Common Banking Across Heterogeneous Regions

Enzo Dia, Lunan Jiang, Lorenzo Menna, Lin Zhang
AUTHORS

Enzo Dia
Università degli Studi di Milano Bicocca
E-mail: enzo.dia@unimib.it

Lunan Jiang
HenU Center for Financial Development and Stability
Henan University
E-mail: jiang.lunan@cfds.henuecon.education

Lorenzo Menna
Banco de Mexico
E-mail: lorenzomenna@yahoo.it

Lin Zhang (Corresponding Author)
HenU Center for Financial Development and Stability
Henan University
E-mail: lin.zhang@cfds.henuecon.education

First version: published March 30th, 2018
Second revised version: published August 29th, 2018

The responsibility for discussion papers lies solely with the individual authors. The views expressed herein do not necessarily represent those of the CFDS. The papers represent preliminary work and are circulated to encourage discussion with the authors. Citation of the discussion papers should account for their provisional character; a revised version may be available directly from the authors. Comments and suggestions on the methods and results presented are welcome.

IMPRESSUM

© CFDS, 2018

HenU Center for Financial Development and Stability
Dongluzhai Building, 85 Minglun Street
Henan University, Minglun Campus
Shunhe, Kaifeng, Henan, China
Tel. +86 (30) 897 89-0
http://cfds.henuecon.education

Papers can be downloaded free of charge from the CFDS website:
http://cfds.henuecon.education/index.php/research
Common Banking across Heterogeneous Regions

Enzo Dia, Lunan Jiang, Lorenzo Menna† Lin Zhang‡

† Università degli Studi di Milano Bicocca.
‡ CFDS, Henan University.
∗ Banco de México.
§ CFDS, Henan University.

Abstract

We describe the existence of a substantial dispersion of interest margins charged by commercial banks among Chinese provinces and find empirically that the main drivers of interest margins are resource costs. We build a parsimonious dynamic stochastic general equilibrium model featuring both banking and production sectors that we calibrate at both the national and provincial levels. Our model can explain a considerable share of the interest margin charged at the provincial level, and we find evidence that when Chinese banks adopt a technology imposing the same capital share across provinces, their productivity becomes substantially lower. Since the differences in wages in Chinese provinces are substantial, the adoption of a common technology implies an inefficient industrial structure for the banking industry and a substantial cost for the economy. The adoption of a standardized technology also generates a stronger response of the loan rate to productivity shocks, and thus the capability of banks to smooth regional idiosyncratic productivity shocks hitting firms declines substantially.

JEL Classification Numbers: E1; G218

Keywords: Interest margins; resource costs; Chinese economy

† The opinions expressed here are those of the author and do not necessarily represent Banco de Mexico’s or its board of governors’ opinions.
‡ Correspondent author. Email: lin.zhang@vip.henu.edu.cn. Address: Dongliuzhai Building, Minglun Campus, Henan University, 85 Minglun Street, Shunhe, Kaifeng, Henan, 475001, China; Phone: 0371-23881606
1 Introduction

Banks operating in large banking markets can reap both diversification benefits and economies of scale. Hence, beginning with the wave of financial deregulation in the 1990s, policymakers have worked to reduce the segmentation among locally differentiated banking markets by, for example, reducing or eliminating regulation that constrains inter-regional or cross-border lending activities. Indeed, market segmentation, both across regions and among banking products, has been drastically reduced in the United States and Eurozone. By contrast, the national regulation of the banking industry in China is common. However, since the economic conditions in the country differ substantially by province, banks’ lending is actually managed at the provincial level. Hence, Chinese banks do not lend across provinces, and most lending is financed locally with core deposits (i.e., the market is in essence segmented).\(^1\)

The People’s Bank of China gathers a rich set of data by province to analyze the banking industry at both the national and the provincial levels. Hence, China represents a perfect case study for analyzing the characteristics of a banking market subject to a common regulatory framework, the same monetary policy, and the same interest rates on securities, but where banks operate across administrative entities that present different underlying economic conditions and business cycles, such as Chinese provinces. Therefore, in this study, we analyze the banking system in China to investigate the implications of a common banking industry on an economy characterized by substantial differences in industrial structure and level of development across regions. Is a common banking industry always efficient in this environment? Are there instances where the segmentation of the industry insulates it against certain adverse shocks? Or does sharing a common banking system end up amplifying the impact of adverse idiosyncratic shocks hitting a single region? These are the questions that we address in this work.

To analyze these problems, it is necessary to model the banking industry in a general

\(^1\)See Han et al. (2017) for evidence of the segmentation of the banking industry in China.
equilibrium framework. We employ a two-sector general equilibrium model composed of perfectly competitive industrial firms and commercial banks. Commercial banks provide loans to goods-producing firms to finance their working capital needs. The interest rate on loans in this set-up therefore works as the price of an intermediate factor that is essential for production. More specifically, we use the two-sector RBC model developed by Dia and Menna (2016) that focuses on the interactions between the banking and other industries generated by resource markets, with impulses transmitted through wages and the cost of capital. The 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 in 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.

This modeling approach is supported by the evidence provided by Dia and Menna (2016) and Dia and VanHoose (2017) that resource costs account for the most relevant share of banks’ cost structures both across the cycle and among countries. Interest costs are the main costs for banks only in countries with high inflation, where however interest revenues are similarly inflated, while loan loss provisions are normally of a much smaller scale than resource costs and only become a comparable size during large-scale banking crises. This evidence should not surprise when considering that large banks have hundreds of thousands of workers. This framework is particularly suitable for describing the Chinese banking industry because the level of loan losses in Chinese banks has been rather stable and low during the past two decades. In line with our modeling assumptions, the empirical evidence provided by Zhou and Wong (2008) suggests that the interest margins of Chinese banks are explained by operating costs rather than by changes in credit quality. Although other factors that we do not model such as loan losses, the cyclical behavior of mark-ups, and the presence of financial frictions and nominal rigidities may alter and influence the processes we describe, the basic underlying dynamics would remain unchanged or would more likely be amplified.
under a more complicated setting, and the available empirical evidence suggests that the factors we identify are the most relevant ones.

In the initial part of this work, we provide further empirical evidence supporting these results and our modeling strategy by analyzing a large sample of Chinese banks. We collect data from both SNL Financial and Orbis (Bankscope) and generate a large panel covering all banks with more than $5 billion of assets that for some banks goes back to 1999. We then estimate the dynamic model of bank interest margins of Dia and Giuliodori (2012) with this panel of Chinese banks and find results similar to theirs regarding the transmission of interest rate and cost shocks on bank interest rates and margins for European and U.S. banks. These results suggest that notwithstanding the extensive regulation of the banking system, Chinese banks set their interest rates as their U.S. and European counterparts do. Indeed, when considering the regulatory framework, it is important to observe that the upper ceiling on the interest rates on loans was removed in 2004 and that the lending rate floor has rarely been binding since then.\footnote{He and Wang (2012) provide evidence that the percentage of loans made at the floor rate has fluctuated between 16% and 32% since 2004. Even more importantly, in line with the theory, we find evidence that both resource costs and loan loss provisions have a strong impact on bank rates and margins. We also find that the impact of resource costs variations is roughly twice as large as that of loan loss provisions.}

We then analyze the institutional framework of the industry. The banking system in China is composed of several classes of banks comprising large commercial banks operating nationally, joint-stock commercial banks operating regionally, and smaller local banks (e.g., city commercial banks and rural commercial banks). Although national banks dominate the market, with the five largest commercial banks accounting for 43.3% of total assets in 2013,\footnote{China Banking Regulatory Commission annual report, 2013}, the market share of the other classes is significant. Hence, we can presume that banks of different sizes and geographical locations adopt different technologies, using labor and physical capital more or less intensively. We provide evidence that this is the case, as we show that the capital intensities of banking firms are widely different across provinces. We then analyze the productivity of the banking system and bank interest rates under two
alternative frameworks, one in which banks adopt a technology specific to each province, so that the relative shares of labor and capital differ by province, and one in which labor and capital are used in the same proportions nationally. We then study the impact on bank interest rates and margins of the adoption of a common technology across provinces in contrast to those of a system in which banks choose 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 substantial across more and less developed provinces.

Our strategy runs in two steps. First, we calibrate the model for each province separately by assuming completely independent industrial and banking sectors (i.e., as if provinces were different countries) for the provinces where reliable data are available, and we obtain endogenously both the values of the relative shares of productive factors and the loan rates for each province and the country as a whole. In this case, banks are modeled as local banks that adopt more or less capital-intensive technologies in different provinces according to the availability of local resources. In the alternative scenario, we take the value for the relative factor shares obtained at the national level in the first stage and impose it on the production function of banks in all the provinces. In this case, we calibrate the model under the assumption that the technology adopted is common; however, since the observed values for the capital intensities of banks display substantial variability across provinces, the respective values of the productivity parameters differ, as, consequently, do the interest rates charged.

By comparing the productivity of the system under these two alternative scenarios, we find that when Chinese banks adopt a common technology across provinces, notwithstanding the different local conditions, the productivity levels that they achieve halve with respect to those achieved by banks adopting differentiated technologies that adapt to different local labor market conditions. These productivity differentials induce a significant difference in the steady-state interest rate on loans and bank margins, ranging from 0.1% to 0.3% of
GDP per year at the regional level. In the final step of the work, we assess the impact of different types of productivity shocks on provinces by analyzing the impulse responses. We find that when banks adopt a common technology, the response to shocks is rather similar across provinces, while it is substantially different when the technology is adapted to local conditions. Moreover, interest rate variations following shocks are much larger when banks adopt a common technology and this effect is particularly strong in China’s less developed regions in which the industrial sector is less capital-intensive.

The presented framework can be extended to expand the role of banks in financing fixed capital investment. However, the basic results would not alter substantially because changes in demand for loans have a small impact on interest rates; therefore, only the magnitude of our results would increase with higher demand for loans. Similarly, in the presence of market power, while the basic mechanism would remain identical, the results would be amplified since banks would charge a mark-up on marginal costs. Our model is based on a banking framework alternative to that of models such as Gertler and Karadi (2011) in which financial frictions play an important role by, for example, making access to equity capital costly. In a more complicated setting including financial frictions, the framework would become more complicated and other shocks would play a relevant role. Nonetheless, the basic mechanism that we individuate would still play an important role. To summarize, our model provides baseline results for the simplest possible environment, but the results would be stronger in a more complicated framework.

The rest of this paper is organized as follows. Section 2 provides the empirical evidence on the Chinese banking system, Section 3 provides the set-up of the model, and Section 4 describes our calibration strategies. In Section 5, we discuss our main findings, while Section 6 concludes.
2 Lending and interest margins across Chinese banks

Obtaining the information needed to calculate the interest margins of commercial loans across China’s provinces is challenging. Although the micro-level databases of local commercial banks in China have data on loan portfolios that would in principle allow us to calculate interest margins, these data are sparse. Further, large and medium-sized national commercial banks only provide financial reports for consolidated accounts and no information is available for provincial-level operations. We therefore use another more direct, but unfortunately rather incomplete data source, which is the regional financial operation reports published by the People’s Bank of China. These provincial-level reports provide detailed information on the proportion of loans falling into different loan rate ranges, thereby allowing us to calculate the weighted average loan rates by province. We further assume the marginal deposit rate to be equal to the one-year SHIBOR rate and define the gap between these two values as the interest margins charged.

Interest margins differ significantly by province because the loan market in China is segmented at the provincial level. Figure 1 displays the interest margins by province, Figure 2 illustrates the statistical distribution of interest margins, and Figure 3 maps the differences in margins by geographical location. We carried out several on-site interviews with regulators and bankers in China and found that most lending in China is made by local banks or local branches of national banks. Lending across provincial borders barely exists, with the only exception the large syndicated loans provided through bank consortiums, but these are uncommon. Lending activities are managed locally because monitoring is much more efficient at this level, particularly since local staff rely on soft information and personal relationships to evaluate the reliability and entrepreneurial skills of potential borrowers. Moreover, by managing the payments of their customers banks obtain real-time information on the cash flows of borrowers, which is important for assessing price risks. Hence, even large banks do not use internal markets to reallocate excess deposits across provinces, and the managerial organization of the branches is strictly based on provincial-level organizations. Therefore, the
Figure 1: Interest margins among Chinese provinces
surplus of deposits is managed at the level of the consolidated national entity and allocated through the interbank market or by purchasing securities, bonds in particular.

Since the labor market is segmented similarly at the provincial level, we can assume that each province in China runs as a quasi-independent economy. Hence, regional economic features play an important role in determining the interest margins. Although the industrial sectors of provinces display all types of interactions (i.e., trade in final goods and intermediate products), most vertically integrated firms operating nationally whose value chains are built across provinces use local, independently managed subsidiaries. Hence, demand for loans is not sensitive to the loan rates charged by the commercial banks in other provinces. Similarly, since commercial banks do not lend to other provinces, their loan supply simply reflects the cost of local resources. To the extent that industrial firms have limited possibilities to substitute bank lending to meet their working capital requirements, given the limited development of the commercial paper market in China, demand for loans is inelastic, and variations in loan rates largely depend on supply-side innovations affecting marginal costs.

Loan market equilibria across provinces are heterogeneous both because of differences in loan demand and because the availability of resources differs among provinces, with labor costs in particular displaying substantial variability.

Interest margins are normally analyzed exclusively at the micro level to reflect the marginal cost of liabilities, the impact of capital regulation, default risk, and so on. Much less attention has been paid to the relationship between interest margins and the macroeconomic cycle. For example, Diallo and Zhang (2017) use panel data for 31 provinces and eight industrial sectors over 2001–2013 to analyze empirically the relationship between bank concentration and economic growth, finding evidence that bank concentration has a negative impact on growth. The empirical analysis of the bank interest margins of Chinese banks by Zhou and Wong (2008), however, provides strong support for our modeling strategy, since they find that the main variable affecting the interest margins of Chinese banks is operating costs, while changes in credit quality have a more modest impact.
Figure 2: Statistical distribution of interest margins among Chinese provinces

Figure 3: Geographic distribution of interest margins among Chinese provinces
2.1 Empirical analysis of bank interest margins across Chinese banks

To analyze the interest margins of Chinese banks, we estimate the dynamic model of bank interest rates and margins developed by Dia and Giuliodori (2012) that is an extension of the dynamic model proposed by Elyasiani et al. (1995) to an environment in which banks benefit from market power.

We estimate a general specification of the interest margins obtained under the assumption of no-portfolio separation:

\[ r_{L_{t+1}} - r_{D_{t+1}} = \zeta_0 + \zeta_1 r_{L_t} - \zeta_2 r_{D_t} + \zeta_3 r_{L_{t-1}} - \zeta_4 r_{D_{t-1}} + \]
\[ + \zeta_5 r_{D_{t+1}}^B + \zeta_6 r_{L_t}^B + \zeta_7 Y_{t+1} + \zeta_8 Y_t + \zeta_9 c_{t+1} + \zeta_{10} f_{t+1} + \zeta_{11} llp_{t+1}. \]  

Here, net interest revenues (NIR in the table) are regressed on the first and second lags of interest revenues (IR) and interest expenses (IE), the contemporaneous and lagged values of either long-term (IRL) or short-term (IRL) interest rates and the rate of growth in GDP (LNY), and the contemporaneous values of fees expressed as a ratio to total assets \( f_{t+1} \) (FEES), costs \( c_{t+1} \) (COSTS), and loan loss provisions \( llp_{t+1} \) (LLP), both again expressed as a ratio to total assets. Because of the dynamic nature of the specification, we use both the fixed-effects OLS and the two-step difference Arellano–Bond GMM estimators.\(^4\)

Columns (1) and (2) of Table 1 show the results of the specification of the model for long-term interest rates, estimated with these two methods, while Columns (3) and (4) of Table 1 show the results for short-term interest rates. Data are obtained from both SNL Financial and Orbis (Bankscope) to extend the sample as much as possible. Our data cover banks with at least $5 billion of assets, and we exclude any bank for which we do not have at least five years of data for all the variables and the three policy banks. A few isolated

\(^4\)We apply both two-step difference and system GMM methods, finding similar results. We only report the estimates of the difference GMM method. All estimates are implemented with the STATA command \texttt{xtabond2}.
missing values are interpolated. We obtain an unbalanced panel of 136 banks. The panel is unbalanced since the sample for some banks goes back until 1999, while we have universal coverage from 2013 to 2017, the final period of the sample. We find that the parameters of the lagged values of \( IR \) and \( IE \) are significant at the 1% level, confirming that a dynamic specification is appropriate. The second lags are statistically marginally significant only for interest expenditure; therefore, the portfolio separation assumption, even if not really supported, provides a good approximation. We find that interest margins respond positively to both long- and short-term market interest rates. Hence, either banks can reprice loans more frequently than deposits, as suggested by Samuelson (1945), or they can exploit their monopolistic power by raising rates on loans more than those on deposits.

The most important result is that both operating expenses (costs) and loan loss provisions have a strong and significant impact on interest margins. The estimated coefficients of these variables are always significant at the 1% level. The results are nearly identical for the different estimation methods, while the absolute size of the parameters is different in the two specifications. In both cases, however, the impact of cost innovations is much larger, nearly double in one case and almost triple in the other.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( IR(-1) )</td>
<td>0.488***</td>
<td>0.375***</td>
<td>0.491***</td>
<td>0.394***</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.078)</td>
<td>(0.045)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>( IR(-2) )</td>
<td>-0.038</td>
<td>0.016</td>
<td>-0.051</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.048)</td>
<td>(0.034)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>( IE(-1) )</td>
<td>-0.527***</td>
<td>-0.469***</td>
<td>-0.529***</td>
<td>-0.501***</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.112)</td>
<td>(0.061)</td>
<td>(0.103)</td>
</tr>
<tr>
<td>( IE(-2) )</td>
<td>0.101*</td>
<td>-0.009</td>
<td>0.111**</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.059)</td>
<td>(0.055)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>( IRL )</td>
<td>0.003***</td>
<td>0.0004***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( IRL(-1) )</td>
<td>-0.001**</td>
<td>-0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( IRS )</td>
<td></td>
<td></td>
<td>0.0008***</td>
<td>0.0007***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>( IRS(-1) )</td>
<td></td>
<td></td>
<td>0.0006***</td>
<td>0.0003***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>( FEES )</td>
<td>0.135</td>
<td>0.241</td>
<td>0.069</td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.050)</td>
<td>(0.114)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>( COSTS )</td>
<td>0.879***</td>
<td>1.069***</td>
<td>0.803***</td>
<td>1.098***</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(0.223)</td>
<td>(0.116)</td>
<td>(0.171)</td>
</tr>
<tr>
<td>( LLP )</td>
<td>0.326***</td>
<td>0.512***</td>
<td>0.332***</td>
<td>0.553</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.098)</td>
<td>(0.084)</td>
<td>(0.088)</td>
</tr>
<tr>
<td>( LNY )</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.013)</td>
<td>(0.001)</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>( LNY(-1) )</td>
<td>-0.000</td>
<td>-0.002</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.011)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>

**Estimation Method**
- OLS
- GMM

**Country Sample**
- Full

**Sample period**
- 1999–2017
- 1999–2017
- 1999–2017
- 1999–2017
- 1999–2007

**Notes**: The dependent variable is NIR. OLS is the within-group estimator with robust standard errors. GMM is the two-step difference GMM estimator with Windmeijer-corrected standard errors. Standard errors are in parentheses. * = significant at 10%, ** = at 5%, and *** = at 1%. The value reported for the Hansen test is the \( p \)-value for the null hypothesis of instrument validity. The values reported for AR(1) and AR(2) are the \( p \)-values for first- and second-order autocorrelated disturbances. Obs = total number of observations and N = number of banks.
3 Model

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

3.1 Households

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

\[
\max_{\{c_t,h_t,k_t,d_s^t\}} E_0 \beta^t \left[ \frac{c_t^{1-\sigma} - 1}{1-\sigma} + \theta \frac{(1-h_t)^{1-\gamma} - 1}{1-\gamma} \right],
\]

s.t. \(c_t + k_t - (1 - \delta)k_{t-1} + d_s^t = w_t h_t + r_t k_{t-1} + r^d d_s^t + \Pi_t^b + \Pi_t^f\) \quad (2)

From the first-order conditions, we obtain

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

(3)

\[
w_t = \theta \frac{(1 - h_t)^{-\gamma} - \gamma}{c_t^{-\sigma}},
\]

(4)

\[
c_t^{-\sigma} = \beta E_t(c_{t+1}^{-\sigma})^d_t.
\]

(5)
3.2 Banks

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

\[ l_t^s = \min \{ d_t^d, z_t (k_t^b)^\kappa (h_t^b)^{1-\kappa} \} \]  \hspace{1cm} (6)

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

\[ \Pi_t^b = d_t^d + r_{t-1}^l l_{t-1}^s - r_{t-1}^d d_{t-1}^d - l_t^s - w_t h_t^b - r_t k_t^b, \]  \hspace{1cm} (7)

and households require banks to maximize the current marginal value of profits, which are thus subject to the discount factor \( \beta_t c_t^{-\sigma} \):

\[ \max_{\{k_t^b, h_t^b\}} \sum_{t=0}^{\infty} \beta_t c_t^{-\sigma} [(r_{t-1}^l - r_{t-1}^d) z_t (k_{t-1}^b)^{1-\kappa} - w_t h_t^b - r_t k_t^b]. \]  \hspace{1cm} (8)

Since banks minimize costs to obtain the optimal combination of inputs, we obtain banks’ demand for labor and physical capital:

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

\[ \beta E_t c_{t+1}^{-\sigma} \kappa z_t \left( \frac{k_t^b}{h_t^b} \right)^{-1} (r_t^l - r_t^d) = c_t^{-\sigma} r_t. \]  \hspace{1cm} (10)

3.3 Firms

Firms in this economy produce output \( y \) according to a Cobb–Douglas production function and standard input requirements; however, they face a liquidity constraint caused by the need to finance working capital in advance:

\[ y_t = a_t (k_t^l)^\alpha (h_t^l)^{1-\alpha}, \]  \hspace{1cm} (11)
\[(w_t h_t^f + r_t k_t^f)\mu = l_t^d. \] (12)

In each period, firms have to borrow from banks a proportion \(\mu\) of their expenses and at the same time pay back what they borrowed from commercial banks in the previous period:

\[\pi_t^f = l_t^d + y_t - r_{t-1}^l l_{t-1}^d - w_t h_t^f - r_t k_t^f, \] (13)

\[\pi_t^f = y_t - r_{t-1}^l (w_{t-1} h_{t-1}^f + r_{t-1} k_{t-1}^f)\mu - (w_t h_t^f - r_t k_t^f)(1 - \mu). \] (14)

Further, since firms need to maximize the current marginal value of profits,

\[
\max \sum \beta^t c_t^{-\sigma} \left[ a_t (k_t^f)^\alpha (h_t^f)^{1-\alpha} - r_{t-1}^l (w_{t-1} h_{t-1}^f + r_{t-1} k_{t-1}^f)\mu - (w_t h_t^f - r_t k_t^f)(1 - \mu) \right]. \] (15)

Hence, from the first-order conditions, we obtain firms’ demand for labor and capital:

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

\[c_t^{-\sigma} \left[ \alpha a_t (k_t^f)^{\alpha-1} (h_t^f)^{1-\alpha} - r_t (1 - \mu) \right] = \beta c_{t+1}^{-\sigma} r_t^l r_t \mu. \] (17)

### 4 Parametrization and quantitative analysis

#### 4.1 Calibration strategy

We use two sets of data to calibrate the model, macroeconomic data and banking industry data. In both, we consider provincial-level and national-level data. The macroeconomic data come from the National Bureau of Statistics of China and the financial statistics yearbooks of each province offer the financial reports of the provincial branches of commercial banks. While the data from these reports are not available in an electronic format, we obtained the figures for physical capital and the number of employees of every bank from the available publications to calculate the average capital intensity of banks for each province. We took
the year for which data on more provincial financial statistics are available as our calibration benchmark, namely 2013, and obtain data for six provinces: Shanghai, Hubei, Zhejiang, Henan, Hebei, and Gansu.

Although this sample is limited, it provides a clear picture of the economic heterogeneity across provinces in China. Shanghai is a mega metropolitan area; its economy is highly developed and much more capital-intensive than any other province in our sample, and its income per capita is far larger. Although Hubei is in mid-west China, its capital city, Wuhan, produces almost half of the province’s GDP and hosts its main economic activities. Zhejiang province is the main location for international trade-oriented manufacturing industries. Henan and Hebei provinces rely on capital-intensive industries such as steel, electricity, coal mining, and aluminum. Gansu is one of the least developed provinces in China with an economy heavily reliant on the mining industry. Hence, the wage differentials among these six provinces are substantial. Figure 4 displays the average wages for the provinces in our sample.

Two groups of targets are employed to pin down all the parameters. First, macroeconomic variables serve as the targets of the standard parameters: \( \{ \beta, \mu, \delta, \theta, \gamma, \sigma, a, \alpha \} \). Second, two sets of ratios that link the industrial and banking sectors, namely the ratio between the productivity factors of industrial firms and banks and the ratio between the capital intensity across these two sectors, are the targets for \( \{ z, \kappa \} \). For more details, see the technical details in Appendix 7.3.

Our model is a two-sector RBC model in which loans work as an essential intermediate input. Indeed, the interest margin in the model serves as the relative price of an intermediate input, namely bank loans. Therefore, the income share parameters \( \alpha \) and \( \kappa \) and technology levels \( z \) and \( a \) are related to the capital intensity ratio, technology level ratio, and relative prices across the industrial and banking sectors.

Accordingly, we develop two types of calibrations. In the first analysis, we obtain endogenously from the model both \( z \) and \( \kappa \) for each province separately as well as at the
Figure 4: Average wages for the sampled Chinese provinces

Table 2: Accounting for the provincial net interest margin difference in China

<table>
<thead>
<tr>
<th>Province</th>
<th>$\alpha$</th>
<th>$\mu$</th>
<th>$K/W$</th>
<th>$K_B/W_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hubei</td>
<td>0.50</td>
<td>0.71</td>
<td>164</td>
<td>245</td>
</tr>
<tr>
<td>Shanghai</td>
<td>0.43</td>
<td>1.47</td>
<td>414</td>
<td>611</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>0.47</td>
<td>1.38</td>
<td>229</td>
<td>279</td>
</tr>
<tr>
<td>Henan</td>
<td>0.50</td>
<td>0.61</td>
<td>164</td>
<td>209</td>
</tr>
<tr>
<td>Hebei</td>
<td>0.49</td>
<td>0.71</td>
<td>183</td>
<td>200</td>
</tr>
<tr>
<td>Gansu</td>
<td>0.49</td>
<td>1.22</td>
<td>107</td>
<td>161</td>
</tr>
<tr>
<td>National</td>
<td>0.46</td>
<td>0.90</td>
<td>170</td>
<td>362</td>
</tr>
</tbody>
</table>

Notes: $\alpha$ is the share of capital in the industrial sector, $\mu$ measures the debt-to-GDP ratio, and $K/W$ and $K_B/W_B$ are capital per worker in the industrial sector and capital per worker in the banking sector, respectively. $K_B/W_B$ is the value of capital per worker in the banking sector predicted by the model under the assumption that the technology is identical across China.

national level, using as targets both the ratios of capital intensity and the technology parameters across the industrial and banking sectors. We thus treat the banking system in each province as independent of those of other provinces under the assumption that the banks in each province are free to choose a more or less capital-intensive technology than the national average, as wages in provincial labor markets differ. We obtain a set of steady-state results for each set of parameters $\{r^d, z, \kappa\}$. We then define this case as the province-specific $\kappa$ technology or flexible technology calibration.

In the second calibration, we assume instead that banks adopt a common technology na-
tionally featuring the same capital share $\kappa$. More specifically, we take the national-level value of $\kappa$ obtained in the first analysis and impose it as an exogenous value in the parametrization for each province. Once the value of $\kappa$ is fixed, by targeting only the capital intensity ratios obtained at the provincial level that reflect the local conditions of market resources, we obtain a new set of results for the ratio $\frac{\bar{z}}{a}$, and consequently also the results for $r^d$ are different.\textsuperscript{5} We can then compare the performance at the provincial level of this version of the model with the one assuming that the $\kappa$ parameter is chosen in every province. As in Dia and Menna (2016), we use $r^d$ to evaluate the performance of the models (i.e., we compare the predicted interest margins with the actual ones). We label this case as the common $\kappa$ technology or identical technology calibration.

4.2 Explanation of the mechanism

Cobb–Douglas production functions describe a world in which production factors are substitutes. If the relative shares of two factors are fixed as normally assumed, the relative quantity of the two inputs, and thus the capital intensity, depends on the relative prices only. In our model, at the steady state, the rental rate of capital is $r_{ss} = \frac{1}{\beta} - 1 + \delta$. Hence, the rental cost of capital is the same as long as depreciation rates are identical across provinces. Although depreciation rates across provinces may differ because of the different composition of the stock of capital in light of the different productive structures of the economies, the rental cost of capital for the same industry across provinces is likely to be similar because any significant difference would drive cross-province investment flows that would restore the equilibrium. Labor, on the contrary, does not move freely across provinces and wage differentials are persistent. Consequently, in our calibration, the differences in the capital intensities of banks among provinces reflect the differences in wages only. When imposing the same $\kappa$ parameter nationally, given the Cobb–Douglas production function, we thus impose that in every province, independently of the total expenditure on resources, a constant

\textsuperscript{5}See equation 45.
and identical share of expenditure is always allocated to capital. On the contrary, under the flexible technology calibration strategy, we assume that banks can choose alternative productive technologies, thereby allowing them to use capital more or less intensively. Indeed, since \( r^d_{ss} = r^d_{ss} + \frac{w_{ss}}{\beta(1-\kappa)z_{ss}} \left( \frac{k^b_{ss}}{h^b_{ss}} \right)^{-\kappa} \), any change in wages is in part offset by a change in hours \( h^b_{ss} \), and then passed through as a change in \( r^d_{ss} \). However, since the productivity parameter \( z_{ss} \) is endogenous in our calibration, it is affected by wage variations. But how far are changes in wages accommodated by variations in capital intensity and how much instead do they generate changes in productivity or loan rates? The answer depends on all the parameter values of the model, but we know that the impact of wages on hours is a decreasing and convex function of \( \kappa \). Indeed, the number of hours is equal to \( h^b_{ss} = \frac{k^b_{ss}}{w_{ss} \left( r^d_{ss} - r^d_{ss} \right) \beta(1-\kappa)z_{ss}} - \frac{1}{\kappa} \), and therefore \( \frac{\partial h^b_{ss}}{\partial w_{ss}} = -\frac{1}{\kappa} k^b_{ss} \left( \frac{w_{ss}}{\left( r^d_{ss} - r^d_{ss} \right) \beta(1-\kappa)z_{ss}} \right) - \frac{1+\kappa}{\kappa} \).

If a bank operating in different provinces is forced to spend the same proportion of resources in capital with a homogeneous rental cost of capital, variations in wages will largely generate corresponding variations in productivity and interest rates on loans. A bank willing to obtain the same productivity across provinces could mitigate somehow the impact of wage differentials only if technologies with different \( \kappa \) values are available, allowing it to allocate expenditure to resources at different proportions. For different \( \kappa \) values, the response of hours to wage innovations varies substantially. If technologies with different values of \( \kappa \) were available, a bank operating in a province with high wages would in this case choose a substantially more capital-intensive technology, thereby spending a larger share of resources on physical capital.

5 Results

5.1 Steady state

Figure 5 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 and
the orange and gray bars the interest margins implied by our model. The orange bar is obtained by assuming that banks from different provinces use different technologies, while the gray one is obtained by assuming that banks use the same technology across provinces. The proportion of total interest margins that can be explained by our model ranges from 20% to 40%, with the model calibrated for the whole country explaining 33% of the margins, which is a considerable share considering that we do not include the impact of default risks or regulatory costs. The impact of the cost of credit on the income of all listed banks for the whole country is 0.60%, while the average for 2013–2015 is 0.93%. After considering the impact of loan loss provisions, the national model explains 48% of the interest margin, and we obtain similar results for most provinces. The remaining wedge is explained by the substantial market power that benefits Chinese banks given the limited availability of deposit substitutes. Table 3 displays the steady-state results of the model. In particular, the table reports the model-generated net interest margins in two columns (NIM-p and NIM-I), while the third to last column displays the actual value of the net interest margin.

Table 3: Accounting for the provincial net interest margin difference in China

<table>
<thead>
<tr>
<th>Province</th>
<th>( K_B/W_B - n )</th>
<th>Ratio1</th>
<th>Ratio2</th>
<th>( \frac{z_t}{TFP} )-p</th>
<th>( \frac{z_t}{TFP} )-I</th>
<th>NIM</th>
<th>NIM-p</th>
<th>NIM-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hubei</td>
<td>392</td>
<td>1.49</td>
<td>2.39</td>
<td>108</td>
<td>0.61</td>
<td>60.5</td>
<td>2.34</td>
<td>0.73</td>
</tr>
<tr>
<td>Shanghai</td>
<td>773</td>
<td>1.48</td>
<td>1.87</td>
<td>85</td>
<td>0.67</td>
<td>61</td>
<td>2.68</td>
<td>0.88</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>516</td>
<td>1.22</td>
<td>2.26</td>
<td>154</td>
<td>0.58</td>
<td>70</td>
<td>2.94</td>
<td>0.58</td>
</tr>
<tr>
<td>Henan</td>
<td>417</td>
<td>1.27</td>
<td>2.54</td>
<td>132</td>
<td>0.56</td>
<td>57</td>
<td>3.28</td>
<td>0.67</td>
</tr>
<tr>
<td>Hebei</td>
<td>466</td>
<td>1.10</td>
<td>2.55</td>
<td>165</td>
<td>0.52</td>
<td>58</td>
<td>3.48</td>
<td>0.59</td>
</tr>
<tr>
<td>Gansu</td>
<td>262</td>
<td>1.50</td>
<td>2.45</td>
<td>112</td>
<td>0.61</td>
<td>64</td>
<td>3.16</td>
<td>0.70</td>
</tr>
<tr>
<td>National</td>
<td>362</td>
<td>2.13</td>
<td>2.13</td>
<td>63</td>
<td>0.72</td>
<td>63</td>
<td>2.85</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Notes: \( K_B/W_B - n \) is the capital intensity for the banking industry predicted by the model when assuming that a national common technology is adopted across all provinces. Ratio1 is the actual ratio between capital per worker in the banking sector and capital per worker in the industrial sector. Ratio2 is the ratio between capital per worker in the banking sector and capital per worker in the industrial sector predicted by the model when assuming an identical production function for banks across China. \( \frac{z_t}{TFP} \)-p is the ratio between the total factor productivity (TFP) of banks and that of the industrial sector when banks adopt the provincial-level technology and \( \frac{z_t}{TFP} \)-I are the values when banks adopt the common \( \kappa \) technology. NIM is the net interest margin, NIM-p is the margin generated by the model when banks adopt the provincial-level technology, and NIM-I is the margin generated by the model with the common \( \kappa \) technology.

These data are from the “Listed banks in China 2015 review and outlook” produced by Ernst & Young.
In all the provinces of our sample, the model assuming a common national share of capital explains a larger share of the observed interest margins. However, we cannot conclude that the model assuming the common technology is necessarily a better framework because the calibration strategies adopted in these two cases fit partially different data. The main result of the analysis is instead that when banks adopt a technology by imposing a common fixed capital share $\kappa$, banks are substantially less productive, and hence they charge higher interest rates on loans. The fifth and seventh columns of Table 3 report the relative productivity data: $\hat{z}_{TFP}^T$ shows the results for the provincial-level technology and $\hat{z}_{TFP}^I$ shows the results for the common $\kappa$ technology. In Shanghai province, where the difference between the two is smaller, when banks adopt the national technology, their productivity declines by 28%, while in Hebei province, productivity is lower by a staggering 65%. In most other provinces, productivity is roughly halved. The common banking technology imposes an equal share of resource expenditure on capital in every province and is much more capital-intensive and much less productive than the alternative technologies that can be adopted at a provincial level if different technologies with different levels of $\kappa$ are available at the local level. Hence, if banks can adopt more or less labor-intensive technologies according to the relative abundance of labor in the local province, they are far more productive in all the provinces in our sample. If local banks adopt different labor-intensive technologies than banks operating nationally, our model predicts that they may have a substantial competitive advantage, as they are far more productive. Symmetrically, lower productivity induces a larger loan rate and the cost of this lower productivity of banks is borne by firms demanding bank credit. Therefore, in a province dominated by banks adopting a locally chosen technology, customers are charged substantially lower loan rates.

To understand the results, note that the national technology that we obtain from the model imposes a larger capital share than those of any of the provincial ones, since $\kappa$ is larger for the national case. The higher $\kappa$ 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 close to the simple average across our sample at the provincial level, which is 309 thousand yuan per worker. Instead, the sixth column of Table 3 shows that the ratio between the capital intensity of banks and that of the economy (Ratio 1) is larger at the national level than in any province, including Shanghai, where banks are more capital-intensive than the average, but where however the industrial system is also extremely capital-intensive, far more than the average. The cost imposed by the fixed share of capital in the banking industry thus largely depends on the distribution of capital across provinces. When the differences in the capital intensity of the industrial sector are substantial, reflecting the large differential in wage costs, as is the case in China, the productivity losses for banks are large in all the provinces that are far from the average.

The productivity of the banking industry and consequently loan rates depend on the capital intensity of the industry relative to that of other industrial sectors, and this is a decreasing function of both Ratio1 and the ratio between $\kappa$ and $\alpha$ (these two ratios are closely linked when $\kappa$ is obtained endogenously). In this case, the higher values of this ratio are associated with lower productivity and higher loan rates for a simple reason. Banks are a rather small part of the economy, and the relative prices of wages and rental cost of capital reflect the availability of resources and demand originating in goods-producing industries, which is strongly influenced by the technology adopted as well as the value of $\alpha$. At the steady state, the rental cost of capital is exogenous and equal for all provinces. It is also substantially larger than $r_d$. A value of $\kappa$ larger than that of $\alpha$ implies that the share of capital in banks is larger than in the rest of the economy, and capital is expensive. The larger this difference, the more banks can benefit when wages are high with respect to the rental cost of capital. In our sample of provinces, banks use too much physical capital, are poorly productive, and loan rates are high because the ratio between the capital intensity of banks and that of goods-producing industries, Ratio1, measured at the national level is larger than in any of the provinces of our sample. Since our sample does not include Beijing, it may thus be tempting to conclude that national values include the cost of a large physical
infrastructure necessary to work at the national level. However, bank capital intensity in Shanghai is much larger than the national average, while Ratio1 is much lower than the national average. This finding implies than in several of the provinces outside our sample, the capital intensity of banks is far larger than that of other industries, particularly in less developed provinces. Our data support this hypothesis since the largest reported value for Ratio1 is that of Gansu, the least developed province in our sample.

Bank productivity levels are similar across provinces, as deviations from the national average are rather small. The potential benefits from adopting local technologies would be huge, as productivity parameters could roughly double. The potential benefits would be larger in provinces where the capital intensity of banks relative to that of other industries is lower, namely Hebei and Zhejiang. These benefits are large whenever the differences in wages are substantial among regions, as for Chinese provinces. Nagayasu and Liu (2008) provide convincing empirical evidence that the wage differentials among provinces in China are substantial.

The adoption of a national technology requires a large amount of physical capital since in more developed regions, where banks generate a disproportionate share of revenues, wages are far higher than in less developed ones. When banks adopt the more capital-intensive national technology, firms in regions in which wages are lower face larger interest costs on loans than would be charged by a segmented local banking system. Adopting a similar technology, however, does not necessarily imply a smoothing of the differences among the interest margins across provinces, as shown in Table 3. Banks do not charge interest rates on loans in Shanghai that are substantially lower than they would be otherwise, while charging more to customers in less developed provinces. Customers would be better off in both more and less developed provinces with a decentralized system because such a system is more efficient in exploiting the relative price differentials. In our sample, all banks are more efficient when they choose a more labor-intensive technology, including those in Shanghai. However, in other provinces, banks might be better off by adopting more capital-intensive
technologies than the national one.

5.2 Impulse responses

We examine two sets of impulse response functions in this subsection. In the first set displayed in Figure 6, we use the fully flexible calibration and compare the responses of regional banks with those of national banks by adopting a production function that reflects the national data. In the second set displayed in Figure 7, we use the common $\kappa$ calibration and compare the responses with those of the fully flexible case.

The adoption of a national rather than a segmented banking market has an important implication since the response to shocks differs in these two cases. In China, the difference in the response is substantial. Indeed, when analyzing the impulse responses, in all the provinces of our sample and for the national-level data, the capital intensity of banks is larger than that of the rest of the economy. This feature is rare in other developed economies, although the Spanish sample studied by Dia and Menna (2016) has this feature. As discussed in Dia and Menna (2016), when this is the case, the response of interest rates to productivity shocks becomes stronger, either when a positive (negative) shock hits the industrial sector, producing higher (lower) loan rates, or when a positive (negative) shock hits banks, producing lower (higher) loan rates.

We conduct a dynamic analysis of our model to investigate the impact of two kinds of productivity shocks on the economy, namely those hitting the industrial sector and those hitting the banking sector, and focus on the impact on interest margins and working hours in the banking sector. Figure 6 illustrates the impulse responses from the model.

The impulse responses we obtain suggest that the response of banks adopting the national technology to a productivity shock hitting the industrial sector is stronger than that of regional banks. A positive shock induces a larger increase in loan rates, which follows the increase in wages produced by the productivity shock. Similarly, a positive bank productivity shock generates a stronger decline in interest rates at the national level than under
regional (i.e., segmented) banking systems. This result suggests that the adoption of a national technology generates two contrasting effects. On the one hand, positive national firm productivity shocks generate negative asymmetric effects that imply negative spillovers for poorer regions that end up paying disproportionately larger loan rates. On the other hand, less productive provinces characterized by lower wages benefit disproportionately from the reductions in loan rates generated by a positive shock hitting banks (but the opposite would hold for a negative one, for example if a regulatory intervention makes banks less productive).

Similarly, the capability of banks to smooth regional idiosyncratic productivity shocks hitting firms is also much reduced when banks adopt a common national technology. Figure 6 displays a set of responses for all the sample provinces to compare the response to local productivity shocks when loan markets are segmented (line) with the national response to the same shock. These responses highlight that all provinces would always be better off with segmented systems, since those whose response to a shock hitting firms is closer to the national one are those whose response to a shock hitting banks is more divergent from the national one, and vice versa. This problem is relevant in view of the substantial degree of heterogeneity in the business cycle dynamics across Chinese provinces and regions found by Poncet and Barthélemy (2008).

Figure 7 displays the impulse responses obtained under the assumption that banks use
Figure 6: Impulse response functions of bank interest margins and bank working hours following a one standard deviation shock to firm and bank TFP.
a technology with the same capital share $\kappa$ across all provinces. The responses are similar to those in the flexible technology case; however, the impact of the shock on loan rates, particularly on interest margins, is much less differentiated across provinces, and this becomes similar to that of the national technology. The impulse responses are thus more similar to those assuming a common national technology than to those assuming regionally differentiated ones. Figures 8 and 9 compare the responses to the two types of productivity shocks for each province under the two possible assumptions for banking technological choices.

Although the magnitude of the differences in the impact differs by province, the size of the productivity shock to firm TFP is always significant. In particular, for Zhejiang province, as was the case for the steady-state results, we find that the immediate impact of the shock is twice as large when banks adopt a common technology. In this case, however, the shock is less persistent and dies out earlier. Further, technology shocks may be more persistent than we assumed, and in the presence of market power, the results would be far stronger because in this case loan rates would be set as a mark-up on the marginal costs we have analyzed, with the mark-up boosting the impact of such shocks.
Figure 7: Impulse response functions of bank interest margins and bank working hours following a one standard deviation shock to firm and bank TFP when banks adopt a technology with the same capital share.
Figure 8: Impulse response functions of loan rates following a one standard deviation shock to firm productivity when banks adopt a technology with the same capital share (blue line) and when banks adopt different technologies across provinces (red line).
Figure 9: Impulse response functions of loan rates following a one standard deviation shock to bank productivity when banks adopt a technology with the same capital share (blue line) and when banks adopt different technologies across provinces (red line).
6 Conclusion

This study provides evidence of the existence of a substantial dispersion among the interest margins charged by commercial banks in Chinese provinces and then builds a parsimonious dynamic stochastic general equilibrium model with both banking and goods-producing firms. Firms require loans from commercial banks to finance their working capital needs, and since both industrial firms and banks use real resources to provide goods and services, resource cost variations affect both sectors and transmit shocks between the two. This model can thus explain a considerable share of the interest margins charged by Chinese banks in different provinces. The results of the model suggest that when Chinese banks adopt a common national technology rather than more or less capital-intensive technologies in different provinces in view of the local dynamics of resource costs, their productivity suffers declines ranging from 28% to 65% in our sample provinces. For the province suffering the worst productivity decline, the cost is around 0.3% of GDP per year. It must be stressed that the results we present are those for a simple model with the smallest possible amount of frictions. In the presence of market power or financial frictions inducing costly access to equity capital, the results would remain qualitatively similar, but would be strongly amplified. Similarly, in this framework, we limit the role of banks to financing the working capital needs of industrial firms; however, if we more realistically imposed a larger role for bank finance in the Chinese economy, the magnitude of the costs induced by these productivity differentials would rise proportionally.

Several noteworthy implications emerge. Since the differences in wages in Chinese provinces are substantial, the adoption of a national technology implies an inefficient industrial structure for the banking industry. In all the provinces we analyzed, banks are more productive when they adopt a more labor-intensive technology. However, it is likely that in other provinces, it might be efficient to adopt a technology that is more capital-intensive than the national one to face the challenge of wage inflation.

Finally, the adoption of a common nationwide technology generates a stronger response
of loan rates to productivity shocks than is the case when banks adopt technologies at the provincial level and the capability of banks to smooth regional idiosyncratic productivity shocks hitting firms declines substantially.

References


7 Appendix I: Market-clearing conditions and steady-state values

7.1 Market-clearing conditions

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

\[ k_{t-1} = k^f_t + k^b_t, \]  
(18)

\[ h_t = h^f_t + h^b_t, \]  
(19)

\[ d^d_t = d^s_t, \]  
(20)

\[ l^d_t = l^s_t, \]  
(21)

\[ c_t + k_t - (1 - \delta)k_{t-1} = y_t. \]  
(22)

7.2 Steady state

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

\[ 1 = \beta(r_{ss} + 1 - \delta), \]  
(23)
\[ w_{ss} = \theta \frac{(1 - h_{ss})^{-\gamma}}{c_{ss}^{-\sigma}}, \tag{24} \]

\[ 1 = \beta r_{ss}^d, \tag{25} \]

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

\[ \beta\kappa z_{ss} \left( \frac{k_{ss}^b}{h_{ss}^b} \right)^{\kappa-1} (r_{ss}^l - r_{ss}^d) = r_{ss}, \tag{27} \]

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

\[ \alpha a_{ss} \left( \frac{k_{ss}^f}{h_{ss}^f} \right)^{\alpha-1} = r_{ss}(\beta r_{ss}^f \mu + 1 - \mu), \tag{29} \]

\[ (w_{ss} h_{ss}^f + r_{ss} k_{ss}^f)\mu = z_{ss}(k_{ss}^b)^\kappa (h_{ss}^b)^{1-\kappa}. \tag{30} \]

### 7.3 Calibration

The labor income share in the model does not correspond to \( \alpha \) directly, and it follows instead

\[ \frac{w_{ss} h_{ss}}{y_{ss}} = \frac{1 - \alpha}{\beta r_{ss}^f \mu + 1 - \mu}. \tag{31} \]

However, given the numerical values of \( \mu, \beta, \) and \( r_{ss}^l - r_{ss}^d, \) since the error is small, we use the labor income share as a proxy of \( \alpha. \) \( \alpha \) is calibrated to reflect the capital income share of each province, or that at the national level, respectively. \( \mu \) is calibrated to match the debt-to-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 estimate the proportion of industrial and business loans of the total to be 75%. To calculate \( \mu, \) we use total loans, but we exclude mortgages by multiplying the figures obtained from the regional financial statistics yearbook by 0.75.
By using (29), the capital intensity in the industrial sector is obtained as

\[
k_{\text{fs}}^f = \left[ r_{\text{ss}} \beta r_{\text{ss}}^f \mu + (1 - \mu) r_{\text{ss}} \right]^{\frac{1}{\alpha - 1}}, \tag{32}
\]

where \( r_{\text{ss}} \) can be obtained from equation 23, \( a_{\text{ss}} \) is normalized to unity, and \( r_{\text{ss}}^f \) is thus the only unknown. By rearranging (27), we can express the capital intensity of banks as

\[
k_{\text{bs}}^b = \left[ r_{\text{ss}} \beta \kappa z_{\text{ss}} \left( r_{\text{ls}} - r_{\text{ds}} \right) \right]^{\frac{1}{\kappa - 1}}, \tag{33}
\]

After substituting the value of \( r_{\text{ds}}^d \) obtained from equation (25) and the value of \( w_{\text{ss}} \) obtained from equation (26) into equation (28), we obtain

\[
\frac{(1 - \alpha) a_{\text{ss}} \left( k_{\text{fs}}^f \right)}{\beta (1 - \kappa) z_{\text{ss}} \left( k_{\text{bs}}^b \right)^{\kappa} \left( r_{\text{ss}}^f - r_{\text{ss}}^d \right)} = 1 - \mu + \mu \beta r_{\text{ss}}^f. \tag{34}
\]

By substituting the values of the capital intensities of firms and banks obtained from equations (32) and (33) into equation (34), we obtain

\[
\frac{(1 - \alpha) a_{\text{ss}} \left( \frac{k_{\text{fs}}^f}{h_{\text{fs}}^f} \right)^{\alpha}}{\beta (1 - \kappa) z_{\text{ss}} \left( \frac{k_{\text{bs}}^b}{h_{\text{bs}}^b} \right)^{\kappa} \left( r_{\text{ss}}^f - r_{\text{ss}}^d \right)} = 1 - \mu + \mu \beta r_{\text{ss}}^f. \tag{35}
\]

From the former expression, we can isolate the value of \( z_{\text{ss}} \),

\[
z_{\text{ss}} = \left( \frac{(1 - \alpha) a_{\text{ss}} \left[ r_{\text{ss}} \beta r_{\text{ss}}^f \mu + (1 - \mu) r_{\text{ss}} \right]^{\frac{1}{\alpha - 1}}} {\beta (1 - \kappa) \left( \frac{k_{\text{bs}}^b}{h_{\text{bs}}^b} \right)^{\frac{\kappa}{\alpha - 1} \left( 1 - \mu + \mu \beta r_{\text{ss}}^f \right)} / \left( r_{\text{ss}}^f - r_{\text{ss}}^d \right)} \right)^{\frac{1}{1 - \kappa}}. \tag{36}
\]

and after substituting the value of \( z_{\text{ss}} \left( r_{\text{ss}}^f - r_{\text{ss}}^d \right) \) obtained from the former equation, (33)
\[
\frac{k^b_{ss}}{h^b_{ss}} = \left( \frac{(1 - \alpha) a_{ss} \left[ r_{ss} \beta r^l_{ss} \mu + (1 - \mu) r_{ss} \right]^{\frac{\alpha}{\alpha - 1}}}{\beta (1 - \kappa) (1 - \mu + \mu \beta r^l_{ss}) \left( \frac{(1 - \alpha) a_{ss} \left[ r_{ss} (\beta r^l_{ss} \mu + 1 - \mu) \right]^{\frac{\alpha}{\alpha - 1}}}{\beta (1 - \kappa) \left( \frac{r_{ss}}{\kappa} \right)^{\frac{\alpha}{\alpha - 1}} (1 - \mu + \mu \beta r^l_{ss})} \right)^{1 - \kappa}} \right)^{\frac{1}{\kappa}}. \quad (37)
\]

We finally express the capital intensity of banks as a function of \( r^l_{ss} \) and a set of exogenous parameters. By multiplying both sides of equation (28) by \( h^f_{ss} \) and both sides of equation (29) by \( k^f_{ss} \), summing the two equations, we obtain

\[
a(k^f)^\alpha (h^f)^{1 - \alpha} = (w_{ss} h^f_{ss} + r_{ss} k^f_{ss}) (\beta r^l_{ss} \mu + 1 - \mu). \quad (38)
\]

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

\[
y_{ss} = (w_{ss} h^f_{ss} + r_{ss} k^f_{ss}) (\beta r^l_{ss} \mu + 1 - \mu). \quad (39)
\]

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

\[
y_{ss} \div l_{ss} = \frac{\beta r^l_{ss} \mu + 1 - \mu}{\mu}. \quad (40)
\]

Substituting the production functions,

\[
\frac{a_{ss} \left( \frac{k^f_{ss}}{h^f_{ss}} \right)^\alpha h^f_{ss}}{z_{ss} \left( \frac{k^b_{ss}}{h^b_{ss}} \right)^\kappa h^b_{ss}} = \frac{\beta r^l_{ss} \mu + 1 - \mu}{\mu}. \quad (41)
\]

We can now substitute the values of the capital intensities of the two sectors that we have obtained in equations 32 and 37 into equation 41. Moreover, assuming, as is standard in the
literature, that total hours worked are 0.3, since then

\[ h_{ss} = h_{ss}^f + h_{ss}^b = 0.3, \]  \hspace{1cm} (42)

it becomes possible to obtain \( h_{ss}^f \) and \( h_{ss}^b \) from the above two equations, and the value of the capital stock of the two sectors since we had previously obtained the values of \( \frac{k_f}{h_{ss}^b} \) and \( \frac{k_b}{h_{ss}} \).

Summing up, from equation 41, we can obtain the values of hours and capital in each of the sectors as a function of \( r_L, z, \kappa \), and a series of exogenous parameters:

\[ \frac{k_{ss}^b}{h_{ss}^b} = \frac{k_{ss}^b}{h_{ss}^b} \left( \frac{k_{ss}^b + k_{ss}^f}{0.3} \right). \]  \hspace{1cm} (43)

After substituting the values of \( k_{ss}^b \) and \( k_{ss}^f \) previously derived in equation 43, we have a first equation linking the three endogenous parameters to a combination of exogenous ones. We need two more equations to obtain a system of three equations for the three unknowns. Since we can obtain data on the capital intensity of banks and on the aggregate economy that we call \( \text{Ratio}1 \), we can use the following relationship:

\[ \frac{k_{ss}^b}{h_{ss}^b} = \text{Ratio}1. \]  \hspace{1cm} (44)

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 obtain from the data. Assuming the exponent of the production function for banks to be the same as that of industrial firms, 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 banks’ production function endogenously. Indeed, one of the crucial aims of this work 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 are more or less capital-intensive. In our calibration, we thus force the ratio between the productivity parameters to respect the corresponding one obtained from the data for an endogenous value of $\kappa$:

$$\frac{z}{a} = z = \frac{L/W_B}{(K_B/W_B)^{\kappa}} TFP,$$  \hspace{1cm} (45)

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 TFP in 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 allows us to solve a system of five equations, namely 41, 42, 43, 44, and 45, in the five unknowns, $r_L$, $h_b$, $h_b$, $z$, and $\kappa$.  
