

Bremus, Franziska; Buch, Claudia M.; Russ, Katheryn N.; Schnitzer, Monika

Conference Paper

Big Banks and Macroeconomic Outcomes: Theory and Cross-Country Evidence of Granularity

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Financial Markets and Macroeconomics, No. B19-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Bremus, Franziska; Buch, Claudia M.; Russ, Katheryn N.; Schnitzer, Monika (2013) : Big Banks and Macroeconomic Outcomes: Theory and Cross-Country Evidence of Granularity, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Financial Markets and Macroeconomics, No. B19-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/80048>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

BIG BANKS AND MACROECONOMIC OUTCOMES:
THEORY AND CROSS-COUNTRY EVIDENCE OF GRANULARITY

Franziska Bremus
(DIW Berlin)

Claudia M. Buch
(University of Tuebingen and CESifo)

Katheryn N. Russ
(University of California, Davis, and NBER)

Monika Schnitzer
(University of Munich and CEPR)

This draft: 23 May 2013

Abstract

Does the mere presence of big banks affect macroeconomic outcomes? In this paper, we develop a theory of granularity (Gabaix, 2011) for the banking sector, introducing Bertrand competition and heterogeneous banks charging variable markups. Using this framework, we show conditions under which idiosyncratic shocks to bank lending can generate aggregate fluctuations in the credit supply when the banking sector is highly concentrated. We empirically assess the relevance of these granular effects in banking using a linked micro-macro dataset of more than 80 countries for the years 1995-2009. The banking sector for many countries is indeed granular, as the right tail of the bank size distribution follows a power law. We then demonstrate granular effects in the banking sector on macroeconomic outcomes. The presence of big banks measured by high market concentration is associated with a positive and significant relationship between bank-level credit growth and aggregate growth of credit or gross domestic product.

Keywords: Granularity, concentration, bank competition, macroeconomic outcomes, bank markups

JEL Codes: E32, G21

Corresponding author: Claudia Buch, University of Tuebingen, Mohlstrasse 36, 72074 Tuebingen, Germany, Phone: +49 7071 2972962. Email: claudia.buch@uni-tuebingen.de

This paper was written in the context of the Priority Programme SPP1578 “Financial Market Imperfections and Macroeconomic Performance” of the German National Science Foundation (DFG) and the Economics Program of the U.S. National Science Foundation (Project No. 201222555), from whom funding is gratefully acknowledged. The authors also thank seminar participants at the University of Mannheim, the Federal Reserve Board of Governors Office of Financial Stability, the 2011 WEAI Summer Meetings, and the Fall 2012 Midwest Macro Meetings as well as Silvio Contessi, Kris Mitchener, Julian di Giovanni, and Andrei Levchenko for helpful comments. Lu Liu provided very valuable research assistance. All errors are our own.

1 Introduction

The purpose of this paper is to determine whether and under what conditions the presence of big banks, in itself, can affect macroeconomic outcomes such as aggregate credit and output. Given the recent interest in regulatory reform, this question has become a focal point both in political debates and in the broader public discourse. A number of prominent policy makers and academics recently have proposed limiting the size of banks or breaking large banks into smaller ones. Yet, the academic literature investigating this potential link is surprisingly small, so our understanding of the implications of bank size for macroeconomic outcomes remains limited. In this paper, we provide a theoretical framework to study this issue in the data. Empirical evidence from more than 80 countries suggests that indeed idiosyncratic shocks to large banks can cause macroeconomic fluctuations.

The idea that bank size can destabilize aggregate credit is not new.¹ Bail out expectations may invite imprudent risk-taking of large banks (“too big to fail”), and close linkages between large banks and highly leveraged shadow financial institutions (“too connected to fail”) may destabilize the entire financial system. The focus on size in policy debate and the media is inspired by some sensational bank failures, but also by the general observation that the banking sector in many countries is indeed very concentrated.² What is more, the largest 15 banks hold about one quarter of commercial bank assets worldwide, so the biggest banks are quite large not only in the relative sense, but also in the absolute.

So far, the literature measuring the influence of bank size has focused on issues of connectedness, spillovers, and exposure to common macroeconomic shocks. Tarashev et al. (2009, 2010) use a Shapley value to map risk associated with individual banks into aggregate risk, while Adrian and Brunnermeier (2009) pioneer the use of a CoVaR model to measure systemic risk in banking. Both papers show that large banks are more systemically important. Work by Hale (2012) identifies connectedness, through interbank lending, as a channel spreading shocks in the lending behavior of large banks from one bank to another with implications for business cycle behavior. Corbae and D’Erasmus (2013) examine the role of large banks in exacerbating or mitigating macroeconomic effects that emerge when banks are exposed to national or regional macroeconomic shocks.

Our approach differs from this research because we study the effects of bank size for macroeco-

¹Boyd and Gertler (1993), for example, point to bad real estate loans by large banks as the primary source of the U.S. banking crisis in the 1980s and subsequent economy-wide credit crunch.

²In a study of trends in financial consolidation, the G10 has found an increase in banking sector concentration in the advanced economies (Group of Ten 2001). Empirical evidence provided by Corvoisier and Gropp (2002) for a group of advanced countries and by Walkner and Raes (2005) for the European countries points into the same direction. Moreover, De Nicolo et al. (2004) find that banking sector concentration tends to increase when looking at the world average. For the US, Berger et al. (1999) find an increase in banking concentration at the national level, while Schargrodsky and Sturzenegger (2000) document increasing local concentration in Argentina. Calderon and Schaeck (2012) and the Organisation of Economic Cooperation and Development (OECD 2010) show that merger activity during the global financial crisis has led to higher concentration in many countries. Other studies point to a rather heterogeneous evolution of banking concentration across the world. These studies find that some countries have experienced increasing concentration, while other countries have seen a decrease in concentration over the last decades (e.g. Hawkins and Mihaljek (2001), De Nicolo et al. (2004), Beck and Demirgüç-Kunt (2009), Davis (2007), and De Bandt and Davis (1999)).

economic outcomes even in the absence of contagion, spillover effects, or shared responses to macroeconomic shocks. Instead, we focus on granular effects as a channel through which large banks can affect macroeconomic outcomes even in otherwise normal times, in addition to times of crisis or common adversity. Generally, the theory of granularity predicts that adverse idiosyncratic shocks to very large (manufacturing) firms do not average out across the population of firms, but rather affect aggregate fluctuations (Gabaix 2011). We apply this concept to the banking sector in two steps. First, we determine whether the banking sector in theory and in practice fits the necessary conditions for granular effects to arise. Second, we test whether there is a statistically significant relationship between the presence of big banks as measured by a high level of market concentration, and macroeconomic outcomes. Our answer to both questions is “Yes.”

Our research builds on Gabaix (2011) who pioneers the concept of granularity in economics, showing that idiosyncratic shocks (the “granular residual”) hitting the largest 100 US firms can explain a significant portion of growth in per capita GDP. The mechanism driving granular effects lies in the unequal distribution of firm sizes. Firm size distributions are usually fat-tailed - there are many small firms but also a few extremely large ones. The fat tail implies that the distribution of firm size resembles a power law. In this case, idiosyncratic shocks to large firms do not cancel out in the aggregate.³ Gabaix provides a theoretical underpinning to calculate macroeconomic outcomes based on a Herfindahl index computed across heterogeneous firms, neatly summarizing the distribution of firm size within an index of market concentration. In his model, markups are constant, so that shocks are passed on fully into prices and thus the equilibrium quantity of output. Di Giovanni and Levchenko (2009) and Di Giovanni et al. (2011) further develop this concept to analyze the link between trade liberalization and macroeconomic fluctuations, also with a theory using constant markups.

We expand the theory of granularity to encompass financial intermediaries of heterogeneous size who charge variable markups. For this purpose, we develop a discrete choice model with a large number of rival banks competing in a Bertrand-like fashion to provide homogeneous loans. We extend the framework developed in De Blas and Russ (2010, 2013) by integrating the concepts of concentration and granularity. Borrowers do not know exactly what interest rate a bank will charge until they apply. In the spirit of Anderson et al. (1987), *ex ante* uncertainty generates market power. Banks also differ in their costs, hence markups may vary across banks depending on the magnitude of the search friction. Into this framework, we incorporate a power law distribution of bank size. The model predicts that macroeconomic outcomes are driven in part by the “banking granular residual”— the product of a measure of idiosyncratic fluctuations and the banking system’s Herfindahl index as a measure of concentration. We characterize the necessary conditions in terms of market concentration for granular effects to emerge: On the one hand, idiosyncratic shocks have

³According to a simple diversification argument, independent idiosyncratic shocks to firms should have an impact of $1/\sqrt{N}$ on aggregate fluctuations (Gabaix 2011, p.735). In an economy with a small number of firms (small N), idiosyncratic shocks would thus be felt in the aggregate. However, if the number of firms is large, as in most economies today, the effect of idiosyncratic firm-level shocks on the aggregate should tend towards zero. Gabaix shows that, under a fat-tailed power law distribution of firm size, macroeconomic volatility arising solely due to firm-level shocks decays much more slowly with $1/\ln(N)$.

to be passed through to firms via changes in lending rates. On the other hand, the distribution of bank size has to follow a fat-tailed power law to be sufficiently dispersed. We show that, under these conditions, the higher the concentration or volatility of idiosyncratic fluctuations in the banking sector, the larger are fluctuations in the aggregate supply of credit and output. Hence, the presence of big banks magnifies the effects of bank-level shocks on aggregate credit and output compared to an economy where the banking sector is less concentrated.

The presence of granular effects in banking hinges crucially upon the size distribution of banks. Thus, we take our model to the data and use Bankscope data to explore whether the distribution of bank sizes exhibits a fat right tail. Maximum likelihood estimates reveal tails for the banking sector in the world's largest economies that are fatter than those found for manufacturing firms by Di Giovanni et al. (2011).⁴ These patterns in the data suggest that shocks hitting large banks may indeed have aggregate effects.

Our work is linked to two strands of literature which study the effects of heterogeneous banks for macroeconomic outcomes. First, among a small number of recent empirical studies, Buch and Neugebauer (2011) show that granularity in banking matters for short-run output fluctuations in a subsample of Eastern European banks. Blank et al. (2009) use data for German banks and find that shocks to large banks affect the probability of distress among small banks. Using industry-level data, Carvalho and Gabaix (2011) show that the exposure of the macroeconomy to tail risks in what is called the "shadow banking system" has been fairly high since the late 1990s. Our analysis is distinct from these studies in that we include a larger set of countries, explicitly test for dispersion in bank size, and investigate the importance of the factors driving granular effects within the framework of a structural model.

A second strand of literature incorporates banks into dynamic stochastic general equilibrium models. Several of these models assume the presence of a representative bank in modeling links between banks and the macroeconomy in the presence of financial frictions (see, e.g., Angeloni and Faia 2009, Meh and Moran 2010, Zhang 2009, and Ashcraft et al. (2011)). Kalemli-Ozcan et al. (2012), van Wincoop (2011), Mandelman (2010), and Ghironi and Stebunovs (2010) show the implications of foreign participation or domestic bank branching for the transmission of shocks overseas in structural models. Several studies introduce nest heterogeneity in bank size by assuming that deposits and loans are CES baskets of differentiated products (Andres and Arce 2012, Gerali et al. 2010), yielding constant markups when banks set interest rates on loans that do not vary by bank size. Two important exceptions are Mandelman (2010), who incorporates heterogeneous bank lending costs into a limit price framework, and Corbae and D'Erasmus (2013), who combine heterogeneous lending costs with Cournot competition. Markups in these two cases are endogenous and, in particular, sensitive to market structure. The focus of these papers is on the impact of bank market structure on the propagation of macroeconomic shocks rather than the feedback between bank-level shocks and macroeconomic outcomes. In contrast, we study how idiosyncratic changes

⁴The tails are truncated in our case because the cost of financing is neither infinitely low nor infinitely high for banks. We demonstrate numerically that this truncation need not prevent granular effects from occurring and empirically that the truncation indeed *does* not prevent granular effect from occurring.

in bank lending behavior can add up to fluctuations in macroeconomic aggregates simply because some banks are very large relative to their competitors.

In the following section, we present a theoretical framework which shows how concentration in banking markets affects the link between idiosyncratic shocks and macroeconomic outcomes. In Section 3, we describe the data that we use to test the predictions of this model, and we provide descriptive statistics on the key features of the model. We demonstrate the link between concentration in the banking sector, idiosyncratic bank-level shocks, and macroeconomic outcomes. Section 4 concludes.

2 Market concentration and macroeconomic outcomes: Theoretical framework

In this section, we develop a model of an economy with a banking sector funded by customer deposits and equity and providing working capital loans to firms. We choose firms as borrowers to provide the simplest link between the credit market and aggregate output.⁵ Our focus is on competition between heterogeneous banks on the loan market. We use this framework to explore the link between bank-specific shocks and macroeconomic outcomes.

2.1 Consumers

The consumer side is captured by a representative consumer. Because the focus of our analysis is on the supply side of the market, we do not explicitly model the labor market. Instead, we follow Obstfeld and Rogoff (1995) by supposing that the utility of the representative consumer is log-linear in aggregate consumption Q and decreasing in the amount of effort expended in the production of aggregate output Y . Thus, the utility function is given by

$$U(Q_t, Y_t) = \ln Q_t - \frac{z}{2} Y_t^2,$$

where z is a parameter reflecting the disutility of effort in producing output. The representative consumer chooses whether to use income to purchase goods for immediate consumption or to save by leaving some of her wealth in a bank. In particular, she maximizes lifetime utility

$$\max \sum_{t=0}^{\infty} \beta^t U(Q_t, Y_t),$$

where β is a constant with $0 < \beta < 1$, subject to the intertemporal budget constraint

$$P_t Q_t + P_t D_{t+1} \leq P_t Y_t + (1 + r^d) P_t D_t.$$

⁵We can derive the qualitative results even if loans are made only to consumers for housing or durable goods, as long as there is a constant (price) elasticity of substitution between the good purchased on credit and other goods.

P denotes the price of a bundle of consumption goods, D is the amount of real wealth deposited in banks as savings, and r^d is the interest rate banks pay on deposits.⁶ Households own the banks. Since bank profits consist of payments from firms, they are embedded in the household budget constraint within firm revenues, PY .

The first-order conditions from the consumer's problem for optimal consumption and deposit holdings yield an endogenous steady-state equilibrium interest rate for deposits (Appendix A): $r^d = (1 - \beta)/\beta$. First-order conditions also yield an expression for consumption as a function of aggregate output, $Q = 1/(zY)$. We focus on comparative statics in steady state and drop time subscripts from this point.

By assumption, the market for deposits is perfectly competitive, i.e. consumers can deposit funds in any bank without cost or other rigidities. Thus, in equilibrium r^d is the deposit rate paid by all banks. We abstract from imperfect competition on the deposit side of the market because deposits are typically guaranteed by the government (implicitly or explicitly), so consumers are indifferent as to where they hold their deposits. Kashyap et al. (2002) have also argued that banks' deposit and lending business are *de facto* two sides of the same coin. Our objective is to emphasize the effects of loan market competition by banks of different efficiency. For this purpose, it suffices to focus on imperfect competition on the lending side. This does not preclude additional investigation into the market for deposits in a more elaborate framework, but it is beyond the scope of this paper.

2.2 Firms

A sector with identical, perfectly competitive firms assembles a homogeneous final good Y . The assembly process for this final good requires a continuum of intermediate goods, $Y(i)$, produced by a continuum of identical manufacturing firms along the $[0,1]$ interval, each of which produces a unique intermediate good i under monopolistic competition. These intermediate goods are bundled as in Dixit and Stiglitz (1977),

$$Y = \left(\int_0^1 Y(i)^{\frac{\mu-1}{\mu}} di \right)^{\frac{\mu}{\mu-1}},$$

where $\mu > 1$ is the elasticity of substitution between the intermediate goods. Producer i sells its good at a price $P(i)$ per unit, with P representing the aggregate price index for intermediate goods, given by

$$P = \left(\int_0^1 P(i)^{1-\mu} di \right)^{\frac{1}{1-\mu}}.$$

Note that the price index P is the cost of all inputs required to produce one unit of the final good, and it is thus the price of the final good as well.

⁶This formulation is consistent with our assumption of a closed economy. In an open economy setting, it would imply domestic ownership of banks, a reflection of home bias in asset holdings which is empirically important despite the ongoing integration of international banking markets (Fidora et al. 2007, Schoenmaker and Bosch 2008).

The demand for any particular good is downward sloping in its price:

$$Y(i) = \left(\frac{P(i)}{P} \right)^{-\mu} Y. \quad (1)$$

Production of intermediate goods requires capital as the sole variable input. Firms produce each good $Y(i)$ with working capital $K(i)$ using the technology

$$Y(i) = \alpha K(i). \quad (2)$$

where α is the productivity of capital. Therefore, the demand for capital is directly proportional to the demand for a firm's output.

Firms face a cash-in-advance constraint. To produce, firms must borrow working capital from financial intermediaries. The need for loans arises because, in steady state, firms cannot accumulate retained earnings but must pay out all profits to consumers in the form of a dividend $\Pi^F(i)$, with Π^F representing total profit from the manufacturing of intermediate goods, summed over all firms i . While, in a dynamic setting, firms can amass retained earnings to provide self-financing, we focus only on the steady-state equilibrium in which positive amounts of cash on hand cannot be optimal. A firm's fiduciary responsibility to its household-shareholder implies a transversality condition in which, ultimately, it must remit any positive amounts of cash holdings to the shareholders. In addition, there is empirical evidence that agency problems compel stockholders to collect dividends and push the firm to seek external finance to benefit from the monitoring capabilities of outside lenders (DeAngelo et al. (2006), Denis and Osobov (2008)).

Let $R(i)$ denote the unit cost of borrowed working capital paid by firm i . Then variable profits for a producer of intermediate goods borrowing at this interest rate is given by

$$\Pi^F(i) = P(i)Y(i) - R(i)K(i). \quad (3)$$

The first-order condition for profit maximization with respect to price yields the usual pricing rule as a markup over marginal cost:

$$P(i) = \frac{\mu}{\mu - 1} \frac{R(i)}{\alpha}. \quad (4)$$

Note that the interest rate $R(i)$ that a particular firm faces affects the price it charges. Setting output equal to demand for good i and substituting in the pricing rule, Eq.(4), yields the firm's demand for capital and thus for loans:

$$L(i) = K(i) = \frac{1}{\alpha} \left[\frac{\frac{\mu R(i)}{(\mu-1)\alpha}}{P} \right]^{-\mu} Y. \quad (5)$$

All else equal, the demand for loans is decreasing in the interest rate and also in the productivity of capital, α , because higher productivity everywhere allows firms to produce more output with less

capital.

2.3 Market concentration and heterogeneity

While we assume that firms are *ex ante* identical, we allow for heterogeneity of banks. The key feature distinguishing banks in our model is their level of efficiency. We are agnostic as to exactly what governs bank efficiency— whether it is better screening, a lower cost of financing, a lower monitoring cost, or conversion of deposits into loans with lower overhead costs. We model efficiency simply as a parameter augmenting the variable cost of lending in the spirit of the Monti-Klein model (Freixas and Rochet 2008), and the more recent Corbae and D’Erasmus (2013).

2.3.1 Bank heterogeneity and loan pricing

In order to examine the effects of market concentration, our model must have banks that differ in size. To keep the focus of our analysis on bank size in a straightforward manner, we model banks’ cost efficiency parameter as a random variable. Cost efficiency is a factor that governs banks’ variable cost of lending. We index banks by the letter j , calling the unspecified outcome for the efficiency of any particular bank $A(j)$ and a particular outcome a . More specifically, if a denotes the cost efficiency of a bank, then an increase in a is associated with a decline in costs.

Suppose that there is a large number of banks J , each of which draws its efficiency parameter a , which lies in some positive range $a_0 < a \leq 1$, from a doubly truncated Pareto distribution, $F(a) = \frac{a_0^{-\theta} - a^{-\theta}}{a_0^{-\theta} - 1}$ with $\theta > 0$. We truncate the distribution from above using $a \leq 1$ such that the funding costs for the bank can never be less than the return required by depositors and equity holders. We truncate it from below at a_0 to capture the fact that banks’ funding costs are never infinitely high, implying that efficiency will not fall below some minimum $0 < a_0 < 1$ due to, for instance, some practical operating constraints.

The profit function for any bank j when making a loan to any firm i is thus given by

$$\max_{R(i,j)} \Pi^B(i,j) = (1 - \delta)R(i,j)L(i,j) - \frac{1}{A(j)} \left[r^d D(i,j) + r^e E(i,j) \right], \quad (6)$$

where δ represents the probability with which the bank anticipates default on its loans, r^e is the return that banks must pay to equity shareholders. Deposits $D(i,j)$ and equity $E(i,j)$ are used to finance a loan of the amount $L(i,j)$. We assume that the equity return is equal to the deposit rate augmented by a tax applied to corporate profits, $r^e = r^d(1 + \tau)$, with $\tau > 0$. The bank is also required to meet a regulatory leverage ratio by keeping equity in the amount of a fraction κ of its loans: $E(i,j) = \kappa L(i,j)$. Given the truncation of the distribution of the efficiency parameter a , we have $\frac{1}{A(j)} \geq 1$, such that the unit cost of lending— i.e. the bank’s funding cost multiplied by its non-interest cost $\frac{1}{A(j)}$ — cannot be less than the bank’s funding cost.

Maximizing profit with respect to the interest rate $R(i,j)$ yields the unconstrained optimal interest rate (see Appendix A). This rate would apply if there were no competition from other

banks, where marginal cost equals marginal revenue:

$$R(i, j) = \left(\frac{\mu}{\mu - 1} \right) \frac{r^d(1 + \kappa\tau)}{(1 - \delta)A(j)}. \quad (7)$$

The unconstrained interest rate varies only with respect to the bank's own efficiency parameter: more efficient banks can charge lower interest rates. Note that the cost of funds, or the marginal cost of lending for a bank with efficiency level a , is $C(a) = \frac{r^d(1 + \kappa\tau)}{(1 - \delta)a}$. Thus, Eq.(7) states that in the absence of head-to-head competition with other banks, the bank sets an interest rate with a constant markup, $\frac{\mu}{\mu - 1}$, over marginal cost. However, we show in the next section that when borrowers can search for a lower-cost lender, banks can compete in a Bertrand-like fashion and this unrestricted constant markup will be an upperbound for loan pricing.

2.3.2 How the threat of search constrains loan pricing

Due to perfect substitutability of loans from different banks in the eyes of the borrower, the bank's markup may be constrained because firms search for the best loan offer across different banks. Banks operate under Bertrand-like competition which is modeled in the following way. The market for loans is not completely transparent, i.e., firms must apply for a loan from a specific bank to get an interest rate quote, incurring a fixed application cost of $v > 0$. This cost can be thought of as a search friction: Firms can apply only to one bank at a time and decide after each offer whether to apply to another bank. In other words, applications for loans take place sequentially.⁷ If a second bank offered a better interest rate, the firm would take out its loan from this bank, otherwise it would stick to the first bank. Let $C(a)$ denote the marginal cost of lending for a bank with an efficiency parameter a . If a firm sends out a second application and finds a better bank, with efficiency parameter $a' > a$, this second bank can charge an interest rate equal to $C(a)$ (or trivially below $C(a)$) to undersell the first bank and to win the customer.

Anticipating the firm's potential search for a better interest rate, the first bank to which a firm applies will attempt to set its interest rate on loans low enough to make the firm's expected gain from applying to another bank no higher than the application fee. This ends the firm's search process after just one application. It is the threat of search which constrains the markup for many banks to be less than the level $\frac{\mu}{\mu - 1}$ seen in Eq.(7).

Therefore, the condition determining the pricing behavior of the first bank with efficiency level a is governed by the probability that a firm's next draw will be some level a' greater than its own level a . We know already that the interest rate a bank sets will depend on its efficiency level even in the unconstrained case. Let $R(a)$ thus denote the interest rate charged on loans by the bank

⁷De Blas and Russ (2013) consider an order-statistic framework where firms apply to multiple banks at once, and they characterize the constrained markup in this scenario. We show below that the distribution of bank size is consistent with a power law in the cumulative distribution. Mathematically, the power law distribution cannot be the result of choosing the best bank from more than one application at a time. This is because there is no distribution with a corresponding distribution of first order statistics from samples of $n > 1$ that is power-law in the cumulative distribution (only in the probability density, which is not sufficient for granularity to emerge). Thus, we assume that firms deal with only one bank at a time, which can be construed as a type of relationship lending.

with efficiency level a and $R(a')$ denote the interest rate charged by a bank with efficiency level $a' > a$ that a firm may find if it sends out a second application. We assume that firms are naive with respect to bank efficiency and randomly choose the banks to which they send applications.⁸ The probability that a firm finds a superior bank if it sends out another application is $1 - F(a)$. So a borrower will stop its search for a lender after one application if the additional profit it expects to gain from a lower interest rate is no greater than the application fee:

$$[1 - F(a)] \{ \Pi^F [R(a') | R(a') = C(a)] - \Pi^F [M(a)C(a)] \} = v. \quad (8)$$

If the first bank to which a borrower applies charges an interest rate so high that the borrower could expect to increase its profits (net of the application fee) by searching for another bank, then the borrower will send out another application and the bank risks losing its customer. The bank sets $R(a)$ to avoid this possibility. It wants to make sure that the borrower is just indifferent to taking the loan and submitting an application to another bank. For this to be the case, the application fee must be at least as great as the extra profit a firm expects from submitting one more application, weighted by the probability that it finds a better bank. We call this the “search closure condition”.

Note that banks could conceivably compete in the way that they set the application fee v . However, there are logistical considerations involved in assembling paperwork and negotiating with the bank, such that we take v as an exogenous parameter to enable a sharper focus on competition in interest rate setting. Similarly, we abstract from competition in the quality and range of services, which is more difficult to quantify than net interest margins.

2.3.3 Deriving the constrained markup and interest rate

To derive the markup a bank will charge its customers based on Eq.(8), we first recall the implication from Eq.(5) that the loan volume depends on the interest rate. This interest rate depends on the efficiency level of its lender, while all borrowers are *ex ante* identical. To simplify notation from this point, we index all firm and bank activity by the bank’s efficiency level a . Substituting Eqs.(1), (4), and (5) into the variable profit function in Eq.(3), variable profit for a manufacturer of an intermediate good borrowing from a bank charging interest rate $R(a)$ can be expressed as

$$\Pi^F(a) = \frac{1}{(\mu - 1)\alpha} R(a)^{1-\mu} \left(\frac{\mu}{(\mu - 1)\alpha} \right)^{-\mu} P^\mu Y.$$

We assume that firms take the aggregate variables P and Y as given, as in Di Giovanni et al. (2011). To find the relationship between the search friction and the restricted markup arising due to head-to-head competition, we substitute this profit function (and its counterpart if the interest rate were from a better bank $R(a') < R(a)$) into the search closure condition in Eq.(8). Let $\tilde{M}(a)$

⁸This assumption is not necessary: we need only assume some noisiness in firms’ perceptions of bank efficiency that is dispelled only by applying and getting a rate quote. Assuming some noise would simply augment banks’ market power by a constant term. The assumption merely helps us keep our analysis more transparent and a sharper focus on the role that dispersion in bank costs may play in bank competition.

denote the markup associated with the interest rate that would just satisfy Eq.(8) for a bank with efficiency level a . Then, the search closure condition becomes

$$\tilde{M}(a) = \left(1 - \frac{v}{[1 - F(a)] \Gamma a^{\mu-1}}\right)^{-\frac{1}{\mu-1}}, \quad (9)$$

where $\Gamma = \frac{1}{(\mu-1)\alpha} \left(\frac{\mu}{(\mu-1)\alpha}\right)^{-\mu} P^\mu Y \left(\frac{r^d(1+\kappa\tau)}{(1-\delta)}\right)^{1-\mu}$ is a constant reflecting residual demand and the firm's production technology.⁹ Because μ is greater than one by assumption, the restricted markup is increasing in the application cost v . It is also increasing in bank efficiency a , and all banks with high enough efficiency that $\tilde{M}(a) > \frac{\mu}{\mu-1}$ can charge the unrestricted markup. Thus, the markup is given by

$$M(a) = \min \left\{ \tilde{M}(a), \frac{\mu}{\mu-1} \right\}.$$

The lending rate is then determined by the product of the endogenous markup and the bank's marginal cost, such that $R(a) = M(a)C(a)$. A summary of the equations which determine the steady state of the model can be found in Appendix C.

2.3.4 Zero profit and free entry conditions for banks

A bank cannot stay in business unless it earns positive profit sufficient to cover a fixed overhead cost, implying a minimum markup $\hat{m} > 1$. The constrained markup over the cost of funds in Eq.(9) is increasing in bank efficiency level a . This is because the most efficient banks gain market power from the fact that additional search is less likely to yield a more efficient new bank for a firm. Thus, there is some minimum level of efficiency \hat{a} for which this minimum markup will bind.

Using Eq.(9), this minimum profit condition is given by

$$M(\hat{a}) = \hat{m},$$

which reduces to

$$\frac{v \left(a_0^{-\theta} - 1\right)}{\Gamma(1 - \hat{m}^{-(\mu-1)})} = \hat{a}^{\mu-1} \left(\hat{a}^{-\theta} - 1\right). \quad (10)$$

If we assume that θ is no smaller than $\mu - 1$, then the right-hand side of this condition is strictly decreasing in \hat{a} . Note that this threshold efficiency is decreasing in the difficulty of search v – greater search costs allow less efficient banks to stay in business – and in the size of the market. It is also increasing in the minimum profit margin, \hat{m} : banks with very low efficiency must charge interest rates low enough to keep customers from searching for a new bank but their high lending costs produce net interest margins that are just too low to stay in business.

Up to now, we have studied the banks' pricing decision for a given number of banks, and we

⁹Given that $\mu > 1$, the expression $1 - \frac{v}{[1-F(a)]\Gamma a^{\mu-1}}$ has to be positive in order for the restricted markup to be a real number. Thus, the application fee has to satisfy $v < [1 - F(a)] \Gamma a^{\mu-1}$ and $F(a)$ must be strictly smaller than 1 in our simulations below.

have not modelled the free entry condition. How many banks enter in equilibrium depends on the free entry condition which stipulates that the expected value of entry equals the fixed cost of entry. Banks must pay a fixed cost f to enter the market, which pins down the total number of banks J . These fixed costs are identical across banks. We think of them as being determined by the documentation that is required to comply with regulatory standards before the issuance of a banking license. For instance, banks have to submit the intended organization chart, financial projections, financial information on main potential shareholders, or information about the sources of capital funds (Barth et al. 2001). An entrepreneur considering entering the banking sector draws an efficiency parameter at the beginning of some period t .

We take the steady state value of bank profits, given in Eq.(6), averaged over all possible efficiency levels as the entrepreneur's expected per-period profit. Discounting this by the probability that an adverse shock generates losses that exceed equity yields the free entry condition:

$$\sum_{t=0}^{\infty} (1 - Pr \{E[\Pi(au)] < -\kappa E[L(a)]\}) E[\Pi(a)] = f, \quad (11)$$

where $E[*]$ represents the expectation operator taken over the distribution of a , $F(a)$, while f is the fixed cost of entry and $\kappa E[L(a)]$ is the level of equity held by the bank. More intuitively, a potential lender decides whether to form a bank by calculating the expected stream of profit, discounted by the probability that it might become insolvent. This determines whether profit is sufficiently large to justify the fixed cost of entry. Insolvency in this context occurs when an adverse shock generates losses that exceed equity holdings so that a bank would not be able to satisfy its deposit liabilities. An increase in the capital requirement, κ , lowers the expected stream of profit for the bank, reducing the level of entry, J . Entrants become active only if they have a sufficiently high level of efficiency a to satisfy the zero profit condition (Eq.(10)) in steady state.

2.4 Macroeconomic outcomes

We now turn to an analysis of the link between idiosyncratic bank risk and macroeconomic outcomes. Idiosyncratic bank risk is modelled as a multiplicative, independently, identically, and lognormally distributed shock u to the bank-specific efficiency parameter a . These idiosyncratic shocks affect macroeconomic outcomes through the loan market: Eq.(5) gives the size of the loan to any firm i as a function of the interest rate it receives. Loan demand by any firm fluctuates with the interest rate it pays, and this interest rate varies with banks' idiosyncratic shock. Thus, bank-specific shocks translate into fluctuations in the interest rates that banks charge and into the loans supplied to (and demanded by) individual firms. When summing over these individual loans, idiosyncratic shocks affect also the total supply of loans in the economy as a whole. The impact of a multiplicative shock to any bank's level of efficiency on the aggregate supply of credit depends on the size distribution of banks – and thus on granular effects.

To model these links between the micro- and the macro-level, we will, in the following, use the steady-state aggregate price level as a numeraire, setting $P \equiv 1$. The size of the loan that a bank

makes to any firm depends on its interest rate. This rate, taking into account the shock to bank efficiency, can be expressed as the product between the bank's markup and costs

$$R(au) = M(au)C(au),$$

where the efficiency of a bank with efficiency parameter a is simply au when augmented by the shock, with $u = 1$ in the steady state. Combining the interest rate rule with loan demand in Eq.(5), multiplied by the probability that any firm i applies to a particular bank, $\frac{1}{J}$, we have an expression for bank size,

$$L(au) = [M(au)]^{-\mu} (ua)^\mu \Phi, \tag{12}$$

where $\Phi = \frac{Y}{\alpha J} \left(\frac{\mu r^d (1 + \kappa \tau)}{\alpha (\mu - 1) (1 - \delta)} \right)^{-\mu}$ is a constant reflecting the marginal cost and the effect of search on loan demand common to all banks.

In Appendix B.1, we show that the restricted markup is a slowly varying function.¹⁰ In Appendix B.2, we can thus show that Eq.(12) is a sufficient condition for bank size in terms of loan volume to be power-law distributed with a fat right tail if the dispersion parameter of the bank size distribution, $\zeta = \frac{\theta}{\mu}$, fulfills the condition $\zeta < 2$. Banks absorb part of any shock to efficiency by charging a higher or lower markup. However, the entire shock is not absorbed in the markup so that the shocks to the largest banks still affect their interest rates and will have measurable impacts on macroeconomic outcomes. In the next section, we explain why in more detail.

2.4.1 Does granularity hold?

Granularity implies that shocks to the largest banks end up generating changes in the aggregate supply of credit. For granularity to emerge, two key conditions are necessary.

First, banks must pass on some portion of cost shocks to the interest rates that they charge borrowers. This would not be the case with a strict limit-pricing framework, where banks always set exactly the same interest rate as their competitors (Mandelman (2010)), but it is the case in our model where the interest rate varies with bank efficiency. Interest rates are never strictly bound by those of a known rival, as in a more traditional Bertrand setting with perfect substitutability among loans (De Blas and Russ (2013)).

Second, bank size must be sufficiently disperse. For this, bank size must be power law distributed, exhibiting a fat right tail. In our framework, two potential problems arise, because – as opposed to other studies – (a) markups are endogenous and (b) the Pareto distribution of efficiency is assumed to be doubly truncated to prevent the lending rate from being smaller than the deposit rate.

Endogenous markups arise in our framework, and these markups vary with the efficiency parameter a . Under *constant* markups, the Pareto distribution of efficiency cleanly generates the

¹⁰As laid out by Gabaix (2011), a function $P(X > x) = x^{-\zeta} f(x)$ with $\zeta \in [0; 2]$ and $f(x)$ slowly varying converges in distribution to a Lévy law with exponent ζ . A function is slowly varying if $\lim_{x \rightarrow \infty} f(tx)/f(x) = 1$ for all $t > 0$ (Gabaix 2011, p.766). The applicability of the Lévy Theorem is needed for granular effects to emerge.

necessary power law distribution for size. This is because markups are a slowly varying function. We have shown that the markup $M(a)$ in our framework is a slowly varying function as well (Appendix B.1). So the endogenous markups need not override the effect of the power law on bank size.

Double truncation of bank efficiency prevents the interest rates charged on loans from being smaller than banks' funding cost. No bank has infinitely high lending costs and no bank has lending costs less than the market return on deposits and equity. Even though bank sizes can follow a power law in a model with endogenous markups, the size distribution might not be sufficiently dispersed. This is because our truncated efficiency distribution for banks necessarily has a finite variance, unlike the standard singly truncated Pareto distribution used in Gabaix (2011) and Di Giovanni et al. (2011). In those studies, the singly truncated Pareto distribution of efficiency yields a power law distribution of firm size with infinite variance, such that the Central Limit Theorem gives way to the Lévy theorem. As a consequence, idiosyncratic shocks do not cancel out in the aggregate, and granularity holds.

The applicability of the Lévy theorem is the sufficient condition for granular effects to emerge. However, in our framework with a doubly truncated Pareto distribution of efficiency and hence finite variance, Lévy's Theorem holds only under the following restriction: not only must bank size be power-law distributed, but we must also have the number of applications that a firm sends out be less than $a_0^{-\theta}$ (Sornette (2006, p. 103)). This condition assures sufficient dispersion in bank size, which is needed for idiosyncratic, multiplicative shocks to bank efficiency *not* to average out too quickly as the number of banks J increases. If dispersion is too low, shocks to large banks would make little quantitative difference in macroeconomic outcomes, as would occur under the Central Limit Theorem. Because we allow firms to apply sequentially and thus to only one bank at a time, firms always stop after one application. Otherwise, due to the properties of order statistics, we can not achieve a power-law distribution in bank size. Hence, the number of applications is always less than $a_0^{-\theta}$, so that the second condition for granularity – sufficient dispersion – is always satisfied in our model. Thus, we have the necessary power law property.

In Appendix B.2, we show numerically that granular effects still emerge in our framework with doubly-truncated Pareto efficiency and endogenous markups. More explicitly, because the variance of our distribution of bank size is finite in the face of the double truncation, our numerical simulations show that Sornette's condition for Lévy's Theorem to apply holds in spite of the truncation. Recall that granular effects arise when idiosyncratic shocks to bank lending do not average out quickly as the number of banks increases as would be the case when the Central Limit Theorem holds. To this end, we set the number of banks, J , to 500 and take one draw for each of these banks from the Pareto distribution. We then calculate the markup and corresponding loan demand for each bank given our calibration described in Appendix B.2. We apply idiosyncratic, identically and lognormally distributed shocks (u) to the efficiency parameter of each bank and repeat the process 1000 times. Figure 1 shows the average results across these 1000 simulations: the standard deviation of the aggregate level of bank loans is not zero in response to the idiosyncratic shock.

Thus, the shocks do not average out, even when summing loans over a rather large number of banks.¹¹

Figure 2 further shows that fluctuations in the aggregate credit supply are positively correlated with the level of concentration in the banking industry. The Herfindahl index measures bank concentration – an increasing Herfindahl indicates an increasing market share for the largest banks (the big are getting bigger). The positive relationship between the Herfindahl and macroeconomic outcomes coincides with Gabaix’s theory of granularity, where shocks to the largest firms drive macroeconomic outcomes. Note that the truncation of our distribution from above dampens the relationship between idiosyncratic shocks and macroeconomic outcomes somewhat. Remarkably, however, it can still result in granular effects.

In an economy where the lower bound of the efficiency spectrum, a_0 , is close to one, so that all banks have a similar efficiency level, granular effects would never occur. We consider this to be a more likely situation in the most developed banking sectors, where banks have access to similar technologies. This reduces dispersion from the bottom end of the efficiency spectrum. Similarly, granular effects are unlikely to occur in an economy with no search costs or where the banking market is sufficiently developed such that the number of loan applications n is always large enough (greater than $a^{-\theta}$) that the Central Limit Theorem would hold and Lévy’s Theorem would not apply.

2.5 Linking idiosyncratic shocks with macroeconomic outcomes

We can calculate the change in the loans extended by a bank in response to the efficiency shock u by taking the total derivative of bank size given in Eq.(12) with respect to u :

$$dL(a) = \frac{dL(au)}{du} du + \frac{\partial L(au)}{\partial M(a)} \frac{dM(au)}{du} du \quad (13)$$

$$= \mu \left[1 - \frac{1}{M(au)} \frac{dM(au)}{du} \right] L(au) du. \quad (14)$$

The term in brackets is the effect of the idiosyncratic shock on the bank’s markup due to the first-order effect on the bank’s marginal cost and a second-order effect on the aggregate variable, residual demand (Γ). If we define the steady state as the state where $u = 1$ for all banks and suppose for a moment that markups are constant, there would be full pass-through of a shock relative to the steady state. Eq.(14) also shows that, net of pass-through effects through the endogenous markup, the change in loans supplied by a particular bank relative to the steady state is increasing in the size of both the bank and the shock ($d\tilde{L}(a) = L(a)du$). The growth rate in an individual bank’s

¹¹As suggested by the theory of granularity, the shocks *do* average out (produce zero volatility in aggregate credit) if we allow multiple loan applications or use a heavy-tailed distribution other than the power law, like the Weibull with a dispersion parameter less than one. The fat tail of the power law is essential.

loan relative to the steady state net of pass-through effects is given by

$$\frac{d\tilde{L}(a)}{L(a)} = \frac{L(a)du}{L(a)} = du, \quad (15)$$

with the variance of loans growth given by $\text{var}\left(\frac{d\tilde{L}(a)}{L(a)}\right) = \sigma_u^2$. Because the amount of working capital that any firm uses is equal to the size of the loan it takes out from a bank, Eq.(2) implies a variance in the growth rate of output relative to the steady state for an individual firm borrowing from a bank with efficiency level a equal to $\text{var}\left(\frac{d\tilde{L}(a)}{L(a)}\right) = \sigma_u^2$.

The relationship between the change in lending and in lending costs implies that the variance of the aggregate credit supply is a function of the Herfindahl index. To see this, we assume for simplicity that the shocks u are uncorrelated across banks. We also want to be as conservative as possible in assessing the role of bank size. Then, again using $E[*]$ to represent the expectations operator, the change in the aggregate credit supply L with respect to the steady state, where $u = 1$ for all banks, is given by

$$\frac{\Delta L}{JE[L(a)]} = \sum_{j=1}^J \frac{d\tilde{L}(a)}{JE[L(a)]}.$$

The variance of aggregate credit supply that is due to the first-order effects of idiosyncratic shocks to bank efficiency is then given by the squared terms:

$$\text{var}\left(\frac{\Delta L}{JE[L(a)]}\right) = \sum_{j=1}^J \left(\frac{d\tilde{L}(a)}{JE[L(a)]}\right)^2 = \sigma_u^2 \sum_{j=1}^J \left(\frac{\tilde{L}(a)}{JE[L(a)]}\right)^2 = h\sigma_u^2, \quad (16)$$

where h represents the Herfindahl index of market concentration. By “first-order,” we mean exclusive of any effects on the markup. Thus, consistent with the discussion above regarding search and the behavior of the markup, this expression is an upperbound for the variance of aggregate output arising due to idiosyncratic shocks to bank lending. Because loan fluctuations are equal to fluctuations of the capital stock, fluctuations in aggregate output are monotonically increasing in fluctuations of the aggregate credit supply. Due to constant returns to scale in working capital, the variance in firm output relative to the steady state is given by the same expression, $h\sigma_u^2$, which again we view as an upperbound.

The following Proposition summarizes the determinants of aggregate fluctuations of credit and aggregate output.

Proposition 1 *Fluctuations in aggregate supply of credit and aggregate output are positively related to both the variance of bank-specific shocks and the Herfindahl concentration index in the banking sector.*

Proof. Equation (16) above shows that the aggregate supply of credit is proportional to both the variance of bank-specific shocks and the Herfindahl concentration index in the banking sector. Recall that (1) loan market clearing implies the amount of capital that firms use in production

equals the size of their loan, and (2) firm size depends only on the interest rate, which is the same for all firms borrowing from a particular bank. Therefore, the variance of production for a firm equals the variance in the amount of the loan it procures and the variance of aggregate output must equal the variance of the aggregate supply of credit. ■

We summarize the empirical prediction that follows from our theoretical framework here:

- The higher the level of concentration in the banking sector, the larger are changes in the aggregate supply of credit. This prediction follows from Proposition 1 and Eq.(16) and is confirmed numerically by the simulation results presented in Appendix B.2 linking concentration with macroeconomic outcomes.

3 Empirical Evidence

We bring the implications of our theoretical model to the data by providing evidence on the validity of our assumptions and by testing the empirical predictions of the model. We next describe our data sources, present evidence on the power law decay in bank sizes, and introduce the measurement of granularity in the banking sector. Finally, we present empirical evidence of the link between idiosyncratic shocks to banks and macroeconomic outcomes.

3.1 Data sources

In order to calculate idiosyncratic shocks to the growth of assets or loans of banks as well as the market shares of these banks, we need bank-level data. We take these data from Bureau van Dijk's proprietary *Bankscope* database, which provides income statements and balance sheets for banks worldwide. A number of standard screens are imposed on the banking data in order to eliminate reporting errors: We keep banks with at least five consecutive observations to make sure that they are included at least for one business cycle; we exclude the bottom 1% of observations for total assets in order to eliminate very small and not very representative banks; and we drop implausible observations where the loans-to-assets or the equity-to-assets ratio is larger than 1 as well as banks with negative values recorded for equity, assets, or loans.

We do not have information on bank mergers. In order to eliminate large (absolute) growth rates that might be due to bank mergers, we winsorize growth rates at the top and bottom percentile. We use banks classified as holding companies, commercial banks, cooperative banks, and savings banks, i.e. we exclude a number of specialized banks which are not representative of the banking industry as a whole.

To compute aggregate real growth, we use data on real GDP per capita from the World Bank's *World Development Indicators (WDI)*. These data are available on an annual basis from the 1970s through 2011. Due to missing data for bank-level variables and because we calculate growth rates, our regression sample includes annual data for the years 1995-2009 ($T = 13$) and 83 countries ($N = 83$). A list of countries can be found in Appendix D. Table 1 presents summary statistics.

We focus on two main macroeconomic indicators. Growth in real domestic credit is defined as the growth rate of log real domestic credit in US dollars taken from the International Monetary Fund’s *International Financial Statistics (IFS)*, with real values obtained by deflating nominal values with the US consumer price index. The growth rate of log real GDP per capita is taken from the WDI. All growth rates are winsorized at the top and bottom percentile in order to eliminate the effect of outliers.

3.2 Power law decay in the distribution of bank sizes

To test whether the size distribution within the banking sectors considered here resembles the power law patterns required for granular effects, we use several methods to measure the tail thickness of bank size. Recall that granularity occurs only when the tail exhibits power law properties, implying a Pareto distribution of bank size with a dispersion or shape parameter less than 2. To check whether this is the case, we estimate the parameter using different methods from the literature.

Table 2 presents estimates of power law coefficients for banks’ total assets, distinguishing a panel of all banks appearing between 1997 and 2009 (Table 2a) and a cross section for the year 2009 (Table 2b). For each specification, we show five different estimates of the power law coefficient.

First, we use a maximum likelihood estimator for the shape parameter, ζ , in a truncated Pareto distribution

$$Pr(L(a) > l) = \frac{L_{min}^{\zeta} \left(L(a)^{-\zeta} - L_{max}^{-\zeta} \right)}{1 - (L_{min}/L_{max})^{\zeta}},$$

where $0 < L_{min} \leq L(a) \leq L_{max} < \infty$, such that L_{min} and L_{max} denote the lower and upper truncation of the distribution of bank size, respectively. The results are given in Columns (1)-(4) of Table 2. We use the methodology proposed by Aban et al. (2006) to estimate the dispersion parameter ζ for a doubly truncated Pareto function of banks’ total assets. Column 2 gives the estimation results for the upper tail of the distribution, while Column 3 displays the r largest order statistics on which this estimator is based. We test the fit of the doubly truncated against the standard Pareto distribution. The null hypothesis of “no upper truncation” is rejected for all countries in the full sample (Column 4) meaning that the doubly truncated Pareto function is the better fit for the tail of the bank size distributions.¹²

Figure 3 provides graphical evidence on truncation in the data. It shows plots of log bank size, measured by banks’ total assets, on the log rank of bank size. Bank size observations are ranked in a decreasing order such that $L_{(1)} > L_{(2)} > \dots > L_{(J)}$ determine bank size rank 1 to J . The graphs in log-log-scale illustrate the upper truncation: as is characteristic of a truncated power law, the graphs curve downwards for the largest banks. In case of a standard (singly truncated) Pareto function, the plot of bank size on bank size rank in logarithmic scale would show a straight line.¹³ For our purposes, the presence of the truncation is less important than the dispersion preceding

¹²In the 2009 cross section, where there are fewer observations, it is rejected in the majority of cases, but not all, at the 5 percent level.

¹³Due to the logarithmic scaling of both axes, a function of the form $F(x) = Cx^{-\zeta}$ would give a straight line on a log-log scale with $-\zeta$ being the slope of that line.

it. Estimating $\zeta = \frac{\theta}{\mu} < 2$ demonstrates a distribution of bank size that is sufficiently disperse for granular effects to emerge in our framework (Column 2).

Second, we estimate the power law coefficient without assuming a truncation, such that

$$Pr(L(a) > l) = L_{min}^{\zeta} L(a)^{-\zeta} . \quad (17)$$

Column 5 in Table 2 shows estimation results using the Hill (1975) estimator. This is a maximum likelihood approach based on the average computed distance between the largest r order statistics, with r determined as the sample where the estimates of ζ become stable.

Third, we employ the Stata code PARETOFIT developed by Jenkins and Kerm (2007) which uses a maximum likelihood approach to estimate ζ over the whole sample of bank sizes (Column 6).

Fourth, we estimate the dispersion parameter using the log-rank method proposed by Gabaix and Ibragimov (2011) where the logarithm of $(Rank_j - 0.5)$ of each bank j is regressed on the logarithm of its total assets (Column 7):

$$\ln(Rank_j - 0.5) = \alpha + \zeta \ln L(a) + \varepsilon_j .$$

Fifth, we estimate the power law coefficient using the cumulative distribution function (CDF) method used by Di Giovanni et al. (2011) (Column 8). This method directly uses the logarithm of Eq.(17) to obtain estimates of the dispersion parameter ζ .¹⁴

All estimates are of the same order of magnitude and all are less than 1, with standard errors implying 95 percent confidence intervals below 1, implying power law properties. In our context, granularity requires $\zeta = \frac{\theta}{\mu} < 2$. In other words, demand for firms' output must be sufficiently elastic. Then, the borrowing firms adjust the amount they borrow in response to differences in the interest rates charged by banks with different efficiency levels. If banks are less disperse (high θ), this requires that firms are more sensitive due to more elastic demand for their goods (high μ).

In Figure 4, we graph the fitted estimates without the truncation against the density from the data for the same countries as in Figure 3, with the top 10% of observations omitted to enhance the visibility of the results. The densities coincide quite closely. The estimated parameter is of the same order of magnitude regardless of the method of estimation. Failing to allow for the truncation increases the size of the estimates for ζ , but not enough to compromise the necessary condition for granular effects to emerge.

Note that previous studies (Gabaix 2011, Di Giovanni et al. 2011) focus on power law properties in sales revenues rather than sales quantities. We focus on loan quantities here, as fluctuations in the aggregate credit supply, rather than bank revenues, are our variable of interest. Our estimates also imply granular properties for bank revenues, since they would in our model be characterized by the dispersion parameter $\zeta + 1$, which is less than two in all cases according to our regressions,

¹⁴We are extremely grateful to these authors for kindly sharing their code to ensure exact replication of their methodology. Estimates of the parameter ζ using their p.d.f. method are very similar to the estimates in Columns (5)-(8) and thus are unreported due to space constraints.

since all estimates of ζ are less than one.

3.3 Computing the banking granular residual (BGR)

According to our theoretical model, the transmission from bank-specific shocks to the real economy runs through banks' provision of loans. For an empirical application, we thus need an estimate of bank-specific, idiosyncratic changes in loan growth that are unrelated to macroeconomic conditions.

For this purpose, we need to compute a conditional measures of idiosyncratic loan growth. The main difference between our data and data used in previous papers calculating idiosyncratic growth across firms is that we have relatively short time strings for each bank included in our dataset. At a minimum, banks are in the sample for 5 years, at the maximum for 12 years. This *a priori* limits the use of regression-based empirical models because (bank) fixed effects would be estimated based on very short strings of data. Also, we need to account for the fact that the banks reside in different countries and thus face different macroeconomic environments. We thus employ and adapt Gabaix's method to calculate idiosyncratic bank-level growth rates.

Using long time strings of data for US manufacturing firms, Gabaix (2011) obtains proxies for idiosyncratic growth rates of firms by subtracting the mean growth rate across all firms from each individual firm's growth rates. In a similar vein, we calculate a banking granular residual (BGR) by taking the difference between bank-level loan growth and the mean growth rate of loans for each country and year. We calculate mean growth rates for each country separately to take into account differences in the macroeconomic environments facing the banks. We include each individual bank from this average because, in some countries, the number of banks is rather small. Thus, we take the difference between each bank j 's loan growth and the country-mean of loan growth across all other banks in country i , i.e. except bank j . Results are robust to a version of the BGR including each individual bank in the country mean. The differences between bank-specific and average loan growth per country and year then serve as a simple measure of idiosyncratic, bank-specific growth: $\frac{d\bar{L}(au)}{L(au)} = \frac{L(au)du}{L(au)} = du$, with $var\left(\frac{d\bar{L}(au)}{L(au)}\right) = \sigma_u^2$.¹⁵

Because we want to avoid a somewhat arbitrary choice when classifying large and small banks, we compute the product of idiosyncratic growth and the market share of each bank and then compute the Banking Granular Residual (BGR) for each country i at time t as the sum of these products across all J banks:

$$BGR_{it} = \sum_{j=1}^J \widehat{du}_{jt} \frac{credit_{ijt}}{credit_{it}}.$$

The BGR thus represents the weighted sum over all banks' idiosyncratic credit growth rates, the

¹⁵Idiosyncratic firm-level growth could alternatively be measured using a regression-based approach as in Bloom et al. (2012). Like Gabaix (2011), they also use data for a large sample of US firms with relatively long time series. Using their approach, one would regress log loan growth of an individual bank on its first lag and on bank- and country-year fixed effects. Due to our short panel, the use of bank-specific fixed effects and lagged bank-level loan growth rates presents a nontrivial issue with Nickell (1981) bias that cannot be corrected without a longer panel. Nickell bias directly impacts the residuals from the regression, which would be the measure of the idiosyncratic shock critical to our analysis.

weights being each bank j 's market share in country i . Note that we do not take any stance about whether the size of shocks is linked to the size of banks: large banks may have more or less volatile loan supply than smaller banks.

3.4 Determinants of macroeconomic growth

Our empirical prediction states that aggregate growth fluctuations are a function of market concentration in banking and fluctuations in loan growth by individual banks. Our main interest in this paper is how idiosyncratic shocks affect macroeconomic outcomes. Thus, we regress aggregate growth on our measure for granular loan growth shocks of banks (the BGR), on time fixed effects, and on log GDP per capita and inflation as additional controls. The model is estimated using a panel fixed effects regression with robust standard errors. We use the log growth rate of domestic credit (Table 3a) and log GDP per capita growth (Table 3b) as alternative dependent variables.

In each Table, we show results using the Banking Granular Residual (BGR) calculated for banks' loans. Columns 1-4 show the results for the BGR based on the difference between banks' loan growth and the country-mean of loan growth as in Gabaix (2011).

We proceed in the following steps. We first estimate the baseline model for the full sample (1997-2009). Second, we estimate the model separately for the periods 1997-2006 and 2007-2009 in order to test whether the global financial crisis affects our results. Finally, we add a set of additional regressors that might affect growth in order to filter out macroeconomic effects embodied in the term Γ , which reflects residual demand and firms' technology (see Eq. (9)). These additional regressors include money and quasi money, domestic credit, stock market capitalization as a percentage of GDP, and trade openness from the *WDI* database. Total foreign assets plus liabilities are taken from the *IFS*, while banking sector concentration, measured as the Herfindahl index is computed from the *Bankscope* database. The banking system's z -score as a measure of the risk of the entire banking system comes from the World Bank Financial Structure Database by Beck et al. (2000).¹⁶

Table 3a shows results using growth in log domestic credit as the dependent variable. We find a positive and significant impact of idiosyncratic loan growth in the full sample for the BGR based on differences (Column 1). The crisis does not seem to affect this result: If we exclude the period 2007-2009 from the regression (Column 2), the effect of the BGR on aggregate credit growth is about the same as in the full sample and it remains statistically significant. Moreover, including additional control variables (Column 4) does not alter the positive impact of the BGR. Estimating the model for the crisis years only (2007-2009) renders the BGR insignificant, but does not change the sign of its coefficient.

The beta coefficient for the BGR is 0.11 – i.e., the BGR accounts for about 11% of the variation in aggregate credit growth in our panel¹⁷. The BGR plus time fixed effects and the control variables that are included in all models explain about 30% of the variation in credit growth across countries

¹⁶We use data from the latest update of the Financial Structures Database by Cihak et al. (2012).

¹⁷The beta coefficient is calculated as the coefficient estimate, multiplied with the standard deviation of the explanatory variable, divided by the standard deviation of the dependent variable.

and across time, depending on the model specification. When dropping the time fixed effects and the control variables (unreported), the BGR remains significant, and the R^2 declines to about 3%. Log GDP per capita has a positive effect on aggregate credit growth, whereas higher inflation leads to less credit growth.

Table 3b shows similar results using growth in GDP per capita as the dependent variable. The economic significance of the BGR is similar to the corresponding model for aggregate credit growth: the beta coefficient reveals that the variation in the BGR contributes about 10% to the variation in GDP growth. In the model for GDP growth, the R^2 falls from 32% to 1% when time fixed effects and controls are dropped.

A higher rate of inflation harms GDP growth. Log GDP per capita has a positive and significant effect on growth which might seem counter intuitive at a first glance. However, in unreported regressions where we include initial log GDP per capita instead of country fixed effects, the coefficient on GDP per capita has the expected negative sign. Thus, the negative effect of a high level of GDP per capita on growth is absorbed by the country fixed effects. The positive effect in our regressions with country fixed effects can be interpreted as the growth-enhancing effect of higher institutional quality in countries with a higher level of GDP per capita.

In unreported regressions, we check whether the positive effect of the BGR is robust to changes in the sample composition. We find that dropping countries or years from the regressions, one at a time, does not affect the results. Moreover, if the BGR is computed based on banks' total assets instead of loans (using differences), we find a positive and significant impact on GDP growth. However, idiosyncratic asset growth is not significantly related to changes in aggregate domestic credit.

4 Conclusions

This paper is a step towards exploring the link between concentration in banking, idiosyncratic shocks in the banking sector, and macroeconomic outcomes. We present a baseline framework which abstracts from channels of propagation through asset price effects or through the interbank market or covariance arising from exposure to a common macro shock. We show that, even without these channels of propagation, the presence of large banks by itself can drive fluctuations in the aggregate supply of credit and output. The reason for this is that, if bank sizes are sufficiently dispersed, idiosyncratic shocks to bank loan growth do not cancel out in the aggregate.

Our contribution to the literature is two-fold. First, we generalize the theory of granularity used in studies of manufacturing firms by Gabaix (2011) and Di Giovanni et al. (2011). These studies assume that firms charge a constant markup. We instead develop a model with endogenous markups. In this model, a large number of rival banks compete in a Bertrand-like fashion to provide homogenous loans. Banks are heterogenous with regard to their efficiency. They can charge markups over their cost of funds, subject to an endogenous upper bound on the markup and on market share. The model predicts that macroeconomic outcomes are driven in part by the

banking granular residual – the product of a measure of idiosyncratic fluctuations and the banking system’s Herfindahl index. Granular effects arise if bank sizes are sufficiently dispersed and follow a power law distribution.

Second, in an empirical application using bank-level data, we find support for our assumption that bank size follows a power law distribution. Our results show that a doubly truncated distribution fits the bank size distribution better than the standard singly truncated one, but also that the truncation needn’t preclude granular effects. Finally, we demonstrate that the banking granular residual is associated with aggregate growth in domestic credit and GDP. Hence, idiosyncratic shocks to large banks may affect macroeconomic outcomes via the concentration of banking markets.

Our findings have implications for the regulation of banks. The current regulatory framework lays a strong emphasis on the stability of individual banks by requiring, most importantly, that banks hold a certain minimum level of capital. Because dealing with the distress and insolvency of large banks through market exit is difficult, regulators often rely on consolidation through mergers. Issues related to systemic risk in banking arising through differences in the size of banks are largely ignored. Because fostering mergers between large players is a common policy response to distress in the banking sector,¹⁸ our results fill an important hole in the existing literature on macroprudential policy and bank regulation. We show that policies that may ultimately increase concentration in the banking sector can increase aggregate volatility in macroeconomic outcomes.

Our analysis presents several new avenues for further research on the topic. It points to the importance of analyzing the effects of capital requirements on concentration in credit markets. Increasing capital requirements for banks may be associated with a lower probability of insolvency for individual institutions, but may also lead to increased concentration which, according to our model, could increase the granular effects shown here. It is difficult to assess the net effect of the tradeoff without a detailed analysis of the impact of bank insolvency on the supply of credit and market concentration, as insolvencies are often followed by takeovers of failing banks by larger, healthier ones. Also, the effects of bank mergers on idiosyncratic risk could be explored in more detail. We consider this fertile ground for future research.

¹⁸Liquidation and consolidation of ailing banks, transferring their assets to more robust incumbents, was a common theme in the regulatory response to the recent financial crisis in the U.S. and Europe. As a result, the big got bigger. In the U.S., for instance, the asset portfolios of the largest three surviving banks in 2009—Wells Fargo, J.P. Morgan Chase, and Bank of America—grew by 43 percent, 51 percent, and 138 percent, respectively, after they acquired large, ailing rivals. Their market share also grew by at least a third in both deposits and some types of loans, more than doubling on both fronts for Wells Fargo to exceed 10 percent of the market (Cho 2009).

Table 1: Descriptive Statistics for the Regression Sample (1996-2009)

Variable	Source	Obs	Mean	Std. Dev.	Min	Max
Growth of log real domestic credit	International Monetary Fund, IFS	744	.090	.187	-.660	.628
Growth of log real GDP per capita	World Bank, World Development Indicators	1059	.024	.036	-.149	.124
BGR (loans, differences)	Bankscope, own calculations	1059	.008	.098	-.480	.782
Herfindahl index (loans)	Bankscope, own calculations	1059	24.62	18.87	1.12	100
Log real GDP per capita	World Bank, World Development Indicators	1059	8.11	1.55	4.96	10.64
Inflation (GDP deflator, annual %)	World Bank, World Development Indicators	1059	.078	.317	-.270	9.87
Money and quasi money (M2) as % of GDP	World Bank, World Development Indicators	997	61.54	43.81	7.40	242.23
(Exports + imports) / GDP	World Bank, World Development Indicators	993	.806	.420	.170	2.53
(Total foreign assets + liabilities) / GDP	International Monetary Fund, IFS	1036	2.14	2.60	.390	30.93
Market capitalization of listed companies (% of GDP)	World Bank, World Development Indicators	882	53.25	55.37	.020	510.38
Bank risk (Z-Score)	World Bank, Financial Structures Database	1055	17.77	11.43	-1.99	118.58

Table 2: Estimates of Power Law Coefficients for Total Assets

(a) Full sample (1996-2009)

Country	Truncated MLE			Hill (MLE)	Paretofit (MLE)	log rank	CDF	
	(1) Obs	(2) $\hat{\zeta}$	(3) r					(4) Trunc. p-value
All countries	128,301	0.639	1,000	0.000	0.967	0.478*** (0.00308)	0.910*** (0.00184)	0.660*** (0.0167)
Austria	1,909	0.482	600	0.000	0.600	0.632*** (0.0393)	0.643*** (0.00683)	0.580*** (0.0189)
Switzerland	3,215	0.542	600	0.001	0.591	0.762*** (0.0441)	0.693*** (0.00649)	0.601*** (0.0293)
China	589	0.499	500	0.000	0.582	0.527*** (0.0478)	0.580*** (0.0123)	0.541*** (0.0137)
Denmark	818	0.489	800	0.000	0.527	0.520*** (0.0403)	0.589*** (0.00871)	0.572*** (0.0201)
Spain	801	0.567	400	0.000	0.677	0.379*** (0.0190)	0.574*** (0.0127)	0.627*** (0.0499)
France	2,438	0.579	400	0.000	0.706	0.468*** (0.0171)	0.634*** (0.00861)	0.611*** (0.0250)
United Kingdom	1,230	0.400	500	0.000	0.502	0.391*** (0.0220)	0.474*** (0.00962)	0.432*** (0.0125)
Germany	17,665	0.593	400	0.000	0.707	0.713*** (0.0132)	0.910*** (0.00266)	0.893*** (0.0316)
Italy	3,012	0.606	600	0.000	0.692	0.480*** (0.0147)	0.589*** (0.00563)	0.712*** (0.0482)
Japan	5,573	0.664	800	0.000	0.786	0.517*** (0.0175)	0.639*** (0.00425)	0.639*** (0.0176)
Norway	599	0.594	300	0.011	0.675	0.681*** (0.0624)	0.668*** (0.00970)	0.678*** (0.0141)
United States	87,586	0.800	1,000	0.000	0.960	0.617*** (0.00555)	0.748*** (0.000708)	0.734*** (0.0126)

(b) Cross section (2009)

Country	(1) Obs	(2) $\hat{\zeta}$	(3) r	(4) Trunc. p-value	Hill (MLE) $\hat{\zeta}$	Paretofit (MLE) $\hat{\zeta}$	log rank $\hat{\zeta}$	CDF $\hat{\zeta}$
All countries	10,541	0.582	1,000	0.000	0.658	0.453*** (0.00308)	0.900*** (0.00184)	0.601*** (0.0129)
Austria	174	0.361	90	0.03	0.546	0.496*** (0.0393)	0.562*** (0.00683)	0.517*** (0.0172)
Switzerland	294	0.373	50	0.33	0.535	0.718*** (0.0441)	0.694*** (0.00649)	0.599*** (0.0334)
China	86	0.321	80	0.04	0.493	0.429*** (0.0478)	0.545*** (0.0123)	0.481*** (0.0142)
Denmark	80	0.380	65	0.24	0.502	0.408*** (0.0403)	0.548*** (0.00871)	0.515*** (0.0193)
Spain	114	0.379	50	0.18	0.796	0.352*** (0.0190)	0.525*** (0.0127)	0.580*** (0.0515)
France	196	0.441	126	0.05	0.577	0.358*** (0.0171)	0.524*** (0.00861)	0.484*** (0.0270)
United Kingdom	105	0.138	60	0.01	0.376	0.328*** (0.0220)	0.382*** (0.00962)	0.308*** (0.00966)
Germany	1,204	0.746	900	0.08	0.769	0.613*** (0.0132)	0.824*** (0.00266)	0.815*** (0.0260)
Italy	447	0.544	300	0.05	0.597	0.451*** (0.0147)	0.598*** (0.00563)	0.681*** (0.0418)
Japan	476	0.542	300	0.009	0.619	0.503*** (0.0175)	0.631*** (0.00425)	0.659*** (0.0281)
Norway	102	0.506	60	0.43	0.669	0.639*** (0.0624)	0.670*** (0.00970)	0.671*** (0.0226)
United States	7,064	0.624	400	0.009	691	0.617*** (0.00555)	0.748*** (0.000708)	0.738*** (0.0128)

This table provides estimates for the power law shape parameter $\zeta > 0$ of the cumulative distribution function of banks' total assets (in 1.000 USD). r is the number of order statistics used to estimate the shape parameter of the truncated Pareto distribution. p -value is the p -value of the Pareto test (H_0 : no truncation). CDF estimates are from the log of the cumulative distribution function, as described in Di Giovanni et al. (2011). Estimates are performed excluding the bottom quartile of observations, using robust standard errors where applicable, and clustering observations at the bank-level. In the log-rank estimations over the full sample (column (7), Table 2a), time fixed-effects are included. Robust standard errors are given in brackets, and *, **, *** indicates significance at the 1%, 5%, and 10% level.

Table 3: Determinants of Aggregate Growth Fluctuations

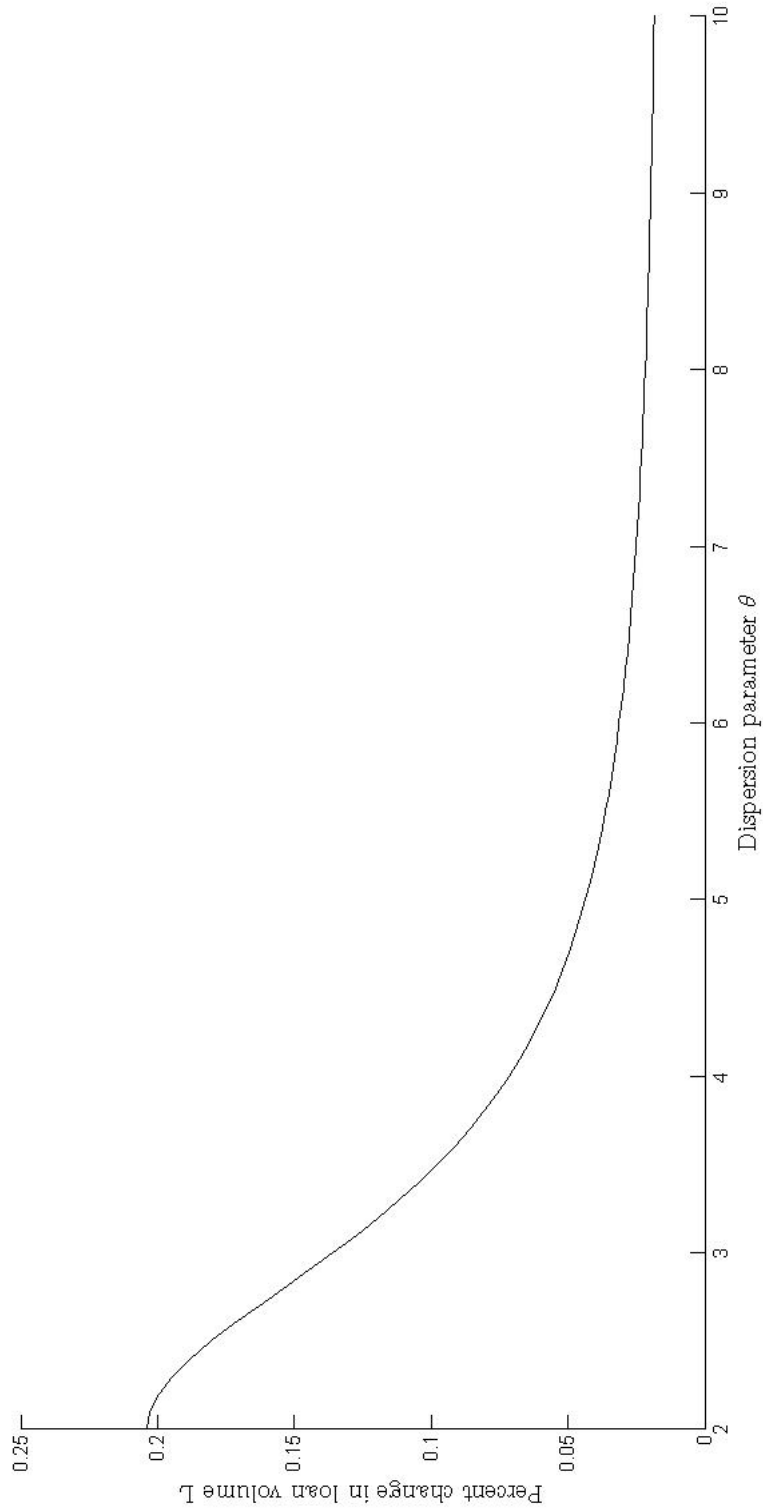
	(a) Growth in log domestic credit			
	(1)	(2)	(3)	(4)
	Full sample	1997-2006	2007-2009	Full sample
BGR (loans, differences)	0.212** (0.082)	0.197** (0.086)	0.121 (0.218)	0.198** (0.093)
Log GDP per capita	0.293** (0.123)	0.462** (0.183)	0.668** (0.333)	0.280** (0.133)
Inflation, GDP deflator (annual %)	-0.094** (0.009)	-0.092** (0.010)	0.842* (0.473)	-0.101** (0.010)
Money and quasi money (M2) as % of GDP				-0.001 (0.001)
(Exports + imports) / GDP				0.012 (0.071)
Domestic credit / GDP (%)				-0.093 (0.089)
HHI (loans)				0.001* (0.001)
(Foreign assets + liabilities) / GDP				-0.000 (0.006)
Market capitalization of listed companies (% of GDP)				0.000 (0.000)
Bank risk (Z-Score)				0.000 (0.001)
Constant	-2.453** (1.051)	-3.901** (1.564)	-5.876* (2.948)	-2.271** (1.130)
Observations	744	602	142	630
R-squared	0.301	0.269	0.679	0.357
Number of countries	62	62	61	60

(b) Growth in log GDP

	(1) Full sample	(2) 1997-2006	(3) 2007-2009	(4) Full sample
BGR (loans, differences)	0.037*** (0.010)	0.024** (0.010)	0.014 (0.024)	0.022* (0.012)
Log GDP per capita	0.072*** (0.014)	0.093*** (0.016)	0.550*** (0.055)	0.073*** (0.020)
Inflation, GDP deflator (annual %)	-0.008*** (0.002)	-0.009*** (0.003)	0.116**	-0.007*** (0.001)
Money and quasi money (M2) as % of GDP				0.000 (0.000)
(Exports + imports) / GDP				-0.017 (0.013)
Domestic credit / GDP (%)				-0.062*** (0.020)
HHI (loans)				-0.000 (0.000)
(Foreign assets + liabilities) / GDP				-0.004* (0.002)
Market capitalization of listed companies (% of GDP)				0.000** (0.000)
Bank risk (Z-Score)				-0.000 (0.000)
Constant	-0.617*** (0.116)	-0.715*** (0.130)	-4.563*** (0.458)	-0.580*** (0.163)
Observations	1,059	815	244	773
R-squared	0.317	0.210	0.799	0.471
Number of countries	83	83	83	72

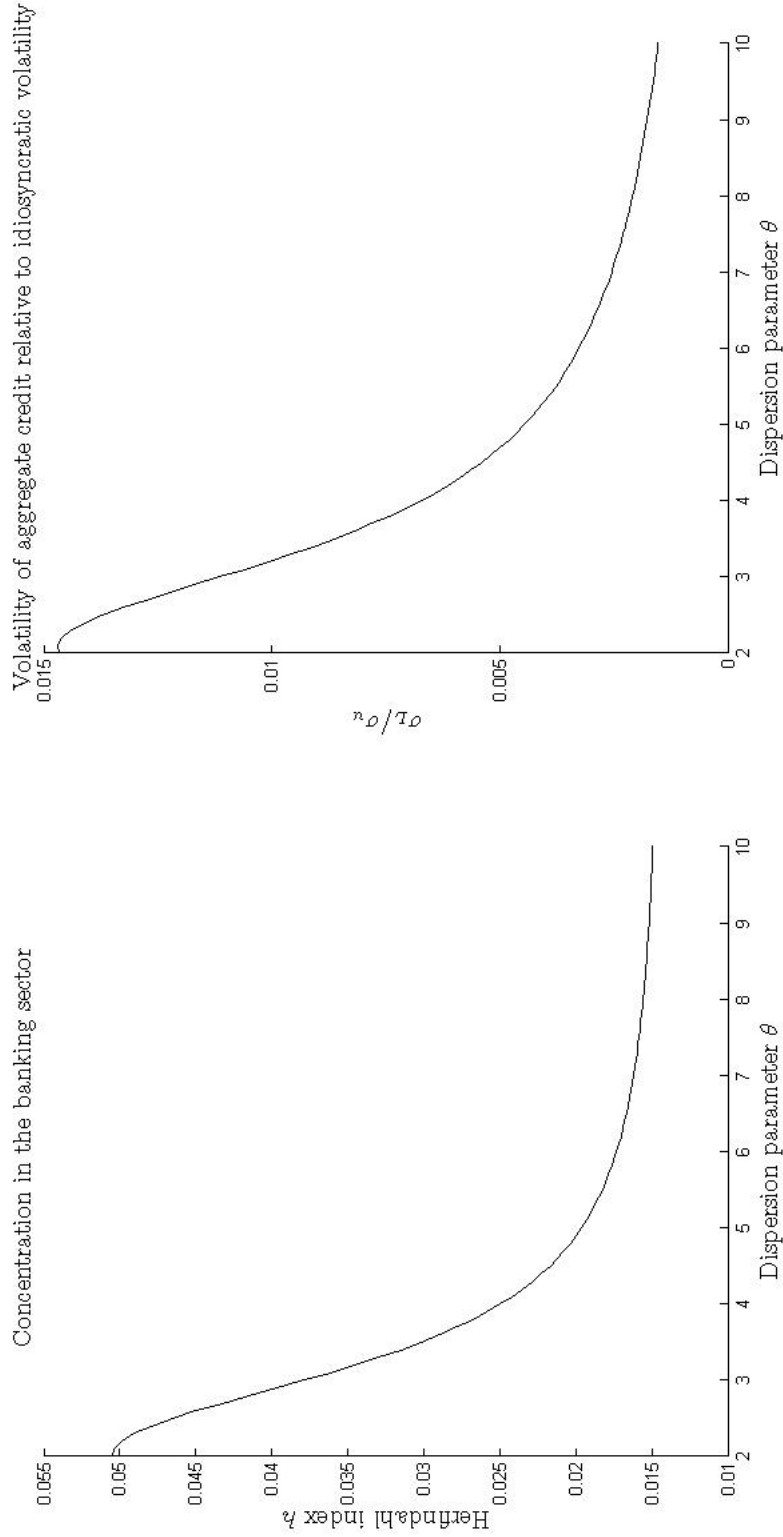
This table reports regressions using the growth of banks' aggregate loans in (a) and GDP per capita growth in (b) as the dependent variables. The sample period is 1996-2009. A set of year dummies and country fixed effects is included in each regression (not reported). BGR is the Banking Granular Residual which is computed using the *differences* between banks' loan growth and the country-mean of loan growth, as described in the main body of the text. Robust standard errors are given in brackets, and *, **, *** indicates significance at the 1%, 5%, and 10% level.

Figure 1: Deviations of aggregate credit from steady state



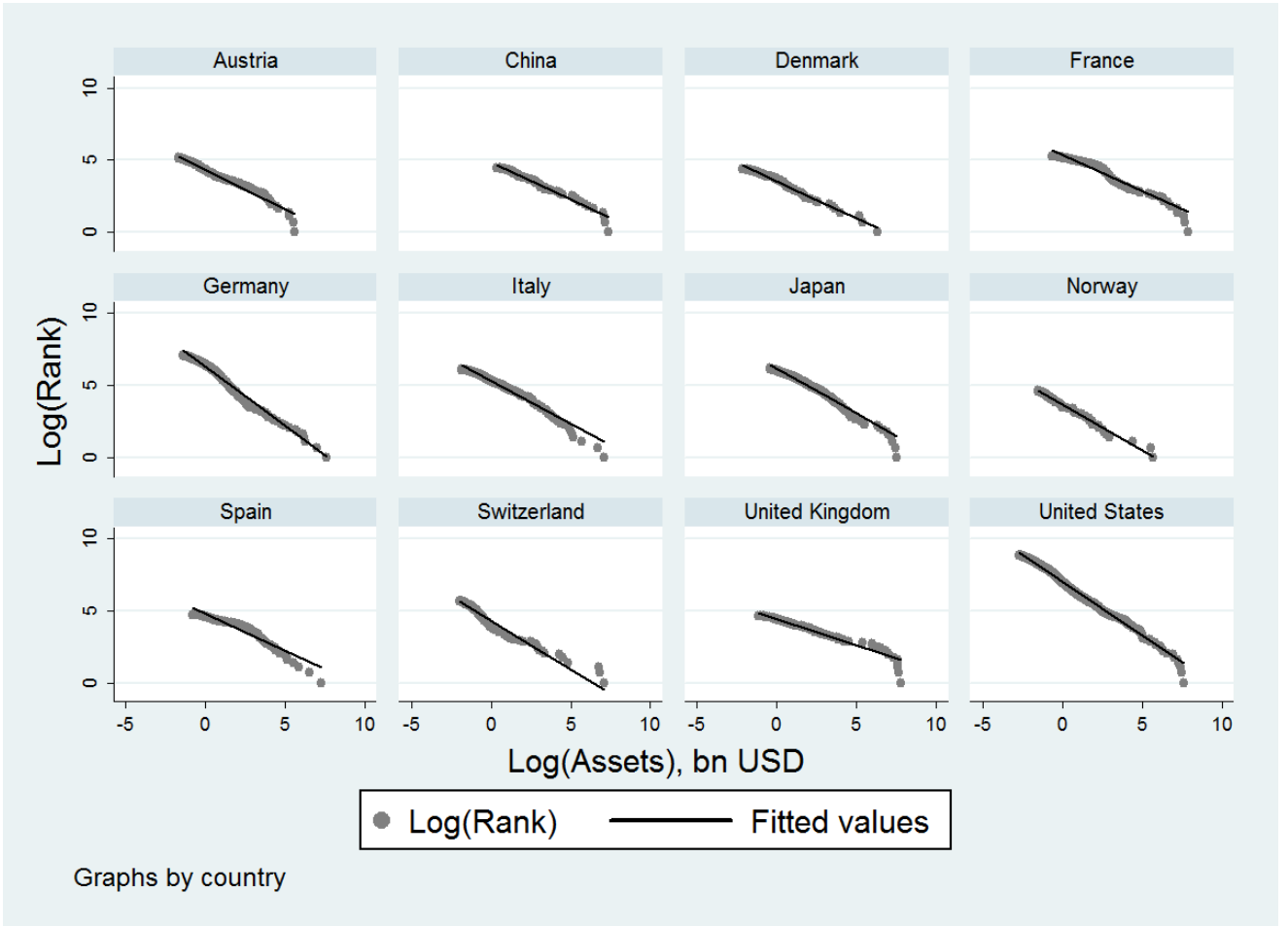
This figure displays the dispersion parameter of banks' efficiency distribution, θ , on the horizontal axis. As θ increases, the dispersion of efficiency parameters falls—the tails of the efficiency-distribution get thinner. On the vertical axis, the Figure shows the deviations of aggregate credit from the steady state after banks each receive an idiosyncratic shock u .

Figure 2: Banking Sector Concentration and Aggregate Volatility



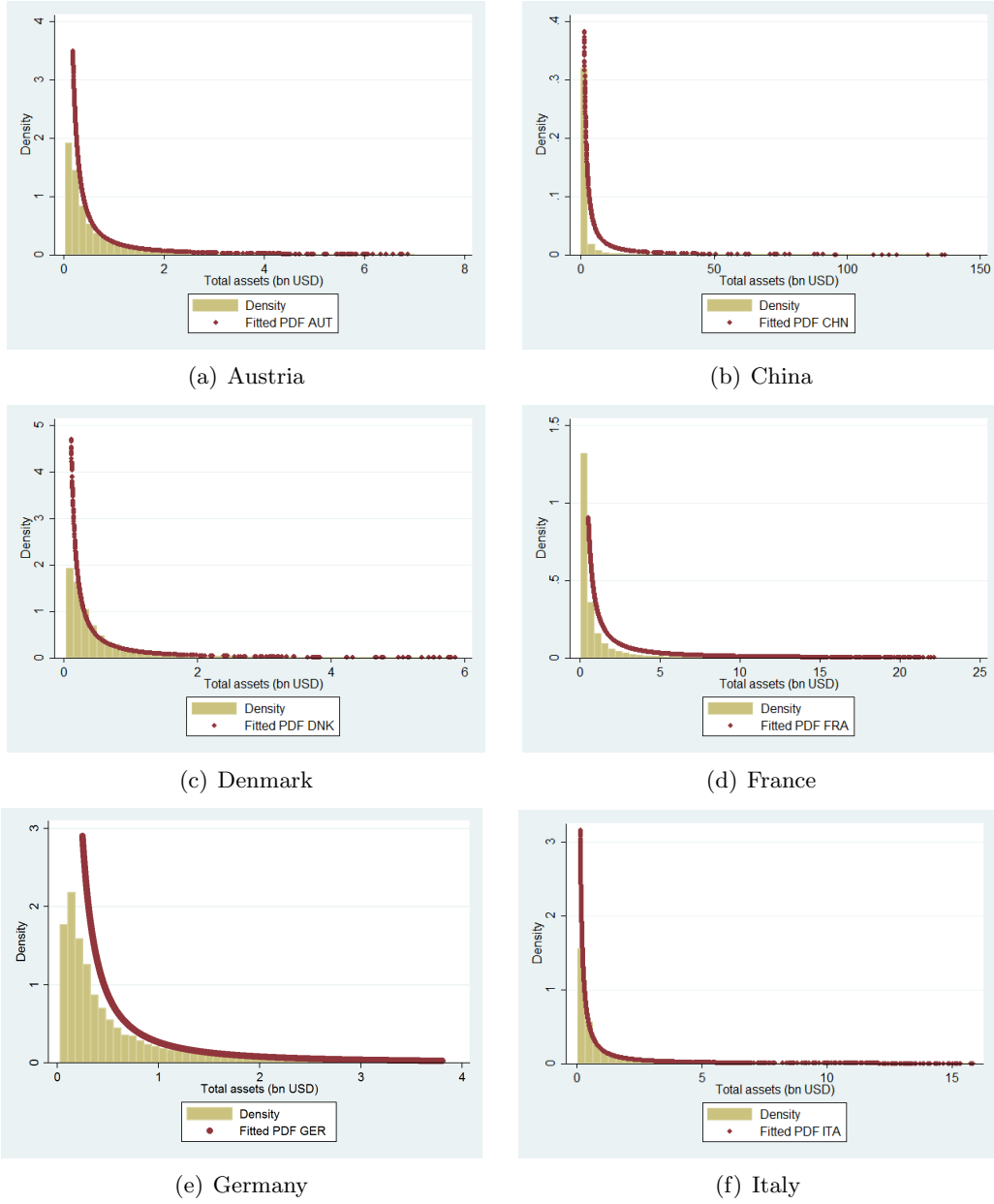
On the horizontal axes, this figure plots the dispersion parameter of the distribution of banks' efficiency parameters, θ . As θ increases, the dispersion of efficiency parameters falls. This means that the tails of the efficiency-distribution get thinner. On the vertical axes, the Figure shows (1) concentration in the banking sector, measured by the square root of the Herfindahl index, and (2) the variance of aggregate credit relative to the variance of the idiosyncratic shock.

Figure 3: Distribution of Ranks and Bank Size by Country



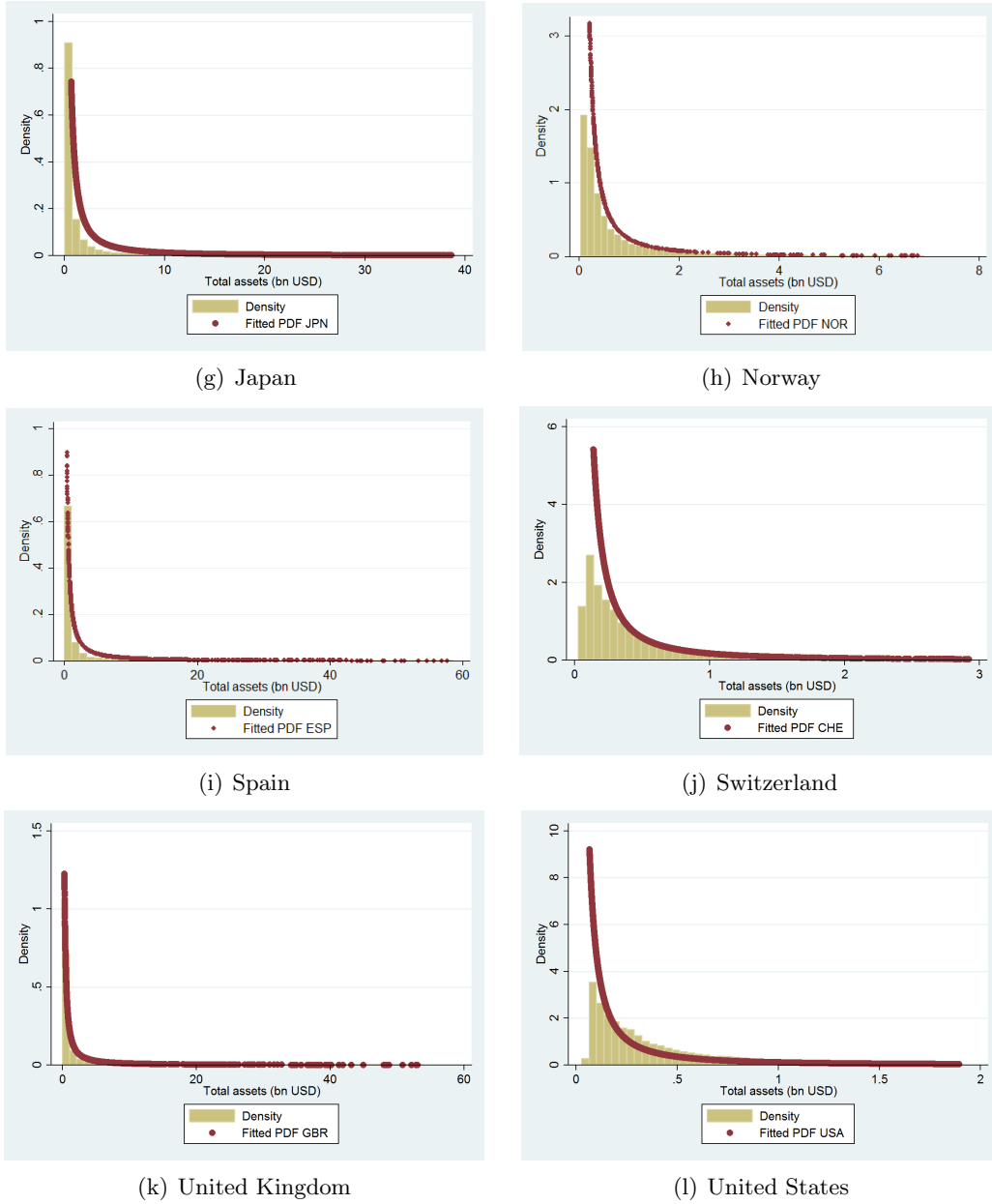
This figure displays the distribution of bank size (log rank of bank size vs. the log of bank size), measured by total assets (in bn USD). Data are for the year 2009, countries with less than 80 banks are excluded.

Figure 4: Distribution of Bank Size by Country



This figure displays the distribution of bank size (A = total assets in billion USD) against the density. The dark line is the estimated Pareto p.d.f value $f(A)$ for each A . Estimates are performed excluding the bottom quartile of observations, using robust standard errors, and clustering observations at the bank-level. In order to enhance visibility, the top 10% of banks in terms of size are not plotted but are included in the estimates of the probability density function.

Figure 4 (continued)



This figure displays the distribution of bank size (A = total assets in billion USD) against the density. The dark line is the estimated Pareto p.d.f value $f(A)$ for each A . Estimates are performed excluding the bottom quartile of observations, using robust standard errors, and clustering observations at the bank-level. In order to enhance visibility, the top 10% of banks in terms of size are not plotted but are included in the estimates of the probability density function.

References

- ABAN, I., M. MEERSCHAERT, AND A. PANORSKA (2006): “Parameter estimation for the truncated Pareto distribution,” *Journal of the American Statistical Association*, 101, 270–277.
- ADRIAN, T. AND M. BRUNNERMEIER (2009): “CoVaR,” *Federal Reserve Bank of New York Staff Report*, 348.
- ANDERSON, S., A. DE PALMA, AND J. THISSE (1987): “The CES is a discrete choice model?” *Economics Letters*, 24, 139–140.
- ANDRES, J. AND O. ARCE (2012): “Banking Competition, Housing Prices and Macroeconomic Stability,” *The Economic Journal*, Forthcoming.
- ANGELONI, I. AND E. FAIA (2009): “A tale of two policies: Prudential regulation and monetary policy with fragile banks,” *Kiel Working Papers*, 1569.
- ASHCRAFT, A., J. MCANDREWS, AND D. SKEIE (2011): “Precautionary reserves and the interbank market,” *Journal of money, Credit, and Banking*, 43, 311–348.
- BARTH, J., G. CAPRIO, AND R. LEVINE (2001): “The regulation and supervision of banks around the world: A new database,” *World Bank Policy Research Working Paper*, 2588.
- BECK, T. AND A. DEMIRGÜÇ-KUNT (2009): “Financial Institutions and Markets Across Countries and over Time: Data and Analysis,” *World Bank Policy Research Working Paper*, 4943.
- BECK, T., A. DEMIRGÜÇ-KUNT, AND R. LEVINE (2000): “A New Database on Financial Development and Structure,” *World Bank Economic Review*, 14, 597–605.
- BERGER, A. N., R. S. DEMSETZ, AND P. E. STRAHAN (1999): “The consolidation of the financial services industry: Causes, consequences, and implications for the future,” *Journal of Banking & Finance*, 23, 135–194.
- BLANK, S., C. BUCH, AND K. NEUGEBAUER (2009): “Shocks at Large Banks and Banking Sector Distress: The Banking Granular Residual,” *Journal of Financial Stability*, 5, 353–373.
- BLOOM, N., M. FLOETOTTO, N. JAIMOVICH, I. SAPORTA-EKSTEN, AND S. TERRY (2012): “Really uncertain business cycles,” *NBER Working Paper No. 18245*.
- BOYD, J. AND M. GERTLER (1993): “US commercial banking: Trends, cycles, and policy,” in *NBER Macroeconomics Annual 1993, Volume 8*, MIT Press, 319–377.
- BRODA, C. AND D. E. WEINSTEIN (2006): “Globalization and the Gains from Variety,” *The Quarterly Journal of Economics*, 121, 541–585.
- BUCH, C. AND K. NEUGEBAUER (2011): “Bank-specific Shocks and the Real Economy,” *Journal of Banking & Finance*, 35, 2179–2187.

- CALDERON, C. AND K. SCHAECK (2012): “Bank bailouts, competitive distortions, and consumer welfare,” *mimeo*.
- CHO, D. (2009): “Banks Too Big to Fail Have Grown Even Bigger: Behemoths Born out of Failure Reduce Consumer Choice, Tempt Moral Hazard,” *Washington Post*, 6175, A01.
- CIHAK, M., A. DEMIRGÜÇ-KUNT, E. FEYEN, AND R. LEVINE (2012): “Benchmarking Financial Development Around the World,” *World Bank Policy Research Working Paper*, 6175.
- CORBAE, D. AND P. D’ERASMO (2013): “A Quantitative Model of Banking Industry Dynamics,” *Working Paper, University of Wisconsin*.
- CORVOISIER, S. AND R. GROPP (2002): “Bank concentration and retail interest rates,” *Journal of Banking & Finance*, 26, 2155–2189.
- DAVIS, K. (2007): “Banking Concentration, Financial Stability and Public Policy,” in *The Structure and Resilience of the Financial System*, ed. by C. Kent and J. Lawson.
- DE BANDT, O. AND E. . DAVIS (1999): “A cross-country comparison of market structures in European banking.” *ECB Working Paper 7*.
- DE BLAS, B. AND K. RUSS (2010): “Understanding Markups in the Open Economy under Bertrand Competition,” *NBER Working Paper*, 16587.
- (2013): “All Banks Great, Small, and Global: Loan Pricing and Foreign Competition,” *International Review of Economics & Finance*, 26.
- DE NICOLO, G., P. BARTHOLOMEW, J. ZAMAN, AND M. ZEPHIRIN (2004): “Bank consolidation, internationalization, and conglomeration: trends and implications for financial risk,” *Financial markets, institutions & instruments*, 13, 173–217.
- DEANGELO, H., L. DEANGELO, AND R. STULZ (2006): “Dividend policy and the earned/contributed capital mix: a test of the life-cycle theory,” *Journal of Financial Economics*, 81, 227–254.
- DENIS, D. AND I. OSOBOV (2008): “Why do firms pay dividends? International evidence on the determinants of dividend policy,” *Journal of Financial Economics*, 89, 62–82.
- DI GIOVANNI, J. AND A. LEVCHENKO (2009): “Trade openness and volatility,” *The Review of Economics and Statistics*, 91, 558–585.
- DI GIOVANNI, J., A. LEVCHENKO, AND R. RANCIERE (2011): “Power laws in firm size and openness to trade: Measurement and implications,” *Journal of International Economics*, 85, 42–52.
- DIXIT, A. AND J. STIGLITZ (1977): “Monopolistic competition and optimum product diversity,” *The American Economic Review*, 297–308.

- FIDORA, M., M. FRATZSCHER, AND C. THIMANN (2007): “Home bias in global bond and equity markets: the role of real exchange rate volatility,” *Journal of International Money and Finance*, 26, 631–655.
- FREIXAS, X. AND J. ROCHET (2008): *Microeconomics of Banking*, MIT Press Cambridge, MA and London, 2 ed.
- GABAIX, X. (2011): “The Granular Origins of Aggregate Fluctuations,” *Econometrica*, 79, 733–772.
- GABAIX, X. AND R. IBRAGIMOV (2011): “Rank-1/2: a simple way to improve the OLS estimation of tail exponents,” *Journal of Business & Economic Statistics*, 29, 24–39.
- GERALI, A., S. NERI, L. SESSA, AND F. SIGNORETTI (2010): “Credit and Banking in a DSGE Model of the Euro Area,” *Journal of Money, Credit and Banking*, 42, 107–141.
- GHIRONI, F. AND V. STEBUNOV (2010): “The Domestic and International Effects of Interstate U.S. Banking,” *NBER Working Paper No. 16613*.
- GROUP OF TEN (2001): “Report on Consolidation in the Financial Sector,” Tech. rep., Bank for International Settlements.
- HALE, G. (2012): “Bank relationships, business cycles, and financial crises,” *Journal of International Economics*, 88, 312 – 325.
- HAWKINS, J. AND D. MIHALJEK (2001): “The banking industry in the emerging market economies: competition, consolidation and systemic stability: an overview,” *BIS Papers*, 1–44.
- HILL, B. (1975): “A simple general approach to inference about the tail of a distribution,” *The Annals of Statistics*, 3, 1163–1174.
- JENKINS, S. P. AND P. V. KERM (2007): “PARETOFIT: Stata module to fit a Type 1 Pareto distribution,” Available online at <http://ideas.repec.org/c/boc/bocode/s456832.html>.
- KALEMLI-OZCAN, S., E. PAPAIOANNOU, AND F. PERRI (2012): “Global Banks and Crisis Transmission,” *NBER Working Paper No. 18209*.
- KASHYAP, A., R. RAJAN, AND J. STEIN (2002): “Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking,” *The Journal of Finance*, 57, 33–73.
- MANDELMAN, F. (2010): “Business cycles and monetary regimes in emerging economies: A role for a monopolistic banking sector,” *Journal of International Economics*, 81, 122–138.
- MEH, C. AND K. MORAN (2010): “The role of bank capital in the propagation of shocks,” *Journal of Economic Dynamics and Control*, 34, 555–576.
- NICKELL, S. (1981): “Biases in Dynamic Models with Fixed Effects,” *Econometrica*, 49, 1417–1426.

- OBSTFELD, M. AND K. ROGOFF (1995): “Exchange rate dynamics redux,” *Journal of Political Economy*, 102, 624–660.
- OECD (2010): “Competition, Concentration and Stability in the Banking Sector,” Tech. rep., Policy Roundtables.
- SCHARGRODSKY, E. AND F. STURZENEGGER (2000): “Banking regulation and competition with product differentiation,” *Journal of Development Economics*, 63, 85–111.
- SCHOENMAKER, D. AND T. BOSCH (2008): “Is the home bias in equities and bonds declining in Europe?” *Investment Management and Financial Innovations*, 5, 90–102.
- SORNETTE, D. (2006): *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools (Springer Series in Synergetics)*, Springer.
- TARASHEV, N., C. BORIO, AND K. TSATSARONIS (2009): “The systemic importance of financial institutions,” *BIS Quarterly Review*, 75, 1–87.
- (2010): “Attributing systemic risk to individual institutions,” *BIS Working Paper*, 308.
- VAN WINCOOP, E. (2011): “International Contagion Through Leveraged Financial Institutions,” *NBER Working Paper No. 17686*.
- WALKNER, C. AND J.-P. RAES (2005): “Integration and consolidation in EU banking - an unfinished business,” *European Commission Economic Papers Number 226*.
- ZHANG, L. (2009): “Bank capital regulation, the lending channel and business cycles,” *Deutsche Bundesbank Discussion Paper Series 1: Economic Studies*, 33.

A Maximization Problems

A.1 Consumers

The first order conditions of the consumer's maximization problem, subject to her budget constraint, are as follows:

$$\frac{\partial \mathcal{L}}{\partial Q_t} : \frac{1}{Q_t} - \lambda_t P_t \equiv 0 \quad (18)$$

$$\frac{\partial \mathcal{L}}{\partial Y_t} : zY_t - \lambda_t P_t \equiv 0 \quad (19)$$

$$\frac{\partial \mathcal{L}}{\partial D_t} : \lambda_t - \beta \lambda_{t+1} (1 + r_t) \equiv 0 \quad (20)$$

$$(21)$$

Solving Eq. (18) for the Lagrange multiplier t and substituting into (19) we can derive consumption as a function of aggregate output,

$$Q_t = \frac{1}{zY_t} \quad (22)$$

Using the expression for t implied by Eq. (18) and substituting it into Eq. (20), one obtains the Euler equation

$$\frac{1 - zC_t}{P_t} = \beta(1 + r_t) \frac{1 - zC_{t+1}}{P_{t+1}} \quad (23)$$

From Eq. (23), we can solve for the steady state interest rate (where $C_t = C_{t+1} = C$ and $P_t = P_{t+1} = P$, and $r_t = r^d$ for all periods t),

$$r^d = \frac{1 - \beta}{\beta} \quad (24)$$

A.2 Firms

Firms maximize profits given by $\Pi^F(i) = P(i)Y(i) - R(i)K(i)$. Demand for the firm's good is downward sloping in its price. Taking the derivative of the demand for any intermediate good with respect to its price yields

$$\frac{\partial Y(i)}{\partial P(i)} = -\mu P(i)^{-\mu-1} P^\mu Y = \frac{Y(i)}{P(i)}.$$

Noting from the firm's technology that $K(i) = (1/\alpha)Y(i)$, the first derivative of the firm profit function with respect to price is given by

$$\frac{\partial \Pi^F(i)}{\partial P(i)} = Y(i) + P(i) \frac{\partial Y(i)}{\partial P(i)} - \frac{R(i)}{\alpha} \frac{\partial Y(i)}{\partial P(i)} \equiv 0.$$

Then, substituting $\partial Y(i)/\partial P(i)$ into the derivative of the profit function yields the pricing equation given in Eq.(7) of the main text.

A.3 Banks

The profit earned by bank j when serving firm i as given in the main text is

$$\Pi^B(i, j) = (1 - \delta)R(i, j)L(i, j) - \frac{1}{A(j)} \left[r^d D(i, j) + r^e E(i, j) \right].$$

We note that a fraction κ of the loan must be financed through equity, allowing the remainder to be financed through deposits. Thus, the profit function can be rewritten as

$$\begin{aligned} \Pi^B(i, j) &= (1 - \delta)R(i, j)L(i, j) - \frac{r^d}{A(j)} [(1 - \kappa) + (1 + \tau)\kappa] L(i, j) \\ &= (1 - \delta)R(i, j)L(i, j) - \frac{r^d(1 + \kappa\tau)}{A(j)} L(i, j). \end{aligned}$$

We note that the external financing assumption and loan market clearing implies $L(i, j) = K(i, j)$, so that $\partial L(i, j)/\partial R(i, j) = -\mu L(i, j)/R(i, j)$. The first-order condition with respect to $R(i, j)$ is then

$$\begin{aligned} \frac{\partial \Pi^B(i, j)}{\partial R(i, j)} &= (1 - \delta) \left(L(i, j) + R(i, j) \frac{\partial L(i, j)}{\partial R(i, j)} \right) - \frac{r^d(1 + \kappa\tau)}{A(j)} \frac{\partial L(i, j)}{\partial R(i, j)} \\ &= [(1 - \delta) - \mu(1 - \delta)] L(i, j) + \frac{\mu r^d(1 + \kappa\tau)}{A(j)} \frac{L(i, j)}{R(i, j)} \equiv 0. \end{aligned}$$

Cancelling $L(i, j)$ in both terms and rearranging yields the unconstrained interest rate rule in the main text.

B Applicability of the Lévy Theorem

B.1 The restricted markup function is slowly varying

$\tilde{M}(a)$ is slowly varying as long as, for any constant t greater than zero,

$$\frac{\lim_{a \rightarrow \infty} \tilde{M}(at)}{\tilde{M}(a)} = \lim_{a \rightarrow \infty} \frac{\left(1 - \frac{v}{[1 - F(at)](at)^{\mu-1}\Gamma} \right)^{\frac{-1}{\mu-1}}}{\left(1 - \frac{v}{[1 - F(a)](a)^{\mu-1}\Gamma} \right)^{\frac{-1}{\mu-1}}} \equiv 1,$$

which is true.

B.2 Bank size is power law distributed

Using Eq.(12) in the main text, the probability that an individual bank's supply of credit is greater than some positive constant l in steady state is given by

$$\begin{aligned} Pr(M(a)^{-\mu} a^\mu \Phi > l) &= Pr\left(M(a)^{-\mu} a^\mu > \frac{l}{\Phi}\right) = Pr\left(M(a)^{-1} a > \left[\frac{l}{\Phi}\right]^{\frac{1}{\mu}}\right) \\ &= l^{\frac{\theta}{\mu}} \Phi^{-\frac{\theta}{\mu}} \psi(l). \end{aligned}$$

If the bank has a large enough efficiency parameter a such that it can charge the unrestricted markup $\frac{\mu}{\mu-1}$, then $\psi(l) = M(a) = \frac{\mu}{\mu-1}$, a constant which is clearly a slowly varying function and thus bank size follows a power law. Because $\lim_{a \rightarrow 1} \tilde{M}(a) = \infty$, then there must be some $a < 1$ above which all banks charge the (constant) unrestricted markup and the far-right tail is power-law distributed up to the right truncation at $a = 1$. The dispersion parameter of the bank size distribution, i.e. the exponent of loan volume l , is thus given by $\zeta = \frac{\theta}{\mu}$. If $\zeta < 2$, the bank size distribution follows a fat-tailed power law.

We simulate the model to demonstrate that granular effects emerge in spite of the right truncation. We set the elasticity of substitution between goods, μ , equal to 3, close to the median estimate in Broda and Weinstein (2006), then simulate data for different values of the dispersion parameter of the efficiency distribution, θ , such that $\mu - 1 < \theta < 2\mu$.¹⁹ For each value of θ , we draw an efficiency parameter a for each of the J banks which are hit by a log-normally distributed shock u with mean one and a standard deviation of one percent. We repeat this procedure 1000 times and average across repetitions. We must discretize the number of banks and choose the number $J = 500$, with 5000 firms sending applications to a randomly chosen bank. In addition, we set $\beta = 0.96$ and $a_0 = 0.1$, and, as scaling factors, $\alpha = 0.36$, $z = 0.01$ and $Y = 10$. The results of the simulation are in Figures 1 and 2.

C Steady State

Representative consumer:

$$\begin{aligned} \text{Aggregate demand} \quad Q &= \frac{1}{zY} \\ \text{Euler equation} \quad r^d &= (1 - \beta)\beta \end{aligned}$$

¹⁹This ensures that $\frac{\theta}{\mu} < 2$, the condition identified by Gabaix (2011) for granular effects to arise.

Firms:

$$\begin{aligned}
\text{Loan demand} \quad L(i, j) &= \frac{1}{\alpha} \left[\frac{\frac{\mu R(i, j)}{\alpha(\mu-1)}}{P} \right]^{-\mu} Y \\
\text{Technology} \quad Y(i) &= \alpha L(i) \\
\text{Optimal price} \quad P(i) &= \frac{\mu}{\alpha(\mu-1)} R(i)
\end{aligned}$$

Banks:

$$\begin{aligned}
\text{Unrestricted loan rate} \quad R(i) &= \frac{\mu}{\mu-1} C(a) \\
\text{Restricted loan rate} \quad R(i) &= \left[1 - \frac{v}{(1-F(a))\Gamma a^{\mu-1}} \right]^{\frac{-1}{\mu-1}} C(a)
\end{aligned}$$

Aggregation and market clearing:

$$\begin{aligned}
\text{Goods market clearing} \quad Y &= Q = \left(\frac{1}{z} \right)^{\frac{1}{2}} \\
\text{Aggregate loans} \quad L &= \int_0^J L(i, j) di \\
\text{Aggregate price} \quad P &= \left(\int_0^1 P(i)^{1-\mu} di \right)^{\frac{1}{1-\mu}} \\
\text{Aggregate production} \quad Y &= \left(\int_0^1 Y(i)^{\frac{\mu-1}{\mu}} di \right)^{\frac{\mu}{\mu-1}}
\end{aligned}$$

D List of Countries

Algeria, Argentina, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia, Brazil, Bulgaria, Cameroon, Canada, Chile, China, Colombia, Costa Rica, Croatia, Czech Republic, Denmark, Dominican Republic, Egypt, El Salvador, Estonia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Honduras, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Kenya, Korea, Kuwait, Latvia, Lithuania, Malawi, Malaysia, Mali, Mauritius, Mexico, Mozambique, Nepal, Netherlands, Nicaragua, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Rwanda, Senegal, Slovak Republic, Slovenia, South Africa, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Thailand, Tunisia, Turkey, Uganda, United Kingdom, United States, Uruguay, Venezuela, Zambia, Zimbabwe.