Common Banking across Heterogenous Regions

Enzo Dia,[†] Lunan Jiang[‡] Lorenzo Menna,[*] Lin Zhang[§]

[†] *Università degli Studi di Milano Bicocca.*

[‡] *Banco de Mexico.*

[§] *Henan University.*

[¶] *Henan University.*

Abstract: In this paper we show there exists a dispersion of interest margins charged by commercial banks among Chinese provinces, and build a parsimonious dynamic stochastic general equilibrium model with both banking and production sectors. Production firms require loans from commercial banks to support their working capital financing. This model can explain a considerable share of interest margin differences by calibrating parameters to the Chinese provincial level economic and financial variables. The segmentation of provincial loan market in China is an important reason for interest margin differences. Thanks to the loan market segmentation, the regional economic features characterized by banking sector productivity and loan demand take effects, and these two factors account for the interest margin differences from supply side and demand side respectively. An extension of our theoretical model also implies the credit control and banking regulation could also bring the influence of loan demand side, which works through a upward sloping loan supply curve.

[†]

[‡] Address: dongliuzhai building, Minglun Campus, Henan University, 85 Minglun street, Shunhe, Kaifeng, Henan, 475001, China; Phone: 0371-23881606; Email:lunan.jiang@vip.henu.edu.cn

[*]

[§] The authors thank the participants of biweekly seminar series of school of economics at Henan University, the econometric society China meeting, and syposium of development and institutional economics for helpful comments and suggestions. This version is only our first draft. All errors in this paper are ours, and comments and suggestions are well taken. We also appreciate the financial support by HENU center for financial development and stability.

Finally, we study how the interest margins response when real and financial shocks hit the economy.

Keywords: interest margins, market segmentation, Chinese economy

JEL Classification Numbers: E1 G21

1 Introduction

Because banks operating in large banking markets can reap both diversification benefits and substantial economies of scale, since the wave of financial deregulation of the 1990s, policy makers have worked to reduce the segmentation among locally differentiated banking markets. The segmentation, both across regions, and among different banking products, has been drastically reduced in the United States first, and progressively also in the euro area. The case of China instead is peculiar, since the industry has always been regulated at the central level, while economic conditions across provinces are substantially different.

Given these characteristics, China represents a perfect case study to analyze the costs and benefits of the introduction of a common banking market across different countries. We thus analyze the banking system of China, to investigate the impact of a common banking industry on an economy characterized by substantial differences in the industrial structure and the level of development across regions. Is a common banking industry always efficient in this environment? Does a common regulatory framework induce the adoption of a common banking technology across different regions? Are there instances where the segmentation of the industry insulates against certain adverse shocks? Or does the sharing of a common regulatory infrastructure end up potentially amplifying the impact of adverse idiosyncratic shocks hitting a single region? These are the questions that we address in this work.

Interest margins are normally analyzed exclusively at micro-level to reflect the marginal cost to run a traditional loan business due to fund raising, capital regulation, default risk and so on. Much less attention has been paid to the relationship between interest margins and the macroeconomic cycle. In this paper, we employ a two-sector general equilibrium model, composed of production firms and commercial banks. The commercial banks sector provides loans to production firm sector, financing their working capital needs. The interest rate on loans in this setup therefore work as the price of intermediate factor which is essential for production. More specifically, we make use of a two-sector RBC model developed by Dia and Menna 2016 that focus on the interactions between the banking and other industries generated by resource markets, with impulses transmitted through wages and the cost of capital. Dia and Menna 2016 and Dia and VanHoose 2017 argue that resource costs are the most relevant share of banks costs both across the cycle, and in different countries. Interests costs are the main cost factor only in countries with high inflation, and loan-loss provisions become of a comparable size to resource costs only in the case of large-scale banking crises. In the current environment, where banks are firms with tens, when not hundreds, of thousands of workers, this is hardly surprising. This framework is particularly suitable for describing the Chinese banking industry, because the interest cost of most Chinese bank liabilities has been regulated up to very recently, and the level of loan losses in Chinese banks has been rather stable and low during the last two decades.

The basic model adopts a simple Cobb-Douglas structure and the crucial underlying assumption is that the same resources can be used either in industrial firms or in banks, so that wages and the cost of capital are equalized across sectors. Since banks use a limited share of the resources of the economy, it then emerges that the dynamics of resource costs are in most instances largely determined by the industrial sectors, and banks have to adapt. Other, alternative DSGE models that include a banking sector take bank interest rates as exogenous variables, assume a parameter value for the exponent of the production function, and obtain the total factor productivity endogenously. The technology of the industry is thus imposed exogenously, and the capital intensity of the banking industry is not different from that of other industrial sectors. As a result, changes in resource costs cannot generate any interesting dynamics, and differences in bank interest margins across countries or regions depend only on substantial differences in the values of the total factor productivity, which is quite implausible in an "old" industry like banking. The models though can become more complicated, for example by introducing decreasing returns to scale induced by time-varying default costs, as in Curdia and Woodford 2010, or by introducing equity capital constraints as in the Gertler and Karadi 2011 approach that has been widely used afterwards. In these more complicated models, variations of bank interest margins are largely induced by loan losses, and differences among the rates that banks charge across regions should differ only because of differentials in loan losses. But the wide variations in margins across Chinese provinces, illustrated in Fig. 1, can hardly be explained by different loan losses only. The empirical analysis by Zhou and Wong 2008 finds that the declining trend of net interest margins of Chinese banks in the first decade of the new millennium is explained by the decline of average operating costs, not by changes in credit quality. These results provide strong support for the underlying assumptions behind our modelling strategy.

Dia and Menna 2016 produce further evidence that is at odds with the standard alternative approach, since they find that capital intensities in the banking industry are substantially different from those of the industrial sectors, although the pattern changes across different countries. It then emerges that banks in different countries adopt more or less capital intensive technologies according to the local availability of resources, as banks adapt their technologies to different wages and rental costs of capital. In view of this evidence, the calibration of the model that they propose differs significantly from the standard one, since they let the interest rate on loans as an endogenous variable, while instead they use the capital intensity of banks as an exogenous parameter. This strategy allows to obtain a calibration of the model where both the total factor productivity and the exponent of the production function in the banking industry are endogenous variables. The steady-state results that they get support the calibration strategy, as it turns out that resource costs explain a very substantial share of interest rate margins in all the countries of the sample analyzed, including the U.S and the main European countries. Furthermore, Dia and Menna 2016 find that the both parameters of the production function for bank loans are substantially different across countries. Consequently, the responses to shocks hitting the economy are substantially different across countries, not only in the case of productivity shocks that hit industrial sectors asymmetrically, but also in the case when shocks are symmetric across both industries.

The result that European banks adopt substantially different technologies across countries is not too surprising, but generates some disturbing implications. Since bank rates on loans respond differently to productivity shocks, any policy measure that aims to respond to these shocks must somehow take into account the differences in the responses. For example, the declared aim of the quantitative easing policy recently conducted on a large scale in the euro area was to stimulate lending, and bank lending in particular, since the relevance of financial frictions made standard tools of monetary policy less effective. However, the results of Dia and Menna 2016 suggest that following a negative productivity shock, either symmetric across sectors, or not, the response of bank interest margins is very different across countries, with some countries benefiting of a substantial reduction of loan rates, while in other the response is much more muted. Following a systemic shock hitting the euro area as a whole, it then follows that the optimal response should differ across countries, and the costs of a one-size-fits-all policy may be substantial. It may then appear obvious that the adoption of similar technologies across countries would tend to solve the problem, and that the development of a common regulatory framework, and a process of cross-country mergers and acquisition leading to a more integrated banking market would be highly beneficial. In contrast, in so far as substantially different conditions remain in the industrial systems of the different countries, or regions of individual countries, the optimal technology that banks may want to adopt may be different, to the extent that wages and the cost of capital do not quickly converge. So if wage differentials are important to generate convergent dynamics across regions, the adoption of different bank lending technologies may significantly contribute to this process, in particular in all those regions and countries where the industrial landscape is characterized mainly by small and medium firms, heavily reliant on bank sources of funds.

We use the Dia and Menna 2016 framework to analyze the Chinese economy, both at the aggregate level, and at the provincial level, for the provinces where reliable data are available. The first and main question we want to ask is if banks across different provinces adopt the same technology or if they instead chose more or less capital intensive technologies according to local conditions dictated by wages and the cost of capital. The Chinese labor market is in fact substantially free to adjust to local conditions, and wage differentials are massive across more and less developed provinces. Our strategy is in several steps. In the first we calibrate the model to analyze the different provinces under the assumption that the choice of the technology is dictated by local conditions. We thus calibrate the model making use of provincial data on the capital intensity of banks and other industries. In the second step we calibrate the model for the whole of the Chinese economy, by assuming that the lending technology is chosen for the whole country, and we thus calculate a weighted average of the capital intensity of banks that we use together with that of other industries to obtain endogenously the exponent of the production function for loans and the correspondent total factor productivity. Next we make an alternative calibration for the model at the provincial level, under the assumption that the technology is common across provinces. In this case we impose the parameters of the production function to be those obtained at the aggregate level, and we instead set the capital intensity of banks free. We then compare the performance of the model under the two different frameworks by comparing the share of interest margins predicted by the model under the two alternative calibrations, with the actual data. We find that the performance of the model assuming a common technology is superior for all the provinces of our sample, as it turns out that for any province the share of the actual interest margin that is explained is larger than the correspondent share calculated when the model is calibrated to match local conditions.

We therefore conclude that Chinese banks adopt a common technology across provinces, notwithstanding the different local conditions. This result is not surprising in light of the dominance of the large state-owned banks, notwithstanding the growth of local private sector banks of the last decade. In the final step of the work we assess the impact of different types of productivity shocks on different provinces, by analyzing the impulse responses. We study both symmetric shocks, affecting symmetrically both industrial firms and banks, and asymmetric ones, hitting one of the sectors only.

The rest of this paper is organized as below: Section 2 provides the setup of the model, and Section 3 describes our calibration strategies. In Section 4 we discuss our main findings, while the final section concludes.

2 Lending and interest margins across Chinese banks

Obtaining the information to calculate the interest margins of commercial loans across different provinces in China is challenging. There indeed exists some micro databases of local commercial banks in China, with data on loan portfolios that in principle allow to calculate interest margins, nevertheless most of these information is available through the financial reports of commercial banks issuing equity or bonds, and it is thus not available for smaller private banks. Moreover, large and medium-size national commercial banks only provide financial reports for the consolidated accounts, and no information is available for the provincial level operations. We therefore make use of another more direct, but unfortunately rather incomplete data source, which is the regional financial market operation reports published by People's Bank of China system. These provincial-level financial operation reports provide detailed information regarding the proportion of total loans falling into different loan rate ranges. Hence, we can calculate the weighted average loan rates for different provinces; we further assume the marginal deposit rate to be equal to the 1-year SHIBOR rate, and finally we define the gap between these two values as the interest margins charged. The interest margins of banks, displayed in Figure 1, are significantly different across provinces, because the loan market in China is segmented at the provincial level. We carried out several on-site interviews with regulators and bankers in China, and we found that most lending in China is supported by local banks or local branches of national banks. Lending across provincial boarders barely exists, the only exception being large syndicated loans provided through bank consortiums, but they are quite uncommon. Furthermore, large banks do not make use of internal markets to reallocate excess deposits across provinces, as the managerial organization of the branches is strictly based on provin-

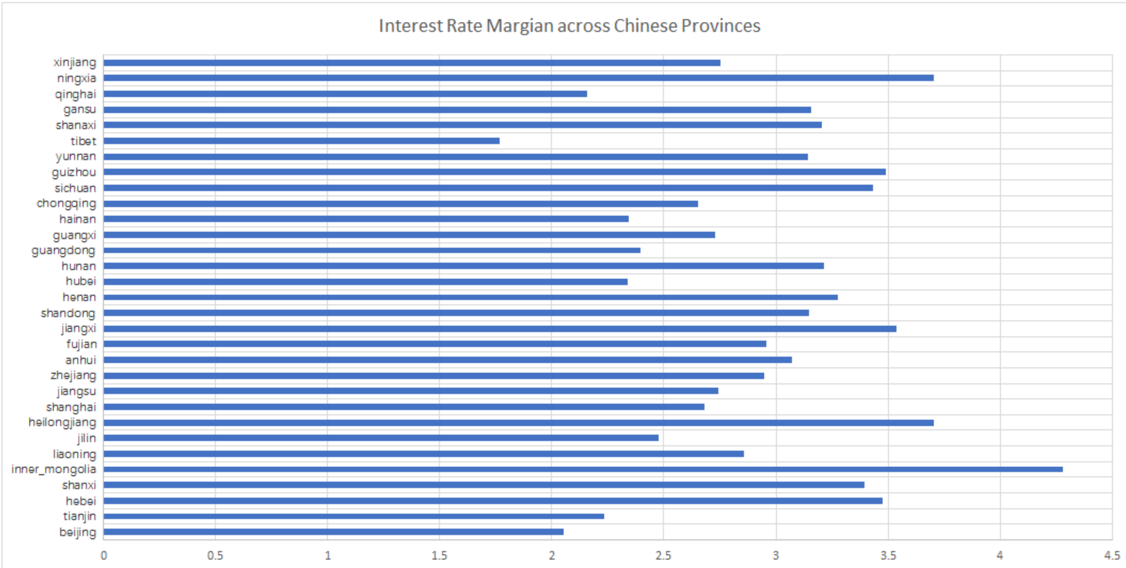

Figure 1: The interest margins among all Chinese provinces

cial level organizations. Surplus of deposits are then managed at the level of the consolidated national entity, and are allocated trough the interbank market or by purchasing securities, bonds in particular. Since the labour market is similarly segmented at the provincial level, we can assume that each province of China runs as a quasi-independent economy. Thus, the regional economic features play an important role in determining the interest margins. Of course, the production sectors of different provinces have all types of interactions, through trade of final goods or intermediate products, but to the extent that lending markets are segmented, any interaction across provinces does not affect our transmission mechanism: As production firms require working capital for their production and such liquidity constraints press the individual firms tightly, the demand for loans will not be sensitive to the loan rates charged by commercial banks in other provinces. Similarly, since commercial banks do not lend to other provinces, their loan supply will simply reflect the local cost of resources. Because we assume that industrial firms do not have substitutes for bank lending to meet their working capital requirements, given the limited development of the commercial paper market in China, the demand for loans if quite inelastic, and variations in the rates on loans largely depend on the changes of marginal costs, and thus on resource costs. Loan market equilibria across provinces differ because across different provinces loan demand changes as the working capital need of different industrial systems are not uniform, because resource costs are different, changing marginal costs and thus loan supply, and because banks may adapt their technology to different market conditions by choosing technologies that are more or less capital intensive. Several recent contributions have analyzed the Chinese banking system in recent years. Diallo and Zhang 2017 is one of the contributions that is most related to our topic. They make use of panel data for 31 provinces and 8 industrial sectors over the period 20012013 to analyze the relationship between bank concentration and economic growth, finding that bank concentration has a negative impact on growth.

3 Model

The economy consists of three classes of agents, households, production firms and banks.

3.1 Households

All the households are homogenous, and have an infinite horizon. They maximize their expected utilities by making decision on their optimal intertemporal consumption, labor supply, and saving choices. Households have two alternative possibilities to allocate their savings. They can either make investment in physical capital or deposit into banks. Accordingly, households earn wage income, capital return, and deposit return. Besides, households also own production and bank sectors, and therefore potential profits of these two sectors benefit households, but they are pushed down to zero by competition.

$$\begin{aligned} &\max_{\{c_t,h_t,k_t,d_t^s\}} E_0\beta^t[\frac{c_t^{1-\sigma}-1}{1-\sigma}+\theta\frac{(1-h_t)^{1-\gamma}-1}{1-\gamma}] \\ &\text{s.t.} c_t+k_t-(1-\delta)k_{t-1}+d_t^s=w_th_t+r_tk_{t-1}+r^d d_t^s+\Pi_t^b+\Pi_t^f \end{aligned} \tag{1}$$

from the first order conditions we obtain:

$$c_t^{-\sigma} = \beta E_t[c_{t+1}^{-\sigma}(r_{t+1} + 1 - \delta)] \tag{2}$$

$$w_t = \theta \frac{(1 - h_t)^{-\gamma}}{c_t^{-\sigma}} \tag{3}$$

$$c_t^{-\sigma} = \beta E_t(c_{t+1}^{-\sigma}) r_t^d \tag{4}$$

3.2 Banks

Banks use capital and labor as inputs of a Cobb-Douglas production function to generate loans from deposit. The following equation gives the loan production:

$$l_t^s = \min\left\{d_t^d, z_t(k_t^b)^{\kappa}(h_t^b)^{1-\kappa}\right\} \tag{5}$$

The profit of banking sector, for a given period is determined according the following equation:

$$\Pi_t^b = d_t^d + r_{t-1}^l l_{t-1}^s - r_{t-1}^d d_{t-1}^d - l_t^s - w_t h_t^b - r_t k_t^b, \tag{6}$$

and household require banks to maximize the current marginal value of profits, which are thus subject to the discount factor $\beta^t c_t^{-\sigma}$:

$$\max_{\{k_t^b, h_t^b\}} \sum_{t=0}^{\infty} \beta^t c_t^{-\sigma}[(r_{t-1}^l - r_{t-1}^d) z_{t-1} k_{t-1}^{\kappa} h_{t-1}^{1-\kappa} - w_t h_t^b - r_t k_t^b]. \tag{7}$$

Since banks minimize cost to obtain the optimal combination of input, we obtain the banks' demand for labor and physical capital:

$$\beta E_t c_{t+1}^{-\sigma}(1 - \kappa) z_t \left(\frac{k_t^b}{h_t^b}\right)^{\kappa} (r_t^l - r_t^d) = c_t^{-\sigma} w_t \tag{8}$$

$$\beta E_t c_{t+1}^{-\sigma} \kappa z_t \left(\frac{k_t^b}{h_t^b}\right)^{\kappa-1} (r_t^l - r_t^d) = c_t^{-\sigma} r_t \tag{9}$$

3.3 Firms

Firms in this economy produce output y according to a Cobb-Douglas production function, and standard input requirements, but the firm faces a liquidity constraint caused by the need to finance its working capital in advance.

$$y_t = a_t(k_t^f)^\alpha(h_t^f)^{1-\alpha}, \tag{10}$$

$$(w_t h_t^f + r_t k_t^f)\mu = l_t^d. \tag{11}$$

In each period, firms have to borrow from banks a proportion μ of their expense and at the same time pay back what they borrowed from commercial banks in the previous period.

$$\pi_t^f = l_t^d + y_t - r_{t-1}^l l_{t-1}^d - w_t h_t^f - r_t k_t^f, \tag{12}$$

$$\pi_t^f = y_t - r_{t-1}^l(w_{t-1}h_{t-1}^f + r_{t-1}k_{t-1}^f)\mu - (w_t h_t^f - r_t k_t^f)(1-\mu). \tag{13}$$

And since firms need to maximize the current marginal value of profits

$$\max \sum \beta^t c_t^{-\sigma}\left[a_t(k_t^f)^\alpha(h_t^f)^{1-\alpha} - r_{t-1}^l(w_{t-1}h_{t-1}^f + r_{t-1}k_{t-1}^f)\mu - (w_t h_t^f - r_t k_t^f)(1-\mu)\right], \tag{14}$$

form the first order conditions we obtain the firms' demand for labor and capital:

$$c_t^{-\sigma}\left[(1-\alpha)a_t(k_t^f)^\alpha(h_t^f)^{-\alpha} - w_t(1-\mu)\right] = \beta c_{t+1}^{-\sigma} r_t^l w_t \mu, \tag{15}$$

$$c_t^{-\sigma}\left[\alpha a_t(k_t^f)^{\alpha-1}(h_t^f)^{1-\alpha} - r_t(1-\mu)\right] = \beta c_{t+1}^{-\sigma} r_t^l r_t \mu, \tag{16}$$

3.4 Market Clearing Condition

There are a total of five markets in this economy. Capital market, labor market, deposit market, loan market, and consumption goods market. The clearing conditions are represented by the following five equations:

$$k_{t-1} = k_t^f + k_t^b, \tag{17}$$

$$h_t = h_t^f + h_t^b, \tag{18}$$

$$d_t^d = d_t^s, \tag{19}$$

$$l_t^d = l_t^s, \tag{20}$$

$$c_t + k_t - (1-\delta)k_{t-1} = y_t. \tag{21}$$

3.5 Steady state

From the first order conditions we obtain the following steady state relationships:

$$1 = \beta(r_{ss} + 1 - \delta) \tag{22}$$

$$w_{ss} = \theta \frac{(1-h_{ss})^{-\gamma}}{c_{ss}^{-\sigma}} \tag{23}$$

$$1 = \beta r_{ss}^d \tag{24}$$

$$\beta(1-\kappa) z_{ss} \left(\frac{k_{ss}^b}{h_{ss}^b} \right)^{\kappa} (r_{ss}^l - r_{ss}^d) = w_{ss} \tag{25}$$

$$\beta \kappa z_{ss} \left(\frac{k_{ss}^b}{h_{ss}^b} \right)^{\kappa-1} (r_{ss}^l - r_{ss}^d) = r_{ss} \tag{26}$$

$$(1-\alpha) a_{ss} \left(\frac{k_{ss}^f}{h_{ss}^f} \right)^{\alpha} = w_{ss}(\beta r_{ss}^l \mu + 1 - \mu) \tag{27}$$

$$\alpha a_{ss} \left(\frac{k_{ss}^f}{h_{ss}^f} \right)^{\alpha-1} = r_{ss}(\beta r_{ss}^l \mu + 1 - \mu) \tag{28}$$

$$(w_{ss} h_{ss}^f + r_{ss} k_{ss}^f)\mu = z_{ss}(k_{ss}^b)^{\kappa}(h_{ss}^b)^{1-\kappa}. \tag{29}$$

4 Parametrization

We require two groups of target to pin down all parameters in this model: macroeconomic variables which serve as the targets of standard parameters, and two sets of ratios that link the industrial and the banking sector of the economy: the ratio between the productivity parameters in the production functions of industrial firms and banks, and the ratio between the capital intensity across the two sectors.

Using (28), the capital intensity in industrial sector is obtained as:

$$\frac{k_{ss}^f}{h_{ss}^f} = \left[\frac{r_{ss}\beta r_{ss}^l \mu + (1-\mu) r_{ss}}{\alpha a_{ss}}\right]^{\frac{1}{\alpha-1}}, \tag{30}$$

where r_{ss} can be obtained from equation 22, a_{ss} is normalized to unity, and r_{ss}^l is thus the only unknown. Rearranging (26) we can express the capital intensity of banks as

$$\frac{k_{ss}^b}{h_{ss}^b} = \left[\frac{r_{ss}}{\beta\kappa z_{ss}(r_{ss}^l - r_{ss}^d)}\right]^{\frac{1}{\kappa-1}}. \tag{31}$$

After substituting the value of r_{ss}^d obtained from equation (24), and the value of w_{ss} obtained from equation (25) into equation (27), we obtain

$$\frac{(1-\alpha)a_{ss}\left(\frac{k_{ss}^f}{h_{ss}^f}\right)^{\alpha}}{\beta(1-\kappa)z_{ss}\left(\frac{k^b}{h^b}\right)^{\kappa}(r_{ss}^l - r_{ss}^d)} = 1-\mu+\mu\beta r_{ss}^l. \tag{32}$$

Substituting the values of the capital intensities of firms and banks obtained from equations (30) and (31) into equation (32) we get

$$\frac{(1-\alpha)a_{ss}\left[\frac{r_{ss}\beta r_{ss}^l\mu+(1-\mu)r_{ss}}{\alpha a_{ss}}\right]^{\frac{\alpha}{\alpha-1}}}{\beta(1-\kappa)\left(\frac{r_{ss}}{\beta\kappa}\right)^{\frac{\kappa}{\kappa-1}}\left[z_{ss}(r_{ss}^l - r_{ss}^d)\right]^{\frac{1}{1-\kappa}}} = 1-\mu+\mu\beta r_{ss}^l. \tag{33}$$

From the former expression we can isolate the value of z_{ss},

$$z_{ss} = \left(\frac{(1-\alpha)a_{ss}\left[\frac{r_{ss}(\beta r^l_{ss}\mu+1-\mu)}{\alpha a_{ss}}\right]^{\frac{\alpha}{\alpha-1}}}{\beta(1-\kappa)\left(\frac{r_{ss}}{\beta\kappa}\right)^{\frac{\kappa}{\kappa-1}}(1-\mu+\mu\beta r^l_{ss})} \right)^{1-\kappa} /(r^l_{ss}-r^d_{ss}), \tag{34}$$

and after substituting the value of $z_{ss}(r^l_{ss}-r^d_{ss})$ obtained from the former equation, (31) becomes: [**Can equation** (31) **directly serve as the solution for** $\frac{k^b_{ss}}{h^b_{ss}}$**?**]

$$\frac{k^b_{ss}}{h^b_{ss}} = \left(\frac{(1-\alpha)a_{ss}\left[\frac{r_{ss}\beta r^l_{ss}\mu+(1-\mu)r_{ss}}{\alpha a_{ss}}\right]^{\frac{\alpha}{\alpha-1}}}{\beta(1-\kappa)(1-\mu+\mu\beta r^l_{ss})\left(\frac{(1-\alpha)a_{ss}\left[\frac{r_{ss}(\beta r^l_{ss}\mu+1-\mu)}{\alpha a_{ss}}\right]^{\frac{\alpha}{\alpha-1}}}{\beta(1-\kappa)\left(\frac{r_{ss}}{\beta\kappa}\right)^{\frac{1}{\kappa-1}}(1-\mu+\mu\beta r^l_{ss})} \right)^{1-\kappa}} \right)^{\frac{1}{\kappa}}. \tag{35}$$

We have finally expressed the capital intensity of banks as a function of r^l_{ss} and a set of exogenous parameters. Multiplying both sides of equation (27) by h^f_{ss} and both sides of equation (28) by k^f_{ss}, summing up the two equations, we obtain

$$a(k^f)^\alpha(h^f)^{1-\alpha} = (w_{ss}h^f_{ss} + r_t k^f_{ss})(\beta r^l_{ss}\mu + 1 - \mu). \tag{36}$$

Remembering that in steady state $y = a_{ss}(k^f_{ss})^\alpha(h^f_{ss})^{1-\alpha}$,

$$y_{ss} = (w_{ss}h^f_{ss} + r_t k^f_{ss})(\beta r^l_{ss}\mu + 1 - \mu). \tag{37}$$

And since the demand for loans is $(w_{ss}h^f_{ss} + r_{ss}k^f)\mu = l^d$, given that the loan market equilibrium condition $l^s_{ss} = l^d_{ss}$ we obtain

$$\frac{y_{ss}}{l_{ss}} = \frac{\beta r^l_{ss}\mu + 1 - \mu}{\mu}, \tag{38}$$

and substituting the production functions,

$$\frac{a_{ss}\left(\frac{k^f_{ss}}{h^f_{ss}}\right)^{\alpha} h^f_{ss}}{z_{ss}\left(\frac{k^b_{ss}}{h^b_{ss}}\right)^{\kappa} h^b_{ss}} = \frac{\beta r^l_{ss}\mu + 1 - \mu}{\mu}. \tag{39}$$

We can now substitute the values of the capital intensities of the two sectors that we have obtained in equations 30 and 35 in equation 39. Moreover, assuming as it is standard in the literature that the total hours worked are 0.3, since then $h_{ss} = h^f_{ss} + h^b_{ss} = 0.3$ it becomes possible to obtain h^f_{ss} and h^f_{ss} from the above two equations, and the value of the capital stock of the two sectors since $k^f_{ss} = h^f_{ss}\left(\frac{k^f_{ss}}{h^f_{ss}}\right)$,$k^b_{ss} = h^b_{ss}\left(\frac{k^b_{ss}}{h^b_{ss}}\right)$ and $k_{ss} = k^b_{ss} + k^f_{ss}$. Summing up, from equation 39 we can obtain the values of hours and capital in each of the sectors as a function of r_L, z, κ, and a bunch of exogenous parameters.

$$\frac{k^b_{ss}/h^b_{ss}}{\frac{k^f_{ss}}{h^f_{ss}}} = \frac{k^b_{ss}/h^b_{ss}}{\left(k^b_{ss} + k^f_{ss}\right)/0.3}. \tag{40}$$

After substituting the values of k^b_{ss} and k^f_{ss} previously derived in equation 40, we have a first equation linking the three endogenous parameters to a combination of exogenous ones. We need two more equations to get a system of three equations for the three unknown. Since we can obtain data regarding the capital intensity of banks and that of the aggregate economy that we call $Ratio1$, we can use the relationship

$$\frac{k^b_{ss}/h^b_{ss}}{\frac{k^f_{ss}}{h^f_{ss}}} = Ratio1. \tag{41}$$

Finally, since in the calibration we normalize the a parameter to unity, we then impose a further restriction involving the ratio among the productivity measures that we get from the data. Assuming the exponent of the production function for banks to be the same as that of industrial firms, by imposing the restriction is straightforward. However, a crucial aspect of this work is that we do not want to impose this equality, and we rather want to obtain the exponent of the banks' production function endogenously. One of the crucial aims of this work, in fact is to study the technology of banks, and we do not want to limit the analysis to the productivity parameter, since banks in different provinces may be willing to adopt technologies that more or less capital intensive. In our calibration we thus force the ratio between the productivity parameters to respect the correspondent one obtained by the data for an endogenous value of κ:

$$\frac{z}{a} = z = \frac{\frac{L/W_B}{(K_B/W_B)^{\kappa}}}{TFP}, \tag{42}$$

where L/W_B is the loan to bank workers ratio, K_B/W_B is the capital intensity of banks, and TFP is the total factor productivity of the economy. An unusual aspect of this calibration is that the target is constrained by one of the endogenous parameters, but the technique we use amounts to solve a system of three equations, equations 40, 41, and 42, in three unknowns, r_L, z, κ.

4.1 Calibration

We use two sets of data to calibrate the model, macroeconomic data and banking industry data. In both cases we use both province level, and aggregate national level data. As a result of data availability, we can study five provinces: Hubei, Shanghai, Zhejiang, Henan, and Hebei provinces. We in fact lack bank balance sheets data for other provinces. The financial statistics year book of each province provide a financial report of the provincial branches of national banks. We can thus obtain the data on the physical capital stock, and the number of employees that allow us to calculate the average capital intensity of banks for each province.

Shanghai is a mega metropolitan area in China, its economy is highly developed, much more capital intensive than that of other provinces in our sample, and the income per capite is far larger. However, Shanghai, Hubei, and Zhengjiang provinces share some common feature in economic structure. Although Hubei is a province located in mid-west China, its capital, Wuhan, produces almost half of the GDP of the province and hosts the main economic activities. Zhejiang province instead is the main area where manufacturing industries oriented for international trade locate. Henan and Hubei provinces have capital-intensive industries such as steel, electricity, coal mining, aluminum, and so on.

Table 1: Accounting for provincial NIM Difference in China

Province	α	μ	K/W	K_B/W_B	$Ratio1$	$\frac{z}{TFP}$	κ	NIM	NIM-p	NIM-I
Hubei	0.50	0.71	164	245	1.49	107.84	0.61	—	0.73	0.87
Shanghai	0.43	1.47	414	611	1.48	84.89	0.67	—	0.88	1.11
Zhejiang	0.47	1.38	229	279	1.22	154.26	0.58	—	0.58	0.98
Henan	0.50	0.61	164	209	1.27	132.95	0.56	—	0.67	0.75
Hebei	0.49	0.53	183	200	1.10	165.40	0.52	—	0.59	0.87
National	0.46	0.90	170	362	2.13	62.70	0.72	—	0.93	0.93

Table 2: calibration results

This model works as a standard two sector models, where loans are an intermediate input, therefore the income share parameters α and κ, and technology levels z_{ss} and a_{ss} are supposed to match the ratio of capital intensity and the relative price of two sectors. The interest margin in this economy indeed serves as the relative price of intermediate input, bank loans.

Labor income share in the model does not correspond to α directly, and it follows instead

$$\frac{w_{ss}h_{ss}}{y_{ss}} = \frac{1-\alpha}{\beta r^l_{ss}\mu + 1 - \mu}, \tag{43}$$

but given the numerical values of μ,β, and $r^l_{ss} - r^d_{ss}$, since the error is very small we use the labor income share as a proxie of α.

α is calibrated to reflect the capital income share of each province, or that at the national level, respectively. μ is calibrated to match the debt-GDP ratio of each province, or the national one. Since we do not have detailed figures for the relative shares of different classes of loans, we have obtained an estimate of the proportion of industrial and business loans out of the total from experienced banking practitioners, and the figure is 75%. To calculate μ

we use total loans, but we exclude mortgages by multiplying the figures obtained from the regional financial statistics yearly book by a value of 0.75 .

Our calibration strategy is repeated at provincial level and at the national level, and we obtain a set of steady-state results for each set of parameters, r_L, z, κ. As in Dia and Menna 2016, we use r_L to evaluate the performance of the model, by comparing the interest margins predicted by the model with the actual one. In the first stage of the analysis we thus obtain endogenously both z and κ for each province separately and at the national level. In the second stage of the analysis we assume that the technology that banks adopt is the same across all provinces, and is set at the national level. As a consequence, we take the value of κ obtained in the first stage of the analysis at the national level and we impose it as an exogenous value in the parametrization for each province. In equation **final4** once the value of *kappa* is fixed, the value of z similarly becomes fixed exogenously and we do not need **final4** anymore. Since we still use the capital intensities obtained at provincial level, which reflect the local conditions in the markets for physical resources, and since we now have a new set of results for the ratio $\frac{z}{a}$, consequently the result for r_L are different. We can then compare the performance at the provincial level of this version of the model, imposing the same technology across the whole nation with the one of a model assuming that the technology, and thus the κ parameter is chosen independently in every province.

5 Results

Figure 2 compares the steady state results of the model with the observed interest margins, with the blue bar representing the actual interest margin obtained from the data, while the orange and grey bars represent the interest margins implied by our model. The orange bar is obtained by assuming that banks from different provinces make use of different technologies, while the grey one is obtained by assuming that banks use the same technology across provinces. The proportion of the total interest margins that can be explained by our model range from 20% to 40%, which is a considerable share of interest margin as we do not exclude the impact of default risks and regulatory costs.

Table 2 displays the steady-state results of the model, and the last two columns show the interest margins implied by our model. In all the provinces of our sample, the model assuming a common national banking technology performs better, as it explains a larger share of the observed interest margins. To understand the results it must be noted that the national technology that we obtain from the model is more capital intensive than any of the provincial ones, since κ is larger for the national case. The higher κ is not just the result of a larger observed bank capital intensity, since the observed capital intensity is much larger in Shanghai province and the national value of 362 thousand yuan per worker is not very far from the simple average across our sample at the provincial level, which is 309 thousand yuan per worker. Instead, as we can observe from Table 2, the ratio between the capital intensity of banks and that of the economy is larger at the national level than in any province, since in Shanghai the industrial system is extremely capital intensive. At the same time, a larger capital intensity is associated with a lower productivity, and a larger interest rate on loans. The picture emerging is that banks operating in the more developed provinces, like Shanghai, need to use capital intensive technologies because they face higher wages than banks in other provinces.

Whenever differences in wages are substantial among regions, as it the case of Chinese provinces, see for example Nagayasu and Liu 2008 which found evidence of substantial divergence in the wage wage ratio among provinces. However, the adoption of a national technology requires a substantial amount of physical capital, since in the more developed regions, where banks generate a disproportionate share of revenues, wages are far higher than in the less developed ones. When the more capital intensive technology is adopted, firms in less developed regions face larger interest costs on loans than would be charged by a segmented local banking system. Following the adoption of these technologies, banks charge interest rate on loans in Shanghai and Beijing that are substantially lower than the average, while poorer regions where wages are lower face larger interest rates because they are relatively inefficient. As it can be noted from Table 2, the adoption of a common technology generates a substantial smoothing of the differences among the interest margins across provinces.

But there's more. The response to shocks is very different in the two cases. In the specific case of China the difference in the response is particularly sharp. When analyzing the impulse responses, in fact, it is important to note that both in all provinces of China, and at the national level, the capital intensity of banks is larger than that of the rest of the economy. This feature is not common in developed economies, as for example in the sample studied by Dia and Menna 2016 Spain only has this feature. As discussed in Dia and Menna 2016, in fact, when this is the case, the response of interest rate to productivity shocks becomes particularly strong, either when a positive (negative) shock hits the industrial sector, producing higher (lower) rates on loans, or when a positive (negative) shock hits banks, producing lower (higher) loan rates.

We conduct the dynamic analysis of our model to investigate the impact of two kinds of shocks on the economy, productivity shocks hitting the industrial sector, and productivity shocks hitting the banking sector, and we focus in particular on the impact on interest margins and working hours in the banking sector. The impulse response from the model are illustrated in Figure 3.

The impulse responses that we obtain suggest that the response of banks adopting the national technology to a productivity shock hitting the industrial sector is much stronger than that of reginal banks. A positive shock induces in fact a much stronger increase of the rate on loans, which follows the increase in wages produced by the productivity shock. Similarly, a positive bank productivity shock generates a much stronger decline of interest rates at the national level than in the case of regional segmented banking systems. This result suggests that the adoption of a national technology generates two contrasting effects: Positive national firm-productivity shocks generate negative asymmetric effects that imply

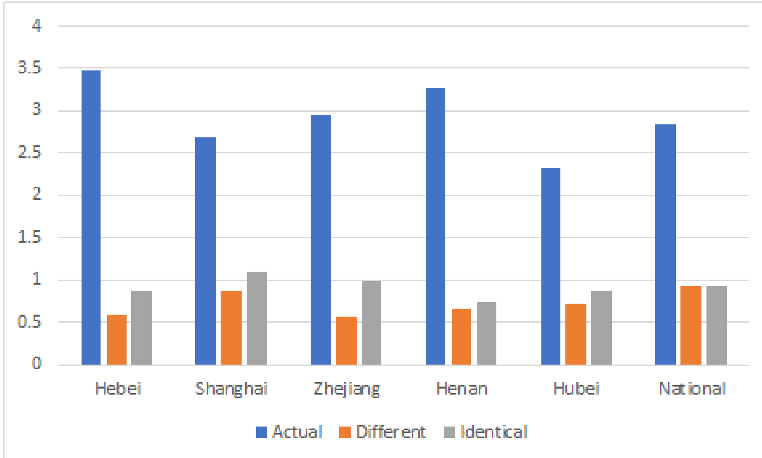

Figure 2: Actual and predicted interest margins

negative spillovers for poorer regions that end paying disproportionately larger rates on loans. On other hand they benefit of the large reductions of rates produced by a positive shock to banks, but conversely the are hard hit by negative shocks, for example if a regulatory intervention makes banks less productive, the burden on poorer regions is much heavier.

Similarly also the capability of banks to smooth regional idiosyncratic productivity shock hitting firms is much reduced when banks adopt a common nation-wide technology. Figure 3 displays a set of responses for all the provinces of our sample comparing the response to a local productivity shocks when loan markets are segmented (line) to the response when the banking technology is uniform at the national level. The responses highlight that the cost generated by the adoption of the common technology is larger in the less developed provinces whose industries are less capital intensive. This problem is relevant, in view of the substantial degree of heterogeneity in the business cycle dynamics across Chinese provinces and regions documented by Poncet and Barthélemy 2008.

6 Conclusion

This paper provides evidence of the existence of a substantial dispersion among interest margins charged by commercial banks in different Chinese provinces, and then builds a par-

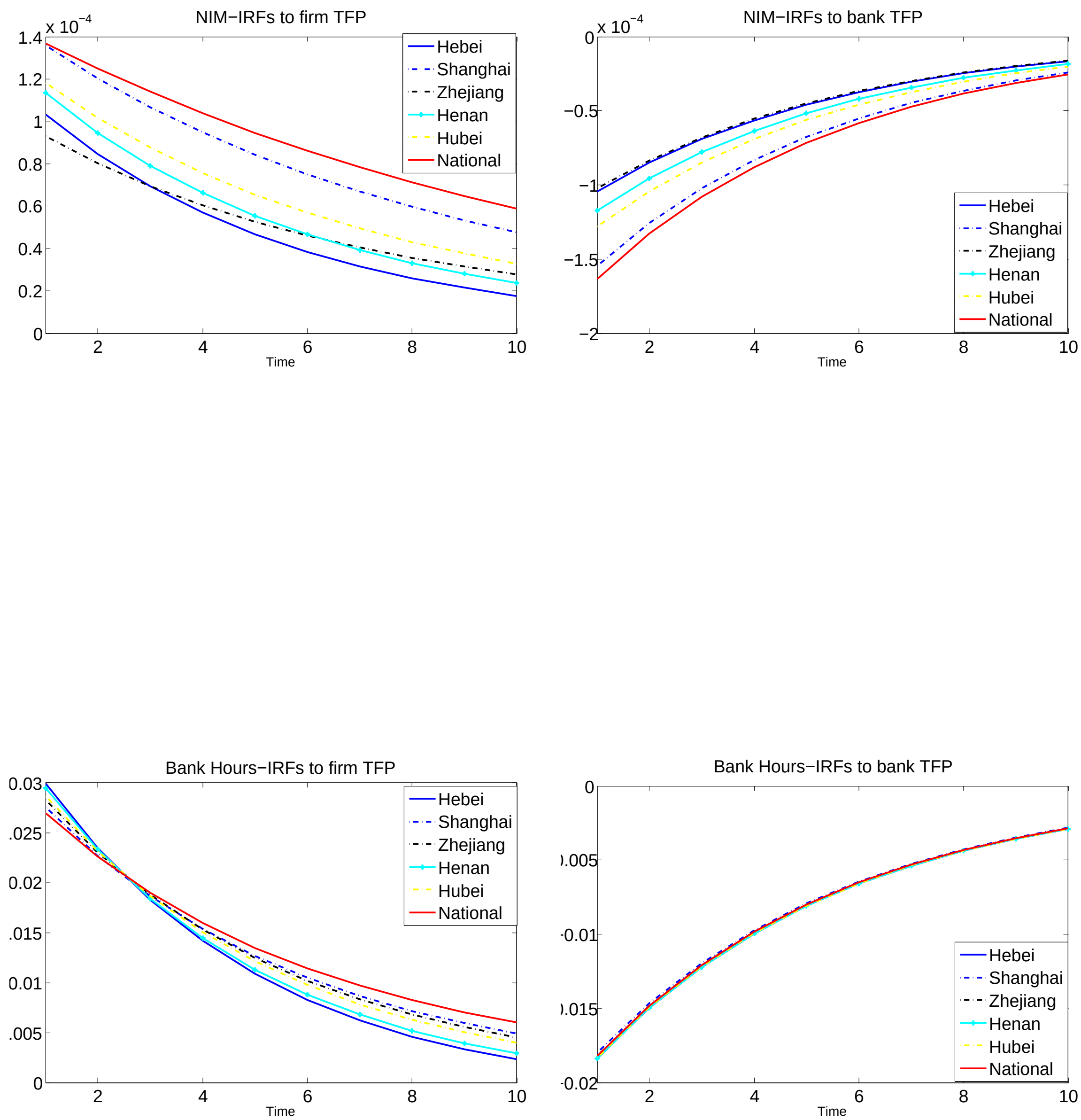

Figure 3: Impulse response functions of bank interest margins and bank working hours following a one standard deviation shock to firm and bank TFP.

simonious dynamic stochastic general equilibrium model with both banking and production sectors. Production firms require loans from commercial banks to support their working capital financing. This model can explain a considerable share of interest margin differences by calibrating parameters to the Chinese provincial level economic and financial variables. The results of the model suggest that Chinese banks adopt a common national technology rather than adopting more or less capital intensive technologies in different provinces in view of the local dynamics of resource costs. Several noteworthy implications emerge: Since in the case of Chinese provinces differences in wages are substantial, the adoption of a national technology implies a substantial amount of physical capital, because in the more developed regions, where banks generate much of their loans, wages are far higher than in the less developed ones. When the more capital intensive technology is adopted, firms in less developed regions face larger interest costs on loans than would be charged by a segmented local banking system, and banks charge interest rate on loans in Shanghai and Beijing that are substantially lower than the average, while poorer regions where wages are lower face larger interest rates because they are relatively inefficient. Moreover, the adoption of a common nationwide technology generates a stronger response of the rate on loans to productivity shocks, and the capability of banks to smooth regional idiosyncratic productivity shock hitting firms decline substantially.

References

[1] Vasco Curdia and Michael Woodford. "Credit spreads and monetary policy". In: *Journal of Money, Credit and Banking* 42.s1 (2010), pp. 3–35.

[2] Enzo Dia and Lorenzo Menna. "Productivity shocks, capital intensities, and bank interest rates". In: *Journal of Macroeconomics* 48 (2016), pp. 155–171.

[3] Enzo Dia and David VanHoose. "Capital intensities and international trade in banking services". In: *Journal of International Financial Markets, Institutions and Money* 46 (2017), pp. 54–69.

[4] Boubacar Diallo and Qi Zhang. "Bank concentration and sectoral growth: Evidence from Chinese provinces". In: *Economics Letters* 154 (2017), pp. 77–80.

[5] Mark Gertler and Peter Karadi. "A model of unconventional monetary policy". In: *Journal of Monetary Economics* 58.1 (2011), pp. 17–34.

[6] Jun Nagayasu and Ying Liu. "Relative prices and wages in China: evidence from a panel of provincial data". In: *Journal of Economic Integration* (2008), pp. 183–203.

[7] Sandra Poncet and Jean Barthélemy. "China as an integrated area?" In: *Journal of Economic Integration* (2008), pp. 896–926.

[8] Kaiguo Zhou and Michael CS Wong. "The determinants of net interest margins of commercial banks in mainland China". In: *Emerging Markets Finance and Trade* 44.5 (2008), pp. 41–53.