(Unobserved) Heterogeneity in the bank lending channel: Accounting for bank-firm interactions and specialization*

Bryan Gutierrez
U Minnesota

Alonso Villacorta
UC Santa Cruz

Lucciano Villacorta
Central Bank of Chile

January 2025

Abstract

Using bank-firm credit data, we develop a framework that estimates the transmission of bank credit supply shocks on firms, incorporating bank-firm interactions in unobserved factors that capture lending specialization. Bank shocks can have heterogeneous effects across firm types, where firm-type is unobserved by the econometrician. We uncover significant heterogeneity in the transmission across estimated firm types, with types predicted by firm observables and bank specialization. Accounting for such heterogeneity and the bank-firm network structure is essential for understanding the bank lending channel and its real effects. We identify a novel bank-firm matching channel: credit shocks are amplified through the bank-firm network.

JEL Classification: C23, C38, G21.
Keywords: Bank lending channel, credit supply identification, unobserved heterogeneity, grouped-fixed effects, bank specialization

*We thank Stephane Bonhomme, Mark Egan, Galina Hale, Victoria Ivashina, Richard Thakor, Manuel Luy, Elena Manresa, Atif Mian, Veronica Rappoport, Daniel Paravisini, Cecilia Parlatore, Monika Piazzesi, Jose-Luis Peydro, Brenda Samaniego, Andres Sarto, Toni Whited, Daniel Wolfenzon and seminar audiences at Princeton, SF Fed, CEMFI, UC Santa Cruz, USC Marshall, U Minnesota, U Pacifico, Central Bank of Chile, Peruvian SBS, PUC Chile, SED Winter 2024, SED Barcelona 2024, SFS Cavalcade 2024, Barcelona GSE Summer Forum, 28th International Panel Data Conference, CEF 2023, AEFIN Finance Forum, Santiago Macro Workshop 2023, and LACEA 2023 for helpful comments and discussions. Contact: avillaco@ucsc.edu.

1 Introduction

The bank lending channel—the impact of bank credit supply shocks on firms' borrowing capacity, investment, and overall real activity—stands at the center of the macrofinance literature.[1] Due to bank-firm interactions and specialization in lending, the bank lending channel may exhibit heterogeneity across banks and firms. In this paper, we provide evidence that the transmission of bank shocks is far from uniform; rather, it hinges on unobserved interactions between banks and firms, highlighting lending specialization. By incorporating these hidden interactions, we uncover that the impact of a credit event is shaped not only by the liquidity or credit supply shock experienced by banks but by the specific ways in which banks and firms are connected.

Empirically identifying the effects of bank shocks is challenging. Credit fluctuations may be driven by demand or supply factors, and events that influence credit supply are likely to simultaneously influence the demand for credit. State-of-the-art methodologies exploit matched bank-firm credit data and, by observing multiple connections between firms and banks, are able to control for firm demand factors. The leading contribution of Khwaja and Mian (2008) (KM hereafter) relies on using firm fixed effects to control for unobserved firm demand factors when assessing the impact of a bank shock, measured as a bank differential exposure to a credit contraction event.[2] Their empirical specification is grounded in banking models where the equilibrium credit growth rate between a firm and a bank can be *linearly* decomposed in two types of heterogeneity: a firm-specific factor capturing changes in firm demand (e.g. productivity shocks) and a bank-specific factor that capture credit supply shocks. This useful methodology has been widely extended to assess the impact of bank shocks in various event studies and firm outcomes (see Related Literature).[3]

While this methodology offers a tractable way to study the bank lending channel controlling for unobserved time-varying firm heterogeneity, it relies on the substantive assumption that the transmission of credit shocks is homogeneous, that is, that shocks to each bank (firm) are propagated equally to all its connected firms (banks), so that the within firm comparison fully absorbs firm-specific changes in credit demand. The absence of *interactions*

[1] Throughout the paper, we refer to bank credit supply shocks (or simply banks shocks) as shifts (contractions/expansions) in the supply of bank credit (e.g., driven by bank-specific liquidity shocks, monetary policy shocks affecting banks ability to attract deposits, or changes in capital regulation) that are unrelated to firm demand for credit or creditworthiness.

[2] They explore the effect of a cross-bank liquidity shock using an unanticipated nuclear test in Pakistan as a natural experiment.

[3] Studies following KM's approach often focus on specific credit events that differentially affect banks (see Related Literature). In cases where explicit bank shocks are not observed, researchers have applied a two-way fixed effects model using matched bank-firm lending data (see Amiti and Weinstein (2018), Huremovic et al. (2020) and Alfaro et al. (2021)).

between bank and firm factors in such a linear framework then does not allow for a heterogeneous transmission, and prevents assessing the relevance of the bank-firm network for aggregate credit.

Bank-firm interactions may be expected when there exists bank specialization or relationship lending. Banks constantly face a portfolio allocation task in which they have to monitor and select their expansion or contraction of credit across borrowers. Such portfolio decision depends on the borrowers' characteristics and on the specific way each bank assesses them, which could differ across banks due to differences in bank information acquisition, lending specialization, or business models. In particular, recent empirical studies provide evidence of lending specialization and an heterogeneous transmission depending on some observable characteristics. Using Peruvian bank-firm credit data, Paravisini et al. (2023) show that bank specialization in export markets is a crucial determinant for the propagation of credit shocks, and Ivashina et al. (2022) show that bank shocks affect differentially firms depending on the type of loan. Blickle et al. (2023) document specialization of U.S. banks across industries. But, what if the econometrician cannot access such specific data about bank or firm specialization? Or, what if the heterogeneity is based on an unobserved margin or multiple margins? How can we learn about the heterogeneity in the transmission of credit shocks and its implications when there exists interactions between unobserved bank and firm factors?

Our first contribution is methodological; we expand upon the current state-of-the-art framework for evaluating the bank lending channel by incorporating bank-firm interactions in *unobserved* factors. These interactions enable a heterogeneous response to credit shocks and can capture specialization in different margins. This framework is particularly useful in situations where researchers are uncertain about the specific dimensions of lending specialization or when specialization is determined by unobservable firm characteristics, allowing them to test and detect the sources of heterogeneity. For example, bank specialization could hinge on factors like export market destinations, industry sectors, loan contract terms, productivity levels, or a combination thereof. In cases where certain variables determining specialization are absent from databases the framework remains adaptable and robust. We propose an estimator for cases in which the relevant interaction occurs between banks and groups (types) of firms. In addition, we discuss interpretation of the identified interactions and potential concerns that may arise when estimating a model with homogeneous effects. Our second contribution regards the novel empirical findings. We uncover significant heterogeneity in our Peruvian dataset: Banks transmit their credit shocks significantly different across the identified firm groups. Importantly, we show that accounting for such heterogeneity and the bank-firm network structure is crucial to learn about the bank

lending channel and its real effects.

In our framework, the effect of a credit shock to each specific bank is allowed to vary across unobserved groups or types of firms, while maintaining the assumption of homogeneous transmission *within groups* of firms. Besides the heterogeneous transmission across groups, the model allows for firm-specific demand shocks that affect equally all banks. Our empirical framework emerges naturally from a banking model in which banks and firms are specialized in certain (discrete) number of activities or markets (a la Paravisini et al. (2023)) and experience a variety of credit shocks (a la KM). We show that credit growth can be dissected into a time-varying firm-specific component (e.g. capturing firm-level productivity shocks) and a time-varying factor that varies across bank-firm groups, capturing interactions between specialized banks and types of firms operating in specific markets/activities.

In our empirical specification the time-varying bank-firm group specific factor, the time-varying firm-specific factor, and the firms' group membership are left unrestricted as in a "fixed-effects" fashion and are estimated from the data. Thus, our methodology decomposes credit growth dynamics into time-varying firm effects and time-varying *bank-firm group effects*, minimizing a least-squares criterion across all possible groupings.

If firm groups were known to the econometrician, we identify bank shocks by applying the KM framework for each group, that is, by measuring the systematic differences in the lending of differently exposed banks to the same set of firms belonging to the same group. However, since groups are unobserved, we combine this idea with machine learning techniques that help in clustering observations. Clustering techniques classify observations into groups depending on some dissimilarity measure based on observable characteristics. Our goal instead is to cluster firms based on an heterogenous unobserved response to an unobserved credit shock. We build on the panel data techniques in Bonhomme and Manresa (2015), which allows for group-fixed effects with unknown grouping, and the bank lending framework of Khwaja and Mian (2008) to propose a bank-firm group effect estimator (BF-GFE). Intuitively, firms whose *differences* in their borrowing patterns from specific banks are most *similar* are grouped together in estimation.

We discuss interpretation of our identified BF-GFE through the lens of our banking model with lending specialization. Such interaction effect may capture heterogeneity in both supply or demand: It could reflect a credit supply shock that banks transmit heterogeneously across firm types, a group-specific demand shock that firms transmit heterogeneously across banks, or a combination of both factors. The challenge of interpretation is mitigated in cases where we can rely on an exogenous credit event that leads to credit supply shifters, as in KM. In such scenarios, we show how to utilize our identified interaction effect to recover a double-heterogeneous (treatment) effect that varies across banks and firm

groups. We underscore the implications of using the standard framework and estimate a homogeneous model in the presence of heterogeneity. Our analysis reveals that under endogenous bank-firm matching based on the heterogeneous effects, the KM estimator fails to provide a consistent estimate of the average bank effect across all banks and groups of firms. Particularly, we find that the estimate of the homogeneous bank effect is a weighted average of the heterogeneous bank effects for each firm group and bank, with weights that may be negative. Negative weights are specially problematic since one could have that the estimate of the homogeneous bank effect (associated with a negative credit event) appears positive, while all bank effects are negative but heterogeneous. Such concern on the double heterogeneity may be expected in cases of endogenous matching arising due to specialization. Intuitively, a specialized bank heavily exposed to the credit event may shield a certain group of firms relative to a less exposed unspecialized bank. Since, due to specialization, the network of firms used in the estimation will be more heavily composed of the group of firms the specialized bank shields, it could give the impression that exposure to the (negative) credit event is positively associated with lending.

We apply our estimation framework to credit registry data from Peru from 2005 to 2017. We observe every bank-firm lending relationship for corporate firms. Additionally, we obtain the financial statements information from the Peruvian Stock Exchange to measure firm investment. Combining these sources of information, we are able to study the real effects of bank shocks on firm investment across time.

Our novel framework allows us to uncover new results on the importance of the bank lending channel that cannot be uncovered without a more flexible treatment of unobserved heterogeneity. First, we provide novel evidence on the heterogeneous effect of credit shocks. We show that the lending patterns of firms connected with each bank shows considerably heterogeneity across our identified groups. Interestingly, for some banks, the estimated BF-GFE may have different signs across types of firms: For certain years, there are banks with an estimated positive effect for a group of firms (e.g. expansion of their credit supply) while a negative effect for other group (e.g. contraction in credit). Such heterogeneity in lending patterns between banks and groups of firms leads to sizable statistical gains of our model relative to the standard homogenous effects model: In-sample and out-of-sample mean square errors improve significantly when the model allows for heterogeneous effects.

We show that certain characteristics help explain the estimated groups. Interestingly, firm exports and collateral predict the group membership. This type of sorting is in line with specialization on export destinations markets and lending contracts as emphasized by Paravisini et al. (2023) and Ivashina et al. (2022). Moreover, building on Paravisini et al. (2023) and Blickle et al. (2023), we measure bank specialization in our firm groups by calcu-

lating the relative importance of each firm group in each bank portfolio. We find that the groupings correlate with the specialization measure, indicating that bank shocks lead to a stronger credit expansion for firms in which banks were more specialized in.

Second, we study the transmission of an observed bank shock across our identified firm groups by relying on the credit event studied in Paravisini et al. (2015). The bank shock is measured as the banks' exposure to the foreign funding shortage experienced during the capital flow reversal in Peru in 2008. We find significant differences in the transmission of such credit shock across our identified firm groups. For instance, the heterogeneous estimates reveal that exposed banks (with high foreign liabilities) decreased (relative to low foreign liability banks) their credit significantly to some group of firms but not to all, there are even some firm groups to which these banks expand their lending during this episode. Moreover, the estimated average of the heterogenous effect across firm groups and banks is around -0.3, indicating exposed banks reduced credit supply by 30% on average relative to non-exposed banks, while the homogenous estimate is -0.18.

Third, we explore the real effects of bank shocks on firm investment. We use our BF-GFE's to analyze whether firms' investment is sensitive to their bank lenders' shocks. For this exercise, the bank effects, that vary by firm group, are aggregated at the firm level by weighting each bank shock by the bank relative importance in the firm borrowing portfolio. Using our identified groups, we control for firm group fixed effects in the investment regression to capture group-specific demand shocks that could correlate with our BF-GFE. We find that when bank shocks are estimated from a model with homogeneous effects, we obtain insignificant effects of bank shocks on investment. Instead, we find a more precise and significant impact on firm investment when we consider the heterogeneity in the impact of bank shocks across firm groups, highlighting the importance of considering heterogeneity to reveal the real effects of bank shocks.

Finally, the presence of heterogeneity in the impact of credit shocks raises the question whether banks and firms create relationships and sorting patterns that amplify or smooth out the aggregate transmission of credit shocks. Our framework allows us to estimate the entire loan growth distribution corresponding to a counterfactual reallocation of relationships between banks and firms. In particular, we estimate as a counterfactual the aggregate credit growth when bank-firm are randomly matched.[4] Our aim is to assess the contribution of the endogenous bank-firm network for the propagation of credit shocks, that is, a *bank-firm matching channel*. We find that, for most years, aggregate credit growth is enhanced by the observed bank-firm credit network relative to a randomly matched network, achiev-

[4]For example, this counterfactual may capture credit allocation in an economy with large information frictions preventing firms to find their optimal specialized bank, as described in our model.

ing up to a 20% higher credit growth rate in 2016 and 2017, an a growth rate that is 3% to 10% higher for most years, with the exception of 2008-09 in which the observed network produces a 5% lower growth rate.[5]

Related Literature. Our paper is mainly related to the empirical literature that studies the bank lending channel. The literature is extensive with many important contributions exploiting matched bank-firm data under the assumption of homogeneous transmission following the seminal contribution of KM. Some examples include Jiménez et al. (2012, 2017), Schnabl (2012), Chodorow-Reich (2014), Paravisini et al. (2015), Amiti and Weinstein (2018), Jiménez et al. (2020), Huremovic et al. (2020), Alfaro et al. (2021), among many others.[6] Most closely, our paper is related to the papers that emphasize *bank-firm interactions* in the transmission of credit shocks. The pioneering contribution of Paravisini et al. (2023) highlights interactions arising due to market specialization and provide an indicator based on export destinations to identify a relevant margin of bank-firm heterogeneity (e.g. banks specialized in US markets treat differently firms exporting to US relative to the treatment of not specialized banks to those types of firms). Gopal (2021) highlights interactions arising due to specialization in collateral choices. These papers correct the estimation by including interactions on the relevant observable margin. Another related paper is Ivashina et al. (2022) which highlights that the propagation of credit shocks may be loan-type specific, inducing a bank-firm interaction when there exist banks specialized in providing certain types of loans, which would bias standard methods assuming transmission homogeneity. They correct the estimation by separately estimating the bank lending channel across each observed type of loan. We show that both the lending specialization model in Paravisini et al. (2023) and the loan-type empirical specification in Ivashina et al. (2022) can lead to our empirical specification with bank-firm group effects. Thus, our paper provides a way of identifying and estimating such types of interactions even when the researcher lacks data access or is uncertain about the specific relevant dimensions. Moreover, we extend their discussion of the potential identification concerns that arise when using current state-of-the-art estimation methods assuming homogeneous effects. In particular, we formalize that the homogenous bank effect estimate is a weighted average of the heterogeneous bank effects for each firm group and bank, with weights that may be negative, which highlights the potential interpretation concerns of standard homogeneous estimates. In addition, we highlight the importance of the endogenous

[5] Without bank-firm interactions (as in the standard linear framework), credit growth is independent of the network so the counterfactual network exercise leads to no change.

[6] Some papers have documented differential transmission of bank shocks across firm observables (e.g. differences by firm size as in KM, Chodorow-Reich (2014)). Rather than relying on a particular observed characteristic, we make the margin of heterogeneity an empirical question. Importantly, we emphasize bank-firm interactions that in general point to heterogeneity in the specific bank-firm match that can happen due to two degrees of heterogeneity: The differential effect across firm types can differ by banks, as would be expected with bank specialization.

creation of bank-firm networks, uncovering and quantifying a novel bank-firm matching channel.

Our econometric approach aligns with the literature that estimates interactions between unobserved factors. We closely follow Bonhomme and Manresa (2015) and extend their grouped-fixed effect estimator to our bank lending framework. An important distinction is that our framework, which utilizes network panel data, encompasses three dimensions compared to the standard panel data model, which has two dimensions. Our framework allows for flexibility in the group membership of firms to change over time, which can be particularly relevant in environments with several years in which firms and banks may alter their business models. Additionally, it enables us to incorporate time-varying firm fixed effects alongside the time-varying bank-firm group effect, which is crucial for controlling for time-varying demand shocks.

Outline. Section 2 outlines the standard framework and identification assumption of homogeneous transmission. Section 3 extends this framework to include bank-firm interactions, covering identification and our estimation algorithm. Section 4 presents a banking model with specialization and uses it to discuss interpretation of bank-firm interactions. Sections 5 and 6 present the empirical results and applications, and Section 7 concludes.

2 Standard framework: a model of homogeneous effects

The empirical model consists of N_F firms and N_B banks that interact in different periods $t = 1....T$. Credit registries contain information about the loan amount $L_{f,b,t}$ of each firm with each bank at time t. If firm f has a positive outstanding balance with bank b during period t and $t-1$, we say a network $D_{f,b,t} = 1$ exists; otherwise $D_{f,b,t} = 0$. We denote with $y^*_{f,b,t}$ the potential loan growth rate between bank b and firm f at time t if the network (f, b, t) exists. We observe realizations of $y^*_{f,b,t}$ only when $D_{fbt} = 1$. Let's define the observed loan growth rate as $y_{f,b,t} = \left\{ y^*_{f,b,t} : D_{f,b,t} = 1 \right\}$.

The main methodology used in the literature, which follows Khwaja and Mian (2008), considers the following linear specification with two sided heterogeneity:

$$y^*_{f,b,t} = \alpha_{f,t} + \beta_{b,t} + \epsilon_{f,b,t}, \tag{1}$$

where $\alpha_{f,t}$ captures firm-specific heterogeneity affecting the demand of credit, $\beta_{b,t}$ captures bank-specific heterogeneity affecting credit supply, and $\epsilon_{f,b,t}$ is an idiosyncratic factor that varies across firms, banks and time (capturing any interaction between bank and firm factors). In the seminal paper of KM, the empirical model in (1) is motivated by a banking model where firms facing heterogeneous productivity shocks negotiate loans with banks fac-

ing heterogeneous funding shocks, and lending can be decomposed linearly in a firm-time and bank-time heterogeneity. In studies examining the bank lending channel, researchers often rely on an observable measuring a bank's exposure to a credit event denoted by $x_{b,t}$ (e.g., exposure to a liquidity shock during a natural experiment). In that case, $\beta_{b,t}$ is modeled as $\theta x_{b,t}$, where θ represents the homogeneous treatment effect of a specific bank shock on credit. KM underscores the importance of incorporating firm-fixed effects $\alpha_{f,t}$ in the regression due to potential correlation with $x_{b,t}$ stemming from sorting mechanisms.

The linear specification in (1) is a model of *homogeneous transmission of credit shocks*. That is, credit demand shocks to firms—e.g. lending changes due to firm-level productivity shocks, changes in factor costs, firm credit constraints—are transmitted *equally* to all banks, so captured by $\alpha_{f,t}$, and credit supply shocks to banks are transmitted *equally* to all firms. Any bank-firm interactions $\epsilon_{f,b,t}$, for example, generated due to either firms with bank-specific demand or banks with firm-specific supply cannot be systematic, so they must be uncorrelated with α_f or θx_b; moreover, they should be orthogonal to the network structure captured by $D_{f,b,t}$. The following states the identification assumption.

Assumption 1. ***Exogenous network.*** *Let D_t, α_t, β_t, x_t denote the entire vector (at time t) of $D_{f,b,t}$, $\alpha_{f,t}$, $\beta_{b,t}$ and $x_{b,t}$ for all f, b. Then, $E[\epsilon_{f,b,t}|D_t, \alpha_t, \beta_t] = 0$ (or $E[\epsilon_{f,b,t}|D_t, \alpha_t, x_t] = 0$)*

Assumption 1 implies the formation of bank-firm networks is exogenous once we condition on α_t and β_t (or x_t). That is, bank-firm interactions $\epsilon_{f,b,t}$ do not affect bank-firm matching, and there are no systematic differences in $\epsilon_{f,b,t}$'s for different partitions of the network D_t.

Assumption 1 allows the network to be endogenous to the factors that we are controlling for in the regression, so it allows for sorting that depends on $\alpha_{f,t}, \beta_{b,t}$ or $x_{b,t}$ (e.g., it could be that firms with high $\alpha_{f,t}$ are more likely matched with banks experiencing a high shock $x_{b,t}$). However, assumption 1 does not allow the network to depend on characteristics that affect loan growth rates that vary at the bank-firm level. For instance, assumption 1 might fail if matching depends on the specific relation of a bank with a type of firms as highlighted by Paravisini et al. (2023) where exporters to a particular market are more likely to start a relationship with a bank that specializes in that market.[7]

Identification. Multiple bank-firm connections for every t allow to absorb firm effects. By taking differences of equation (1) for the same firm in two banks b_0, b we get

$$y_{f,b,t} - y_{f,b_0,t} = \left(\beta_{b,t} - \beta_{b_0,t}\right) + \left(\epsilon_{f,b,t} - \epsilon_{f,b_0,t}\right). \tag{2}$$

[7] KM raise the identification concerns in the presence of bank-firm interactions. KM emphasize "... perhaps a firm's loan demand is bank-specific and is correlated with shocks to the bank's liquidity." They provide some evidence that this is not too relevant for their event study.

Assumption 1 plays a key role to separate $(\beta_{b,t} - \beta_{b_0,t})$ and $(\epsilon_{f,b,t} - \epsilon_{f,b_0,t})$ in (2). It implies that $E_{f \in I(b,b_0)}\left[\varepsilon_{f,b,t}\right] = E_{f \in I(b,b_0)}\left[\varepsilon_{f,b_0,t}\right]$, where $I(b, b_0) = \{f : D_{f,b} = D_{f,b_0} = 1\}$ is the set of firms that borrow from both banks b and b_0. Therefore, by comparing the average differential borrowing from the same set of firms connected with two different banks we can identify the bank-specific shocks using the following moment condition:

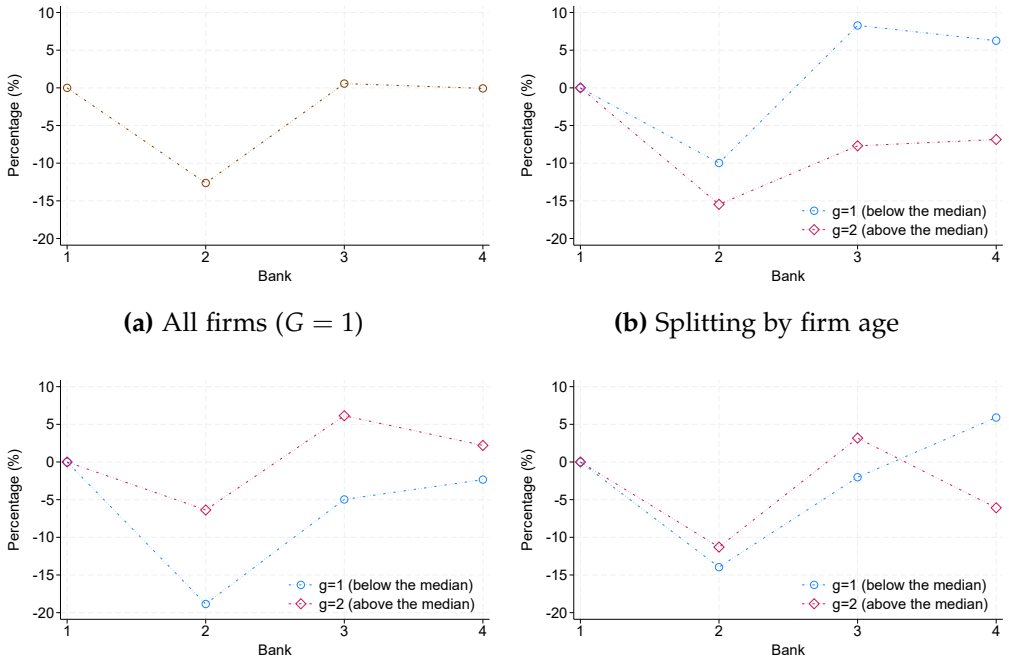
$$E_{f \in I(b,b_0)}\left[y_{f,b,t} - y_{f,b_0,t}\right] = \beta_{b,t} - \beta_{b_0,t} = \theta\left[x_{b,t} - x_{b_0,t}\right]. \tag{3}$$

Such moment condition thus identifies the effect of a specific credit supply shock to bank b relative to bank b_0: $\beta_{b,t} - \beta_{b_0,t}$ (or analogously $\theta\left[x_{b,t} - x_{b_0,t}\right]$). Intuitively, under this framework of homogeneous transmission of credit shocks, any systematic difference in the lending of two banks to the *same* set of firms cannot be driven by firm demand, so it is attributed to the (relative) bank shock. Moreover, this average lending to firms from two banks b, b_0 should be the same across different sets/groups of firms, that is, the left hand side in equation (3) should be the same for any subset of $I(b, b_0)$.

Suggestive evidence of heterogeneity. Figure 1 presents the sample analog of the left hand side of (3) in our data set for different subsets of firms common to the four largest banks in our data set in year 2017. We show the results for the entire set of common firms and when we split firms in two groups by: i) firm age, ii) their posted collateral size, iii) total debt size in previous years.

Panel 1a shows the average loan growth rate (relative to bank 1) across all firms. According to (3), such differential borrowing captures the effects of each specific (homogeneous) bank shock $\beta_{b,t} - \beta_{b=1,t}$. We would conclude that bank 2 experienced a negative bank shock of about -12% (relative to bank 1), while banks 3 and 4 experienced very small banks shocks. However, the next panels show that such differential borrowing across banks changes considerably across the different firm groupings. Panel 1b shows the differential average growth rates for firms above/below the median firm age. We can see that on average firms in both groups borrow less from bank 2 than from bank 1, but the magnitude is considerably different (around -9% vs -16%). More strikingly are the results for bank 3 and 4: firms in the blue (below-median) group experienced an increase in their borrowing from these banks, while firms in the red (above-median) group experienced decrease in their borrowing (relative to bank 1). Panels 1c and 1d also show differential patterns when splitting firms by their collateral or debt size. Interestingly, 1d shows that bank 3 reduces lending to the blue group and increases it to the red group; while the opposite pattern is shown by bank 4.

All panels indicate there exists heterogeneity in the responses by groups, suggesting that the additive effects of bank and firm shocks implied by the linear specification in section 2



(a) All firms ($G = 1$)

(b) Splitting by firm age

(c) Splitting by collateral size

(d) Splitting by past firm debt size

Figure 1: Mean differential loan growth rate for $t = 2017$ (relative to bank 1), by firm classification. Panel 1a shows average credit growth for all firms in the sample. The next panels split the sample in two groups by: Firm age (1b), the firm posted collateral (panel 1c), and total firm debt size in previous years (panel 1d).

is not satisfied in the data. Moreover, they suggest that heterogeneity may depend on the specific match between banks and firm groups.

3 A model with interactions and heterogeneous effects

A general framework that accounts for interactions can be written as:

$$y^*_{f,b,t} = \alpha_{f,t} + \beta_{b,t} + \psi_{f,b,t} + \epsilon_{f,b,t}, \tag{4}$$

where $\psi_{f,b,t}$ accounts for any interactive factors that can affect credit growth and, as opposed to the idiosyncratic $\epsilon_{f,b,t}$, can be systematically correlated with $\alpha_{f,t}, \beta_{b,t}$ or the probability of creating a relationship $D_{f,b,t}$. $\psi_{f,b,t}$ may capture credit supply (or demand) shocks that are heterogeneously propagated across firms (banks). We discuss interpretation of our identified interactions through the lens of a banking model with lending specialization in section 4. Interactions affecting credit growth $y^*_{f,b,t}$ that are not systematically correlated either with α, β or the network D are instead accounted by $\epsilon_{f,b,t}$.

Without imposing some structure we cannot identify specification (4). There are differ-

ent methods that help uncovering interactions across unobserved factors. In our preferred approach, we model interactions occurring across a discrete number of firm types or groups, in which heterogeneity is left unrestricted across groups. Grouping firms allows us to consider some commonalities in firms credit behavior to average out noise and uncover true heterogeneity.[8] Importantly, group heterogeneity arises naturally in banking models with specialization in markets or activities, as we describe in section 4. In a second specification, described in Online Appendix B, we use an interactive fixed effects model that allows for heterogeneity to be continuous but restricted to a factor structure.

Group interactions. Let G be the number of groups. For each f, we have $g_t(f) \in \{1, ..., G\}$ which maps firms into groups. Let $\mathcal{G} = \{g_t(1), ..., g_t(N_F)\}$ be any grouping of firms into the G groups in year t. We consider the following specification with bank-firm group specific effects as an extension of KM:

$$y^*_{f,b,t} = \alpha_{f,t} + \beta_{b,g_t(f),t} + \epsilon_{f,b,t}, \tag{5}$$

where $\alpha_{f,t}$ is an unobserved time-varying firm effect, $\beta_{b,g(f),t}$ is an unobserved time-varying *bank-firm group effect*, and $\epsilon_{f,b,t}$ is the unit specific error term. The firm effect $\alpha_{f,t}$ and bank-firm group effect $\beta_{b,g_t(f),t}$ are allowed to be arbitrarily correlated: The specification in (5) allows for *unrestricted interactions* between banks' and firm-types' characteristics. We replace assumption 1 for the following one:

Assumption 2. ***Exogenous network within groups***. *Let β^G_t denote the entire vectors of $\beta_{b,g(f),t}$ for all f, b for given t. Then, $E[\epsilon_{f,b,t}|D_t, \alpha_t, \beta^G_t] = 0$ in (5).*

Assumption 2 states that the formation of networks is exogenous once we conditioned on the firm-specific and the bank-*firm group* unobserved heterogeneity. That is, within-group bank-firm interactions $\epsilon_{f,b,t}$ do not affect bank-firm matching. Importantly, as opposed to assumption 1, assumption 2 allows for endogenous matching that depends on interactions occurring across groups. For example, it allows firms exporting to particular market (belonging to a specific group g) to have a higher/lower probability to form links with different banks (specialized in different markets), as Paravisini et al. (2023) shows.

Specification (5) can capture heterogeneity in the transmission of credit shocks that arises due to "market-specific" specialization as highlighted in Paravisini et al. (2023) and Blickle et al. (2023) or "loan-type" specialization as emphasized by Ivashina et al. (2022). Paravisini et al. (2023), using data from Peru, show that there is lending specialization based on the firms' export destination market and that the transmission of credit shocks depend on such

[8] Bonhomme et al. (2019) use a similar dimension reduction technique to model firm heterogeneity in matched employer-employee data.

margin. Blickle et al. (2023), using data from US, show that bank lending specialization occurs at the industry level. Ivashina et al. (2022), using data from Peru, show that bank shocks affect differentially firms depending on their type of loan contract used (e.g. asset based loans vs cash-flow based loans). Thus, firms would be grouped according to their export destination market, their industry, or the type of loans they use, respectively. Importantly, in our framework, the grouping is left unrestricted and treated as an additional unobserved function. This is useful in contexts in which the econometrician cannot access the specific data about bank or firm specialization (lack of data about export destination markets, industries, or loan-type used), or, when the heterogeneity is based on an unobserved margin or multiple margins. As shown below, as long as interactions happen at some discrete group level, we will be able to consistently identify the effects in model (5).

3.1 Identification and estimation

Known groups. If groups were known, a comparison between the loan growth rate of firms in the same group g borrowing from two banks b and b_0, allows us to identify the differential $\beta_{b,g,t}$ from the following moment condition (analogous to (3)):

$$E_{f\in g_t, f\in I(b,b_0)}\left[y_{f,b,t} - y_{f,b_0,t}\right] = \beta_{b,g_t,t} - \beta_{b_0,g_t,t},\ \forall g_t \in \{1,...,G\}. \tag{6}$$

Under a normalization (e.g. $\beta_{b_0,g_t,t} = 0$), we can recover each of the $\beta_{b,g,t}$ from the moment condition in (6). Finally, we can calculate for each firm its loan growth rate relative to the $\beta_{b,g,t}$ and average over all its connected banks, denoted $I(f)$, which from (5) allows to identify the firm-specific factor $\alpha_{f,t}$ as:

$$E_{b\in I(f)}\left[y_{f,b,t} - \beta_{b,g,t}\right] = \alpha_{f,t}. \tag{7}$$

If groups were known, then a standard estimation of time-varying bank and firm fixed effects for each group would provide consistent estimates of these objects.

Unknown groups. When the classification of firms into groups is unknown to the econometrician, we rely on clustering techniques to group firms based on their heterogeneous responses to bank lending. Typically, clustering methods group data based on some dissimilarity measure derived from observable characteristics. In contrast, our approach groups firms according to their dissimilarities in unobserved responses to unobserved shocks, as the heterogeneous credit shocks are not directly observable. To implement our framework, we combine the method in Bonhomme and Manresa (2015), which allows for time-varying group effects in panel data, with the two-sided heterogeneity framework that utilizes bank-firm network data from KM.

Lets define the parameter spaces A and B which are subsets of $\mathbb{R}$, and $\alpha_{f,t} \in \mathrm{A}$ and $\beta_{b,g_t(f),t} \in B$. Also, $\mathcal{G} \in \Gamma_G$ where Γ_G is the set of all groupings of $\{1, ..., N_F\}$ into at most G groups. Following Bonhomme and Manresa (2015), for a pre-defined number of groups G we define our estimator, which we refer to as the "bank-firm grouped fixed-effects" (BF-GFE), as the solution of the following problem.[9] For each $t = 1, ..., T$,

$$\left(\widehat{\alpha}_t, \widehat{\beta}_t, \widehat{\gamma}_t\right) = \arg \min_{(\alpha,\beta,\gamma) \in A^{N_F} \times B^{N_B G} \times \Gamma_G} \sum_{f=1}^{N_F} \sum_{b=1}^{N_B} \left(y_{f,b,t} - \alpha_{f,t} - \beta_{b,g_t(f),t}\right)^2, \tag{8}$$

where the minimum is taken over all possible groupings $\gamma = \{g_t(1), ..., g_t(N_F)\}$ of the N_F firms into G groups, firm-time effects, and bank-group-time effects.

For given $\alpha_{f,t}$ and $\beta_{b,g_t(f),t}$, the optimal group assignment for each firm is:

$$\widehat{g}\left(f|\alpha_t, \beta_t\right) = \arg \min_{g} \sum_{b=1}^{N_B} \left(y_{i,t} - \alpha_{f,t} - \beta_{b,g_t(f),t}\right)^2, \tag{9}$$

where we take the minimum g in case of a non-unique solution. The estimator of $\alpha_{f,t}$ and $\beta_{b,g(f),t}$ in (8) can be written as:

$$\left(\widehat{\alpha}_t, \widehat{\beta}_t\right) = \arg \min_{(\alpha,\beta) \in R^{N_F} \times R^{N_B G}} \sum_{f=1}^{N_F} \sum_{b=1}^{N_B} \left(y_{f,b,t} - \alpha_{f,t} - \beta_{b,\widehat{g}_t(f|\alpha_t,\beta_t),t}\right)^2, \tag{10}$$

where $\widehat{g}_t\left(f|\alpha_t, \beta_t\right)$ is given by (9), and the BF-GFE estimate of $g_t(f)$ is $\widehat{g}_t\left(f|\widehat{\alpha}_t, \widehat{\beta}_t\right)$.

A couple of key differences with the model in Bonhomme and Manresa (2015): First, in their model, the individual effects must be fixed in time, our framework allows for both individual and the bank-firm group effects to vary across time by using the multi-year bipartite network. The time-varying individual firm effects are crucial in this framework to account for time-varying demand shocks. Second, our framework allows for the grouping γ_t to change every period, while it must be fixed in time in theirs. We are able to estimate time-varying firm effects and time-varying grouping because we have a panel model with three dimension and we can exploit the multiple connections as in KM.[10]

[9] For the estimator, we fix G and treat it as known. In our data analysis, we present results for different values of G.

[10] We also estimate the model fixing the grouping in certain time windows, in which case the estimator in (8) is replaced by the following. Let $\mathbb{T} \equiv \{t_1, ..., t_m\}$ be the time window, then for each $t \in \mathbb{T}$:

$$\left(\left\{\widehat{\alpha}_t, \widehat{\beta}_t\right\}_{t \in \mathbb{T}}, \widehat{\gamma}\right) = \arg \min_{(\alpha,\beta,\gamma) \in A^{mN_F} \times B^{mN_B G} \times \Gamma_G} \sum_{t \in \mathbb{T}} \sum_{f=1}^{N_F} \sum_{b=1}^{N_B} \left(y_{f,b,t} - \alpha_{f,t} - \beta_{b,g_t(f),t}\right)^2.$$

The estimation approach in (8) jointly estimates the unobserved grouping structure of the data $g(f)$ and the unobserved time-varying effects $\alpha_{f,t}$ and $\beta_{b,g_t(f),t}$. The problem in (8) can be split into two *unfeasible* estimation problems. The problem in (9) is a classification problem as in the unsupervised learning literature. However, the classification here depends on two unobserved components $\alpha_{f,t}$ and $\beta_{b,g(f),t}$ as opposed to the standard classification problem in unsupervised learning where the classification is based on a dissimilarity measure that is observable. The problem in (10) is a two-way fixed-effects regression for the group classification given by (9). In fact, the estimator in (10) is the sample analogue of the moment conditions in equations (6) and (7) for $g_t(f) = \widehat{g}_t\left(f|\alpha_t, \beta_t\right)$ when $N_B = 2$.

The following proposition indicates that the BF-GFE estimator is consistent and asymptotically equivalent to the estimator under known groups. It also implies that N_B can increase polynomially more slowly than N_F which is a crucial condition for empirical applications since standard bank-firm data contains many more firms than banks. Online Appendix C discusses the problem of group misclassification. It shows that firms' group classification is learned from their connections with different banks (see also the example in figure 2). Thus, there exists a probability of misclassification in panels with small N_B, which leads to a small sample bias in the BF-GFE estimator. Importantly, Proposition 1 indicates the rate of misclassification decreases very rapidly as N_B increases.[11] Through simulations, Online Appendix D illustrates that for moderate values of N_B, including calibrations to our dataset, the estimator performs well and is centered around the true value. This finding holds significant relevance given that typical databases often contain a limited number of banks.

Proposition 1. *(Consistency) Assume a fully connected network with number of groups G and a grouping function $g_t(f)$ for which assumption 2 holds. Then, under weak cross-sectional dependence in the errors of the model in (5) and well-separated groups, the estimator $\left(\widehat{\alpha}_t, \widehat{\beta}_t\right)$ in (8) provide consistent estimates of $(\alpha_{f,t}, \beta_{b,g_t(f),t})$ as N_F and N_B tend to infinity, and for all $\delta > 0$:*

$$\widehat{\beta}_{b,\widehat{g}_t(f),t} = \widehat{\beta^u}_{b,g_t(f),t} + o_p\left(N_B^{-\delta}\right) \qquad \text{for all } b, f, t.$$

where the unfeasible estimator $\widehat{\beta^u}_{b,g_t(f),t}$ is the solution of (10) when the grouping is fixed to its population counterpart rather than being estimated.

Computation. Following the discussion in subsection 3.1, our proposed algorithm decomposes the estimator in (8) into two distinct problems. We suggest an iterative approach

[11] This phenomenon resembles the well-known incidental parameter problem in panel data for small T setups (Neyman and Scott (1948)). Bonhomme and Manresa (2015) shows that the group fixed effect estimator in standard panel data suffers from the incidental parameter problem for fixed T but the problem decreases very fast with T, as the group membership in their framework learns from the time series.

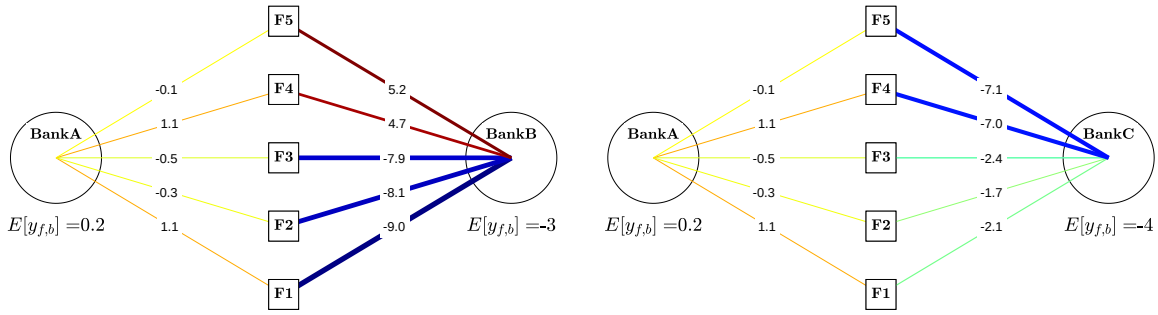
that alternates between addressing the classification problem in (9) and solving the two-way fixed effect estimation problem in (10). The algorithm is detailed in Appendix D.

Performance. Online Appendix D evaluates the performance of our proposed algorithm in different simulations, with exogenous and endogenous networks calibrated to our dataset. As our asymptotic analysis predicts, the simulations show that increasing the number of banks reduces exponentially the probability of misclassifying firms into groups and decreases the small-sample bias in the BF-GFE estimator. For moderate N_B and large N_F, the BF-GFE estimator provides consistent estimates of the interaction effect. We also show that our estimator perform well even when group classification depends on a combination of firm characteristics such as those presented in Section 2.

Number of groups. For our applications, we provide results for different number of groups. Online Appendix G, building on Almagro and Manresa (2021), proposes an N-fold cross-validation procedure for our algorithm that helps determining the optimal number of groups based on an out-of-sample forecasting performance.

Inside the black box. In order to provide insight into our algorithm's group identification, Figure 2 presents a simulated example of a model according to specification (5) with five firms ($N_F = 5$). Each circle in the middle denotes each firm, the line between each bank and the firm denotes the credit growth $y_{f,b}$, with the color and width representing the value of $y_{f,b}$. In the example, panel 2a shows the average credit growth from bank B to the five firms is around -3%, while for bank A is close to 0. The standard framework suggests that such difference in average credit growth from the two banks represents the effect of a bank shock $\beta_B - \beta_A$. However, we can visually detect that the difference in credit growth for the top two firms is around 5%, while for the three bottom firms is around -8%, suggesting they belong to different groups. Still, as seen in (5), such differential credit could be explained either by a different $\beta_{b,g(f)}$ or by the $\epsilon_{f,b}$. How can we disentangle between these two factors? This is achieved by the algorithm by looking at how other banks treat these firms. For instance, Panel 2b shows the credit growth from an additional bank C. If the differences in credit were generated by the idiosyncratic factor $\epsilon_{f,b}$, then the group of firms treated differently by bank B should not predict the group of firms treated differently by bank C. However, Panel 2b shows that bank C also treats differently the top two firms relative to the bottom three, suggesting that the two top firms belong to one group, while the three bottom firms belong to another group. In this particular example, bank C lowers credit more to the top firms while bank B expands their credit, which could indicate bank B has an special relationship with them.

The example illustrates that the algorithm learns the group classification by considering



(a) Bank A vs Bank B

(b) Bank A vs Bank C

Figure 2: Group Identification Illustration: Example with $N_F = 5$

how banks treat systematically different some firms relative to others. This is further discussed in Online Appendix C where we show that the probability of misclassifying tends to zero at an exponential rate as N_B increases.

3.2 Observed Credit Supply Shifter and Average Bank Effect (ABE)

Our framework captures any variation in credit due to bank-firm interactions at the bank-firm group level. Any type of bank specialization (e.g. by industry, export destination, type of loan contract) that leads to a differential transmission of credit shocks across firm groups $\beta_{b,g_t(f),t}$ can be identified by our method. However, these estimated parameters require careful interpretation, as we discuss in the next section through the lens of a banking model with specialization: $\beta_{b,g_t(f),t}$ may reflect a credit supply shock to bank b that is transmitted differently across each group of firms $g_t(f)$, or a *common* credit demand shock to firms in group $g_t(f)$ that leads to a differential borrowing across banks (and therefore not captured by $\alpha_{f,t}$); but this issue of interpretation is mitigated when researchers rely on credit supply shifters, as it is common in the literature.

In case we observe a variable $x_{b,t}$ capturing bank-specific exposure to a credit supply event, we consider the following extension to KM model that allows for interactions:

$$y^*_{f,b,t} = \alpha_{f,t} + \theta_{b,g(f),t} x_{b,t} + \epsilon_{f,b,t}, \tag{11}$$

where $\theta_{b,g_t(f),t}$ captures an heterogeneous response of firms to the specific bank shock. In this model, we leave the relationship between $\alpha_{f,t}$, $x_{b,t}$ and $\theta_{b,g(f),t}$ totally unrestricted allowing for correlation between the bank "treatment" shock ($x_{b,t}$) and the bank "treatment" effect ($\theta_{b,g_t(f),t}$), as in the "potential outcome framework" with heterogeneous treatment

effects used in the micro-econometric literature. Allowing for unrestricted dependence between $\theta_{b,g_t(f),t}$ and $x_{b,t}$ is important under endogenous bank-firm matching.

$\theta_{b,g_t(f),t}$ can be interpreted as the heterogeneous elasticity to the credit supply shock $x_{b,t}$ as long as any demand factor correlated with the bank-specific shock $x_{b,t}$ is accounted by the firm-specific factor $\alpha_{f,t}$. Thus, any heterogeneous demand factor captured by $\epsilon_{f,b,t}$ must be uncorrelated to bank exposure x_b and its heterogeneous effects, as stated by:[12]

Assumption 3. ***Exogenous network within groups conditional on*** x_b. *Let* θ_t^G *denote the entire vector* $\theta_{b,g(f),t}$ *for all* f, b *for given* t. *Then,* $E[\epsilon_{f,b,t}|D_t, \alpha_t, x_t, \theta_t^G] = 0$ *in (11).*

How can we identify the heterogenous bank (treatment) effects $\theta_{b,g_t(f),t}$? Can we use our estimated interactions $\beta_{b,g_t(f),t}$ to consistently estimate $\theta_{b,g_t(f),t}$? Would the homogeneous bank estimate $\hat{\theta}^{\text{Homo}}$ in (1) identify any interesting *average bank effect*?

Let's start this discussion emphasizing that there are two degrees of heterogeneity in the bank treatment effect of $\theta_{b,g_t(f),t}$. First, the bank effect may differ across firm groups. Second, the effect for each group may be different by banks. This double heterogeneity is expected under bank specialization, since a bank specialized in firms in a particular group g is likely to treat those firms differently compared to a non-specialized bank.

For notational convenience, we eliminate the subindex t. Let $x_{b_0} = 0$ denote the case for a bank not exposed to the shock (e.g. a bank with low foreign liabilities in the credit event studied in section 6.1). We define the average bank-specific response of a bank b with bank shock x_b as:

$$\Delta_{x_b} = E\left[\sum_{g=1}^{G} \frac{\sum_{f \in I_{b,b_0,g}} \left[y_{f,b} - y_{f,b_0}\right]}{N_{b,b_0}\left[x_b - x_{b_0}\right]} \mid x_b, D\right], \tag{12}$$

where $I_{b,b_0,g} = \{f : g(f) = g\} \cap I_{b,b_0}$ is the set of firms in group g that belong to I_{b,b_0}, $N_{b,b_0,g} = \sum_{f \in I_{b,b_0,g}} 1$ is the number of firms in this set, and $N_{b,b_0} = \sum_{g=1}^{G} N_{b,b_0,g}$ is the total number of firms in I_{b,b_0}. Then, given assumption (3):

$$\Delta_{x_b} = E\left[\sum_{g=1}^{G} \frac{N_{b,b_0,g}}{N_{b,b_0}} \theta_{b,g(f)} \mid x_b, D\right]. \tag{13}$$

Δ_{x_b} represents the average response in credit growth for all groups of firms connected to a bank with exposure x_b, relative to the non-exposed bank, in the specific b, b_0 joint network, that is, weighted by the number of firms in each group connected to both banks. (13) shows this bank-specific effect is a function of the heterogeneous parameter $\theta_{b,g(f)}$.

[12] Similar to assumption 2, assumption 3 allows for endogenous matching based on the specific bank-group response $\theta_{b,g_t(f),t}$.

Researchers are often interested in the overall response to the credit event. We define the Average Bank Effect (θ_{ABE}) across banks (relative to the non-exposed) as:

$$\theta_{\text{ABE}} \equiv \sum_{b \neq b_0}^{N_B} \frac{N_{b,b_0}}{N} \Delta_{x_b} = \sum_{b \neq b_0}^{N_B} \sum_{g=1}^{G} \frac{N_{b,b_0,g}}{N} E\left[\theta_{b,g(f)} \mid x_b, D\right],$$

where $N = \sum_{b \neq b_0}^{N_B} N_{b,b_0} = \sum_{b \neq b_0}^{N_B} \sum_{g=1}^{G} N_{b,b_0,g}$. The second equality follows from (13) and shows θ_{ABE} is a weighted average across banks and groups of the expected treatment effects $\theta_{b,g(f)}$, where weights capture the importance of each group across the joint networks between exposed and unexposed banks, reflected by $N_{b,b_0,g}$.

Having defined the objects of interest, we discuss how to estimate them consistently. We propose using our BF-GFE estimates $\hat{\beta}_{b,g}$ to construct a consistent estimators of θ_{ABE}. We first project $\hat{\beta}_{b,g}$ onto the credit supply shifter to estimate each $\theta_{b,g}$:

$$\hat{\theta}_{b,g} \equiv \frac{\hat{\beta}_{b,g} - \hat{\beta}_{b_0,g}}{x_b}, \ \forall g \in \{1, ..., G\}. \tag{14}$$

Then, we estimate the average bank effect as:

$$\widehat{\theta}_{\text{ABE}} = \sum_{b \neq b_0}^{N_B} \sum_{g=1}^{G} \frac{N_{b,b_0,g}}{N} \hat{\theta}_{b,g}. \tag{15}$$

Using (6), proposition 1 and that $x_{b_0} = 0$, we have that $\hat{\beta}_{b,g} - \hat{\beta}_{b_0,g}$ is a consistent estimator of $E\left[y_{f,b} - y_{f,b_0}\right] = \theta_{b,g(f),t} x_{b,t}$, implying that $\widehat{\theta}_{\text{ABE}}$ is consistent.

A critical question arises regarding the consistency of the estimator of the homogeneous effect in specification (1), where a common parameter is estimated using a firm fixed-effect regression. We refer to such estimate as $\hat{\theta}^{\text{Homo}}$. Proposition 2 demonstrates that the expectation of the fixed effect estimator in (1) is also a weighted average of the heterogeneous effects $\theta_{b,g(f)}$, but with different weights than θ_{ABE}.

Proposition 2. *Assume a bank-firm bipartite network data generated from (11). If assumption 2 holds, then:*

$$E\left[\hat{\theta}^{Homo} \mid x, D\right] = \sum_{b \neq b_0}^{N_B} \sum_{g=1}^{G} \omega_{g(f),b} \frac{N_{b,b_0,g}}{N} E[\theta_{b,g(f)} | x_b, D], \tag{16}$$

where $\omega_{g(f),b} = \frac{\sum_{f \in I_{b,b_0,g}}^{N_{b,b_0,g}} v_{f,b} x_b}{\frac{N_{b,b_0,g}}{N} \sum_{b=1}^{N_B} x_b \sum_{g=1}^{G} \sum_{f \in I_{b,b_0,g}}^{N_{b,b_0,g}} v_{f,b}}$ *are weights that add up to one and* $v_{f,b}$ *term represents the residual of a firm fixed effect regression of* x_b *on firm-specific fixed effects, utilizing the network*

data for firms with connections with at least two banks.[13]

As opposed to x_b, $v_{f,b}$ varies at the firm level since the network is unbalanced so each firm f are connected with different number of banks. Proposition 2 implies that in general the homogeneous estimator $\hat{\theta}^{\text{Homo}}$ is a biased estimator of the θ_{ABE}. This result is similar to the bias that appears in a typical panel model with random coefficients when the random coefficient is not independent of the regressors of the model (see Chamberlain (1992) and Arellano and Bonhomme (2012)), and the bias that appears in the two-way fixed effect estimator in a dif-dif model with heterogenous treatment effects (see De Chaisemartin and d'Haultfoeuille (2020), Sun and Abraham (2021)).[14]

Corollary 1. *If and only if* $E[\theta_{b,g(f)} \mid x_b, D] = E[\theta_{b,g(f)}]$, *then* $\hat{\theta}^{Homo}$ *is consistent of* θ_{ABE}.

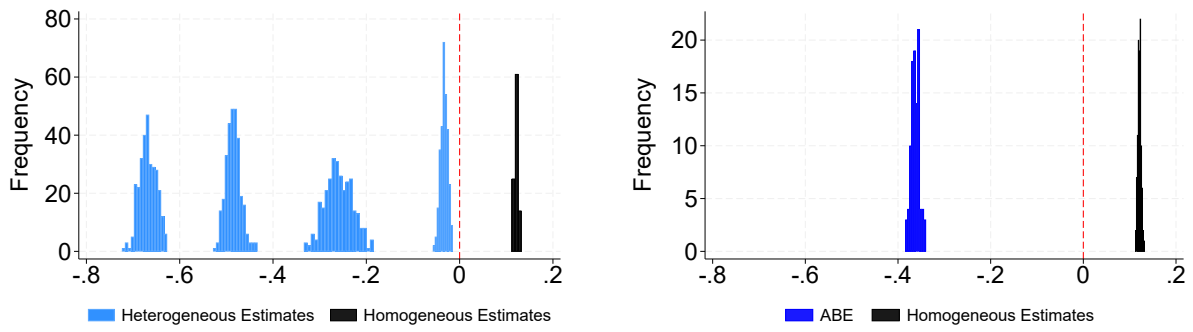
Corollary 1, which its proof is a direct direct consequence of (16), states that only under the absence of endogenous matching in the heterogeneous treatment effect (i.e., $E[\theta_{b,g(f)} \mid x, D] = E[\theta_{b,g(f)}] = \theta$) or in absence of heterogeneous effects (i.e., $\theta_{b,g(f)} = \theta$), $\hat{\theta}^{\text{Homo}}$ is a consistent estimator of θ_{ABE}. This result is analogous to Arkhangelsky et al. (2021) in which a random network in our framework plays a similar role than the random assignment in their dif-in-dif model.

Proposition 2 states that the weights $w_{g,b}$ add up to 1 but they may be negative, in a similar fashion to the result in De Chaisemartin and d'Haultfoeuille (2020). Negative weights make the interpretation of $\hat{\theta}^{\text{Homo}}$ problematic. For instance, one could have that $\hat{\theta}^{\text{Homo}}$ is positive even when all $\theta_{b,g(f)}$ are negative. Endogenous matching on the treatment effect and the issue of negative weights may be expected in a context of bank specialization, as the following example illustrates.

Consider a simulated example with $N_B = 7$, $N_F = 10000$, and $G = 2$. We assume banks $b = 1, 2, 3$ are specialized in lending to firms in $g = 1$, banks $b = 4, 5, 6$ are not specialized, and bank 7 is specialized in $g = 2$. Banks experience a (negative) credit event with the following exposures: $x_{b=1:3} = 1$, $x_{b=4:6} = 0.8$, and $x_{b=7} = 0$. We assume the shock is propagated according to: $\theta_{b=1:3,g=1} = -0.05$, $\theta_{b=4:6,g=1} = -0.7$, so that specialized banks in $g = 1$ protect these firms relative to others. For $g = 2$, we assume $\theta_{b=1:3,g=2} = -0.5$, $\theta_{b=4:6,g=2} = -0.3$. The likelihood of matching with firms increases with specialization: We assume Prob $\left(D_{f,b} = 1 | g(f) = 1, b\right)$ equals 0.8 for $b = 1:3$ due to their specialization in $g = 1$, while it equals 0.5 for $b = 4:6$ and 0.1 for $b = 7$, while Prob $\left(D_{f,b} = 1 | g(f) = 2, b\right)$ equals 0.1 for $b = 1:3$, 0.5 for $b = 4:6$, and 0.8 for $b = 7$.

[13] The KM fixed effect estimator uses the sample of firms for which I_{b,b_0} is not null for all b.

[14] Under endogenous matching, the two-way fixed estimator (used in the absence of a supply shifter) is also a bias estimator of $E\left[\beta_{b,g(f)}\right]$ as we show in Appendix F.



(a) Heterogeneous vs homogeneous estimates

(b) ABE vs homogeneous estimates

Figure 3: Subfigure (a): Distribution of the heterogeneous $\widehat{\theta}_{b,g}$ (light blue) and homogeneous $\widehat{\theta}^{\text{Homo}}$ (black) estimates. Subfigure (b): Histogram of $\widehat{\theta}_{\text{ABE}}$ (blue) and $\widehat{\theta}^{\text{Homo}}$ (black).

We conduct 100 simulations for this example, estimating $\widehat{\theta}^{\text{Homo}}$ as the fixed-effect regression of (1), and using our algorithm to calculate $\widehat{\theta}_{b,g}$ and $\widehat{\theta}_{\text{ABE}}$ according to (14) and (15). Figure 3 shows the results of the simulation. Panel 3a shows the histogram from the estimates of $\widehat{\theta}_{b,g}$ using our algorithm (light blue) and $\widehat{\theta}^{\text{Homo}}$ (black). Panel 3b shows the histogram for $\widehat{\theta}_{\text{ABE}}$ (blue) and $\widehat{\theta}^{\text{Homo}}$ (black). We can see that $\widehat{\theta}_{\text{ABE}}$ is negative as it is a weighted average of the estimated $\widehat{\theta}_{b,g}$'s, which are all negative. Finally, we can see that the whole distribution of $\widehat{\theta}^{\text{Homo}}$ is positive. The intuition is as follows. The homogenous estimate compares the credit growth of banks across the common firms to calculate the average effect. In this case, banks 1:3 protect a certain group of firms, which leads to have the most exposed banks 1:3 to lend more to them relative to the less exposed banks 4:6 which are not specialized. Moreover, since banks 1:3 are specialized, the joint network will be weighted towards firms in their specialized group, aggravating the problem. The homogenous estimate then finds that being more exposed is associated to more credit.

4 Interpretation of heterogenous bank-firm group effects.

In a similar spirit of KM, we motivate our econometric model for the transmission of credit shocks using a simple banking model where firms negotiate loans with banks; however, as opposed to that paper, we consider bank specialization. We model specialization by following the model in Paravisini et al. (2023) in which banks are heterogenous in their lending capabilities for specific activities (e.g. a bank may have an advantage in lending to firms when exporting to a specific country). We show that our empirical specification with bank-firm group effects arises if groups of firms are defined by the relative importance of activities (e.g. a group of firms has a higher presence in US markets while other group may have higher presence in Europe). We describe the ingredients of the model and leave the

derivations for the appendix.[15]

4.1 A model of bank specialization.

Consider a two period model $t = 0, 1$ with N_F firms belonging to G groups, and N_B banks. Each firm f is defined by a collection of activities/markets $c = 1, ..., C$ (e.g. export destination country), and the grouping $g(f)$ determines the relative importance of each activity for the firm. Firms use credit to fund each of those activities, and choose at $t = 0$ the bank b that minimizes the cost of credit for the corresponding activity.

For each activity c, a firm f faces a CES demand structure with elasticity of substitution $\sigma > 1$ and group-activity demand shifter $M^c_{g(f)}$. That is, firm f faces a demand function for activity c: $q^c_f = M^c_{g(f)}(p^c_f)^{-\sigma}$, where $M^c_{g(f)}$ captures market-wide variables exogenous to the firm (e.g. $M^c_{g(f)}$ captures the market size of country c for specific products produced by firms specialized in exports to country c).

Following Paravisini et al. (2023), we focus on the choice of credit and assume credit is the single factor of production. Specifically, assume a production function $q^c_f = A_f L^c_{f,b}$, where A_f is a technology shock and $L^c_{f,b}$ is the amount of credit from a chosen bank b.

Banks lend to firms at an activity-specific rate $r^c_b = r_b/\gamma^c_b$, where $\gamma^c_b > 0$ represents bank specialization in activity c. For example, the bank may have expertise in securitizing loans associated with a particular activity allowing it to fund such loans at a lower cost, or may have superior information and can assess better the risk associated with the activity; alternatively, $\gamma^c_b > 0$ may be interpreted as value-added services attached to the issuance of credit. We assume $r_b = \bar{r}\delta_b$ so that $\bar{r}$ captures the baseline rate common to all banks and δ_b is a bank-specific risk-premium (related to each bank's balance sheet health).

At $t = 0$, firms choose a bank for every activity to minimize the cost of credit but facing some uncertainty about bank abilities and pricing. Specifically, each firm chooses:

$$b = \min_{b'} \left\{ r^c_b e^{-\mu \nu^c_f} \right\} \tag{17}$$

$$= \min_{b'} \left\{ (\ln \bar{r} + \ln \delta_b - \ln \gamma^c_b) - \mu \nu^c_{f,b} \right\}, \tag{18}$$

where $\nu^c_{f,b}$ is an idiosyncratic factor that can be interpreted as information noise experienced by the firm about banks' specialization γ^c_b, for example, due to marketing, location, or misinformation. As in Paravisini et al. (2023), we assume $\left\{ \nu^c_{f,b} \right\}_f$ are i.i.d. Gumbel across

[15] In appendix A we also show that the model with loan-type specific shocks considered by Ivashina et al. (2022) leads to our empirical specification with bank-firm group effects, when loan-type specific demand happens at a firm group level.

firms, so the probability of a firm choosing b for activity c is:

$$\pi_b^c \equiv \frac{\left(\frac{\gamma_b^c}{r_b}\right)^{1/\mu}}{\sum_{b'}\left(\frac{\gamma_{b'}^c}{r_{b'}}\right)^{1/\mu}}. \tag{19}$$

Given a chosen bank b for activity c, each firm chooses the credit amount $L_{f,b}^c$ from that bank to maximize: $p_f^c q_f^c - r_{b,f}^c L_{f,b}^c$. The (potential) optimal amount of credit equals:

$$\ell_{f,b}^c = M_{g(f)}^c \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} \left(A_f\right)^{\sigma-1} \left(r_{b,f}^c\right)^{-\sigma}. \tag{20}$$

Thus, the amount of credit borrowed depends on the importance of the activity for the firm (measured by $M_{g(f)}^c$), the firm technology A_f, and the cost of credit $r_{b,f}^c$.

We can write the expected total borrowing of a firm f from bank b as:

$$E\left[L_{f,b}\right] = E\left[\sum_c \mathbb{I}_{f,b}^c \times \ell_{f,b}^c\right] = \sum_c \pi_b^c \left(M_{g(f)}^c \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} \left(A_f\right)^{\sigma-1} \left(\bar{r}\delta_b\right)^{-\sigma} \left(\gamma_b^c\right)^{\sigma}\right),$$

where the first equality sums credit from bank b across all activities and $\mathbb{I}_{f,b}^c$ equals one if bank b is chosen for activity c and zero otherwise.

At $t = 1$, firms and banks experience different types of shocks: Firm productivity shocks $A_f^{'} = A_f e^{a_f}$, demand shocks $M_{g(f)}^{c'} = M_{g(f)}^c e^{d_{g(f)}^c}$, interest rate shocks $\bar{r}^{'} = \bar{r}e^z$, and bank shocks $\delta_b^{'} = \delta_b e^{s_b}$ which may be propagated heterogeneously across activities $\gamma_b^{c'} = \gamma_b^c e^{s_b^c}$ (e.g. changes in banks' securitization abilities, risk assessment of specific activities, or strategic incentives to tilt their portfolios toward particular sectors), where we use primes to denote $t = 1$ variables.

We assume the firm chosen bank persists for both $t = 0$ and $t = 1$. Then, the loan growth rate $y_{f,b}$ of firm f with bank b can be approximated as:

$$y_{f,b} = \bar{c} + (\sigma-1)a_f - \sigma\left(s_b - \sum_c w_{g(f),b}^c s_b^c\right) + \sum_c w_{g(f),b}^c d_{g(f)}^c + \epsilon_{f,b}, \tag{21}$$

where $\bar{c}$ is a constant capturing common aggregate shocks (z) and a Jensen's inequality term,

$\epsilon_{f,b}$ is an idiosyncratic factor, and

$$w^c_{g(f),b} = \frac{\pi^c_b \left(M^c_{g(f)} \left(\gamma^c_b \right)^\sigma \right)}{\sum_c \pi^c_b \left(M^c_{g(f)} \left(\gamma^c_b \right)^\sigma \right)} \tag{22}$$

is a weight capturing the relative (expected) importance of the activity c in the bank-firm relationship. Since the importance of the activity depends on the firm specific group, we have that the weights are bank-firm group specific.

Therefore, our empirical specification (5) captures model equation (21) with $\alpha_f = \bar{c} + (\sigma - 1)a_f$ and $\beta_{b,g(f)} = -\sigma \left(s_b - \sum_c s^c_b w^c_{g(f),b} \right) + \sum_c d^c_{g(f)} w^c_{g(f),b}$. That is, our identified empirical interaction factor $\beta_{b,g(f)}$ may capture either bank supply shocks s_b and s^c_b heterogeneously propagated to firms (due to differences in the importance of bank-specific lending advantages across firm groups), firm-group demand shocks $d^c_{g(f)}$ heterogeneously propagated to banks, or a combination of both forces.

4.2 Interpretation of *bank-firm group effect*.

As presented above the coefficients $\beta_{b,g_t(f),t}$ will capture any factors that vary at the bank-firm group level. In general, such bank-firm group variation may be interpreted as either credit supply shocks transmitted differently across each group of firms $g_t(f)$, or demand shocks leading to a differential borrowing across banks (and therefore not captured by the firm-fixed factor $\alpha_{f,t}$).

In this subsection, we present different cases to help clarify under what conditions bank or bank-firm fixed effects can be interpreted as indicators of credit supply factors.

Case 1. Absence of bank specialization. In the absence of bank specialization, i.e. $\gamma^c_b = \overline{\gamma}$ and $s^c_b = \bar{s}$, we have that there would not be any bank-firm systematic sorting/matching that correlate with activities or the demand or supply shocks. In such a case, we have that weights $w^c_{g(f),b} = w^c_{g(f)}$ would be independent of banks (see (22)), and (21) becomes:

$$E[y_{f,b}] = \bar{c} + (\sigma - 1)a_f - \sigma s_b + \sum_c w^c_{g(f)} \bar{s} + \sum_c w^c_{g(f)} d^c_{g(f)}, \tag{23}$$

That is, even if there were heterogeneous demand shocks to activities those would not lead to a systematic differential borrowing from banks, so all demand factors would be captured by a firm-fixed effect; and the credit supply effect can be identify using the moment condition in (3).[16]

[16] In fact, the KM model is a special case when $d^c_g = 0$ and s_b is observable. In case s_b is unobserved, the model is identified with a two-way fixed effects.

Case 2. Bank specialization. Bank specialization may create bank-firm sorting by the relative activities captured by the weight $w^c_{g(f),b}$, which could lead to a systematic differential firm borrowing from banks. In general, a firm fixed effect would not be able to control demand factors, and our identifying moment condition (6) would lead to:

$$E_{f\in g, f\in I(b,b_0)}[y_{f,b} - y_{f,b_0}] = -\sigma(s_b - s_{b_0}) + E\left[\sum_c (w^c_{g(f),b} s^c_b - w^c_{g(f),b_0} s^c_{b_0})\right] + E\left[\sum_c (w^c_{g(f),b} - w^c_{g(f),b_0}) d^c_{g(f)}\right]. \quad (24)$$

Identifying credit supply shocks in this case requires a credit supply shifter x_b that do not correlate with heterogenous demand factors as assumption (3) in section 3.2 states. For instance, assume that we have an observable factor x_b capturing bank exposure to a liquidity shock affecting bank funding costs ($E[z_b|x_b] = \phi x_b$ and $E[s^c_b|x_b] = \phi_{c,b} x_b$) while not capturing heterogenous demand $E[w^c_{g(f),b} d^c_{g(f)}|x_b] = E[w^c_{g(f),b_0} d^c_{g(f)}|x_{b_0}]$.[17] Note that assumption does not restrict the event to generate demand factors across firms (a_f or $d^c_{g(f)}$) just that the demand shock to activities is not sorted systematically with banks exposure x_b, so that conditional on x_b, demand factors can be controlled (on average) by the firm-fixed effect. In such a case, we have that the moment condition:

$$E_{f\in g, f\in I(b,b_0)}[y_{f,b} - y_{f,b_0}|x_b, x_{b_0}] = -\sigma(z_b - z_{b_0}) + E\left[\sum_c (w^c_{g(f),b} \phi_{c,b} x_b - w^c_{g(f),b_0} \phi_{c,b_0} x_{b_0})\right]$$

identifies the causal effect of the credit supply shock, as assumed in (12).

For example, Paravisini et al. (2023) shows how bank specialization might affect the estimation of the credit supply event used in Paravisini et al. (2015). They show, using an observable measure of specialization in country destinations, a significant interaction effect between bank country-specialization and firms' market-specific exports in the Peruvian data (see table 5 in Paravisini et al. (2023)). This could reflect heterogeneous firm-bank demand that might lead to bias if we only rely on firm-fixed effects, as shown in equation (24). However, in that application, the firm-bank demand is not sorted by the specific credit event x_b, as reflected by their estimated effect of x_b which remains unchanged when controlling for such firm-bank demand. Thus, while their firm-bank demand significantly affects lending, it appears uncorrelated to x_b. In this case, the projection of $\beta_{b,g_t(f),t}$ on x_b will capture the heterogeneous effect of supply shocks as shown by (14). In section 6.1 we use our BF-

[17] An obvious case is when there is no demand shock to activities at the time the credit supply event occurs $E[w^c_{g(f),b} d^c_{g(f)}|x_b] = 0$.

GFE estimator to recover the heterogeneous treatment effects and the ABE associated to the supply event studied in Paravisini et al. (2015).

In absence of credit supply shifters, we show in section 6.2 how our identified groups can alleviate endogeneity concerns in firm real effects regressions. In section 6.3, we stress that detecting bank-firm interactions (driven either by supply or demand factors) help understanding the role of bank-firm networks in the aggregate transmission of shocks.

5 Estimation Results

5.1 Data

Credit Registry Data: The source of information is named *Registro de Credito de Deudores* (RCD) and belongs to the administrative registries of the Peruvian financial regulator, *Superintendencia de Banca, Seguros y AFPs* (SBS). We observe the loan balance of every corporate firm at a given bank from 2005 to 2018. We consider that a firm is corporate if according to SBS is classified as "*Corporativo*" or "*Gran Empresa*". Considering the definitions of such classifications, our sample of firms correspond to those that obtained at least 5 million dollars in annual sales. Corporate firms represent approximately 55% of the total amount of commercial loans. Additionally to the information of loan balance, we can observe the total amount of collateral and a credit rating. Firms are identifed through their tax id. Credit registry data from Peru has been previously used by Schnabl (2012), Paravisini et al. (2015), Ivashina et al. (2022), and Paravisini et al. (2023).

Financial Statements: We have the financial statements from the Peruvian Stock Exchange from 2007-2017, *Superintendencia del Mercado de Valores.* We observe the balance sheet for all the firms that list on the Peruvian Stock Exchange and also a group of firms that does not list, but report voluntarily to such institution. Our main variable of interest from the balance sheet is the value of fixed capital. Firms are identifed through their tax id.

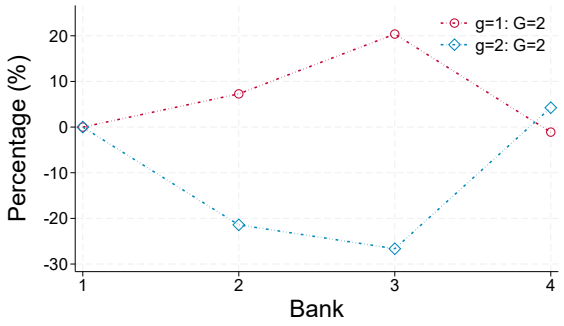
Export Value: We use export information from the Peruvian Tax Bureau, *Superintendencia Nacional de Aduanas y de Administracion Tributaria.* We observe the total value of exports at the firm level for the period 2004-2012. Firms are identifed through their tax id.

5.2 Estimated group heterogeneity

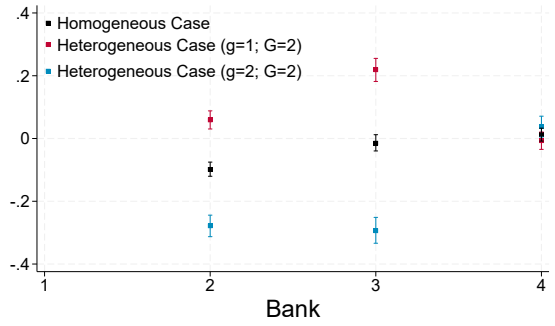
This section present evidence of the significance of the estimated interactions.

Evidence of interactions. For illustration, we initially present estimation results for $G = 2$; however, our applications include results for various values of G.

We start our analysis by examining how credit growth varies across banks for the dif-



(a) Mean credit growth, by estimated group



(b) Estimated bank effects $\widehat{\beta}_{g(f),b,t}$

Figure 4: Panel 4a: Average credit growth $E\left[y_{f,b,t}|b,t\right]$ for the four largest banks splitting the sample based on estimated groups for $G = 2$. Panel 4b: $\widehat{\beta}_{g(f),b,t}$ (normalized bank 1 to 0). Standard errors calculated by non-parametric bootstrap with 1000 repetitions. Data corresponds to 2017.

ferent grouping estimated by our algorithm. Figure 4a displays a similar plot as Figure 1, but splitting firms based on the estimated grouping instead on an arbitrary group choice. The figure displays the average credit growth $E\left[y_{f,b,t}|b,t\right]$ of the four main banks splitting firms across the estimated grouping, for year 2017. We can see that there exists significant heterogeneity in the patterns of credit growth across groups. For example, for bank 3 we can see that the average credit growth for $g = 1$ is around 20%, while for $g = 2$ is close to -25%. For the remaining banks, the differences are also significant. These differences are reflected in the BF-GFE estimator displayed in Figure 4b. The estimated bank effect for the homogeneous case ($G = 1$) is displayed as a confidence interval in black and the BF-GFE estimator for the heterogeneous case ($G = 2$) as blue and red intervals for each of the two groups. Our visual representations are normalized with respect to bank 1. Similar to the findings in Panel 4a, the estimates vary significantly across groups. For example, for bank 3, we find a non-significant bank effect under the homogeneous case, while when we allow for heterogeneity, we find the bank effect is positive for group 1 and negative for group 2. Thus, the homogeneity assumption hides variation in the transmission of credit shocks, overestimating the bank shocks for some firms and underestimating them for others.

Statistical gain of model with interactions. Table 1 compares the R-squared from our model with interactions relative to the standard homogenous model. Our model with interactions considerable improves the goodness of fit relative to the homogenous case.

Out-of-sample performance. We conduct an out-of-sample analysis following the cross-validation procedure described in Online Appendix G which builds in Almagro and Manresa (2021). The model is estimated in training samples and used to do forecasts on testing samples. Online Appendix G shows that the model with interactions outperforms the model

Table 1: Goodness of fit across specifications

Group Specification	R^2	Adjusted-R^2
Model without interactions ($G = 1$)	0.410	0.111
Interaction with our estimated groups ($G = 2$)	0.548	0.318
Interaction with our estimated groups ($G = 3$)	0.642	0.458
Interaction with our estimated groups ($G = 4$)	0.709	0.560

Note. R^2 and adjusted-R^2 of the following specification: $y_{fbt} = \alpha_{ft} + \beta_{g_t(f)bt} + \epsilon_{fbt}$, where $g_t(f)$ denotes the estimated grouping.

with homogenous effects in the testing sample.

Bank specialization. Building on Paravisini et al. (2023) and Blickle et al. (2023), we measure the extent to which banks specialize in each of our firm groupings. We calculate specialization in a group as the share of bank b's portfolio invested in group g compared to the relative share lent to that group by all banks for a given period. Letting $L_{b.g.t} = \sum_{f:g(f)=g} L_{f,b,t}$ be the total lending of bank b to group g, we define:

$$\text{Group Specialization}_{b,g,t} = \frac{L_{b.g.t}}{\sum_{g'} L_{b.g'.t}} \left(\frac{\sum_b L_{b.g.t}}{\sum_b \sum_{g'} L_{b.g'.t}} \right)^{-1}. \tag{25}$$

Table 2 shows the specialization measure significantly predicts the bank-firm interaction effects. This indicates that banks propagate more favorably credit shocks specifically to firms in which they were specialized in the previous period.

Table 2: Bank specialization and bank shock

Dep. Variable	$\sum_{f \in g_t} \widehat{\beta}_{b,g_t(f),t} / N_f$	
	(1)	(2)
Portfolio share in a group$_{t-1}$	0.673*	
	(1.73)	
Group specialization measure$_{t-1}$		0.334*
		(1.73)
Bank-Year FE	Yes	Yes
Group-Year FE	Yes	Yes
R-squared	0.23	0.33
N	142	142

Note. In this table we estimate the relationship between an average of our estimated bank-group interaction estimates and measures of relevance of each estimated group in the bank portfolio. Portfolio share in a group represents the percentage of the corporate portfolio of the bank assigned to the given group: $\sum_{f \in G} L_{G(f)b,t-1} / \sum_f L_{fb,t-1}$. Group specialization measure denotes a relative portfolio share measure in a given group: $\frac{\sum_{f \in G} L_{G(f)b,t-1} / \sum_f L_{fb,t-1}}{\sum_b \sum_{f \in G} L_{G(f)b,t-1} / \sum_b \sum_f L_{fb,t-1}}$. The sample corresponds to a balance panel from 2006 to 2017.

6 Applications

Having identified the relevant bank-firm interactions and the firms' grouping $\mathcal{G}$, we conduct some exercises to uncover the importance of heterogeneity in the bank lending channel. First, we estimate the heterogeneous impact of an observed bank liquidity shock across our firm groups, using as credit supply shifter the event studied in Paravisini et al. (2015) which exploits the differential exposure of banks to the Global Financial Crisis (GFC). Second, we estimate the effect of bank shocks on firm investment. Third, we measure the importance of the bank-firm network for the transmission of credit shocks: We quantify a *bank-firm matching channel* by estimating how the overall credit changes under a counterfactual in which banks and firms are randomly matched.

6.1 Event study: Transmission of bank shocks during GFC

We follow the empirical setting in Paravisini et al. (2015), which builds on KM, and measure bank-specific credit supply shocks x_b with bank-level heterogeneity in the exposure of Peruvian banks to the 2008/09 financial crisis.

Consider the following specification for log-credit as in Paravisini et al. (2015):

$$\ln L_{f,b,t} = \eta_{f,b} + \tilde{\alpha}_{f,t} + \theta x_b \times Post_t + \epsilon_{f,b,t},$$

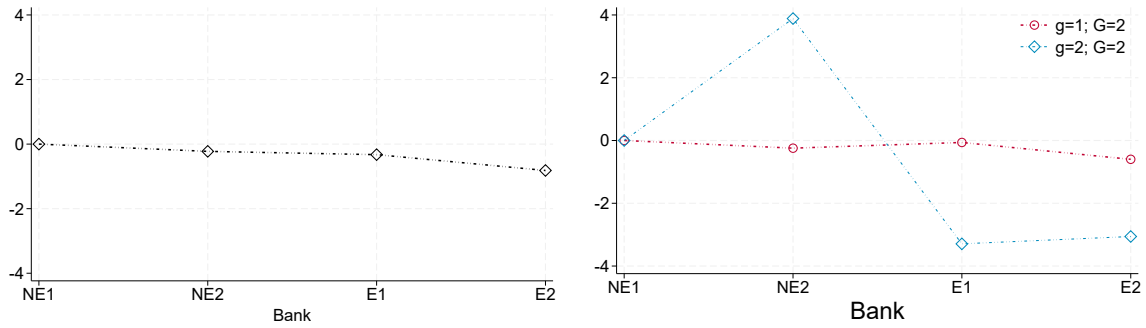
where $L_{f,b,t}$ is the average outstanding debt of firm i with bank b during the intervals $t = Pre, Post$, where Pre and $Post$ periods correspond to the 12 months before and after July 2008. $Post$ is a dichotomous variable equal one when $t = Post$. $\eta_{f,b}$ is a time-invariant firm-bank fixed effect that controls for time-invariant unobserved bank-firm characteristics, $\tilde{\alpha}_{f,t}$ is a time-varying firm fixed effect, x_b is a bank-level observable variable and takes the value of one if a bank b has high foreign liabilities in the pre-period and zero otherwise. The specification leads to the following model for credit growth:

$$y_{f,b} \equiv \ln L_{f,b,Post} - \ln L_{f,b,Pre} = \alpha_f + \theta x_b + \nu_{f,b}.$$

Using our identified groups, we allow for heterogeneous effects, and estimate:

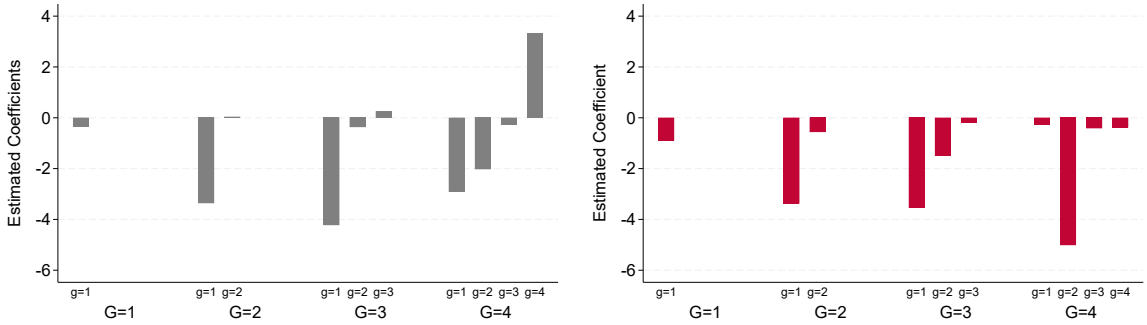
$$y_{f,b} = \alpha_f + \theta_{b,g(f)} x_b + \nu_{f,b}.$$

As in Paravisini et al. (2015), θ is identified by comparing the differential lending of high foreign liabilities banks ("exposed") and low foreign liabilities banks ("non-exposed") to the same set of firms. For instance, Figure 5 illustrates such differences in the average credit growth for two non-exposed banks (banks NE1 and NE2 in the graph) and two exposed



(a) All firms ($G = 1$) **(b)** Splitting firms by estimated grouping ($G = 2$)

Figure 5: Panel 5a shows average (before and after) credit growth, $\ln L_{f,b,Post} - \ln L_{f,b,Pre}$, for exposed and non-exposed banks. Panel 5b splits firms by the estimated group by our algorithm, for $G = 2$.



(a) Estimates $\widehat{\theta}_{g,\tilde{b}}$ for exposed bank $\tilde{b}$ **(b)** Estimates $\widehat{\theta}_{g,\tilde{\tilde{b}}}$ for exposed bank $\tilde{\tilde{b}}$

Figure 6: Estimated $\theta_{g,b}$ for each group $g \in \{1 : G\}$ and two exposed banks: $\tilde{b}$ and $\tilde{\tilde{b}}$.

banks (banks E1 and E2) in the data, for all the joint firms in the first panel and splitting firms across the groups estimated by our algorithm (for case $G = 2$) in the second panel. The figure shows that exposed banks are lending less than non-exposed banks to the same set of firms, but there is a differential lending pattern when splitting the firms across the estimated groups: exposed banks lend much less (than non-exposed) to the blue group ($g = 2$) while the difference is smaller for the red group ($g = 1$). Such differential lending will be reflected in the estimates of $\theta_{b,g}$.

Figure 6 displays the estimated heterogeneous coefficients for two (anonymous) exposed banks (which we name $\tilde{b}$ and $\tilde{\tilde{b}}$), for cases $G = 1$ to $G = 4$. Table 3 shows the number of firms in each group. The results point to considerable heterogeneity in the transmission of this credit shock as it reveals that exposed banks (relative to non-exposed) decreased their credit significantly to some group of firms but not to all. Interestingly, we can see in some

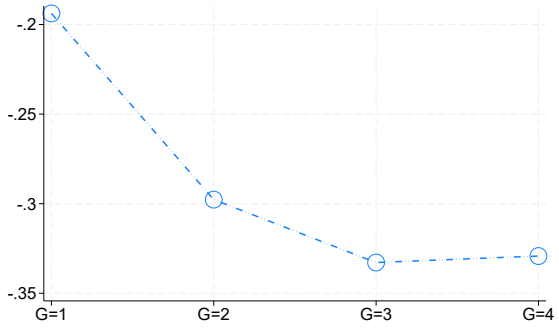


Figure 7: Average effect $\hat{\theta}_{\text{ABE}}$ defined in (15) for different G.

cases that each bank treats differently each specific group. For instance, bank $\tilde{b}$ protects certain groups and even increased their lending during this episode, while this is not the case for bank $\tilde{\tilde{b}}$.

We move into calculating the average effect across all banks and groups of firms, captured by our average bank effect $\widehat{\theta}_{\text{ABE}}$ defined in (15). Figure 7 illustrates the estimates $\widehat{\theta}_{\text{ABE}}$ for $G > 1$, the case $G = 1$ replicates the homogeneous case as obtained in Paravisini et al. (2015). The results show that, when considering heterogeneity, the estimate of the average bank effect increases in magnitude, implying a more severe negative effect of the bank liquidity shock on firms. The estimated average of the heterogenous effect across firm groups and banks is about -0.33, indicating that on average exposed banks reduced credit supply by 33% relative to non-exposed banks, while the estimate under the homogenous model is -0.18.

Group characteristics. Table 3 presents some observable characteristics behind the groupings identified by our algorithm during this event. Groups show differences in firms' posted collateral and exports. This type of sorting is in line with specialization on export destinations markets and lending contracts as emphasized by Paravisini et al. (2023) and Ivashina et al. (2022). The group index in the table corresponds to the one in figure 6. For example, it suggests that bank $\tilde{b}$ (on average) contracted lending more to firms with higher exports, collateral and debt size ($g = 1$ vs $g = 4$).

Table 3: Observable characteristics across groups during event study

$G = 2$	Mean $(g = 1)$	Mean $(g = 2)$	Mean $(g = 3)$	Mean $(g = 4)$
Collateral	12.18	9.82	-	-
Debt Size	16.69	12.02	-	-
Exports	26.81	23.97	-	-
# Firms	156	2,207	-	-
$G = 3$				
Collateral	7.34	11.65	8.24	-
Debt Size	12.64	14.58	9.01	-
Exports	31.71	29.66	13.78	-
# Firms	226	1,258	879	-
$G = 4$				
Collateral	12.89	9.84	9.64	10.46
Debt Size	21.72	11.21	11.88	11.37
Exports	56.88	15.23	24.43	11.21
# Firms	140	260	1,638	325

Note. Descriptive statistics across identified firm groups g. We measure the variables before the shock in the event study. Variables definition in Online Appendix H.

6.2 The impact of credit supply shocks on real outcomes

We examine the real effects of credit supply shocks on firm investment. We first calculate the credit supply shock at the firm level by weighting the BF-GFEs estimates by the share of each bank in the firm borrowing portfolio in the previous period: [18]

$$\text{Supply}_{f,t} = \sum_b \lambda_{f,b,t-1} \hat{\beta}_{b,g(f),t} \text{ with } \lambda_{f,b,t-1} = \frac{L_{f,b,t-1}}{\sum_b L_{f,b,t-1}}, \tag{26}$$

where $L_{f,b,t-1}$ denote borrowing by firm f from bank b at time $t-1$.

We estimate the following regression for investment:

$$\text{Investment}_{f,t} = c + \phi_f^F + \phi_{g(f),t}^G + \beta_1 \hat{\alpha}_{f,t} + \beta_2 \text{Supply}_{f,t} + \epsilon_{f,t}. \tag{27}$$

Note that in this firm-level regression, we cannot control for firm-demand using the bank dimension (we cannot rely on multiple connections to absorb demand shocks). For instance, Paravisini et al. (2015) highlights the importance of product-destination-time fixed effects to absorb demand shocks when estimating the impact of a supply shock on firm-level real outcomes regressions. This concern is even stronger when we do not rely on a credit supply shifter, as $\hat{\beta}_{b,g(f),t}$ may reflect demand shocks affecting the group of firms, as noted

[18] Weighting with the lagged portfolio is common in the literature. The novelty here is that bank effects $\beta_{b,g(f)}$ are heterogeneous across firms.

earlier. To address this, we use our identified groups to include time-varying fixed effects at the group level to control for demand factors common to firms in the same group. Similar to the product-destination fixed effects in Paravisini et al. (2015), our regression compares firms within the same group, exposed to the same group demand factors, but with varying exposures to credit supply shocks based on their network connections from the previous period. Importantly, we can identify β_2 even in the presence of $\phi^G_{g(f),t}$ due to variability in $\lambda_{f,b,t-1}$ across firms within the same group. We also include $\hat{\alpha}_{f,t}$, however as we highlighted in section 3, this is a noisy estimator for fixed N_B.[19]

Remark. Why do firms within the same group have ex-ante different $\lambda_{f,b,t-1}$? Several explanations are plausible. There could exist idiosyncratic factors influencing the ex-ante bank-firm network structure, such as the information frictions $\nu^c_{f,b}$ described in the model in section 4.1 which prevent all firms in the same group to matched only with their optimal specialized bank. Also, firms may evolve their business models. For example, some firms within a group exporting to the U.S. may have previously exported to China and established relationships with banks specializing in China, which they retain. Alternatively, firms may have varying risk assessments, even within the same group, with more risk-averse firms establishing more relationships, including those with banks not specialized in their type. In the end, this is a testable assumption and the fact that we find a significant estimate of β_2 reveals that there is within-group variation in $\text{Supply}_{f,t}$.

Figure 8 plots the estimated confidence interval for the elasticity of the supply shock on investment ($\widehat{\beta}_2$), under the estimation for different number of groups (as represented in the X-axis). Table 5 in Online Appendix H shows all the estimated coefficients. The results reveal that when banks shocks are estimated under the standard framework with homogeneous effects ($G = 1$), the estimated effect is very imprecise, the point estimate is negative and not statistically significant. Instead, when assuming heterogeneous effects ($G > 1$), the estimated impact of banks shocks on firm investment is statistically significant. The estimates for the heterogeneous cases are similar across $G = 2$ to 5, and suggest that a 1% change in the credit supply increases investment by around 4-6%. We cluster standard errors at the leading bank level as in Huremovic et al. (2020).

The results are consistent with the pronounced heterogeneity found in our data set. For instance, in Figure 4a, we observed that, in some cases, the estimated bank effect is positive for firms belonging to one group while it is negative for firms in other group. If credit supply actually matters for investment, one would expect that, investment increases for firms in the former group while decreases for the latter. When $G = 1$ is assumed, the estimated bank

[19] While including $\hat{\alpha}_{f,t}$ may help in controlling for demand effects, $\hat{\alpha}_{f,t}$ may suffer from incidental parameter bias as it is estimated from the number of connection each firm has, which is typically small.

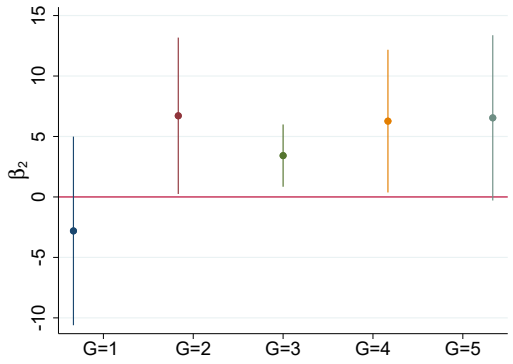


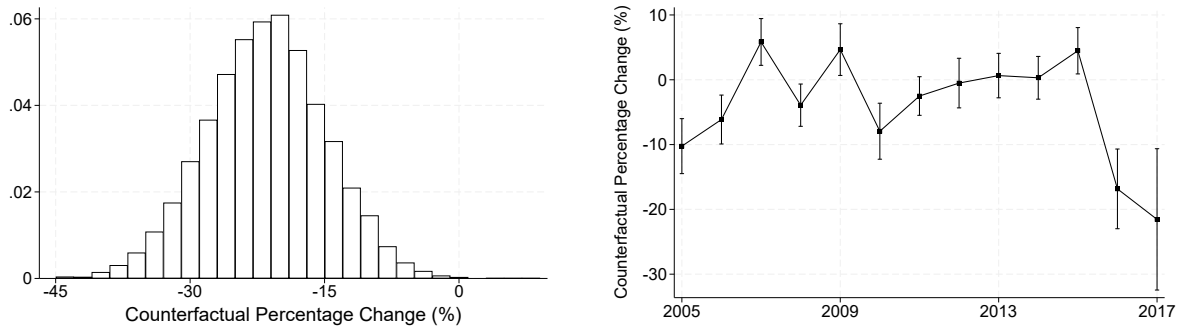
Figure 8: Estimated $\widehat{\beta}_2$ in (27) for $G = 1$ to 5. Clustered standard errors at the leading bank.

effect is homogenous to all of these firms so the correlation with such varying investment levels would be imprecise. Thus, missing the heterogeneity makes hard to uncover the elasticity, even when the average bank effect may be well estimated.

6.3 *Bank-firm matching channel:* The value of bank-firm networks

When the transmission of credit shocks is heterogeneous, the network of relationships may become key for the overall effect of credit shocks: We should expect a stronger (weaker) effect of a shock, when banks create more relationships with firms for which the transmission is stronger (weaker). In this section, we ask do banks and firms create relationships in a way that amplify or smooth out credit shocks? In order to answer this question, we estimate as a counterfactual the aggregate credit growth when banks and firms are randomly matched. For instance, according to the model in section 4.1, this counterfactual represents an economy in which information frictions $\nu^c_{b,f}$ are so large that firms cannot predict their complementarities with banks arising from specialization, so they end up matching with a random bank instead of their optimal specialized bank. We find that for most of the years aggregate credit growth is enhanced by the observed bank-firm credit network relative to the counterfactual random matched network. We refer to such overall effect from bank-firm complementarities as the *bank-firm matching channel.*

We conduct a random reallocation exercise for every firm-bank observation in our data: First, we use our empirical strategy to decompose the average credit growth as the sum of the firm effect $\widehat{\alpha}_{f,t}$ and the heterogeneous bank-firm group effects $\widehat{\beta}_{b,g(f),t}$. Second, keeping the same number of firm connections, we randomize without replacement the banks lending to each firm. For example, if a firm is connected with three banks in our data, we keep the three connections but match the firm randomly with three banks. Third, we calculate the credit growth for each of the counterfactual connections using our model, that is, as the sum of the estimated firm effect and the bank-firm group effect associated with the



(a) $t = 2017$ **(b)** Over different years

Figure 9: Percentage change between the average credit growth rate under random network and the observed network. Panel 9a shows the histogram for 2017. Panel 9b shows 90% confidence intervals for all years.

random network. Fourth, we calculate the average credit growth across all the observations. Finally, we repeat this exercise simulating 10,000 counterfactual networks to calculate the distribution of the average credit growth under random matching.[20]

Figure 9 shows the distribution across simulations of the percentage change in average credit growth relative to the observed network. Panel 9a presents it as an histogram for 2017. The median of the distribution suggests that aggregate credit growth falls by 20% under random matching. This results highlight the importance of the value of bank-firm relationships: The observed network of bank-firm relationships enhances credit growth. Note that, under a model with homogeneous effects ($G = 1$), this exercise would lead to no change in credit growth by construction.

Panel 9b explores the bank-firm matching channel across all years, and present the confidence interval for the simulated change in credit growth for every year. For most of the years the change is negative, meaning that the endogenous matching in the data enhances credit growth relative to a random network. Only for 2007, 2009 and 2015 we find the opposite result. This could be driven by adverse macroeconomic factors, which were in turn amplified by the network. More specifically, 2009 is influenced by the GFC, indicating that more exposed banks were more connected with firms belonging to the group for which the transmission of the (negative) shock was stronger, while 2015 is also associated with adverse external conditions (large drop in commodity prices and exchange rate depreciation).

[20] In this counterfactual we are assuming the absence of general equilibrium effects, so that the estimates of $\widehat{\alpha}_{f,t}, \widehat{\beta}_{b,g,t}$ do not change under a reallocation of connections.

7 Conclusions

We develop an empirical framework to identify and estimate heterogeneous effects of bank shocks exploiting bank-firm credit data. This framework allows us to study the importance of bank-firm interactions and heterogeneity in the bank lending channel, which is particularly relevant in the presence of relationship lending or specialization in markets as stressed by Paravisini et al. (2023) or in lending forms as stressed by Ivashina et al. (2022). To allow for interactions and heterogeneity in a flexible yet parsimonious way, we rely on the idea that from the perspective of the banks and their relationship with firms, there is a discrete number of "types" of firms, and hence, the propagation of shocks depends on the firm's type.

We combine state-of-the-art panel data techniques that allow for time-varying group fixed effects (Bonhomme and Manresa (2015)) with the two-sided fixed effects framework used to disentangle demand and supply shocks from bank-firm credit data (Khwaja and Mian (2008)). We show theoretically and in simulations that our proposed estimator consistently estimates the bank-firm interactions. And, we discuss potential concerns that may arise when estimating a model with homogeneous effects.

We apply our framework to credit registry data from Peru. We uncover significant heterogeneity in the bank lending channel: The patterns of bank lending and the estimated bank-firm group effects show significant differences across our identified firm groups. Importantly, we show that accounting for such heterogeneity and the bank-firm network structure is crucial to learn about the bank lending channel and its real effects.

References

Alfaro, L., M. García-Santana, and E. Moral-Benito (2021). On the direct and indirect real effects of credit supply shocks. *Journal of Financial Economics 139*(3), 895–921.

Almagro, M. and E. Manresa (2021). Data-driven nests in discrete choice models. *Slides*.

Amiti, M. and D. E. Weinstein (2018). How much do idiosyncratic bank shocks affect investment? evidence from matched bank-firm loan data. *Journal of Political Economy 126*(2), 525–587.

Arellano, M. and S. Bonhomme (2012). Identifying distributional characteristics in random coefficients panel data models. *The Review of Economic Studies 79*(3).

Arkhangelsky, D., G. W. Imbens, L. Lei, and X. Luo (2021). Design-robust two-way-fixed-effects regression for panel data. *arXiv preprint arXiv:2107.13737*.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4).

Blickle, K., C. Parlatore, and A. Saunders (2023). Specialization in banking. Technical report, National Bureau of Economic Research.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. *Econometrica 87*(3), 699–739.

Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica 83*(3), 1147–1184.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica: Journal of the Econometric Society*, 567–596.

Chodorow-Reich, G. (2014). The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis. *The Quarterly Journal of Economics 129*(1), 1–59.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

Gopal, M. (2021). How collateral affects small business lending: The role of lender specialization. Working paper.

Huremovic, K., G. Jimenez, E. Moral-Benito, F. Vega-Redondo, and J.-L. Peydro (2020). Production and financial networks in interplay: Crisis evidence from supplier-customer and credit registers. *CEPR Discussion Paper No. DP15277*.

Ivashina, V., L. Laeven, and E. Moral-Benito (2022). Loan types and the bank lending channel. *Journal of Monetary Economics 126*, 171–187.

Jiménez, G., A. Mian, J.-L. Peydró, and J. Saurina (2020). The real effects of the bank lending channel. *Journal of Monetary Economics 115*, 162–179.

Jiménez, G., S. Ongena, J.-L. Peydró, and J. Saurina (2012). Credit supply and monetary policy: Identifying the bank balance-sheet channel with loan applications. *American Economic Review 102*(5), 2301–26.

Jiménez, G., S. Ongena, J.-L. Peydró, and J. Saurina (2017). Macroprudential policy, countercyclical bank capital buffers, and credit supply: Evidence from the spanish dynamic provisioning experiments. *Journal of Political Economy 125*(6), 2126–2177.

Khwaja, A. I. and A. Mian (2008). Tracing the impact of bank liquidity shocks: Evidence from an emerging market. *American Economic Review 98*(4), 1413–42.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 1–32.

Paravisini, D., V. Rappoport, and P. Schnabl (2023). Specialization in bank lending: Evidence from exporting firms. *The Journal of Finance 78*(4), 2049–2085.

Paravisini, D., V. Rappoport, P. Schnabl, and D. Wolfenzon (2015). Dissecting the effect of credit supply on trade: Evidence from matched credit-export data. *The Review of Economic Studies 82*(1), 333–359.

Schnabl, P. (2012). The international transmission of bank liquidity shocks: Evidence from an emerging market. *The Journal of Finance 67*(3), 897–932.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics 225*(2), 175–199.

Appendix (Proof of Propositions)

Proof. **Proof of Proposition 1.**

Consider the following data generating process for the "bank difference" version of model (5) for all (f, b, t) with $D_{f,b,t} = 1$:

$$y_{f,b,t} - y_{f,b_0,t} = \beta^0_{b,g^0_t(f),t} - \beta^0_{b_0,g^0_t(f),t} + \epsilon_{f,b,t} - \epsilon_{f,b_0,t}$$

where 0-superscripts refer to true parameters. We normalize $\beta_{b_0,g^0_t(f),t} = 0$.

Consider the following additional assumptions. Assumption 4 states $\epsilon_{f,b,t}$ has finite moments and limits the dependence structure of the errors $\epsilon_{f,b,t}$ in the cross section. It is fulfilled when $\epsilon_{f,b,t}$ is independent across units, but also allows for weak spatial dependence in $\epsilon_{f,b,t}$. The time series dependence is left unrestricted so shocks can exhibit strong time series correlation. Assumption 5 implies that the population groups have a large number of firms and the groups are well-separated.

Assumption 4. ***(Errors)****: For each t, there exist a constant C_t such that $\epsilon_{f,b,t}$ is a strong mixing process in the cross sectional dimension with:*

i) $E\left(\epsilon_{f,b,t}\right) = 0$, $E\left(\epsilon^4_{f,b,t}\right) < C_t$,

ii) $\frac{1}{N_F}\sum_{f=1}^{N_F}\sum_{f'=1}^{N_F} \mid \frac{1}{N_B}\sum_{b=1}^{N_B} E\left(\epsilon_{f,b,t}\epsilon_{f',b,t}\right) \mid < C_t$,

iii) $\mid \frac{1}{N_F^2 N_B}\sum_{f=1}^{N_F}\sum_{f'=1}^{N_F}\sum_{b=1}^{N_B}\sum_{b'=1}^{N_B} Cov\left(\epsilon_{f,b,t}\epsilon_{f',b,t}, \epsilon_{f,b',t}\epsilon_{f',b',t}\right) \mid < C_t$,

iv) $\mid \frac{1}{N_F N_B}\sum_{f=1}^{N_F}\sum_{b=1}^{N_B}\sum_{b'=1}^{N_B} E\left(\epsilon_{f,b,t}\epsilon_{f,b',t}\right) \mid < C_t$, *and*

v) the time series dependence of $\epsilon_{f,b,t}$ is left unrestricted.

Assumption 5. ***(Group effects)****: For each t:*

i) $\forall g_t \in \{1, \ldots, G\} : plim_{N_F\to\infty}\frac{1}{N_F}\sum_{f=1}^{N_F} 1\left\{g^0_t(f) = g_t\right\} = \pi_g > 0$,

ii) $\forall\left(g_t, \tilde{g}_t\right) \in \{1, \ldots, G\}^2$ *s.t.* $g_t \neq \tilde{g}_t$: $plim_{N_B\to\infty}\frac{1}{N_B}\sum_{b=1}^{N_B}\left(\beta^0_{b,g_t,t} - \beta^0_{b,\tilde{g}_t,t}\right) = c_{g_t,\tilde{g}_t} > 0$,

iii) $\beta^0_{b,g_t,t}$ *is a strongly mixing process in the cross sectional dimension,*

iv) the time series dependence of $\beta^0_{b,g_t,t}$ is left unrestricted.

Assumption 2 implies that $\beta^0_{b,g_t(f),t}$ is contemporaneously uncorrelated with $\epsilon_{f,b,t}$ and $\epsilon_{f,b_0,t}$. Using assumption 2, 4, 5 and full connections between bank and firms, we apply the-

orems 1 and 2 in Bonhomme and Manresa (2015) (for a model without regressors) replacing the time-series dimension in their framework for the bank dimension for a given t in our network data. Then, for each t separately and for all $\delta > 0$ as N_F and N_B tend to infinity:

$$\frac{1}{N_F N_B} \sum_{f=1}^{N_F} \sum_{b=1}^{N_B} (\widehat{\beta}_{b,\hat{g}_t(f),t} - \beta^0_{b,g^0_t(f),t})^2 \quad \xrightarrow{p} \quad 0, \tag{28}$$

$$\text{Prob}\left(\underset{(f \in \{1,\ldots,N_F\})}{sup} \mid \hat{g}_t(f) - g^0_t(f) \mid > 0 \right) \quad = \quad o(1) + o(N_F N_B^{-\delta}), \tag{29}$$

and $\widehat{\beta}_{b,g,t} = \widehat{\beta^u}_{b,g,t} + o_p\left(N_B^{-\delta}\right)$ for all b, g.[21] □

Proof. **Proof of Proposition 2**

From the Frisch-Waugh theorem, the KM fixed effect estimator can be express as:

$$\hat{\theta}^{\text{Homo}} = \left(\sum_{b=1}^{N_B} \sum_{f=1}^{N_F} D_{f,b} v_{f,b} y_{f,b} \right) \left(\sum_{b=1}^{N_B} \sum_{f=1}^{N_F} D_{f,b} v_{f,b} x_{f,b} \right)^{-1},$$

where $x_{f,b}$ equals x_b for all the firms in I_{b,b_0}. For each firm f, $\sum_{b=1}^{N_B} v_{f,b} = 0$ since $v_{f,b} = x_b - \frac{\sum_{b=1}^{N_B} D_{f,b} x_b}{\sum_{b=1}^{N_B} D_{f,b}}$. The expected value conditional on observables is:

$$E\left[\hat{\theta}^{\text{Homo}} \mid x, D\right] = \left(\sum_{b=1}^{N_B} \sum_{f=1}^{N_F} D_{f,b} v_{f,b} E[\alpha_f + \theta_{b,g(f)} x_b \mid x, D] \right) \left(\sum_{b=1}^{N_B} \sum_{f=1}^{N_F} D_{f,b} v_{f,b} x_{f,b} \right)^{-1}$$

where this equality comes from assumption 3. The fixed effect regression implies that $\sum_{f=1}^{N_F} \sum_{b=1}^{N_B} v_{f,b} \alpha_f = 0$, thus

$$E\left[\hat{\theta}^{\text{Homo}} \mid x, D\right] = \frac{\sum_{b=1}^{N_B} x_b \sum_{f=1}^{N_F} D_{f,b} v_{f,b} E[\theta_{b,g(f)} \mid x, D]}{\sum_{b=1}^{N_B} \sum_{f=1}^{N_F} D_{f,b} v_{f,b} x_{f,b}} = \frac{\sum_{b=1}^{N_B} \sum_{g=1}^{G} E[\theta_{b,g(f)} \mid x, D] \sum_{f:g(f)=g} D_{f,b} v_{f,b} x_b}{\sum_{b=1}^{N_B} \sum_{f=1}^{N_F} D_{f,b} v_{f,b} x_{f,b}}.$$

Using $\sum_{f:g(f)=g} D_{f,b} v_{f,b} x_b = \sum_{f \in I_{b,b_0,g}}^{N_{b,b_0,g}} v_{f,b} x_b$ and $\sum_{b=1}^{N_B} \sum_{f=1}^{N_F} D_{f,b} v_{f,b} x_{f,b} = \sum_{b=1}^{N_B} x_b \sum_{f \in I_{b,b_0}}^{N_{b,b_0}} v_{f,b}$:

$$E\left[\hat{\theta}^{\text{Homo}} \mid x, D\right] = \frac{\sum_{b=1}^{N_B} \sum_{g=1}^{G} E[\theta_{b,g(f)} \mid x, D] \sum_{f \in I_{b,b_0,g}}^{N_{b,b_0,g}} v_{f,b} x_b}{\sum_{b=1}^{N_B} x_b \sum_{f \in I_{b,b_0}}^{N_{b,b_0}} v_{f,b}} = \sum_{b=1}^{N_B} \sum_{g=1}^{G} \frac{E[\theta_{b,g(f)} \mid x, D] \sum_{f \in I_{b,b_0,g}}^{N_{b,b_0,g}} v_{f,b} x_b}{\sum_{b=1}^{N_B} x_b \sum_{g=1}^{G} \sum_{f \in I_{b,b_0,g}}^{N_{b,b_0,g}} v_{f,b}}.$$

Define $\omega_{g(f),b} = \frac{\sum_{f \in I_{b,b_0,g}}^{N_{b,b_0,g}} v_{f,b} x_b}{\frac{N_{b,b_0,g}}{N} \sum_{b=1}^{N_B} x_b \sum_{g=1}^{G} \sum_{f \in I_{b,b_0,g}}^{N_{b,b_0,g}} v_{f,b}}$. Using $\omega_{b_0} = 0$, we have: $E\left[\hat{\theta}^{\text{Homo}} \mid x, D\right] = \sum_{b=1}^{N_B} \sum_{g=1}^{G} \omega_{g(f),b} \frac{N_{b,b_0,g}}{N} E[\theta_b$

[21] Online Appendix C intuitively discusses consistency when shocks are iid normally distributed.

$x, D] = \sum_{b \neq b_0}^{N_B} \sum_{g=1}^{G} \omega_{g(f),b} \frac{N_{b,b_0,g}}{N} E[\theta_{b,g(f)} \mid x, D]$. □

ONLINE APPENDIX

(Unobserved) Heterogeneity in the bank lending channel: Accounting for bank-firm interactions and specialization

Bryan Gutierrez Alonso Villacorta Lucciano Villacorta

Section A provides details of the model with bank specialization and shows that our main specification with bank-firm group interactions also emerges in the empirical model with firm loan-type specific demand as in Ivashina et al. (2022). Section B reports estimates of the interaction factors for the interactive fixed effects model described in section 3, which allows for continuous but factor-structure-restricted heterogeneity. Section C discusses the group classification of the estimator under unknown groups. Section D describes the estimation algorithm, presents simulation results including a calibration to our dataset, and compares the estimation of average bank effects of the BF-GFE vs a two-way fixed effects under endogenous matching. Section E outlines a method to evaluate out-of-sample performance of the estimator. Section F provides additional variable definitions and tables.

In this document, all numbering of equations and sections are preceded by a letter, while all references to equations without letters correspond to equations in the main paper.

A Model with lending specialization

First, in A.1, we present derivation details int the banking model in section 4. Second, in A.2, we show that the empirical specification with loan-type specific shocks considered by Ivashina et al. (2022) leads to our empirical specification with bank-firm group effects, when loan-type specific demand happens at a firm group level.

A.1 Model

Consider the model setup described in section 4.

Given a chosen bank b for activity c, each firm chooses the output level q_i^c and credit amount $L_{f,b}^c$ for every activity to maximize:

$$\max_{q_f^c, L_{f,b}^c} p_f^c q_f^c - r_{b,f}^c L_{f,b}^c$$

subject to their demand $q_f^c = M_{g(f)}^c (p_f^c)^{-\sigma}$ and production function $q_f^c = A_f L_{f,b}^c$, which leads to standard optimal choices on price and quantity:

$$\begin{aligned} p_f^c &= \frac{\sigma}{\sigma - 1} \text{mc}_f^c, \\ q_f &= M_{g(f)}^c \left(\frac{\sigma}{\sigma - 1} \text{mc}_f^c \right)^{-\sigma}, \end{aligned}$$

where marginal cost of production $\text{mc}_f^c = r_{b,f}^c / A_f$. Then, the (potential) optimal amount of credit equals:

$$\ell_{f,b}^c = M_{g(f)}^c \left(\frac{\sigma}{\sigma - 1} \right)^{-\sigma} (A_f)^{\sigma - 1} \left(r_{b,f}^c \right)^{-\sigma}. \tag{A.1}$$

The total amount of credit borrowed by the firm thus depend on the importance of the activity for the firm (measured by $M_{g(f)}^c$), the firm technology A_f, and cost of credit $r_{b,f}^c$.

From (A.1) and (19), the expected total borrowing of a firm f from bank b:

$$\begin{aligned} E\left[L_{f,b}\right] = E\left[\sum_c L_{f,b}^c\right] &= E\left[\sum_c \mathbb{I}_{f,b}^c \times \ell_{f,b}^c\right] \\ &= \sum_c \pi_b^c \left(M_{g(f)}^c \left(\frac{\sigma}{\sigma - 1} \right)^{-\sigma} (A_f)^{\sigma - 1} (\bar{r}\delta_b)^{-\sigma} (\gamma_b^c)^{\sigma} \right), \end{aligned}$$

where the first line sums the credit from bank b across all activities and $\mathbb{I}_{f,b}^c$ is an indicator function equal to one if bank b is chosen for activity c and zero otherwise.

To think about credit growth and the influence of credit supply and demand shocks,

we assume that, at $t = 1$, firms and banks may experience different types of shocks: Firm productivity shocks $A_f^{'} = A_f e^{a_f}$, demand shocks $M_{g(f)}^{c'} = M_{g(f)}^{c} e^{d_{g(f)}^{c}}$, interest rate shocks $\bar{r}^{'} = \bar{r}e^{z}$, bank shocks $\pi_b^{'} = \pi_b + s_b$ or bank shocks specific to their lending abilities $\gamma_b^{c'} = \gamma_b^c e^{s_b^c}$ (e.g. changes in their securitization abilities, risk assessment of specific activities/sectors), where we use primes to denote $t = 1$ variables.

We assume the firm chosen bank persists for both $t = 0$ and $t = 1$. Letting $g_x = x'/x - 1$ be the growth rate of any variable x. We have

$$g_{E[L_{f,b}]} \approx \sum_c w_{f,b}^c \left(d_{g(f)}^c + (\sigma - 1)a_f - \sigma(z + s_b - s_b^c) \right),$$

with weights $w_{f,b}^c = \frac{\pi_b^c\left(\ell_{f,b}^c\right)}{\sum_c \pi_b^c\left(\ell_{f,b}^c\right)}$. Note that the weights capture the relative (expected) importance of the activity c in the bank-firm relationship. Since the importance of the activity depends on the firm specific group, we have that the weights are bank-firm group specific; specifically, we have $w_{f,b}^c = w_{g(f),b}^c$ with

$$w_{g(f),b}^c = \frac{\tilde{\gamma}_b^c \left(M_{g(f)}^c \left(\gamma_b^c\right)^\sigma \right)}{\sum_c \tilde{\gamma}_b^c \left(M_{g(f)}^c \left(\gamma_b^c\right)^\sigma \right)}.$$

Therefore, we can write (up to a growth rate approximation):

$$y_{f,b} = \bar{c} + (\sigma - 1)a_f - \sigma \left(s_b - \sum_c w_{g(f),b}^c s_b^c \right) + \sum_c w_{g(f),b}^c d_{g(f)}^c + \epsilon_{f,b}, \tag{A.2}$$

where $y_{f,b}$ is the loan growth rate, $\bar{c}$ is a constant capturing common aggregate shocks (z) and a Jensen's inequality term, and $\epsilon_{f,b}$ is an idiosyncratic factor.

Comparing (A.2) with our empirical specification (5), we have that $\alpha_f = \bar{c} + (\sigma - 1)a_f$ and $\beta_{b,g(f)} = -\sigma \left(s_b - \sum_c s_b^c w_{g(f),b}^c \right) + \sum_c d_{g(f)}^c w_{g(f),b}^c$. As discussed in section 4, our interaction factor $\beta_{b,g(f)}$ can be driven either by bank supply shocks s_b or s_b^c heterogeneously propagated to firms (due to differences in importance of the bank specific lending advantage across firm groups) or by a firm-group demand shock $d_{g(f)}^c$ heterogeneously propagated to banks.

A.2 Empirical model with firm loan-type specific shocks.

Ivashina et al. (2022) argues that the transmission of credit shocks could depend on the specific type of loan contract l used. They consider the following empirical model with firm-loan specific shocks:

$$\Delta \ln L_{f,b,l,t} = \alpha_{f,t} + \eta^{l}_{f,t} + \beta_{b,t} + \epsilon^{l}_{f,b,t}, \tag{A.3}$$

where $\eta^{l}_{f,t}$ captures a loan-type specific shock. The assumption is that conditional on the observed loan type, credit shocks are transmitted homogeneously. So, when loan-type information is available to the econometrician, they can estimate bank and firm fixed effects by including firm-loan specific fixed effects or conditioning the estimation on the specific firm-loan type.

It is common that lending by loan type is not observed by the econometrician, instead some credit registries provide only information about the total lending from each bank $L_{f,b,t} = \sum_l L_{f,b,t,l}$. Estimating the standard fixed-effects specification, as in section (2), ignoring the differential effects by loan-type could lead to bias estimates. For example, this would happen if a firm demand shock is loan-type specific and different banks specialize in providing different types of loans.

We show below that if the loan-type specific demand happens at a firm group level, our empirical framework with bank-firm group interactions arises as an aggregation of the empirical model (A.3) at the bank-firm level.

If we aggregate credit growth at the firm level (dropping subscript t for convenience), we get $y_{f,b} \equiv \Delta \ln L_{f,b} \approx \sum w^{l}_{f,b} \Delta \ln L_{f,b,l}$, where $w^{l}_{f,b} = \frac{L^{l}_{f,b}}{\sum_l L^{l}_{f,b}}$ denotes the importance of loan-type l in the firm-bank relationship. To capture that banks specialize in providing a certain type of loan, we assume the relative importance of a loan-type across all loans given by the bank $w^{l}_{f,b} = h(\gamma^{l}_{b})$ is a function of the bank specialization parameter γ^{l}_{b} (analogous to the model above). Finally, we assume that the firm loan-specific demand depends on the firm group, so that $\eta^{l}_{f} = \bar{\eta}^{l}_{g(f)}$. For example, a group of firms experiencing a liquidity need may demand a short-term loan, firms planning to take a fixed long-term investment may demand long-term loans, or the group of firms wanting to finance equipment purchases may demand asset-based loans. Each type of firm then will have a preference to borrow differently from each bank.

Then, the aggregation of model (A.3) leads to:

$$y_{f,b,t} = \alpha_{f,t} + \sum_{l} h(\gamma^{l}_{b,t}) \bar{\eta}^{l}_{g(f)} + \beta_{b,t} + \epsilon_{f,b,t}, \tag{A.4}$$

with $\epsilon_{f,b,t} = \sum_l h(\gamma^{l}_{b,t}) \epsilon^{l}_{f,b,t}$. Comparing (A.4) with our empirical specification (5), we have that $\beta_{b,g(f)} = \sum_l h(\gamma^{l}_{b,t}) \bar{\eta}^{l}_{g(f)} + \beta_{b,t}$.

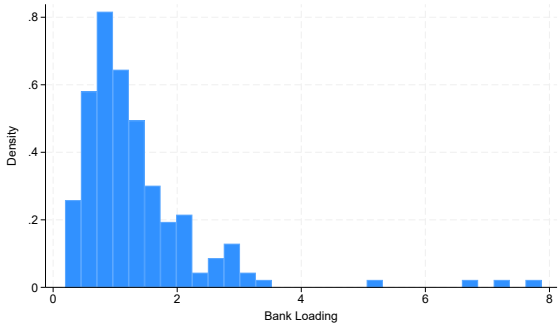
B Interactive fixed effects

Consider the following interactive fixed effect model in the spirit of Bai (2009) which allows for continuous heterogeneity in the transmission of shocks:

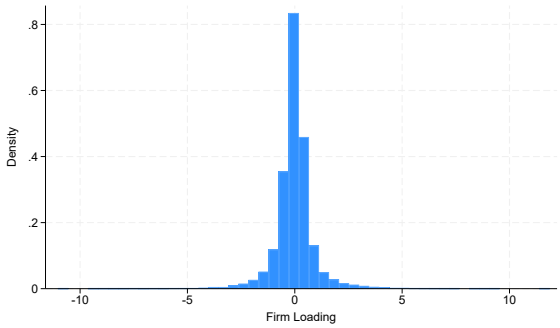
$$\psi_{f,b,t} = \phi_{f,t}\gamma_{b,t}, \tag{B.1}$$

where $\phi_{f,t}$ is an unobserved time-varying firm-specific factor and $\gamma_{b,t}$ is an unobserved time-varying bank-specific factor.

We report the distribution of our estimated firm and bank loadings in figure B.1. Additionally, we investigate whether the firm loadings are explained by observable characteristics in Table 4.



(a) Distribution of the bank loading



(b) Distribution of the firm loading

Figure B.1: Distribution of the firm and bank loadings, $\phi_{f,t}$ and $\gamma_{b,t}$, from a regression of the form: $y_{f,b,t} = \beta_{bt} + \phi_{f,t}\gamma_{b,t} + \epsilon_{f,b,t}$. We have winsorized at 1% for visualization.

Figure B.1 shows that there is significant heterogeneity in both the estimated bank component $\hat{\gamma}b,t$ and the firm component $\hat{\phi}f,t$ of the interactive effect model. This suggests that any bank shock $\hat{\gamma}b,t$ has different effects across firms, as measured by the loading $\hat{\phi}f,t$. Table 4 shows that $\hat{\phi}f,t$ correlates with some firm characteristics available in our database, but there remains considerable variability in $\hat{\phi}f,t$ that is unexplained by observable factors.

Table 4: Firm Loadings from Interactive Models and Observable Characteristics

Dep. $\widehat{\phi}_{f,t}$	Interactive model with additive bank-time FE
ln(Collateral)	-0.0152***
	(0.58)
ln(Debt Size)	0.296***
	(0.79)
Risk Score	-0.123***
	(-1.08)
Fixed Effects	Yes
R-squared	0.48
N	21,357

Note. Regression between the firm loading, $\widehat{\phi}_{f,t}$, and firm observable characteristics. In column (1) we report the estimates of the firm loading estimated using from a regression of the form: $y_{f,b,t} = \beta_{bt} + \phi_{f,t}\gamma_{b,t} + \epsilon_{f,b,t}$. Regressions include other observables as fixed effects: firm, sector-year, financial age-year, firm size-year, firm age-year. Sample goes from 2006 to 2017. Variables definition in Online Appendix H.

There are two main issues with this model. First, the estimation of $\phi_{f,t}$ is based on the number of banks connected to each firm f, so the fixed-effect estimator of $\phi_{f,t}$ is $\sqrt{N_B}$-consistent. Thus, the limited number of banks in the dataset may make the estimation of $\phi_{f,t}$ very noisy and part of the heterogeneity found might be just spurious noise rather than true heterogeneity in effects. Moreover, in this multiplicative model the bias in $\phi_{f,t}$ will contaminate the estimation of $\gamma_{b,t}$ (our object of interest), as it happens in nonlinear panel models.[22] A more important issue is that this model restricts the bank shock transmission to a monotonic pattern. For instance, $\gamma_{b,t}$ can capture a bank shock that affects differently firms with different loadings $\phi_{f,t}$. Firms with high $\phi_{f,t}$ will experience an amplification of shocks $\gamma_{b,t}$ to any bank b, while low $\phi_{f,t}$ will experiencing a mitigation from all banks. In contrast, models with specialization point towards a transmission that depends on the specific match between each bank and firm.[23] Grouping firms allows us to consider some commonalities in firms credit behavior to average out noise and uncover true heterogeneity.[24] Importantly, group heterogeneity arises naturally in banking models with specialization in markets or activities, as we describe in section 4.

[22] In the case of a linear specification, as in equation (1), the bias in $\alpha_{f,t}$, which is also $\sqrt{N_B}$-consistent because it is estimated based on the number of banks connected to each firm f, does not affect the estimation of θ or $\beta_{b,t}$. Both θ and $\beta_{b,t}$ are $\sqrt{N_F}$-consistent, as $\alpha_{f,t}$ can be differentiated out due to the linearity.

[23] One could potentially capture such specific match interaction by saturating the model including a large number of J factors (i.e. $\psi_{f,b,t} = \sum_j^J \phi_{f,t}^j \gamma_{b,t}^j$) at the cost of a more noisy estimation.

[24] Bonhomme et al. (2019) use a similar dimension reduction technique to model firm heterogeneity in matched employer-employee data.

C Properties of the BF-GFE under unknown groups

In this section, we discuss the problem of group misclassification for fixed N_B, and show that the probability of misclassifying tends to zero at an exponential rate.

Lets define the unfeasible group fixed effect estimator $\left(\widehat{\alpha^u}, \widehat{\beta^u}\right)$ as the solution of (10) when $g_t(f)$ is fixed to its population counterpart $g_t^0(f)$ rather than be estimated. Under known groups the group fixed effect estimator $\left(\widehat{\alpha^u}, \widehat{\beta^u}\right)$ coincides with the sample analogue of the moment conditions in (6) and (7),[25] which under the normalization $\beta_{b_0,g,t} = 0$:

$$\widehat{\beta^u}_{b,g_t(f),t} = \frac{\sum_{f \in g_t^0, f \in I(b,b_0)} \left[y_{f,b,t} - y_{f,b_0,t}\right]}{\sum_{f \in g_t^0, f \in I(b,b_0)} 1},$$

and

$$\widehat{\alpha^u}_{f,t} = \frac{\sum_{b \in I(f)} \left[y_{f,b,t} - \beta_{b,g,t}\right]}{\sum_{b \in I(f)} 1}.$$

For simplicity, assume full connections and balanced groups so that: $\sum 1_{f \in g, f \in I(b,b_0)} = N_F/G$ and $\sum 1_{b \in I(f)} = N_B$. In the case of fixed known groups, we have $\widehat{\beta^u}_{b,g_t(f),t} \xrightarrow{p} \beta_{b,g_t(f),t}$ as N_f tend to infinity and $\widehat{\alpha^u}_{f,t} \xrightarrow{p} \alpha_{f,t}$ as N_B tend to infinity for all t. In the case of unknown groups, the group fixed effect estimator will coincide with the unfeasible group fixed effect estimator $\left(\widehat{\alpha^u}, \widehat{\beta^u}\right)$ under correct group classification $\widehat{g}_t\left(f|\widehat{\alpha}, \widehat{\beta}\right) = g_t^0(f)$. However, in finite samples, there is a non-zero probability that estimated and population group membership will not coincide. To illustrate the point, we follow Bonhomme and Manresa (2015) and consider a simple version of the model with just group effects with two groups ($G = 2$) and normal errors:

$$y_{f,b,t} = \beta^0_{g_t^0(f),t} + \epsilon_{f,b,t}$$

where $\beta^0_{g_t(f),t}$ are the true parameters, $g_t^0(f) \in \{1, 2\}$ is the true grouping assignment and $\epsilon_{f,b,t} \sim N(0, \sigma^2)$. From (9), we can see that the probability of wrongly assigning into group

[25] This is the case for $N_B = 2$, in a similar manner to how the within-group estimator and the first-difference estimator coincide in a standard panel data with $T = 2$. For $N_B > 2$, the unfeasible estimator under known groups will combine all the valid moment conditions across all possible bank first-differences.

2 a firm that belongs to group 1 is:

$$\begin{aligned}\Pr\left(\hat{g}_t(f)=2 \mid g_t^0(f)=1\right) &= \Pr\left(\sum_{b=1}^{N_B}\left(\beta_{1,t}^0+\epsilon_{f,b,t}-\beta_{2,t}^0\right)^2 < \sum_{b=1}^{N_B}\left(\beta_{1,t}^0+\epsilon_{f,b,t}-\beta_{1,t}^0\right)^2\right) \\ &= \Pr\left(\frac{\sum_{b=1}^{N_B}\epsilon_{f,b,t}}{N_B} > \frac{\beta_{2,t}^0-\beta_{1,t}^0}{2}\right) \\ &= 1-\Phi\left(\sqrt{N_B}\left(\frac{\beta_{2,t}^0-\beta_{1,t}^0}{2\sigma}\right)\right), \text{ for all t,} \end{aligned} \tag{C.1}$$

where Φ is the standard normal CDF. For fixed (and small) N_B, there is a non-negligible probability of misclassifying firms, so $\hat{g}_{(f),t}$ is inconsistent even if N_F tend to infinity. Intuitively, according to (9) the estimator is classifying firms based on their similar responses to different banks, so we need to observe firms connected to several banks to be able to separate the effect $\beta^0_{g_t^0(f),t}$ from the idiosyncratic error $\epsilon_{f,b,t}$. As a consequence, $\widehat{\beta}$ will suffer from an incidental parameter bias and it will be inconsistent for fixed N_B even when N_F tend to infinity. The latter is in contrast to the case when groups are known where $\widehat{\beta}$ is $\sqrt{N_F}$-consistent. Note that the firm fixed effects $\hat{\alpha}$ is always inconsistent for fixed N_B even when groups are known or even when the true model has homogenous effects as in Khwaja and Mian (2008). However, equation C.1 indicates that the probability of misclassifying tends to zero at an exponential rate which implies that the bias generated by the incidental parameter problem goes to zero very fast as N_B increases, similarly to the group fixed effect estimator of the standard panel model when T increases (see Bonhomme and Manresa (2015)). As we show in simulations, the performance of our estimator increases very rapidly with N_B and is centered at the true value for moderate N_B.

D BF-GFE estimation algorithm

D.1 Algorithm.

We use the following iterative strategy for estimation: For each $t = 1, ..., T$,

1. Set the number of groups: G.
2. Set $s = 0$. Guess initially some group assignment $g^{(s=0)}(f) \in \{1, ..., G\}$.[26]
3. For given $g^{(s)}(f)$, estimate firm and bank-firm-group fixed effects $\hat{\alpha}^{(s)}_{f,t}, \hat{\beta}^{(s)}_{b,g(f),t}$:

$$\left(\hat{\alpha}^{(s)}, \hat{\beta}^{(s)}\right) = \arg\min_{\alpha_{f,t}, \beta_{b,g^{(s)}(f),t}} \sum_{f=1}^{N_F}\sum_{b=1}^{N_B}\left(y_{f,b}-\alpha_{f,t}-\beta_{b,g^{(s)}(f),t}\right)^2. \tag{D.1}$$

[26] We propose starting by grouping firms using k-means over $\widehat{\alpha}_{f,t}$'s obtained from the homogeneous model.

4. For given $\widehat{\alpha}^{(s)}, \hat{\beta}^{(s)}$, select optimal group assignment: For all $f = 1, ..., N_F$

$$g^{(s+1)}(f) = \arg \min_{g \in \{1,...,G\}} \sum_{b=1}^{N_B} \left(y_{f,b} - \hat{\alpha}_{f,t}^{(s)} - \hat{\beta}_{b,g,t}^{(s)} \right)^2 \tag{D.2}$$

5. Set $s = s + 1$ and go to Step 3 until numerical convergence.

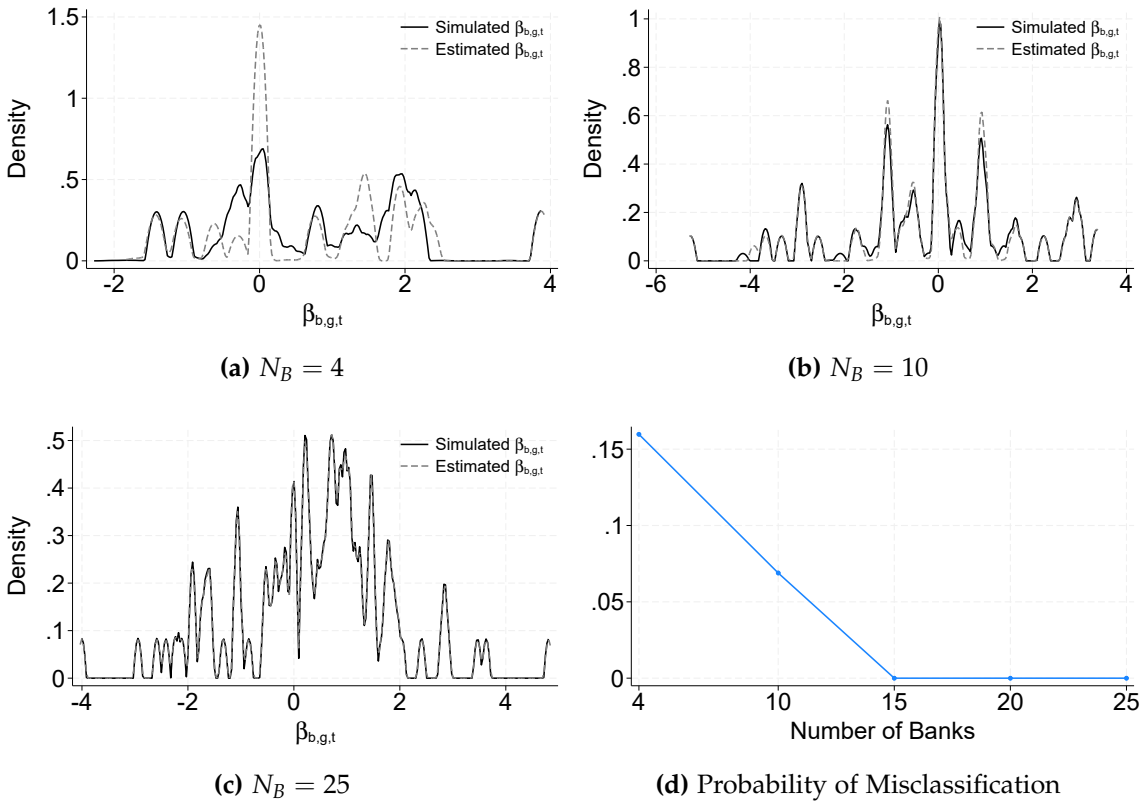
D.2 Properties of our estimator.

To illustrate the performance of our proposed algorithm, we simulate a model based on (5). In an initial simulation, we assume an exogenous network with full connections. The primary objective of this exercise is to emphasize our algorithm's performance across various values of N_B. We fixed $G = 4$ and $N_F = 5000$ and conducted 100 replications of $y_{f,b}$ based on (5). The distribution of the components are defined as follows: $\beta_{b,g,t} = N(3, 1.5)$, $\alpha_f \sim N(5, 1)$, and $\varepsilon_{f,b} \sim N(0, 0.35)$.

For each replication, we estimate $\beta_{b,g(f)}$ using our algorithm. Figure D.1 illustrates the estimator accuracy for different values of N_B: 4, 10, and 25. Since there are many ($G \times N_B$) credit supply shocks $\beta_{b,g(f)}$ that are simulated, the figure presents a k-density over these $G \times N_B$ generated parameters and the k-density of the estimated ones. Panels (a)-(c) display the density of each of the true parameter $\beta_{b,g(f)}$ (the black solid line) and the density of each of the estimates $\hat{\beta}_{b,g(f)}$ across replications for different values of N_B. Panel (d) displays the probability of misclassification for different values of N_B. As our asymptotic analysis predicts, increasing the number of banks reduces the probability of misclassifying firms into groups and decreases the bias in the estimator. For a small number of banks ($N_B = 4$), the probability of misclassification is 16% due to the incidental parameter problem inherent in small samples. The latter translate in a small sample bias in $\hat{\beta}_{b,g(f)}$. However, the probability of misclassification decreases rapidly as we increase the number of banks. It reduces to 6.89% for $N_B = 10$ and it is close to 0 for $N_B = 15$. That is, the bias generated by the incidental parameter problem disappears.

Next, we present the simulation results for an example specifically calibrated to our dataset.

Calibrated simulation. We evaluate the performance of our estimator in a simulation calibrated to our database. We fix $G = 4$, $N_F = 5000$, and $N_B = 10$ (note that there are 17 banks and 5000 firms in our dataset). We set the standard deviation of the error term to $\sigma_\epsilon = 0.35$ (the one estimated with our data) and we assume partial connections with a probability of matching of 40%. The distribution of the components are defined as follows: $\beta_{b,g,t} = N(3, 1)$, $\alpha_f \sim N(5, 1)$, and $\varepsilon_{f,b} \sim N(0, 0.5)$. The exogenous matching probability



(a) $N_B = 4$ **(b)** $N_B = 10$

(c) $N_B = 25$ **(d)** Probability of Misclassification

Figure D.1: Properties of the estimator using simulations. Panels (a)-(c) show the simulated and estimated parameters' k-density across all β_{bg}'s (for $N_B = 4, 10, 25$). Panel (d) displays the probability of misclassification.

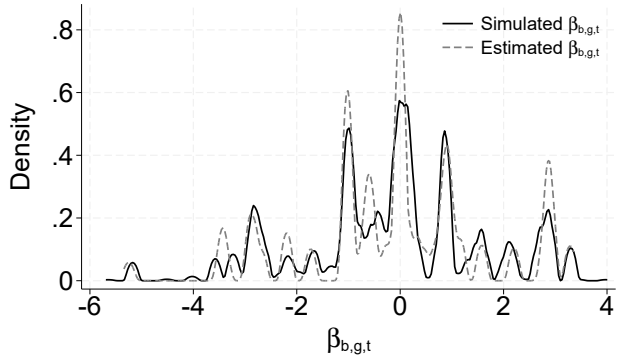
follows $\text{Prob}(D_{f,b} = 1 \mid \alpha_f, \beta_{b,g(f)}) = 0.4$. We conduct 100 replications of $y_{f,b,t}$ based on (5); for each replication, we estimate $\beta_{b,g(f)}$ using our algorithm. Figure D.2 displays the density of the true parameter $\beta_{b,g(f)}$ and estimates $\hat{\beta}_{b,g(f)}$ across replications. The figure shows that the estimation closely tracks the true parameters.

D.3 Average bank effect under endogenous matching: BF-GFE vs two-way fixed effects

Even when the true data-generating process includes interactions as in specification (5), the decomposition of credit growth into (homogeneous) firm-specific and bank-specific factors as in equation (1) is indeed possible, but the estimated parameters require careful interpretation.

Let us denote the estimate of the homogeneous effect $\beta_{b,t}$ in (3) as $\hat{\beta}^{\text{Homo}}_{b,t}$. In the case in which there exists endogenous matching based on interactions, so that the probability of creating a relationship $D_{f,b,t}$ depends on the interaction $\beta_{b,g,t}$, then the estimate $\hat{\beta}^{\text{Homo}}_{b,t}$ may not even converge to the average of the heterogeneous effects. For instance, consider the case of two banks b, b_0, and let I_{b,b_0} be the joint network; we have that the standard fixed

Figure D.2: Estimator's properties: calibrated Monte Carlo simulation



Note. Illustration of estimator's properties for a Monte Carlo simulation calibrated to our dataset. We assume $G = 4$, $N_B = 10$, $N_F = 5000$, $\sigma_\epsilon = 0.35$, and matching probability of 40%. The results are generated after 100 replications. The dark solid line displays the k-density across all simulated β_{bg}'s, while the gray dashed line displays the k-density across estimated $\widehat{\beta}_{bg}$'s.

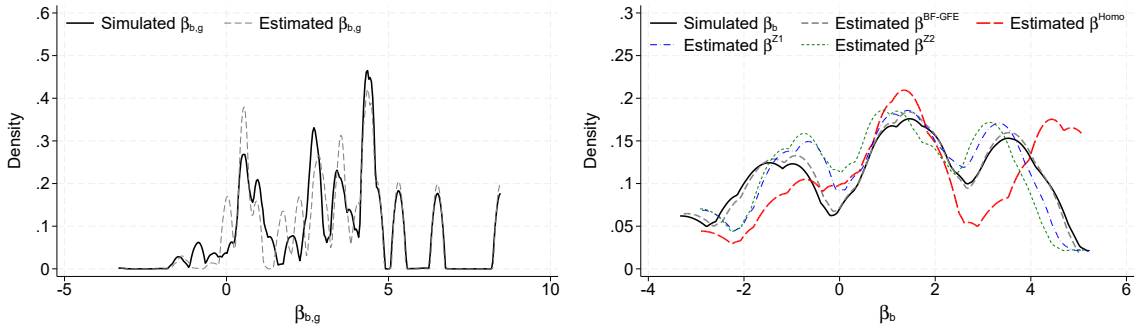
effect estimator (normalizing the bank effect of b_0 to zero):

$$\hat{\beta}_{b,t}^{\text{Homo}} \overset{a}{=} E\left[y_{f,b,t} - y_{f,b_0,t} \mid f \in I_{b,b_0}\right] = E\left[\beta_{b,g(f)} \mid f \in I_{b,b_0}\right] \neq E\left[\beta_{b,g(f)}\right], \tag{D.3}$$

where the first equality follows from (3), the second equality uses (5) and assumption (2), and the third inequality arises because of the conditioning on the joint network. For instance, bank b and b_0 may each be specialized in lending to different types of firms, affecting the composition of their joint network.

Simulation with endogenous matching and groups defined by observables. To illustrate the bias of $\hat{\beta}_b^{\text{Homo}}$ when the network of connections depends on the interaction effect $\beta_{b,g}$, we conduct a Monte Carlo simulation with a matching probability between banks and firms as a function of $\beta_{b,g}$. Following the evidence in figure 1, we model the firm group's heterogeneous response as a function of a linear combination of two observable characteristics in the data: firm age (Z_1) and collateral (Z_2). We simulate 5000 firms distributed into $G = 4$ groups based on the variable Group_f as the quartile of $\frac{1}{2}Z_1 + \frac{1}{2}Z_2$. We choose $N_B = 10$ and define $\beta_{b,g,t} = \text{Group}_f \times \psi_b$ with $\psi_b \sim N(0,1)$, and set $\alpha_f \sim N(0,1), \varepsilon_{f,b} \sim N(0, 0.05)$. The endogenous matching probability follows $\text{Prob}\left(D_{f,b} = 1 \mid \alpha_f, \beta_{b,g(f)}\right) = \pi\left(\alpha_f + \beta_{b,g_t(f)}\right)$ where $\pi(.)$ is an increasing function. We generate 100 replications of $y_{f,b,t}$ according to (5).

Panel (a) of Figure D.3 displays the k-density of the $G \times N_B$ interaction effects $\beta_{b,g,t}$ and our estimates $\hat{\beta}_{b,g(f)}$. Similar to the simulations in Figure D.1, our BF-GFE estimator closely tracks the true parameters. In Panel (b) of Figure D.3, we present estimators of bank-specific *average* effects: $\beta_b = \sum_{g=1}^{G} \frac{N_{b,g}}{N_b} \beta_{b,g(f)}$ for $b = \{1 \dots 10\}$, where $N_{b,g}$ is the number of



(a) Estimates of $\beta_{b,g(f)}$ **(b)** Estimates of β_b by method

Figure D.3: Properties of estimators of $\beta_{b,g(f)}$ and β_b using Monte Carlo simulations.

firms in group g connected to b and $N_b = \sum_{g=1}^{G} N_{b,g}$. We compare the distribution across simulations of: $\hat{\beta}_b^{\text{BF-GFE}} = \sum_{g=1}^{G} \frac{N_{b,g}}{N_b} \hat{\beta}_{b,g(f)}$, where $\hat{\beta}_{b,g(f)}$ is the BF-GFE estimator with unknown groups, and $\hat{\beta}_{b,t}^{\text{Homo}}$ the two-way fixed effect estimator (assuming homogeneity). The BF-GFE estimator tracks closely the true average β_b for all banks, while the homogeneous estimator is a biased estimator of the bank-specific average effect. Additionally, we show the results for a group estimator defined by using only one of the observable characteristics: $\hat{\beta}_b^{Z_j} = \sum_{g=1}^{G} \frac{N_{b,g}}{N_b} \hat{\beta}_{b,g(f)}^{Z_j}$, where $\hat{\beta}_{b,g(f)}^{Z_j}$ is a within-group estimator (a two-way fixed effect for each group) with groups defined by the observable Z_j, for $j = \{1, 2\}$. The example shows that controlling heterogeneity using only one observable variable is insufficient.

E Out-of-sample performance and number of groups

Building on Almagro and Manresa (2021), we propose the following N-fold cross-validation procedure for our algorithm to evaluate the out-of-sample performance of the model with number of groups G.

1. For a given year t, set the number repetitions: M.

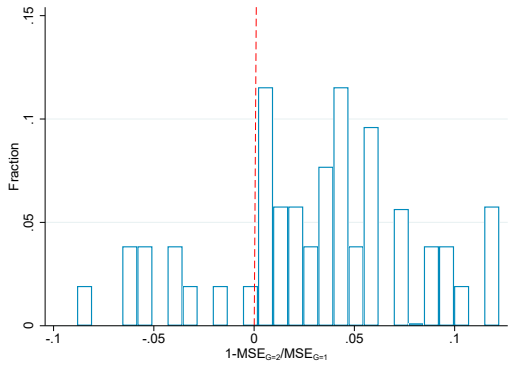
2. For $\forall m \in \{1, ..., M\}$, split the sample in N parts (folds):

$$P_t(m) \equiv \{P_{t,1}(m), P_{t,2}(m), ...P_{t,N}(m)\}.$$

3. Take $P_{t,k}(m)$ as the testing sample, and the remaining parts of the sample, $P_{t,-k}(m)$, as the training sample.

4. Using the training sample, $P_{t,-k}(m)$, estimate the group structure and the parameters of interest. After that, compute the out-of-sample mean squared error (MSE) for the

Figure E.1: Estimated $MSE_{G=2}$ vs $MSE_{G=1}$



Note. Out-of-sample mean squared errors for $G = 2$ vs $G = 1$. A positive value of the measure indicates that the out-of-sample MSE of $G = 2$ is lower than $G = 1$. The red dashed line indicates the case where both measures are equal. The information covers 2005 through 2017.

testing sample, $P_{t,k}(m)$, which number of observations is denoted by J, as follows:

$$MSE_t(m, P_{t,k}) = \frac{1}{J} \sum_{j \in P_k} \left(y_{f,b,t,j} - \widehat{\alpha}_{f,t,-j}(m) - \widehat{\beta}_{b,g(f),t,-j}(m) \right)^2. \tag{E.1}$$

5. Take the average for each of the folds and the repetitions in the sample:

$$\overline{MSE_t} = \frac{1}{N} \frac{1}{M} \sum_{m=1}^{M} \sum_{k=1}^{N} MSE_t(m, P_{t,k}).$$

In Figure E.1, we display the distribution of the out-of-sample MSE for each of the folds and repetitions for $G = 2$ relative to to $G = 1$ (homogenous case). A positive value implies that the heterogeneous case model has a better out-of-sample explanatory power than the homogeneous case. In general, we observe the heterogeneous model with $G = 2$ is most likely to perform better than the traditional homogeneous case with $G = 1$ in the testing sample.

F Additional variable definitions and tables

Variable definitions in Table 3

We define the variables used in Table 3:

- Collateral denotes one plus the total value of the collateral across all firm loans.

- Debt size denotes the total value of firm debt across all its loans.

- Exports is the total value of firm exports conditional on being a exporter.

- Firm size is a dummy equal to one if the firm belongs to the group of large corporate firms, 0 otherwise.

- Financial age are the years of the firm in the banking system. Firm age are the years of the firm since its constitution.

Additional tables.

Table 5: Firm-level investment regression

Dependent variable: $\frac{\text{Investment}_{f,t}}{\text{Capital}_{f,t-1}}$	(1)	(2)	(3)	(4)	(5)
	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
Bank Shock	-2.809	6.711*	3.419**	6.271*	6.539
	(-0.73)	(1.86)	(2.31)	(1.82)	(1.63)
Firm FE	-0.145	0.810**	0.804	1.161	1.289
	(-0.42)	(2.58)	(1.50)	(1.40)	(0.99)
Ln(Asset)	5.937*	6.737	6.354	5.380	5.947
	(2.18)	(1.54)	(1.49)	(1.07)	(1.50)
Fixed Effects:					
Group-Time	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes
R-squared	0.21	0.24	0.25	0.29	0.29
N	785	785	785	785	785

Note. Standard errors are cluster at the main leading bank. The information spans from 2005 to 2017.