared sector exposures, each expressed as a fraction of a bank's total exposure. The index ranges from $1/|S|$ for a perfectly diversified portfolio to 1 for a perfectly focused portfolio. I use the same sector classification as for the *relationship to sector* variable above. For each bank in each inequality, I calculate the HHI for its actual loan portfolio and its counterfactual portfolio, which results from one actual counterparty being swapped with a hypothetical one. To calculate the counterfactual portfolio I assume that the principal of the counterfactual loan contract is the same as that for the actual loan contract for which it is swapped.

The contribution of the HHI to bank's match valuations might be hard to identify.

[26] In the finance literature, the HHI is a commonly used measure of diversification. See, for instance, Acharya et al. (2006) and Degryse and Ongena (2005).

There is no variation in the index when a bank's actual and counterfactual counterparty are active in the same sector, and even when they are in different sectors, reclassifying a single loan will often not change the HHI by much.

4.3 Summary statistics

Table 1 summarises the data. In an average year in the Japanese corporate loan market, some 140 active banks and 2000 active firms agree to about 12,500 new loan contracts. But panels (A) to (C) in Figure 1 show that these averages mask clear trends. Most importantly, the number of lending banks has steadily declined since the mid 1990s. As Woo and Kanaya (2000) explain, this is because the government was no longer able to find "white knight" institutions to support distressed banks and thus decided to let these banks fail. The number of borrowing firms is more stable, albeit it has been on a downward trend since the beginning of the millennium. The series jumps in 1996 because since March of that year, NEEDS also covers JASDAQ listed companies. The pattern of new loans mirrors both of these developments; the jump in 1996 is driven by the newly covered JASDAQ firms and the subsequent decline by the declining number of lending banks.

Bank and firm degree count the average number of new loan contracts that banks and firms enter into with different counterparties during a year. An average bank lends to 93 different firms, while an average firm borrows from 4 different banks. This latter number is in line with those reported in studies that look at the number of banking relationships of Japanese firms and that are summarised in Degryse et al. (2009). Panel (D) in Figure 1 shows that the average HHI of all lending banks is quite stable over time and hovers around 0.35. The minimum value is 0 because some banks do not grant new loans in some years. The maximum of 1 indicates that there are some banks that perfectly specialise by lending to firms from a single sector only.

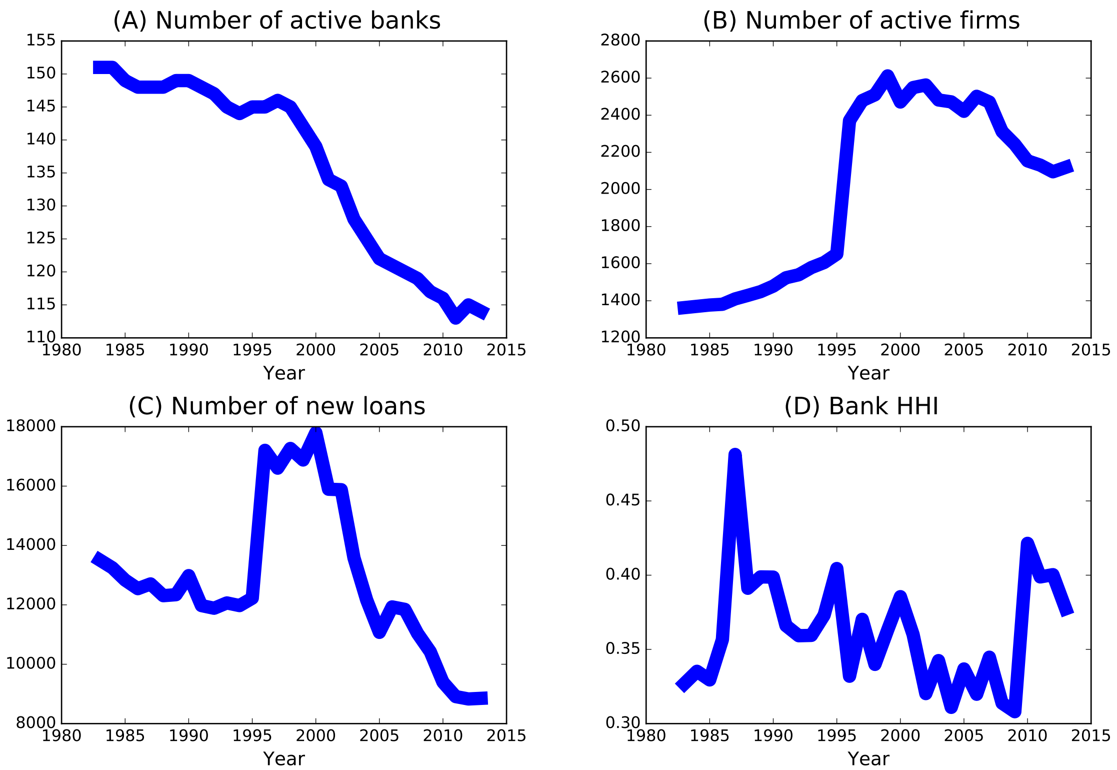
The large majority of new loans are between parties that either have a direct relationship, or where a bank has previously lent to a firm in the same sector as the newly borrowing firm (the mean of *relationship to sector* is 1 due to rounding, it is 0.99).

Table 1: Summary statistics

	Mean	Median	Min.	Max.	Std.
Active banks	135.7	142.0	114.0	150.0	13.3
Active firms	1,978.3	2,092.0	1,336.0	2,679.0	517.2
New loans	12,844.5	12,309.0	8,830.0	17,800.0	2,525.2
Bank degree	92.8	27.0	0.0	1,695.0	180.3
Firm degree	3.8	2.0	0.0	103.0	5.6
Bank HHI	0.4	0.3	0.0	1.0	0.3
Relationship to firm	0.8	1.0	0.0	1.0	0.4
Relationship to sector	1.0	1.0	0.0	1.0	0.1
Rank difference	29.6	24.0	0.0	99.0	23.6
Distance	437.9	265.4	0.0	3,743.6	687.9

Note: Statistics for active banks and firms, and for new loans are based on annual data. Remaining statistics are calculated based on pooled data for all years. A bank is active in a given year if, at the end of a year, it is the creditor of at least one firm; a firm is active if, at the end of a year, it is the debtor of at least one bank. Bank and firm degree are the number of new loans that banks and firms agree to in a year. *Relationship to firm, Relationship to sector, Rank difference*, and *Distance* are calculated based on matched bank-firm pairs.

Figure 1: Time trends



Note: A bank is active in a given year if, at the end of a year, it is the creditor of at least one firm; a firm is active if, at the end of a year, it is the debtor of at least one bank. The number of firms jumps in 1996 because the NEEDS dataset also covers JASDAQ-listed companies from March of that year onwards. A new loan is an instance where the outstanding loan volume between a bank and a firm has increased within a given year. Bank HHI is the annual average of the HHI of all lending banks, calculated based on their loan exposure to different sectors. The HHI ranges from 1 for a perfectly focused portfolio to 0 for a perfectly diversified one.

Average rank difference is larger than what one would expect if there was a strong pattern of positive assortative matching (for perfect positive assortative matching, rank difference would be zero). Average differences in the order of 27 indicate that banks and firms frequently match with counterparties that are of different relative size. Distances between banks and their borrowers are large. Degryse and Ongena (2005) find that for a large Belgian bank, the median distance to its borrowers is 2.25 kilometres, while Petersen and Rajan (2002) find that the median distance between US banks and small borrowing firms is about 6.4 kilometres. In comparison, the median distance in my data is 265 kilometres. Almost certainly, this is at least partially because I lack information on what local bank branches firms interact with and thus have to calculate distances between headquarters.

5 Hypotheses and results

I present two sets of results: first, a baseline specification in which both firms and banks have substitutable preferences and in which match sorting patterns are driven by information-based characteristics (Subsection 5.1); second, a specification where banks have complementarities in preferences and sorting patterns are also driven by banks' aim to focus or diversify their loan portfolios (Subsection 5.2).

5.1 Information-based characteristics

Banks mitigate problems resulting from asymmetric information between borrowers and lenders.[27] If we think of banks as delegated monitors in the sense of Diamond (1984), where banks channel funds from lenders to borrowers and monitor borrowers on behalf of lenders, then all characteristics that make it cheaper for a bank to monitor a firm and for a firm to borrow from a bank should make it more likely that they agree to a new loan contract. Four such characteristics are: similar relative size of the two parties, the geographical distance between them, and previous lending by the

[27] Banks are unique compared to other financial institutions because by screening and monitoring firms they obtain private information about those firms. See Degryse et al. (2009) for a discussion.

bank either directly to the firm or to other firms in the same sector. This subsection discusses why and how theory suggests each of these characteristics should matter, and tests whether these predictions are borne out in the data.

Interpretation of estimates: To test the predictions, I assume that the same set of variables X determines the observable match valuation (5) of both banks and firms: *rank difference*, *distance*, *relationship to firm*, and *relationship to sector*. After some rearranging (see Appendix B) the matching maximum score objective function (9) can be written as

$$Q(\beta) = \sum_{m \in M} \sum_{g \in G_m} 1\left[\sum_{x \in X} Z_x(\beta_x^b + \beta_x^f) > 0 \right], \tag{10}$$
$$\text{with } Z_x = \sum_{\{\omega_1, \omega_2\}} x(\omega) - \sum_{\{\omega_3, \omega_4\}} x(\omega),$$

where Z_x is the difference between the sum of the values of variable x for the two actual matches ω_1 and ω_2 and the sum of its values for the two counterfactual matches ω_3 and ω_4. β_x^b and β_x^f are the structural parameters from the match valuation functions of banks and firms.

The definition of Z implies that it is bank-firm (or pair) characteristics that drive match sorting patterns in the model. Individual characteristics of banks and firms that appear on both sides of the inequality would sum to zero and remain unidentified.[28]

Another implication of (10) is that for pair-characteristics that enter the valuations of both banks and firms, as is the case for all four information-based characteristics, the matching maximum score estimator identifies a characteristic's relative contribution to the joint valuation of a matching pair (as defined in (6)), which is the characteristic's combined contribution to the valuation of banks (captured by β^b) and firms (captured by β^f).

This is the interpretation of the first four sets of estimates reported in Table 2: the

[28] This is the reason why I do not include, for instance, any measure of counterparty risk. Assuming that the counterparty risk of any given bank or firm is correctly and identically assessed by all other parties, it is a constant firm-specific characteristic, which remains unidentified.

coefficient of each characteristic is an estimate of its contribution to the joint valuation of a matching bank-firm pair, relative to the contribution of a previous relationship between the two parties, which is normalised to 100. None of the four characteristics accounts for complementarities in preferences of banks, so that this specification is a baseline case where both banks and firms have substitutable preferences.

Table 2: Main results

Variable		No Complementarities	Complementarities
Relationship to firm	Est.	**100**	**100**
	95% CI	Superconsistent	Superconsistent
Relationship to sector	Est.	**15.70**	**13.25**
	95% CI	(5.69, 90.09)	(2.96, 90.69)
Rank difference	Est.	**0.13**	**0.07**
	95% CI	(0.01, 0.33)	(0.02, 0.07)
Distance	Est.	**-0.01**	**-0.005**
	95% CI	(-0.02, -0.01)	(-0.02, -0.01)
Bank HHI	Est.		**162.55**
	95% CI		(56.41, 194.18)
Number of inequalities		620,000	620,000
Objective function value		559,688	560,970
Percent correctly predicted		90.27	90.48

Note: Coefficients are estimates of a variable's importance to match valuation relative to the importance of *relationship to firm*, which is normalised to 100. Results in the two columns are produced by optimising the objective functions (10) and (11), respectively, using the differential evolution algorithm in Mathematica. I randomly select 20,000 inequalities out of all possible inequalities G_{N_m} for each year from 1984 to 2013, and estimate the coefficients based on the 620,000 pooled inequalities. Confidence bands are calculated using subsampling, based on 100 subsamples using data from a subsample of 30 banks. Confidence bands based on subsampling need neither be symmetric nor contain the point estimate. The estimate of a parameter that can take only two values is superconsistent, so I do not report a confidence interval for *relationship to firm*.

Previous lending relationship: If screening and monitoring are costly, then it is cheaper for a bank to lend to a firm it has lent to before because it already has acquired some information about that firm. In particular, it has both firm-specific information (management quality, the business model) as well as information about the sector in which that firm operates (growth prospects, competitive landscape), both of which make it cheaper to lend to the firm again and will increase the likelihood of a new match between the bank and the firm. The sector-specific knowledge the bank has acquired also makes it cheaper to lend to other firms in that same sector, albeit to a lesser extent because the bank still needs to obtain firm-specific information about those firms. Thus, both a previous loan to a firm itself and to other firms in the same sector should increase the willingness of the bank to lend to a firm, but the former should increase its willingness more.

For a firm, changing lenders can be costly. If the firm is a high-quality borrower, Sharpe (1990) and von Thadden (2004) suggest that borrowing from a new (uninformed) bank can be more costly than borrowing from a bank that has already learned of the firm's quality. The firm essentially faces a "lemon problem" (Akerlof 1970), where the uninformed bank charges a loan rate equal to its expected loss from lending to all firm types, which is higher than the expected loss from lending to a high-quality borrower. Kim et al. (2003) find evidence that switching costs are substantial, which provides firms with an incentive to borrow from banks they have borrowed from in the past. Incentives for banks and firm to match up with past partners thus reinforce each other and suggest that

Hypothesis 1: *A match between a bank and a firm is more likely if the bank has lent to that firm in the past and – to a lesser extent – if the bank has lent to another firm in the same sector in the past.*

The results in Table 2 support the hypothesis. The estimates for both variables are positive and significant at the 95 percent level, and suggest that a past relationship between a bank and a firm, and between a bank and a firm's sector, increase joint match valuation and thus make a match more likely. As predicted, a previous relationship to a firm's sector does increase match valuation, but it does so by less than a previous relationship to a firm directly. The difference is statistically significant, given that the upper bound of the confidence interval for *relationship to sector* is below 100.

Similar relative size: Grossman and Hart (1986) and Hart and Moore (1990) show that in a world of incomplete contracts, agent's ex-ante incentives are shaped by the degree to which they have control over physical assets.[29] Stein (2002) builds on these

[29]The intellectual history of the argument goes back all the way to Coase (1937), who tries to understand what determines the (vertical) boundaries of firms. The question of whether banks of

insights and develops a model that suggests that large banks are better suited to lend to large (and well established) firms, while small banks are better suited to lend to small (and less well established) firms – there should be positive assortative matching on size. The model has three main assumptions. The first is that screening a large firm involves gathering "hard information" (company reports, track records, and public rating) while screening a small firm involves gathering "soft information" (meeting with the founder of a startup to assess her diligence, work ethic, and prudence). The second is that soft information is more costly to acquire and – because it is not verifiable – harder to transmit. The third assumption is that loan officers are motivated to invest in acquiring information only if they know that capital will be allocated based on that information.

Why do these assumptions lead to positive assortative matching on size? Because there is a difference in control structure between large and small banks: in a single agent bank, the loan officer is also the director so that expertise and authority are completely aligned and the loan officer can allocate capital at will. This gives him a strong incentive to acquire soft information about potential borrowers because while the information might be costly to acquire, it is immediately relevant in deciding how to allocate capital. In a very large bank, by contrast, the loan officer has no say in the decision of how to allocate capital so that expertise and authority are completely separated. Because the loan officer can neither use soft information to allocate capital himself nor – because it is not verifiable – transmit it to his superiors, he has no incentive to acquire it. In contrast, hard information is easily verifiable and easily transmitted, and so the loan officer in a large bank will acquire it to increase the likelihood that capital is allocated to his projects. Thus

Hypothesis 2: *A match between a bank and a firm is more likely if they are of similar relative size.*

The estimates in Table 2 reject the hypothesis. The coefficient of rank difference is positive and significant, which suggests that – all else equal – the higher the difference

different sizes should specialise into serving customers of different sizes can be thought of as applying similar reasoning to understand the horizontal boundaries of firms.

in relative size between a bank and a firm, the higher the match value and thus the higher the likelihood that they match. This contradicts earlier findings by Berger et al. (2005), Chen and Song (2013), Cole et al. (2004), and Hubbard et al. (2002), all of which find that similar size of a bank and a firm makes it more likely that they agree to a new loan contract.

One possible explanation for why there is no positive assortative matching in the data are the results of Fricke and Roukny (2017). The authors find that both banks and firms with many counterparties (what they call "generalists") tend to interact with counterparties of all sizes, while borrowers and lenders with few counterparties ("specialists") tend to interact with generalists. Rather than to positive assortative matching, this behaviour gives rise to a matching pattern where firms of all sizes borrow from large banks and banks of all sizes lend to large firms. And if there are more specialists than generalists, then the sign of *rank difference* should indeed be positive.

Distance: If we assume that a bank collects at least some soft information when monitoring publicly listed firms, then monitoring a firm that is geographically close is easier and cheaper for the bank, which should make it more likely that the bank grants a loan to such a firm. At the same time, if we assume that firms need to visit the bank frequently and bear the cost of those visits, then it is more attractive for firms to borrow from nearby banks. Given these incentives for banks and firms,

Hypothesis 3: *A match between a bank and a firm is more likely if they are geographically close.*

The results in Table 2 support this hypothesis. The estimate for the distance parameter is negative and statistically significant in both columns, which means that a reduction in the distance between a bank and a firm increases joint valuation and thus the probability of a match. These findings are broadly in line with Petersen and Rajan (2002) and Berger et al. (2005).

5.2 Portfolio diversification

Banks do not just lend to any set of firms. Instead, they focus or diversify their loan portfolio in a particular way. As a result, a bank's valuation from lending – and thus its willingness to lend – to a particular firm depends on the set of other firms it can also lend to. In the language of the matching literature, the bank has complementarities between its borrowers. Their ability to account for these complementarities is what makes the framework of Azevedo and Hatfield (2015) and Fox (2016) such valuable tools to study matching in the loan market.

Interpretation of estimates: To take these complementarities into account, I add the HHI variable to the set of characteristics that drive the match valuation of banks. This leads to a matching maximum score objective function similar to (10), with the additional term for the HHI:

$$Q(\beta) = \sum_{m \in M} \sum_{g \in G_m} 1\left[\sum_{x \in X} Z_x(\beta_x^b + \beta_x^f) + \gamma^b Z_{HHI} > 0 \right], \tag{11}$$

$$\text{where } Z_{HHI} = \sum_{\{\omega_1, \omega_2\}} HHI(\Psi_{b(\omega)}) - \sum_{\{\omega_3, \omega_4\}} HHI(\Psi_{b(\omega)}).$$

The interpretation of Z_{HHI} is similar to the other characteristics: it is the difference in the HHI of the two banks given their portfolios under the actual two matchings ω_1 and ω_2 and their HHIs under the counterfactual portfolios once one of their counterparties has been swapped. Because the HHI enters the valuation function of banks only, the relative contribution of the HHI to overall match valuation is driven by its contribution to the valuation of banks only, which is captured by the coefficient γ^b. Apart from that, the interpretation of the estimate is the same as for the information-based characteristics: it is the relative contribution of the HHI to the joint valuation of a matching bank-firm pair.

Complementarities can arise both from focusing and from specialising. The standard result of modern portfolio theory by Markowitz (1952) suggests that investors should diversify as much as possible, and perfectly diversified loan portfolios maximise gains from delegated monitoring in Diamond (1984). In contrast, Stomper (2006) suggests that building up industry expertise by screening and monitoring firms from a particular industry is valuable to banks, and Winton (1999) finds that the optimal degree of diversification depends on a bank's riskiness. Whether to focus or specialise thus involves a trade-off. Theory is not conclusive on what banks should do, and empirical contributions, such as Acharya et al. (2006), Hayden et al. (2007), and Tabak et al. (2011), show that the degree of diversification does indeed vary across banks.

Irrespective of whether a bank focuses or diversifies, how lending to a particular firm fits into that objective should play an important role in the bank's decision whether or not to lend to that firm, which suggests that

Hypothesis 4: *A match between a bank and a firm is more likely if lending to that firm helps the bank achieve its portfolio focus or diversification objective.*

The last row of results in Table 2 supports that hypothesis. The coefficient on the HHI is positive and significant. The positive sign of the coefficient suggests that a higher HHI increases match valuation, which means that – in the aggregate – banks focus rather than diversify their loan portfolio.

A bank's choice of whether to focus or specialise might partially be affected by regulation. In Japan, the conduct of banks is governed by the Banking Act of 1981 and its amendments, which cap a bank's maximal loan exposure to a single counterparty at 25 percent of its (non-consolidated) regulatory capital (Miyamoto 2016). To the extent that this limit was binding, the magnitude of the HHI coefficient might be a lower bound of the valuation that banks derive from specialisation. The generalist-specialist interaction patterns from Fricke and Roukny (2017) that help explain why we do not observe positive assortative matching might also explain the tendency of banks to focus their loan portfolio: given that generalists diversify and specialists focus their portfolios and that there are more specialists than generalists, this is the pattern we would expect to find. Furthermore, the authors show that even generalists tend to concentrate the bulk of their lending to a small number of industries, which further reinforces the overall tendency to focus.

6 Discussion

The model fits the data well. The share of correctly predicted inequalities is a measure of statistical fit, and the last line in Table 2 indicates that in about 90 percent of all inequalities considered, the model correctly predicts that parties match with their actual rather than the counterfactual counterparties.

Table 3 gives a sense of the relative importance of different characteristics in driving joint valuations and thus match sorting patterns. For each of the four information-based variables, the third column in the table reports the change in joint valuation of a matched bank-firm pair, $\Delta\Pi_\beta$ (defined in (6)), for a one-standard-deviation change in the variable. So, the values are calculated as $Std. \times Est. = \Delta\Pi_\beta$. Because a single loan contract is very unlikely to ever change a bank's HHI by 0.3 (the standard deviation of the HHI of all banks in the sample), I use the standard deviation of Z_{HHI} instead, which, for each inequality considered, measures the difference in the two bank's HHIs calculated based on their actual and based on their counterfactual portfolios (see Subsection 5.2). The subsequent columns in the table show the changes in joint valuations based on the lower and upper bounds of the coefficient's confidence intervals.

Table 3: Relative importance of characteristics

		Point estimate		CI lower bound		CI upper bound	
Variable	Std.	Est.	$\Delta\Pi_\beta$	Est.	$\Delta\Pi_\beta$	Est.	$\Delta\Pi_\beta$
Relation to firm	0.4	100	**40**	100	**40**	100	**40**
Relation to sector	0.1	13.25	**1.33**	2.96	**0.30**	90.96	**9.10**
Rank difference	23.6	0.07	**1.65**	0.02	**0.47**	0.07	**1.65**
Distance	687.9	-0.005	**-3.44**	-0.02	**-13.67**	-0.01	**-6.88**
Bank HHI	0.03	162.55	**4.88**	56.41	**1.96**	194.18	**5.83**

Note: Standard deviations (Std.) are from Table 1, estimates (Est.) from Table 2. $\Delta\Pi_\beta$ is the change in the joint valuation of a matched bank-firm pair, as defined in (6), for a one-standard-deviation change in the variable. It is calculated as $Std. \times Est. = \Delta\Pi_\beta$. For the HHI I use the standard deviation of Z_{HHI}.

The table has two main messages. First, among the characteristics considered, a previous relationship between a bank and a firm has the largest effect on joint valuations and is thus the strongest driver of match sorting patterns. This finding is consistent with the results of Chen and Song (2013), who consider the same set of information-based variables and find that among them, a previous relationship is the strongest driver of joint valuations. The finding is plausible, given that a previous relationship will considerably lower monitoring (and screening) costs for banks in the event of a new loan, and given that firms seem to face quite substantial switching costs when they change lenders.

One possible explanation for why the effect of distance is not stronger is that the variable is calculated based on locations of headquarters, which, as discussed in Section 4, is a noisy measure of the distance between firms and the bank branches that they interact with. Another is that borrowing from nearby banks involves a trade-off for firms: Degryse and Ongena (2005) find that banks charge higher interest rates to nearby borrowers, presumably because they exploit their market power that results from proximity; if firms were to borrow from other banks that were located further away, then they would incur higher costs when visiting those banks. To the extent that banks exploit this market power, a firm's value of borrowing from nearby banks is lower, which should reduce the likelihood of a match.

The second main message is that complementarities in preferences of banks are an important driver of match sorting patterns. The point estimates suggest, for instance, that a one-standard-deviation increase in Z_{HHI} adds about 1.5 times as much to the joint valuation of the two matching parties as a one-standard-deviation reduction in the distance between them. Complementarities remain equally important when we focus on the lower or upper bounds of the estimates instead. Moreover, as discussed in Subsection 5.2, the HHI's contribution to joint valuation is driven by its contribution to the match valuations of banks (γ^b) only, while for each information-based variable, the coefficient estimate is the sum of the variable's contribution to the valuations of banks and firms combined ($\beta^b + \beta^f$). This means that for the matching decisions of banks, the relative importance of the HHI is higher than suggested by the comparisons in Table 3.

Caveats: My results come with the caveat that the relative importance of information-based characteristics might be underestimated for two reasons. First, as discussed in Section 4, bank-firm lending in Japan is driven by motives other than profit maximisation, especially between firms and their main bank. To account for this, I drop loans from a firm's main bank from my sample. Yet another particular feature of the Japanese corporate loan market is that non-main banks delegate the monitoring of firms to main banks. So, to the extent that non-main banks do monitor less, monitoring cost and thus the information-based characteristics that affect these costs should be less of a concern to them.

Second, the rationale for why information-based factors matter relies on the need for banks to acquire soft information about borrowing firms. The financial intermediation literature generally assumes that this need is strongest for lending to small and young firms. Because my dataset only provides information on publicly-listed firms, it is biased towards large and more established firms. To the extent that banks rely more on hard rather than soft information when monitoring such firms, the relative importance of information-based characteristics might be lower than it is for lending decisions to firms of all sizes.

Garbade and Silber (1978) note that "diversity of institutional detail is an advantage rather than a disadvantage. Certain types of economic behavior are sufficiently powerful to transcend a particular historical framework" (p. 820). Understanding whether my results transcend the historical framework of Japan is an important next step in understanding match sorting patterns; main bank relationships and close bank-firm ties in general, the particular sample period of my data, or the bias of the sample towards more established and larger firms might all limit the explanatory power of my results in explaining match sorting patterns in the corporate loan markets of other countries and in other contexts. However, that many of my results with regard to information-based variables are consistent with those of Chen and Song (2013), who use US data from 2000 to 2003, suggests that my attempts to limit the degree to which Japan specific institutional features influence my results were at least partially successful. It also suggests that there might indeed be universal drivers of match sorting patterns.

Extensions: There are a number of ways in which the results of this paper can be extended. First, it would be interesting to see how the results change for different types of banks and firms, for loans with different durations, and across a number of different countries. Findings that in all these cases information-based variables and complementarities in preferences of banks are important drivers of match sorting patterns would corroborate the argument above that – in the words of Garbade and Silber (1978)– the conclusions of this paper do indeed transcend the particular historical and institutional context from which they are inferred.

Second, I assume in this paper that firms do not have complementarities between lenders. The validity of that assumption can be tested: one of the innovations of the framework of Fox (2016) is that it allows all agents to have complementarities in preferences, so that testing for the significance of a suitably defined HHI index for firms (based on the riskiness or size of their lenders, for instance) could lend empirical support to that assumption.

Third, and maybe most interestingly, the availability of three decades of data could be exploited to test whether the developments in information technology during that time have reduced the degree of asymmetric information between banks and firms. If access to such technology has made it easier for banks to obtain soft information about potential borrowers, monitoring costs should have fallen and variables determining monitoring cost – such as *distance*, *rank difference*, and *previous relationship* – should become less important drivers of bank's decision of what firms to lend to.[30] Assuming that banks' choice of whether to focus or diversify its loan portfolio is independent of the degree of asymmetric information, one can test this hypothesis by testing whether the relative importance of the information-based variables relative to the HHI variable has decreased over time. When estimating the matching maximum score objective

[30]There is some evidence for that: Petersen and Rajan (2002) find, for instance, that the average distance between banks and their borrowers has increased over time and attribute that increase to a productivity increase of banks that they link to their increased use of computer and information technology.

functions in (10) and (11) separately for data of each year (or for pools of years), a declining importance of information-based variables would manifest itself in two ways: the magnitude of the parameter estimates, relative to that of the HHI, should decline, and the share of inequalities that is correctly predicted by the baseline specification without the HHI should fall over time.

As part of my doctoral work I plan to explore all three of these areas.

7 Conclusion

In this paper I show that both a set of information-based characteristics, such as geographical distance and previous relationships, as well as banks' objective to focus or diversify their loan portfolio in a particular way drive match sorting patterns in the corporate loan market.

The finding that complementarities in bank preferences are a driver of match sorting patterns opens up an exciting and largely unexplored field of inquiry. The degree to which loan portfolios of banks overlap has important consequences for financial stability because it shapes the structure of networks of interdependencies among banks and between banks and the real economy. A nuanced understanding of the role of complementarities for the portfolio choice of different types of banks, and maybe in different periods of economic and financial cycles, might therefore help explain why loan portfolios overlap the way they do, and thus help us better understand and measure the fragility of the financial system over time.

Appendix A Computational details

A.1 Differential evolution:

The differential evolution (DE) algorithm is due to Storn and Price (1997) and works through the following steps:

1. Randomly select an initial population of N points $\mathbf{x} \in \mathbb{R}^n$ from the search space. (For my baseline specification with three real-valued parameters to be estimated, each point $\mathbf{x}$ corresponds to a parameter vector $\tilde{\beta} \in \mathbb{R}^3$.)

2. Then, for each $\mathbf{x}$ pick three other points, $\mathbf{a}, \mathbf{b}, \mathbf{c}$, that are distinct from each other and from $\mathbf{x}$.

3. Define $\mathbf{x}$'s potentially new position as $\mathbf{y} = (y_1, \ldots, y_i, \ldots, y_n)$. To compute each element i of $\mathbf{y}$, do the following: draw a number r_i from $r \sim U(0, 1)$. If $r_i < p_c$, construct y_i as $y_i = a_i + S(b_i - c_i)$, where $p_c \in [0, 1]$ is called the *crossover probability*, and $S \in [0, 2]$ is a *scaling factor* or weight. If $r_i \geq p_c$, set $y_i = x_i$. Intuitively, the potentially new position of $\mathbf{x}$ is a crossover of point $\mathbf{x}$ with a linear combination of the other three points.

4. If $Q(\mathbf{y}) > Q(\mathbf{x})$, replace $\mathbf{x}$ with $\mathbf{y}$. Otherwise leave $\mathbf{x}$ in the population.

5. Repeat steps (2) to (4) for the specified number of iterations, K. If that number of iterations is reached, pick from the set of points $\mathbf{x}$ the one for which $Q(\mathbf{x})$ is maximal.

A.2 Estimation procedure and parameter settings

To produce my results, I do the following:

- For each of my 31 markets (which correspond to the 31 years of data from 1983 to 2013), I randomly sample 20,000 inequalities out of the set of all possible inequalities. In each of these markets, the set of all possible inequalities is the set of all possible pairs of bank-firm matches such that (i) in neither of the two matches the bank is the firm's main bank, and (ii) the counterfactual matchings created by swapping partners do not already exist in the data.

- In the objective function, I add $1e-10$ to the right hand side of each inequality to ensure that inequalities that are evaluated as $0 > 0$ are consistently evaluated as satisfied, and not driven by Mathematica internal approximation errors.
- To produce the main results in Table 2, I pool the $31 \times 20,000$ inequalities to optimise the objective function (9) based on the data from all 620,000 inequalities.
- I optimise (9) using Mathematica's built-in DE routine, using 250 initial points $\mathbf{x}$, and a maximum of 15,000 iterations through the steps (2) to (4) described above. For the crossover probability p_c and the scaling factor S I use the default values, which are 0.5 and 0.6, respectively.
- I run each optimisation procedure 20 times with a different set of initial points. The results are identical in all cases.
- To choose whether to normalise *relationship to firm* to -100 or 100, I estimate the results based on both normalisations and then choose the specification that generates a higher objective function value.

Appendix B Derivations

For each loan contract ω, the vector $x(\omega)$ contains the variables *distance*, *rank difference*, *relationship to firm*, and *relationship to sector* for the bank and the firm associated with that contract. Using the assumption that firms have substitutable preferences between their lenders and banks have complementarities between their borrowers (captured by the HHI variable in their valuation), we can write the valuation functions from (5) for firms and banks, respectively as

$$X(\Phi)'\beta^f = \beta^{f\prime} \sum_{\omega \in \Phi} x(\omega) \qquad (12)$$

$$X(\Psi)'\beta^b = \beta^{b\prime} \sum_{\omega \in \Psi} x(\omega) + \gamma^b HHI(\Psi). \tag{13}$$

The matching maximum score inequality is

$$\begin{aligned}
&X(\Psi_{b(\omega_1)})'\beta^b + X(\Phi_{f(\omega_1)})'\beta^f \\
&+X(\Psi_{b(\omega_2)})'\beta^b + X(\Phi_{f(\omega_2)})'\beta^f \\
&> X(\Psi'_{b(\omega_1)})'\beta^b + X(\Phi'_{f(\omega_1)})'\beta^f \\
&+X(\Psi'_{b(\omega_2)})'\beta^b + X(\Phi'_{f(\omega_2)})'\beta^f.
\end{aligned}$$

Using the above valuation functions we get

$$\begin{aligned}
&\left(\beta^{b\prime} \sum_{\omega \in \Psi_{b(\omega_1)}} x(\omega) + \gamma^b HHI(\Psi_{b(\omega_1)})\right) + \left(\beta^{f\prime} \sum_{\omega \in \Phi_{f(\omega_1)}} x(\omega)\right) \\
+&\left(\beta^{b\prime} \sum_{\omega \in \Psi_{b(\omega_2)}} x(\omega) + \gamma^b HHI(\Psi_{b(\omega_2)})\right) + \left(\beta^{f\prime} \sum_{\omega \in \Phi_{f(\omega_2)}} x(\omega)\right) \\
>&\left(\beta^{b\prime} \sum_{\omega \in \Psi'_{b(\omega_1)}} x(\omega) + \gamma^b HHI(\Psi'_{b(\omega_1)})\right) + \left(\beta^{f\prime} \sum_{\omega \in \Phi'_{f(\omega_1)}} x(\omega)\right) \\
+&\left(\beta^{b\prime} \sum_{\omega \in \Psi'_{b(\omega_2)}} x(\omega) + \gamma^b HHI(\Psi'_{b(\omega_2)})\right) + \left(\beta^{f\prime} \sum_{\omega \in \Phi'_{f(\omega_2)}} x(\omega)\right).
\end{aligned}$$

Using the fact that except for the swap, all other contracts remain the same, we can drop common elements on both sides of the inequality to get

$$\begin{aligned}
&x(\omega_1)'\beta^b + \gamma^b HHI(\Psi_{b(\omega_1)}) + x(\omega_1)'\beta^f \\
&+x(\omega_2)'\beta^b + \gamma^b HHI(\Psi_{b(\omega_2)}) + x(\omega_2)'\beta^f \\
&> x(\omega_3)'\beta^b + \gamma^b HHI(\Psi'_{b(\omega_1)}) + x(\omega_3)'\beta^f \\
&+x(\omega_4)'\beta^b + \gamma^b HHI(\Psi'_{b(\omega_2)}) + x(\omega_4)'\beta^f
\end{aligned}$$

and then rearrange to get

$$\begin{aligned}
&\Big(x(\omega_1)' + x(\omega_2)'\Big)\beta^b - \Big(x(\omega_3)' + x(\omega_4)'\Big)\beta^b \\
+\gamma^b\Big(HHI(\Psi_{b(\omega_1)}) + HHI(\Psi_{b(\omega_2)})\Big) &- \gamma^b\Big(HHI(\Psi'_{b(\omega_1)}) + HHI(\Psi'_{b(\omega_2)})\Big) \\
&+\Big(x(\omega_1)' + x(\omega_2)'\Big)\beta^f - \Big(x(\omega_3)' + x(\omega_4)'\Big)\beta^f \\
&> 0
\end{aligned}$$

$$\begin{aligned}
&\left(\sum_{\omega\in\{\omega_1,\omega_2\}} x(\omega)' - \sum_{\omega\in\{\omega_3,\omega_4\}} x(\omega)'\right)(\beta^b + \beta^f) \\
+\gamma^b&\left(\sum_{\omega\in\{\omega_1,\omega_2\}} HHI(\Psi_{b(\omega)}) - \sum_{\omega\in\{\omega_3,\omega_4\}} HHI(\Psi_{b(\omega)})\right) > 0
\end{aligned}$$

$$\begin{aligned}
&\left(\sum_{\omega\in\{\omega_1,\omega_2\}} x_1(\omega) - \sum_{\omega\in\{\omega_3,\omega_4\}} x_1(\omega)\right)(\beta_1^b + \beta_1^f) \\
&\left(\sum_{\omega\in\{\omega_1,\omega_2\}} x_2(\omega) - \sum_{\omega\in\{\omega_3,\omega_4\}} x_2(\omega)\right)(\beta_2^b + \beta_2^f) \\
&\qquad \ldots\text{[all elements in } x(\omega)]\ldots \\
+\gamma^b&\left(\sum_{\omega\in\{\omega_1,\omega_2\}} HHI(\Psi_{b(\omega)}) - \sum_{\omega\in\{\omega_3,\omega_4\}} HHI(\Psi_{b(\omega)})\right) > 0
\end{aligned}$$

Letting, for $x \in x(\omega)$,

$$Z_x = \left(\sum_{\omega\in\{\omega_1,\omega_2\}} x(\omega) - \sum_{\omega\in\{\omega_3,\omega_4\}} x(\omega)\right),$$

we can write

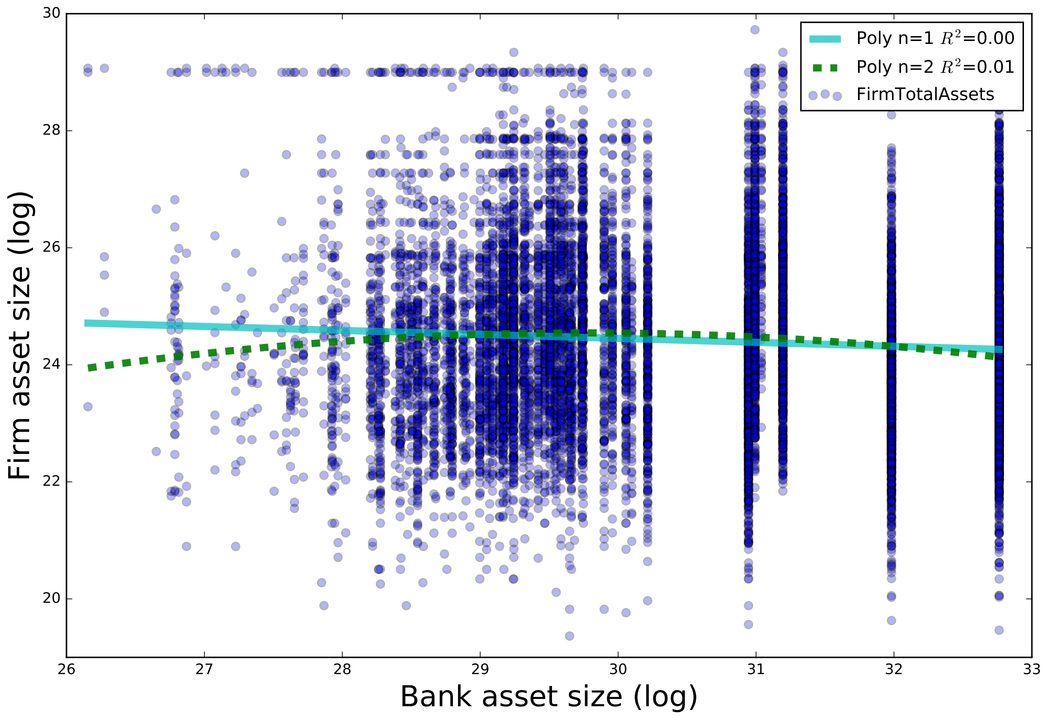
$$\sum_{x\in x(\omega)} Z_x(\beta_i^b + \beta_i^f) + \gamma^b\left(\sum_{\omega\in\{\omega_1,\omega_2\}} HHI(\Psi_{b(\omega)}) - \sum_{\omega\in\{\omega_3,\omega_4\}} HHI(\Psi_{b(\omega)})\right) > 0.$$

Appendix C Robustness checks

C.1 Positive assortative matching

Figure 2 plots the predicted values of polynomial regressions of firm size on bank size, where size corresponds to the log of total assets. The linear prediction shown by the solid cyan line (as well as unreported OLS estimation results) corroborates the finding from Table 2: to the extent that there is a relationship between bank and firm size, that relationship is negative. The dashed green line is produced by adding the square of bank size to the right hand side of the regression and suggests that the relationship might be slightly non-linear, although not in a way that would support the hypothesis of positive assortative matching.

Figure 2: Correlation between bank and firm size



Note: Bank and firm size are measured as the log of total assets. Lines are predicted values from a linear regression of firm size on bank size including an intercept. In addition to bank size, the model plotted by the dashed green line also includes the square of bank size as a right-hand side variable.

C.2 Further robustness checks

The size of my dataset, the time it takes to compute HHI indices for banks, and especially the time required to compute confidence intervals based on subsampling means that producing my results takes several days: producing the data to feed into the objective functions (10) and (11) takes about 2 days, while producing estimation results takes between 3 and 9 days, depending on server availability at the University of Oxford's Advanced Research Computing facility (http://www.arc.ox.ac.uk). As a result, I cannot produce more robustness checks before the submission deadline of the MPhil thesis. In this section I list a number of robustness checks I intend to carry out in later versions of the paper.

Estimates for full sample and non-main bank sample: The results presented in the main text are based on loan contracts between firms and banks that are not their main bank only. It would be instructive to compare these results with results based on data that includes all loans or loans from main banks only. As pointed out in Section 2, main banks extend loans based on criteria that cannot be captured by the variables in my model so that, when including those loans, the share of correctly predicted inequalities should be lower. Yet at the same time, non-main banks delegate the monitoring of firms to main banks, so that for main banks, the set of information-based variables that drives monitoring costs should be more important, and the model should thus predict more inequalities correctly. Which of these effects dominates is not clear a priori.

Sample without city banks: As discussed in Section 4, *distance* is calculated based on the location of headquarters of banks and firms, rather than on the locations of the bank and firm local branches that actually interact with each other. This will lead to especially high and misleading distance measures for loans involving city banks, which are headquartered in Tokyo or Osaka but operate branches nationwide. For regional banks and trusts this will be less of a problem, because the former focus on regional lending and the latter might have no branches at all. Reproducing the main results based on data that involves no loans from city banks would thus be interesting for two reasons: to see whether distance becomes more important once it is more accurately measured, and to see whether the relative importance of the HHI is robust.

Alternative time horizon of previous relationship dummies: Defining previous lending relationships to firms and sectors based on three year windows (see Section 4) is arbitrary. To make sure that my results are robust to that choice, I will produce robustness test with 5-year windows as well.

Alternative dataset: All my results are based on a single random sample from my dataset. Given the size of the samples – I sample 20,000 inequalities for each year – it is unlikely that my results are an artefact of the specific sample. But to make sure this is not the case, I want to reproduce my results with a second sample.

Results for pre and post 1996 period: Panel (C) in Figure 1 shows a clear break in the pattern in the number of loans in 1996. While this time trend can be explained by the number of firms covered in the NEEDS data and the declining number of banks (as discussed in Section 4), I want to compare results based on data from these two periods separately.

Eliminate loan repayments: As discussed in Section 4, my definition of the *new loan* variable can lead to cases where the counterfactual matching in the matching maximum score inequality is not between a bank and a firm that are not matched but between a bank and a firm where the firm has repaid a loan to the bank over the course of the year. This is the case in 313,307 bank-firm-year observations, or about 2 percent or all observations. To demonstrate that these cases do not affect my results, I want to present results where I exclude such cases from the data.

Comparison to probit and logit models: As argued in Section 3.3, the matching maximum score estimator has a number of advantages over estimates based on logit and probit models. To see whether those advantages manifest themselves in different results, I want to compare my results with results obtained from these estimators.

Correct for large firm bias: As discussed in Section 6, my sample of publicly-listed companies is biased towards larger and more established firms. To get a sense of how much this drives my results, I want to conduct two robustness checks. One where I use loans to the largest firms in my sample only, and another where I use loans to the smallest firms in my sample only. The latter will be informative under the assumption that the smallest publicly-listed firms are more similar to unlisted firms than larger publicly-listed firms.

References

Acemoglu, Daron, Asuman E Ozdaglar, and Alireza Tahbaz-Salehi (2015). "Systemic Risk in Endogenous Financial Networks". In: *Available at SSRN 2553900*.

Acharya, Viral V., Iftekhar Hasan, and Anthony Saunders (2006). "Should Banks Be Diversified? Evidence from Individual Bank Loan Portfolios". In: *The Journal of Business* 79.3, pp. 1355–1412.

Agarwal, Nikhil (2015). "An Empirical Model of the Medical Match". In: *American Economic Review* 105.7, pp. 1939–78.

Ahlin, Christian (2016). "Group lending, matching patterns, and the mystery of microcredit: Evidence from Thailand". Tech. rep. Working Paper.

Akerlof, George A (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism". In: *The Quarterly Journal of Economics*. The Quarterly Journal of Economics 84.3, pp. 488–500.

Akkus, Oktay, J. Anthony Cookson, and Ali Hortaçsu (2016). "The Determinants of Bank Mergers: A Revealed Preference Analysis". In: *Management Science* 62.8, pp. 2241–2258.

Allen, Franklin and Douglas Gale (2000). "Financial Contagion". In: *Journal of Political Economy* 108.1, pp. 1–33.

Aoki, Masahiko and Hugh Patrick (1995). *The Japanese main bank system: Its relevance for developing and transforming economies*. Oxford University Press.

Azevedo, Eduardo M and John William Hatfield (2015). "Existence of equilibrium in large matching markets with complementarities". Working Paper. Wharton School, University of Pennsylvania.

Baccara, Mariagiovanna, Ayşe İmrohoroğlu, Alistair J. Wilson, and Leeat Yariv (2012). "A Field Study on Matching with Network Externalities". In: *American Economic Review* 102.5, pp. 1773–1804.

Beck, Thorsten, Andrea Colciago, and Damjan Pfajfar (2014). "The Role of Financial Intermediaries in Monetary Policy Transmission". In: *Journal of Economic Dynamics and Control* 43.C, pp. 1–11.

Becker, Gary S. (1973). "A Theory of Marriage: Part I". In: *Journal of Political Economy* 81.4, pp. 813–846.

Berger, Allen N., Nathan H. Miller, Mitchell A. Petersen, Raghuram G. Rajan, and Jeremy C. Stein (2005). "Does function follow organizational form? Evidence from the lending practices of large and small banks". In: *Journal of Financial Economics* 76.2, pp. 237–269.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist (1999). "The financial accelerator in a quantitative business cycle framework". In: *Handbook of Macroeconomics*. Ed. by J. B. Taylor and M. Woodford. Vol. 1. Handbook of Macroeconomics. Elsevier. Chap. 21, pp. 1341–1393. URL: https://ideas.repec.org/h/eee/macchp/1-21.html.

Boot, Arnoud W. A. (2000). "Relationship Banking: What Do We Know?" In: *Journal of Financial Intermediation* 9.1, pp. 7–25.

Caballero, Ricardo J., Takeo Hoshi, and Anil K. Kashyap (2008). "Zombie Lending and Depressed Restructuring in Japan". In: *American Economic Review* 98.5, pp. 1943–77.

Che, Yeon-Koo, Jinwoo Kim, and Fuhito Kojima (2013). "Stable matching in large economies". In: *Manuscript, Stanford University*.

Chen, Jiawei and Kejun Song (2013). "Two-sided matching in the loan market". In: *International Journal of Industrial Organization* 31.2, pp. 145 –152.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque (2012). "Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market". In: *Journal of Political Economy* 120.4, pp. 659–695.

Choo, Eugene and Aloysius Siow (2006). "Who Marries Whom and Why". In: *Journal of Political Economy* 114.1, pp. 175–201.

Coase, R. H. (1937). "The Nature of the Firm". In: *Economica* 4.16, pp. 386–405.

Cole, Rebel A., Lawrence G. Goldberg, and Lawrence J. White (2004). "Cookie Cutter vs. Character: The Micro Structure of Small Business Lending by Large and Small Banks". In: *Journal of Financial and Quantitative Analysis* 39.2, pp. 227–251.

Degryse, Hans and Steven Ongena (2005). "Distance, Lending Relationships, and Competition". In: *Journal of Finance* 60.1, pp. 231–266.

Degryse, Hans, Moshe Kim, and Steven Ongena (2009). *Microeconometrics of banking: methods, applications, and results*. Oxford University Press, USA.

Dell'Ariccia, Giovanni, Enrica Detragiache, and Raghuram Rajan (2008). "The real effect of banking crises". In: *Journal of Financial Intermediation* 17.1, pp. 89–112.

Diamond, Douglas W. (1984). "Financial Intermediation and Delegated Monitoring". In: *The Review of Economic Studies* 51.3, pp. 393–414.

Dupuy, Arnaud and Alfred Galichon (2014). "Personality Traits and the Marriage Market". In: *Journal of Political Economy* 122.6, pp. 1271–1319.

Echenique, Federico, Sangmok Lee, Matthew Shum, and M. Bumin Yenmez (2013). "The Revealed Preference Theory of Stable and Extremal Stable Matchings". In: *Econometrica* 81.1, pp. 153–171.

Eisenberg, Larry and Thomas H. Noe (2001). "Systemic Risk in Financial Systems". In: *Management Science* 47.2, pp. 236–249.

Farboodi, Maryam (2015). "Intermediation and voluntary exposure to counterparty risk". Working Paper. URL: https://ideas.repec.org/p/red/sed014/365.html.

Fleiner, Tamás, Zsuzsanna Jankó, Akihisa Tamura, and Alexander Teytelboym (2015). "Trading Networks with Bilateral Contracts". Working Paper. URL: https://ideas.repec.org/p/arx/papers/1510.01210.html.

Fox, Jeremy T. (2010). "Identification in Matching Games". In: *Quantitative Economics* 1.2, pp. 203–254.

— (2016). "Estimating Matching Games with Transfers". Working Paper. Houston: Rice University. URL: http://fox.web.rice.edu/working-papers/fox-matching-maximum-score.pdf.

Fox, Jeremy T and David Santiago (2015). "A Toolkit for Matching Maximum Score Estimation and Point and Set Identified Subsampling Inference". Working Paper. URL: fox.web.rice.edu/computer-code/matchestimation-452-documen.pdf.

Freixas, Xavier and Jean-Charles Rochet (2008). *Microeconomics of banking.* MIT press.

Freixas, Xavier, Bruno M Parigi, and Jean-Charles Rochet (2000). "Systemic Risk, Interbank Relations, and Liquidity Provision by the Central Bank". In: *Journal of Money, Credit and Banking* 32.3, pp. 611–38.

Fricke, Daniel and Tarik Roukny (2017). "Generalists and Specialists in Credit Markets". Tech. rep. Said Business School Working Paper 2016-11.

Gabaix, Xavier and Augustin Landier (2008). "Why Has CEO Pay Increased so Much?" In: *The Quarterly Journal of Economics* 123.1, pp. 49–100.

Gale, David and Lloyd S Shapley (1962). "College Admissions and the Stability of Marriage". In: *The American Mathematical Monthly* 69.1, pp. 9–15.

Garbade, Kenneth D. and William L. Silber (1978). "Technology, Communication and the Performance of Financial Markets: 1840-1975". In: *The Journal of Finance* 33.3, pp. 819–832.

Glasserman, Paul and H. Peyton Young (2016). "Contagion in Financial Networks". In: *Journal of Economic Literature* 54.3, pp. 779–831.

Goldsmith, Raymond (1969). *Financial Structure and Development.* New Haven: Yale University Press.

Grossman, Sanford J and Oliver D Hart (1986). "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration". In: *Journal of Political Economy* 94.4, pp. 691–719.

Hachem, Kinda (2011). "Relationship lending and the transmission of monetary policy". In: *Journal of Monetary Economics* 58.6–8, pp. 590 –600.

Hall, Robert E. (2010). "Why Does the Economy Fall to Pieces after a Financial Crisis?" In: *Journal of Economic Perspectives* 24.4, pp. 3–20.

Harada, Kimie, Takeo Hoshi, Masami Imai, Satoshi Koibuchi, and Ayako Yasuda (2015). "Japan's financial regulatory responses to the global financial crisis". In: *Journal of Financial Economic Policy* 7.1, pp. 51–67.

Hart, Oliver and John Moore (1990). “Property Rights and the Nature of the Firm”. In: *Journal of Political Economy* 98.6, pp. 1119–1158.

Hatfield, John William, Scott Duke Kominers, Alexandru Nichifor, Michael Ostrovsky, and Alexander Westkamp (2013). “Stability and Competitive Equilibrium in Trading Networks”. In: *Journal of Political Economy* 121.5, pp. 966–1005.

Hayden, Evelyn, Daniel Porath, and Natalja Westernhagen (2007). “Does Diversification Improve the Performance of German Banks? Evidence from Individual Bank Loan Portfolios”. In: *Journal of Financial Services Research* 32.3, pp. 123–140.

Hitsch, Gunter J., Ali Hortaçsu, and Dan Ariely (2010). “Matching and Sorting in Online Dating”. In: *American Economic Review* 100.1, pp. 130–63.

Hoggarth, Glenn, Ricardo Reis, and Victoria Saporta (2002). “Costs of banking system instability: Some empirical evidence”. In: *Journal of Banking Finance* 26.5, pp. 825 –855.

Hoshi, Takeo and Anil K Kashyap (2010). “Will the U.S. bank recapitalization succeed? Eight lessons from Japan”. In: *Journal of Financial Economics* 97.3, pp. 398–417.

Hoshi, Takeo and Anil Kashyap (2001). *Corporate financing and corporate governance in Japan: the road to the future.*

Hubbard, R Glenn, Kenneth N Kuttner, and Darius N Palia (2002). “Are There Bank Effects in Borrowers’ Costs of Funds? Evidence from a Matched Sample of Borrowers and Banks”. In: *The Journal of Business* 75.4, pp. 559–581.

Hüser, Anne-Caroline (2015). “Too interconnected to fail: A survey of the interbank networks literature”. Working Paper. URL: https://ideas.repec.org/p/zbw/safewp/91.html.

Jackson, Matthew O (2008). *Social and Economic Networks.* Princeton: Princeton University Press.

Jackson, Matthew O. (2014). “Networks in the Understanding of Economic Behaviors”. In: *Journal of Economic Perspectives* 28.4, pp. 3–22.

Jackson, Matthew O, Brian W Rogers, and Yves Zenou (forthcoming). “The economic consequences of social network structure”. In: *Journal of Economic Perspectives.*

Kim, Moshe, Doron Kliger, and Bent Vale (2003). "Estimating switching costs: the case of banking". In: *Journal of Financial Intermediation* 12.1, pp. 25–56.

Markowitz, Harry (1952). "Portfolio Selection". In: *The Journal of Finance* 7.1, pp. 77–91.

Miyamoto, Koichi (2016). *Global legal insights: Banking regulation in Japan.* URL: https://www.globallegalinsights.com/jurisdictions/japan.

Navarro, Noémie and Fabio Castiglionesi (2016). "(In)Efficient Interbank Networks". Working Paper. URL: https://ideas.repec.org/p/grt/wpegrt/2016-13.html.

Nikkei, Nikkei Digital Media Inc. (2008). *NEEFDS-FinancialQUEST Codebook: Corporate Attribute Database.*

— (2012). *NEEFDS-FinancialQUEST Codebook: Corporate Borrowings from Financial Institutions Database.*

— (2013). *NEEFDS-FinancialQUEST Codebook: Corporate Financial Data (TANSHIN / YUHO) Database.*

Peek, Joe and Eric S. Rosengren (2005). "Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan". In: *American Economic Review* 95.4, pp. 1144–1166.

Petersen, Mitchell A. and Raghuram G. Rajan (2002). "Does Distance Still Matter? The Information Revolution in Small Business Lending". In: *Journal of Finance* 57.6, pp. 2533–2570.

Reinhart, Carmen M. and Kenneth S. Rogoff (2013). "Banking crises: An equal opportunity menace". In: *Journal of Banking Finance* 37.11, pp. 4557 –4573.

Romano, Joseph P. and Azeem M. Shaikh (2008). "Inference for identifiable parameters in partially identified econometric models". In: *Journal of Statistical Planning and Inference* 138.9, pp. 2786 –2807.

Roth, Alvin E. (2002). "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics". In: *Econometrica* 70.4, pp. 1341–1378.

— (2015). *Who Gets What—and Why: The New Economics of Matchmaking and Market Design*. Houghton Mifflin Harcourt.

Sharpe, Steven A. (1990). "Asymmetric Information, Bank Lending and Implicit Contracts: A Stylized Model of Customer Relationships". In: *The Journal of Finance* 45.4, pp. 1069–1087.

Stein, Jeremy C. (2002). "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms". In: *The Journal of Finance* 57.5, pp. 1891–1921.

Stomper, Alex (2006). "A Theory of Banks' Industry Expertise, Market Power, and Credit Risk". In: *Management Science* 52.10, pp. 1618–1633.

Storn, Rainer and Kenneth Price (1997). "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces". In: *Journal of Global Optimization* 11.4, pp. 341–359.

Tabak, Benjamin M., Dimas M. Fazio, and Daniel O. Cajueiro (2011). "The effects of loan portfolio concentration on Brazilian banks' return and risk". In: *Journal of Banking & Finance* 35.11, pp. 3065–3076.

Terviö, Marko (2008). "The Difference That CEOs Make: An Assignment Model Approach". In: *The American Economic Review* 98.3, pp. 642–668.

Uchida, Hirofumi and Gregory Udell (2014). "Banking in Japan". In: Oxford Handbook of Banking. Oxford: Oxford University Press, pp. 903 –940.

Wasmer, Etienne and Philippe Weil (2004). "The Macroeconomics of Labor and Credit Market Imperfections". In: *American Economic Review* 94.4, pp. 944–963.

Winton, Andrew (1999). "Don't Put All Your Eggs in One Basket? Diversification and Specialization in Lending". Working Paper. URL: https://ideas.repec.org/p/wop/pennin/00-16.html.

Woo, David and Akihiro Kanaya (2000). "The Japanese Banking Crisis of the 1990's; Sources and Lessons". IMF Working Papers 00/7. International Monetary Fund. URL: https://ideas.repec.org/p/imf/imfwpa/00-7.html.

von Thadden, Ernst-Ludwig (2004). "Asymmetric information, bank lending and implicit contracts: the winner's curse". In: *Finance Research Letters* 1.1, pp. 11 –23.