

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/126799/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Matthews, Kent and Zhiguo, Xiao 2020. Rational cost inefficiency and convergence in Chinese banks. *Economic Modelling* 91 , pp. 696-704.
10.1016/j.econmod.2019.07.008

Publishers page: <https://doi.org/10.1016/j.econmod.2019.07.008>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Rational Cost Inefficiency and Convergence in Chinese Banks

Kent Matthews (Cardiff Business School, Cardiff University, UK, and School of Public Finance and Taxation, Zhongnan University of Economics & Law, Wuhan, PR China)

and

Zhiguo Xiao (School of Management, Fudan University School, Shanghai, PR China)

July 2019

Abstract

We argue in this paper that the widely held view that the high level of cost inefficiency in Chinese banks is indicative of management deficiency is questionable. Using a non-parametric bootstrapping method, we decompose cost-inefficiency into technical-inefficiency and allocative-inefficiency and argue that allocative inefficiency should be viewed as rational in the sense that they are the outcomes of political and administrative constraints. Published studies of Chinese banks tend to place cost-inefficiency in the region of 50%. Such estimates would suggest that the management of Chinese banks was grossly inefficient. Over-utilization of labour in Chinese banks meets the social goals of the semi-planned economy. This paper finds evidence of over-utilization of labour and suggests that allowing for rational allocative inefficiency, Chinese banks have converged on a low level of technical inefficiency indicating a high level of management efficiency

Corresponding Author: Kent Matthews
Cardiff Business School
Colum Drive
Cardiff CF10 3EU
matthewsk@cardiff.ac.uk
+44(0)7792 421053

Keywords: Allocative inefficiency, Banking; China;

JEL codes: D23, G21, G28

This paper has benefited from helpful comments from Jaap Bos and Mette Asmild, members of the La Pour-Societe of Fudan University, and participants of the 2018 Vietnam Symposium of Banking. Zhiguo Xiao would like to thank research support from the National Natural Sciences Foundation of China (Grant #: 71473040, 11571081, 71661137005).

1. Introduction

Bank efficiency has always been a popular subject of research in China, and a large number of studies have been published in Chinese scholarly journals (e.g., Xue and Yang, 1998, Wei and Wang, 2000, Zhao, 2000, Zhao, Zhou and Jiang, 2001 and Qing and Ou, 2001 with non-parametric methods, and Qian, 2003, Liu and Song, 2004, Sun, 2005 and Zhang, Gu and Di, 2005 with parametric methods), and there is also a growing literature available to the non-Chinese reader, the first of these is perhaps Chen et al. (2005), but other studies include Matthews et al. (2007), Fu and Heffernan (2007; 2009), Yao et al. (2008), Lin and Zhang (2009), Berger et al. (2009), and Asmild and Matthews (2012). The consensus of findings among Chinese scholars is that the state-owned commercial banks (SOCBs) tend on average to exhibit the lowest levels of efficiency and the joint-stock commercial banks (JSCBs) show a faster growth in performance and efficiency.

Cost inefficiency relative to “best practice” is usually blamed on bad management and poor motivation. Following Leibenstein (1966) this efficiency gap is sometimes termed “X-inefficiency”. In an oft cited study of bank efficiency Berger et al. (1993) argue that 20% of bank costs is due to X-inefficiency. Studies of bank efficiency in China have estimated historic cost inefficiency in the region of 50%. Such figures imply that either Chinese bank management is grossly inefficient or that the estimates of cost efficiency have failed to take into account policy objectives and/or policy constraints that enter the decision making process. More recently Asmild and Matthews (2012), use multi-directional efficiency analysis (MEA) to confirm the conventional finding that SOCBs are less efficient than JSCBs. But importantly they also find differences in the patterns of inefficiencies due to labour input. Specifically, they find JSCBs are more efficient in labour utilization than SOCBs but that this inefficiency difference declines over time.

Given the vast literature on Chinese bank efficiency, a pertinent question is, what does this paper add to the existing research? Briefly, our contribution is threefold. First, we use a relatively underutilized modelling technology to measure bank efficiency in China. To our knowledge there have been no studies that have applied this technology to the national mainstream banks in China and previous estimates can be challenged on statistical grounds. Second, we use the concept of rational inefficiency first espoused by Bogetoft and Hougaard (2003) to identify a major source of inefficiency and provide evidence of misallocation of factor inputs by backing out the vector of factor shadow input prices. Third, we examine the convergence properties of our estimates to identify if market pressure and learning drive banks to a common level of efficiency. We show in this paper that in contrast to previous findings, a more rigorous measure of efficiency than what has hitherto been used in the literature, places Chinese banking at a relatively high level of management efficiency.

This paper has three objectives. First, it aims to measure cost inefficiency of Chinese banks using the familiar non-parametric method of Data-Envelopment-Analysis (DEA) by bootstrapping methods to provide estimates that lend themselves to statistical inference. Here we differ from the vast number of studies that use conventional DEA to study Chinese bank efficiency. Simar and Wilson (1998; 2000) pioneered the usage of bootstrap in deriving the sampling characteristics (standard deviation and bias, etc.) of the nonparametric DEA estimators. The intuitive appeal of the bootstrap DEA has attracted a lot of followers (for recent reviews and applications, see Simar and Wilson, 2008; 2013; Zhang and Matthews, 2012; Wijesiri et al., 2015; Wanke and Barros, 2016; Hosseinzadeh et al., 2016; etc.).

Second, it decomposes the measure of cost inefficiency into its constituent parts of technical inefficiency and allocative inefficiency. While we are not the first to undertake such

a decomposition for Chinese banks (Chen et al., 2005)¹, here we provide an interpretation for the existence of allocative inefficiency and evidence of its source. This paper argues that while the underutilization of factor inputs is consistent with the notion of X-inefficiency, the wrong factor-mix is indicative of long-standing employment constraints imposed on the banking system in the pre-reform period. Insofar as allocative inefficiency can be explained as the result of officially sanctioned constraints, the implied cost inefficiency cannot be viewed wholly as a management deficiency but rather as a rational outcome of constrained optimizing behavior. In this respect the term ‘rational’ is fully consistent with the classification of Bogetoft and Hougaard (2003) who argue that that inefficiency in some cases may be the outcome of a rational choice (see also Tsionas and Izzeldin, 2018). The decomposition of cost inefficiency into its constituent parts allows us to examine their evolution over the sample period, and the speed of decline of this respective inefficiency.

Third, the bootstrap estimates of inefficiency are used to test various hypotheses regarding the levels, trends and convergence in technical inefficiency and allocative inefficiency. Hypothesis testing using conventional DEA can be challenged because the efficiency score estimates may be biased due to the random errors in the input/output data with finite sample size (Simar and Wilson, 1998, 2000; Sunil Dharmapala, 2018). Bootstrapping produces estimates of the bias enabling a more rigorous inference. Specifically, we test three hypotheses.

Hampered by its history of over-employment, banks that are directly controlled by the state will have higher allocative efficiency than the JSCBs. The corresponding hypothesis is:

Hypothesis 1: SOCBs will exhibit higher allocative inefficiency than JSCBs.

A *Corollary* of Hypothesis 1 is that the allocative inefficiency is mostly due to the over-utilization of labour. However, as the profit motive replaces other (social and economic)

¹ It could be argued that Chen et al. (2005) do not fully comply with the underlying assumption of homogeneity in DEA by mixing banks that operate at the national level with banks that operate within geographically limited urban areas.

imperatives the levels of technical and allocative inefficiency should decline over time. Regulatory and market pressure, and the long-term threat of foreign banks entry may have sparked improved management, resulting in improved technical efficiency and lower cost-inefficiency as incumbent banks attempt to cut costs and consolidate their balance sheets. The corresponding hypothesis is:

Hypothesis 2: Both SOCBs and JSCBs will exhibit declining inefficiency over time.

But given the history of the state-owned banking system in maintaining the socialist plan, the SOCBs will reduce inefficiency at a slower rate than the JSCBs. The corresponding hypothesis is:

Hypothesis 3: SOCBs will reduce both technical inefficiency and allocative inefficiency at a slower rate than JSCBs.

To test the veracity of results we distinguish between two types of models. In model 1, the outputs of the banks are what has been typically used in the literature. In model 2, and given the history of China's non-performing loans (NPL) history, we treat NPLs as an undesirable output. This paper is organized on the following lines. The next section provides a brief description of the Chinese banking system as a contextual background. Section 3 reviews the literature and provides a motivation for the concept of rational cost inefficiency. Section 4 outlines the bootstrap methodology and discusses the data. Section 5 tests the hypotheses and discusses the results. Section 6 concludes.

2. Chinese banking

Several good descriptions of the Chinese banking system exist (e.g., Chen et al. , 2005, and Garcia-Herreto et al., 2006) and what follows is a recap and an update of the elements relevant to the issue of efficiency. Up until the mid-1990s, the principal function of the Chinese banking sector was to direct domestic savings to support state-owned-enterprises. Control of

the banking system remained firmly under the government and its agencies². Under state control, the banks in China mainly served the plan of directing credit to specific projects dictated by political preference rather than commercial imperative. The reform process signaled by the Banking Law of 1995, saw the gradual evolution of the banking system towards full commercialization. This gradual process has involved recapitalization, foreign strategic investment, and improved governance, modern methods of risk management and limited liberalization of loan rate setting.

Some of the big changes were the divestment of NPLs of the SOCBs in the period 1999-2003 by the asset management companies set up for that purpose. The process of NPL divestiture continued during 2003-2005 (Li, 2013). Up until 2004 lending rates were strictly controlled within a narrow range by the Peoples Bank of China (PBOC). The legacy of the policy of strict control of lending rates along with policy directed lending was the underpricing of risk and the well-known history of China's non-performing loans (NPLs). After 2004 the upper limit on interest rates were lifted and banks had the capability to risk price marginal lending. Since 2001 foreign banks and financial institutions were allowed to take a stake in selected Chinese banks. While control of individual Chinese banks remain out of reach for the foreign institution³, the pressure to reform management, consolidate balance sheets, improve risk management and reduce unit costs has increased with greater foreign exposure.

According to China Bank Regulatory Commission's Annual Report of Year 2017, the Chinese commercial banking market is made up of five large SOCBs and a dozen JSCBs. In 2016, together they account for 78% of the commercial banking market and operate on a nationwide scale. The remaining 22% of the market is made up of 134 city commercial banks

² According to La Porta, et. al (2002), 99% of the 10 largest commercial banks were owned and under the control of the government in 1995.

³ There is a cap of 25% on total equity held by foreigners and a maximum of 20% for any single investor, except in the case of joint-venture banks.

that normally operate within geographic boundaries, smaller rural cooperatives and other entities, and 68 foreign bank branches (constituting less than 1.5% of the market).

Like many economies that have undeveloped financial and capital markets, the banking sector in China plays a pivotal role in financial intermediation. Table 1 below shows that the ratio of total bank assets to GDP has doubled in the two decades in 1997 to 2016. The market remains dominated by the state-owned banks, although their share of the market has decreased steadily through competition from the other nationwide banks (and some City Commercial Banks).

Table 1: The Chinese banking Market

Variable	1997	2007	2016
Total Assets to GDP	125.6%	213.4%	253.2%
Market share SOCB % assets	88.0	53.2	54.9*
NPL ratio	52.0%	8.1%	1.9%
ROA	0.2%	0.9%	1.0%
NIM	2.4%	2.7 %	2.6%
Cost-Income Ratio	52.7%	42.8%	27.8%

Sources: IMF *International Financial Statistics*, Individual Bank Annual Accounts, China Regulatory Banking Corporation Annual Report (various issues), *Almanac of China's Finance and Banking*, Fitch-Bankscope data base, National Bureau of Statistics of China, *Includes Bank of Communications which was included as a SOCB in 2007.

Return on assets (ROA) was low in the early part of the sample largely due to the poor asset quality left on the balance sheets and a high level of non-performing loans (NPL). The PBOC created four asset management companies in 1999 to manage the NPLs of the SOCBs. As a result of the activities of the asset management companies in the period 1999-2002, the NPL ratio fell from over 50% in 1997 to under 2% in 2016. Net-interest margins (NIM) are respectable by Western standards but are well below levels that would be consistent with economies in a similar stage of development (as for example India where NIM would be in the region of 3%). Part of the problem as stated earlier, was that interest rates were heavily

controlled during much of this period. The non-performing loans (NPL) ratio of the commercial banks has fallen from 52% in 1997 to under 2% in 2016.

It is arguable that a mixture of regulatory and market pressure to reform provided the background for the major banks to begin the process of restructuring and reducing unit costs. Recapitalization by the state, reformed corporate governance, strategic foreign investment, improved profitability and efficiency of the SOCBs was a prelude to international listing⁴ (Lin et al., 2015). Other markers that acted as gateposts of the reform process were accession to WTO in 2001 and the roll-out of the conditions to banking in 2006, the setting up of the China Bank Regulatory Commission in 2003, limited loan rate reform 2004 and 2013, and limited deposit rate reform in 2014.

3. Literature review and the concept of rational allocative inefficiency

A number of efficiency studies of Chinese banks have emerged in recent years, using both DEA and stochastic frontier analysis. Other studies by Chinese scholars that have used non-parametric and parametric techniques include Fang et al. (2004), Liu and Liu (2004), Sun (2005), Qian (2003), Chi, Sun and Lu (2005), Yao, Feng and Jiang (2004). The consensus of finding from the DEA studies is threefold. First, because of the continued banking reform programme technical inefficiency has been declining over time. Second, average bank efficiency is lower in the state-owned banks (SOCBs) than in the joint stock banks (JSCBs). Third, the gap between the two has been narrowing in recent years.

Studies of bank efficiency have used the terms technical efficiency and X-efficiency interchangeably as if they were the same thing. While similar in concept they are not

⁴ Bank of Communications was listed on Hong Kong Stock Exchanges in June 2005, and on Shanghai Stock Exchange in May 2007; China Construction Bank was listed on Hong Kong Stock Exchanges in Oct 2005, and on Shanghai Stock Exchange in Sep 2007; Bank of China was listed on Hong Kong Stock Exchanges in June 2006, and on Shanghai Stock Exchange in July 2006; ICBC was listed on Hong Kong and Shanghai Stock Exchanges Oct, 2006; Agricultural Bank of China was listed on Hong Kong and Shanghai Stock Exchanges in July, 2010.

necessarily the same. The concept of technical efficiency derives its basis in the neo-classical theory of the firm and assumes profit maximising behaviour. A firm or a bank may be technically inefficient for technical reasons such as low training or low human capital levels of managers and workers, or the use of inferior or out-of-date technology. The diffusion of new technology is not instantaneous, and some firms or banks may lag behind others in the acquisition and utilisation of new technology. With further training and updating of capital, the firm or bank can expect to move towards the efficient frontier. X-inefficiency is not caused by the variability of skills or the time variability of technology diffusion but by the use and organisation of such skills and technology.

Whereas X-inefficiency is viewed by Leibenstein (1966) as the outcome of non-maximising behaviour, Bogetoft and Hougaard (2003) suggest that the existence of X-inefficiency in production is the result of a rational decision-making process that represents on-the-job compensation to managers or optimisation subject to extra constraints. To quote:

“.....we see inefficiency and in particular the allocation of inefficiency (slack) among different inputs as the result of a rational choice made by that firm.”....”The bureaucrat may well use a cost-minimizing mix of input but the level of resource use is socially inefficient”
Bogetoft and Hougaard (2003), pp. 244-245.

In other words, allocative inefficiency may be rational.

Berger, Hunter and Timme (1993) have argued that X-inefficiency constitutes 20% or more of bank costs. X-inefficiency arises as a result of low pressure for performance. Some institutions would be protected by government regulation that would reduce the external pressure of competition. But even with a higher degree of pressure from the environment, firms may have organisational deficiencies so that management signals and incentives are lost in the hierarchy of the organisation.

In the earlier period the banks in China operated under stringent state control. After 1995 the state relaxed its grip, but an important but relatively unknown feature of the pre-reform banking system was that the banks were compelled to employ all banking graduates of the

universities set up by the People's Bank of China. Similarly, the banks were compelled to employ party officials and retirees of the People's Liberation Army who had completed their tour of duty. The overhang of Party officials and former PLA officers employed in the banks during the pre-reform period contributes to the overall picture of overstaffing⁵. One example is the following quote from an anonymous risk manager:

“Many heads of sub-branches are not professional bankers but are retired army officers of 10-15 years' experience. They tend to be appointed at high levels even as President or Head of HRM and even on the credit committee” **Risk Manager - China Construction Bank (Dalian 2007)**⁶

Other forms of inefficiency arise from government imperatives for the state-owned banks to locate branches in unprofitable areas. Former studies suggest exceptionally high levels of cost inefficiency in Chinese banks⁷. However, since listing of the SOCBs the compulsion to employ retired personnel from the PLA has been relaxed.

“Historically this was the case but since listing it is no longer true. Social objectives were important in the past but now the profit objective is most important. But even now the profit objective is not the only one. Loss making branches are maintained if there is a social need”. **HRM Manager Industrial and Commercial Bank of China - Shenzhen 2009**⁸.

Figure 1 shows an isoquant qq producing a given output with factor inputs D and L and isocost ww , which traces the ratio of factor prices. The efficient cost minimising position is shown at e where ww is tangential to qq . However, employing a factor combination shown by

⁵ This practice has received a recent boost with the announcement that state-owned enterprises will have to absorb an extra 300,000 laid-off members of the PLA, undermining the pressure on state-owned companies to become more efficient. *Financial Times*, December 29 2015, China orders SOEs to hire former soldiers.

⁶ One of a number of interviews with managers in Chinese banks conducted as part of a British Academy funded project. See Matthews (2013).

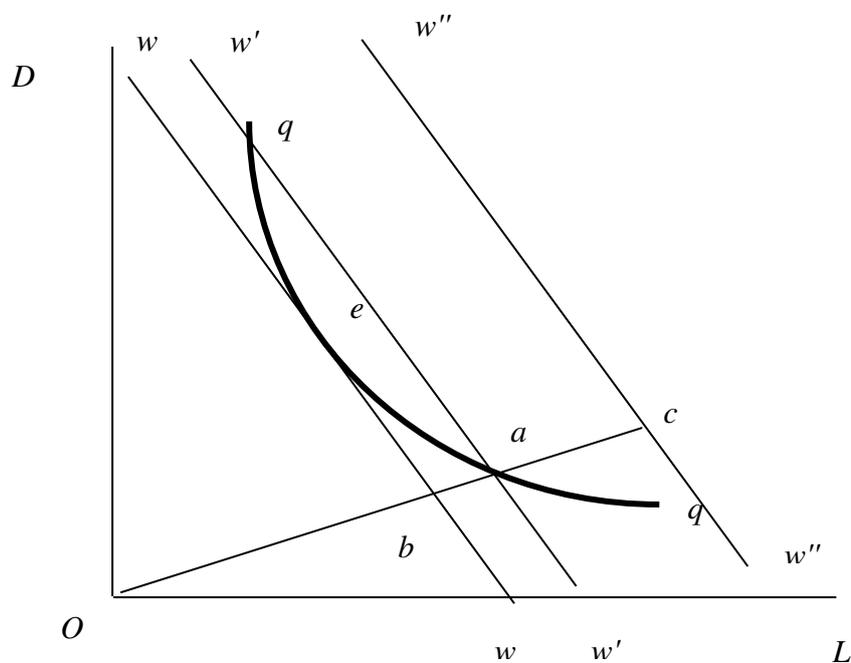
⁷ Fu and Heffernan (2007) find average inefficiency levels of 50%-60% for 1993-2002; Chen et al. (2005) estimate it as 44%-66% in 1993-2000, and Matthews et al. (2007) estimate it at 50%-60% in 1997-2004.

⁸ While this practice has been relaxed evidence is that it still continues. A recent announcement that state-owned enterprises will have to absorb an extra 300,000 laid-off members of the PLA, undermines the pressure on state-owned companies including banks to become more efficient. *Financial Times*, December 29 2015, *China orders SOEs to hire former soldiers*.

point c , which is to the right of the isoquant qq indicates that the firm is technically inefficient. Efficiency is decomposed into technical efficiency (TE) and allocative efficiency (AE).

Technical efficiency is measured by the ratio Oa/Oc (Technical inefficiency is given by ac/Oc). The cost to the firm is shown by $w''w''$ which is parallel to ww and passes through point c . Cost efficiency (CE) is measured by Ob/Oc and Ob/Oa gives AE^9 .

Figure 1: Technical Efficiency and Allocative Efficiency



To understand the perceived high level of cost inefficiency in Chinese banks we use an illustrative model of allocative inefficiency based on staffing targets provided by the central authorities. Assume that the bank produces a single earnings asset (A). In reality this will consist of a combination of commercial loans, mortgages, government bonds, short-term bills,

⁹ It can be seen from this decomposition that under the assumption of constant returns to scale that $AE = CE/TE$.

etc. We assume that this earning asset is produced by the inputs deposits (D), labour (L) and fixed capital assets (\bar{K})¹⁰:

$$A = f(D, L, \bar{K}) = D^\alpha \left(L_1^\gamma L_2^{(1-\gamma)} \right)^\beta \bar{K}^{(1-\alpha-\beta)} \quad (1)$$

The prices of inputs are, the cost of deposits (r), the cost of labour (w) and the cost of fixed assets (ρ). The bank can hire two types of labour $\{L_1$ and $L_2\}$. The first type (L_1) are bank workers who have a higher marginal product than the second type (L_2) who are bureaucrats. However, the bank is constrained to pay the same wage to both types of workers. The objective of the bank manager is to minimise costs subject to an output target:

$$\text{Min } \rho\bar{K} + wL_1 + wL_2 + rD - \lambda \left(D^\alpha \left(L_1^\gamma L_2^{(1-\gamma)} \right)^\beta \bar{K}^{(1-\alpha-\beta)} - A \right). \quad (2)$$

The bank is constrained to employ some type 2 labour but clearly in an unconstrained world the bank would only employ type 1 workers. From the FOC of (1) the ratio of the marginal products of the two types of labour is given by $\frac{\gamma(A/L_1)}{(1-\gamma)(A/L_2)} = 1$. Since this contradicts the assumption of type 1 labour having a higher marginal product than type 2 labour it follows that $\gamma = 1$ and $L_2 = 0$. The solution for output in the unconstrained case is given by:

$$A = \left(\frac{\alpha}{\beta} \right)^\alpha w^\alpha r^{-\alpha} \left(\frac{1-\gamma}{\gamma} \right)^{(1-\gamma)\beta} \gamma^{-\alpha} L_1^{\alpha+\beta} \bar{K}^{(1-\alpha-\beta)}. \quad (3)$$

In the constrained case, the bank has to employ a certain number of type 2 labour given by the central government so that $L_2 = \bar{L}_2$. The objective function is now:

$$\text{Min } \rho\bar{K} + wL_1 + wL_2 + rD - \lambda_1 \left(D^\alpha \left(L_1^\gamma L_2^{(1-\gamma)} \right)^\beta \bar{K}^{(1-\alpha-\beta)} - A \right) - \lambda_2 (L_2 - \bar{L}_2). \quad (4)$$

The first order conditions are:

¹⁰ This uses the assumption of the intermediation approach that recognises that the outputs are the interest earning assets while deposits and borrowed funds are included with capital labour as inputs (Sealey and Lindley, 1977).

$$\begin{aligned}
r - \lambda_1 \alpha \left(\frac{A}{D} \right) &= 0, \\
w - \lambda_1 \beta \gamma \left(\frac{A}{L_1} \right) &= 0, \\
w - \lambda_1 \beta (1 - \gamma) \left(\frac{A}{L_2} \right) - \lambda_2 &= 0.
\end{aligned}$$

The marginal wage premium for type 2 labour is given by λ_2 .

The output function is now:

$$A = \left(\frac{\alpha}{\beta} \right)^\alpha w^\alpha r^{-\alpha} \gamma^{-\alpha} L_1^{\alpha+\beta\gamma} \bar{L}_2^{(1-\gamma)\beta} \bar{K}^{(1-\alpha-\beta)}. \quad (5)$$

Denoting the input of type 1 labour in the unrestricted case as L_{1U} and the same for the restricted case as L_{1R} , from (5) and (3) we have the relationship described by:

$$\left(\frac{1-\gamma}{\gamma} \right)^{(1-\gamma)\beta} L_{1U}^{\alpha+\beta} = L_{1R}^{\alpha+\beta} \bar{L}_2^{(1-\gamma)\beta}. \quad (6)$$

The allocative inefficiency generated by the additional constraint in the restricted case is described by:

$$\frac{L_{1R} + \bar{L}_2}{L_{1U}}. \quad (7)$$

Expression (7) must be strictly greater than unity for an allocative inefficiency to exist. Because the marginal productivity of the type 2 labour is less than type 1 labour, the type 1 labour displaced by the constraint of having to employ a fixed amount of type 2 labour is less than one-for-one if the target level of output (earnings assets) is to be maintained.

In reality, banks are multi-output enterprises and one of the conventional ways of modelling the efficiency of banks is the non-parametric method of Data Envelopment Analysis. Technical efficiency (TE) is measured as the ratio of projected output (on the efficient frontier) to actual input used. There are a number of papers that describe the methodology of DEA as

applied to banking, and therefore will not be elaborated here (Drake, 2004; Fethi and Passourias, 2010).

DEA constructs a non-parametric frontier of the best practices amongst the decision-making units (DMUs).¹¹ An efficiency score for each DMU is measured in relation to this frontier. DEA is relatively insensitive to model specification (input or output orientation) and functional form; however, the results are sensitive to the choice of inputs and outputs. The weakness of the DEA approach is that it assumes data are free from measurement errors. Furthermore, since efficiency is measured in a relative way, its analysis is confined to the sample used. This means that an efficient DMU found in the analysis cannot be compared in a straightforward way with other DMUs outside of the sample.

4. Bootstrap methodology and data

As pointed out by Simar and Wilson (1998, 2000a, 2011), the scores produced by DEA methods are not deterministic but the estimates of underlying true scores. It is therefore natural to consider the sensitivity of the estimators to sampling variation. Due to the complex nature of the modelling framework and the computation of the DEA estimators, there are in general no readily available analytical results for the sampling distributions of the DEA estimators (Kneip et.al. 2008; Simar and Wilson, 2008). However, without the capability for statistical inference, the non-parametric DEA method is viewed as a weak alternative to parametric methods of estimating efficiency. Meanwhile, the conventional DEA efficiency scores estimators are biased by construction, and though in certain cases they are consistent in theory, their convergence rates suffer from the “curse of dimensionality” (i.e., low number of DMUs

¹¹ Stochastic frontier analysis (SFA) is another popular way of modelling technological frontier and relative efficiency/productivity. For a recent study and review of SFA, see Tsionas and Mallick (2019).

relative to number of input and output variables) and their accuracy relies on the relative sample size. In light of the above concerns, a number of researchers (Simar and Wilson 1998, 2000a, 2000b, 2007, 2009; Thanassoulis et al., 2008; Kneip et al., 2011) recommend the bootstrap method to obtain the sampling distribution of the DEA estimators and to reduce their potential biases by bootstrap correction.

The bootstrap procedures for DEA models are set out in Simar and Wilson (1998, 2000a, 2000b, 2007). We assume that the observed inputs-outputs pairs data sample $\mathcal{X} = \{ (x_i, y_i) \mid i = 1, \dots, n \}$ are generated by a Data Generating Process (DGP) \mathcal{P} , which satisfies a number of regularity conditions. Let (x, y) be a fixed point, which may not necessarily be one of the points in \mathcal{X} , and $\hat{\theta}(x, y)$ be the estimated DEA efficiency score of the true unknown efficiency score $\theta(x, y)$. As we mentioned above, the sampling distribution of $\hat{\theta}(x, y) - \theta(x, y) \mid \mathcal{P}$ is hard to derive. Let $\hat{\mathcal{P}}$ be a consistent estimator of \mathcal{P} from the observed data \mathcal{X} . Now suppose we draw a random sample of data $\mathcal{X}^* = \{ (x_i^*, y_i^*), i = 1, \dots, n \}$ from $\hat{\mathcal{P}}$. The new data can be then used to compute an efficiency score $\hat{\theta}^*(x, y)$. According to the bootstrap theory, we have that;

$$\hat{\theta}^*(x, y) - \hat{\theta}(x, y) \mid \hat{\mathcal{P}} \text{ approx} \sim \hat{\theta}(x, y) - \theta(x, y) \mid \mathcal{P}.$$

Therefore we can use the distribution of $\hat{\theta}^*(x, y) - \hat{\theta}(x, y)$ to approximate the distribution and calculate the small sample bias of $\hat{\theta}(x, y) - \theta(x, y)$, the former can be obtained by resampling from the distribution $\hat{\mathcal{P}}$.

There are three popular bootstrap methods to obtain $\hat{\mathcal{P}}$, the estimator of \mathcal{P} : the naive bootstrap, the homogeneous bootstrap and the heterogeneous bootstrap. The naive bootstrap use the empirical distribution to estimate \mathcal{P} . However, since \mathcal{P} has bounded support and that the efficiency scores are bounded by definition, the naive bootstrap gives inconsistent estimates (Simar and Wilson, 2000b, 2011). Both the homogeneous bootstrap and the heterogeneous bootstrap were devised to deal with the inconsistency of the naive bootstrap. The major steps

of the homogeneous and the heterogeneous bootstrap are as follows (see Simar and Wilson 1998; 2000a for detailed implementation algorithms):

- First, the density function $f(x, y)$ of $\{(x_i, y_i) : i = 1, \dots, n\}$ is translated into its polar coordinates version $f(\theta, \eta, y)$, where θ denotes efficiency and η denotes angle (of x) respectively. The homogeneous bootstrap assumes that $f(\theta|\eta, y) = f(\theta)$, while the heterogeneous bootstrap does not maintain this restriction. As acknowledged by Simar and Wilson (2011), the generality of heterogeneous bootstrap comes at the price of increased complexity and computational burden.
- Second, B pseudo-samples $\{(\theta_{ib}^*, \eta_{ib}^*, y_{ib}^*) : i = 1, \dots, n; b = 1, \dots, B\}$ are drawn from a smooth, consistent and nonparametric estimate of $f(\theta, \eta, y)$.
- Third, the B pseudo-samples $\{(\theta_{ib}^*, \eta_{ib}^*, y_{ib}^*) : i = 1, \dots, n; b = 1, \dots, B\}$ are translated into the corresponding pseudo-samples $\{(x_{ib}^*, y_{ib}^*) : i = 1, \dots, n; b = 1, \dots, B\}$.

After the B sets of bootstrap samples $\{(x_{ib}^*, y_{ib}^*) : i = 1, \dots, n; b = 1, \dots, B\}$ are obtained, we can use each of the data sample to compute the efficiency scores $\hat{\theta}_b^*(x, y), b = 1, \dots, n$. The bias-corrected estimate of $\theta(x, y)$ is defined as

$$\begin{aligned} \hat{\hat{\theta}}(x, y) &= \hat{\theta}(x, y) - \widehat{bias} = \hat{\theta}(x, y) - \left[\frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*(x, y) - \hat{\theta}(x, y) \right] \\ &= 2\hat{\theta}(x, y) - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*(x, y). \end{aligned}$$

In this paper we show the results from the homogeneous bootstrap approach to obtain $\hat{\mathcal{P}}$, but we also use the heterogeneous approach as a robustness check. The results are qualitatively similar but less precise. This is not surprising since the finite sample performance of the bootstrap procedures depend on a number of factors such as the sample size, the return to scale assumption, and the curvature of the frontier. Simar and Wilson (2011) find that the homogeneous bootstrap generally has stable and satisfactory performance using a number of numerical simulations. The results from the heterogeneous bootstrap is shown in the appendix.

We follow the Simar-Wilson (2000b) method to obtain 1000 bootstrap values of the individual DMU for the efficiency scores in each year. Recent bootstrapping applications to DEA have been conducted by Casu and Molyneux (2003); and in the case of Chinese rural credit cooperatives, Dong and Featherstone (2006).

This study employs annual data (1997-2016) for 14 banks: the five SOCBs, and nine JSCBs. The choice of sample is based on the argument that the banks chosen are national based and not restricted to city or regional operation. The total sample consisted of 280 bank year observations. The main source of the data is Fitch/Bankscope, and individual annual reports of banks. The full data set is available on request.

The most common approach for the choice of inputs and outputs is to treat bank assets as outputs and liabilities as inputs (see Fethi and Pasiouras, 2010). An increasing number of studies take a hybrid approach that uses the conventional three inputs of employees (*LAB*), fixed assets (*FA*) and total deposits (*DEP*) but include non-interest income (*NII*) as an additional output to loans, other earning assets (*OEA*). In this study we use the conventional set of inputs but use two sets of outputs. Specifically, we consider two models. In Model 1 the outputs consist of total performing loans, other earning assets, and non-performing loans as a 'bad' output. Following Thanassoulis et al., 2008, the inverse of NPL was included as an output as a means of measuring bad output. In Model 2 the outputs consist of performing loans (*PLOAN*) plus other earning assets (*OEA*) to give total earning assets, non-performing loans (*NPL*) as a bad output, and non-interest earnings as an additional output. Fethi and Pasiouras (2010) list six recent studies that have used NII with stock variables as an output in DEA studies of banks. Non-interest income is a growing area of business for Chinese banks rising from around 2% of earnings in 1997 to 17% in 2012. As Chinese banks have had a troubled history of non-performing loans, it is appropriate to include NPLS as a bad output produced alongside good outputs of performing loans and other earning assets.

The inputs for the construction of cost-efficiency additionally require the factor prices of the relevant inputs above. We distinguish between the price of labour (PL), price of fixed capital (PK) and the price of funds (PF). Conventionally, the price of labour is obtained as the ratio of personnel expenses to the number of employees. The price of fixed capital is obtained as operating expenses less personnel expenses divided by fixed assets (less depreciation). The price of funds is obtained from the ratio of interest paid to total funds.

Table 2 presents the summary statistics of the input and output data for 1997 and 2016 as a snapshot indicator of the scale of the variables used. The high standard deviation is an indication of the dominance of the 5 state owned banks. The table shows how fast earnings assets have grown over this period. The total stock of performing loans has grown at an average of 19 per cent a year. Other earning assets have also grown at an average rate of 19% a year, in part reflecting the activities of the asset management companies that swapped a proportion of the NPLs of the big 4 state-owned banks for bonds in 1999, 2001 and 2004. The most remarkable growth is in non-interest earnings which have grown at an average rate of 28% a year, reflecting an increasing source of profit for banks that have traditionally depended on the banking book for the generation of income.

Table 2: Output-Input Variables 1997 - 2016

Variable	Description	Mean 1997	SD 1997	Mean 2016	SD 2016
<i>PLOANS</i> RMB mill	Total stock of performing loans	217558	320219	4687964	4107095
<i>NPL</i> RMB mill	Non-performing loans.	212500	353886	81769	73518
<i>OEA</i> RMB mill	Investments	205103	301626	3123224	2015631
<i>NII</i> RMB mill	Non-interest income	862	1922	57234	37929
<i>LAB</i>	Total Employed	105138	175233	155184	171382
<i>DEP</i> RMB mill	Total stock of Deposits	604013	891353	7902297	6452333
<i>FA</i> RMB mill	Fixed assets	12831	19398	68385	68183

<i>PL</i>	Unit price of labour	.0631	.0380	.3359	.0791
<i>PF</i>	Unit price of funds	.0502	.0202	.0202	.0039
<i>PK</i>	Unit price of fixed assets	.6528	.5282	.8408	.4283

Sources: Bankscope and author calculations.

Other points to note are that employment has grown by an average of 2% a year but average labour cost has grown by a remarkable 11% a year, reflecting the increasing skill premium paid to workers in this sector. A further point to note is reduced relative dispersion of the variables (coefficient of variation) which also indicates an increased convergence of the nationwide banks on each other.

Before moving to the empirical results and their implications, let's make some comments on the data quality issue. The specific question is whether information contained in our limited data sample is sufficient enough to reliably identify the difference in efficiencies across individual DMUs. We stated above that DEA with small sample size may suffer the "curse of dimensionality". We want to emphasize that such an issue is actually immaterial if not completely irrelevant here for the following observations. First, a number of existing studies suggest that for general cases the efficiency estimates of DEA are reliable if the relative information ratio $n/(p + q) \geq C$, where n, p, q are the number of DMUs, the number of inputs, and the number of outputs, respectively, and C is some threshold. Golany and Roll (1989) propose that $C = 2$, while Banker et al. (1989), Friedman and Sinuany-Stern (1998) and Cooper et al. (2007) propose that $C = 3$. In our case the ratio of $n/(p + q)$ is 2.3 for Model 1 and 2 for Model 2, which is close to the proposed theoretical threshold. Second, a number of studies (Wang, 2009; Xu and Chen, 2012; Liu, Miao and Zhu, 2012) have demonstrated that the national wide banks in China, as those used in our sample, have very similar operations structure, hence there is strong homogeneity among their input/output characteristics. The similarity of the SOCBs are even more prominent, as there are frequency alternations of senior

officials among them. The homogeneity of the DMUs' characteristics largely obviates the sample size requirement for the validity of DEA estimates, in that the reliability of DEA estimators, as of any other statistical estimators, is inversely related to the variation across individual units. Third, to minimize the potential bias of the DEA estimators caused by limited sample size, this paper uses the bootstrap corrected efficiencies instead of the original DEA estimates. Chernick (2007, pps. 173-174) pointed out that though bootstrap is theoretically justified asymptotically, it can work well with small sample.¹² Chernick suggested that in general applications sample size should be greater than 10 for bootstrap to deliver reliable estimates. This condition is satisfied in all our bootstrap calculations. Meanwhile, the strong homogeneity of the population of China's national wide banks again further substantiates the validity of bootstrap with small sample size, as when the population characteristics is of low variation, a small number of observations might be representative of the population and sufficient to approximate the true unknown population distribution. To summarize, the accuracy of the DEA estimators and their bootstrap bias corrections depend both on the effective sample size and the degree of variation across individual units. The strong homogeneity among China's national wide banks relaxes the demand for large sample size in both DEA and bootstrap analyses.

5. Empirical results

Table 3 provides the summary statistics of technical and allocative inefficiency for the two models for the full sample, and sub-samples. Cost inefficiency (CI) and technical inefficiency (TI) estimates were obtained as $CI = (1 - CE)$ and $TI = (1 - TE)$ respectively. The allocative inefficiency estimate was obtained as the residual of the cost inefficiency CI and TI

¹² To quote Chernick (2007), pp. 173: "Although we have good reasons not to trust the bootstrap in very small samples and theoretical justification is asymptotic, the results were surprisingly good even for sample size as small as 14 in the two-class problem."

$(AI = CI - TI)^{13}$. Given the higher propensity for overstaffing and wider branch network in the SOCB banks, a higher level of allocative inefficiency would be expected for those banks. Table 3 shows that the mean allocative inefficiency for the JSCB banks is consistently and significantly lower than that of the SOCB. In keeping with the mainstream finding of Chinese scholars, there is also evidence that the JSCBs have on average, a lower level of technical inefficiency than the SOCBs.

Table 3: Mean inefficiency 1997 – 2016

Statistic	Model 1		Model 2	
	TI	AI	TI	AI
Mean	.089	.162	.117	.131
Std Dev	.094	.178	.130	.139
Mean SOCB	.119	.263	.145	.205
Mean JSCB	.078	.121	.105	.101
Z test	-0.35	-7.42***	-0.65	-5.54***

*** significant at the 1%,

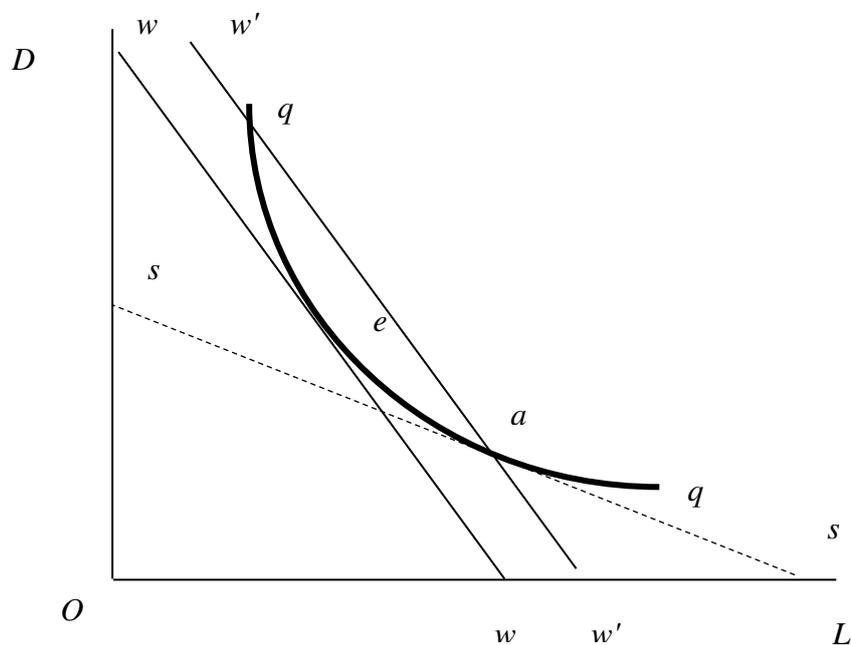
The results of the non-parametric tests indicate a significant difference between the average level of allocative inefficiency between the SOCBs and JSCBs but also significant differences between the two bank types for technical inefficiency. Therefore *Hypothesis 1* is confirmed. Average allocative inefficiency is higher in SOCBs than JSCBs in both models.

However, allocative inefficiency does not necessarily imply that the misallocation is due to excess utilization of labour. We provide evidence for our case by extracting the shadow price of labour from the technically efficient *but not cost-efficient banks*. The vector of shadow prices

¹³ Strictly $AE = CE/TE$ but as we are dealing with the mean values of an unknown distribution it was convenient to define $AI = CI - TI = TE - CE$. The alternative measure is $AI^* = (1 - AE) = (TE - CE)/TE$.

consonant with an allocative inefficiency is shown in Figure 2. At point “a” on the isoquant, the bank is technically efficient but allocatively inefficient. However, it is allocatively efficient relative to the shadow input price ratio. The ratio of shadow prices is given by the dashed line “ss” and is tangent to the isoquant qq at point “a” (see Kumbhakar and Knox-Lovell (2003), pps. 221-222). At this point $\left. \frac{\partial f/\partial L}{\partial f/\partial D} \right|_a = \left. \frac{w}{r} \right|_a < \left. \frac{w}{r} \right|_e$ which says that the bank is allocatively inefficient with respect to the benchmark ratio of input prices but allocatively efficient with respect to the shadow prices.

Figure 2: Shadow Price Vector



To test whether the allocative efficiency is induced by the over-utilisation of labour, each year we set up a benchmark bank that is cost efficient and use the virtual input weights of factors inputs as the efficient factor allocation. For each year we pick out those banks that are

technically efficient but allocatively inefficient (not cost efficient) and compare their factor input weights with the benchmark. Lower factor input weights in labour imply that the bank is under-pricing labour (shadow price is lower than actual) and over-utilizing labour. We can isolate all technically efficient but allocatively inefficient banks that have the wrong factor mix, for all years. We can then test if the distribution of labour-over-utilized and labour-under-utilized banks is random or systematic.

Let N be the number of technically efficient but allocatively inefficient banks across all years, and X be the number of labour-over-utilized banks among those N banks. Define $p = X/N$, and p_0 be the theoretical proportion in the population. We test $H_0: p_0 \leq 0.5$ vs $H_1: p_0 > 0.5$. The test statistic for this problem is $t = \frac{p-0.5}{\sqrt{p(1-p)/N}}$. The p value for this test is $\Pr(N(0,1) > t)$. Table 4 presents our results.

Table 4: Over-utilization of labour 1997-2016

	Model 1	Model 2
No.: TE banks	121	129
No.: TE but not CE banks	55	57
No.: TE over-utilizing labour	35	38
t-statistic	18.4***	2.7**

*** significant at the 1%,; ** significant at the 5%

The results show that the distribution of technically efficient but cost inefficient banks is not random, and the hypothesis that more banks over-utilize labour than under-utilize labour is supported by empirical data. The *Corollary of Hypothesis 1* is confirmed.

The next important question is, is inefficiency in Chinese banks as a whole declining and if so, what can be said about the speed at which the inefficient banks are converging on the

benchmark efficient banks? Using the concept of *beta-convergence* from the growth convergence literature (Barro, 1991), we can obtain a measure of the speed of convergence to a common level of inefficiency by regressing the change in the level of inefficiency on the lag of inefficiency (see also Fung, 2006). The absolute value of the negative coefficient on the lagged inefficiency indicates the speed of convergence. An interactive term between a zero-one dummy variable identifying the SOCBs and the lagged inefficiency variable is used to identify differences in the speed of decline in inefficiency between the SOCBs and the JSCBs.

As section 2 describes, Chinese banks have been in a perpetual state of reform since 1996 and subject to continuous regulatory changes. All these continuous innovations and reforms may have affected average cost efficiency and are difficult to model. Hence, we use time dummy variables to account for the many reforms and changes to Chinese banks over time. Each year is identified as ‘1’ for the year and ‘0’ for all other years. The definition of the two dependent variables lend itself to estimation by Seemingly Unrelated Regression (SURE) as $CI=TI+AI$. We didn’t use the two-step DEA procedure of Simar and Wilson (2007) in the light of the following two concerns. First, our dependent variables is the change in the efficiency scores, not the efficiency scores *per se*; therefore, we don’t face the bounded dependent variable issue as in the two-step DEA. Second, our original efficiency scores were calculated year by year, which indicates that the issue of the serial correlation of efficiency scores is also not very serious. Table 5 presents our results.

Table 5: SURE Estimates. Dependant Variable is the year-on-year change in inefficiency, ‘p’ values in parenthesis 1998-2016

	Change in technical-inefficiency = ΔTI_t		Change in allocative-inefficiency = ΔAI_t	
	Model 1	Model 2	Model 1	Model 2
Intercept	.015 (.302)	.075*** (.000)	.021 (.98)	-.076 (.217)
TI_{t-1}	-.599*** (.000)	-.532*** (.004)	-	-
AI_{t-1}	-	-	-.418*** (.000)	-.489*** (.000)
$SOCB* TI_{t-1}$.172***	.042***	-	-

<i>SOCB* AI_{t-1}</i>	(.006)	(.005)		
	-	-	.122*** (.003)	.156*** (.007)
<i>Time dummies</i>	Yes	Yes	Yes	Yes
<i>R²</i>	.7555	.7384	.8863	.6834
<i>Chi Square</i>	$\chi^2(20)=823^{***}$	$\chi^2(20)=2083^{***}$	$\chi^2(20)=745^{**}$	$\chi^2(20)=568^{***}$
<i>Obs</i>	266	266	266	266

*** significant at the 1%, ** significant at the 5%, * significant at the 10%.

A consistent result is the negative coefficient on the lagged measure of inefficiency which indicates a decline in both types of inefficiency over time. The estimated coefficients suggest that the JSCBs are reducing both technical (X-) inefficiency and allocative efficiency at roughly the same rate in both models. Hence, the results confirm *Hypothesis 2* and support the findings of Asmild and Matthews (2012). Importantly the interaction term of lagged inefficiency with the SOCB dummy variable suggests that the state-owned banks reduce both types of inefficiency at a slower speed than the JSCBs. Hence *Hypothesis 3* is confirmed.

6. Conclusion

This paper has used non-parametric methods to conduct an analysis of cost inefficiency in a sample of Chinese banks. Two models were considered. Both had a common set of inputs but outputs in the first model were performing loans, other earning assets and NPLs as a bad output, and in the second model were total earnings assets, non-performing loans, as a bad output, and non-interest earnings. The estimates of bank inefficiency were obtained using a homogeneous bootstrapping method to enable statistical inference and finite sample bias correction. Cost inefficiency was partitioned into technical (X-) inefficiency and allocative inefficiency. We find evidence to suggest that the allocative inefficiency is dominated by

excess labour. This confirms the findings of Asmild and Matthews (2012), but our findings also show that Chinese banks have been improving performance by reducing both types of inefficiency. However, the state-owned banking sector has been reducing both types of inefficiency at a slower rate than joint-stock commercial banks. This suggests that state-owned banks are more constrained by social and political objectives in their downsizing strategy than JSCBs.

In contrast to the earlier findings we find that average cost inefficiency in Chinese banks are in the region of 27-28% for the full sample period, and declining. We have argued in this paper that given the social and political constraints that Chinese banks have had to operate under, allocative inefficiency is symptomatic of rational decision making dictated by social employment objectives and not management deficiency. However, the results must be interpreted with caution. Not all allocative inefficiency can be attributed to over-staffing. It is also possible that the same poor management decisions that have contributed to technical inefficiency may also have contributed to allocative inefficiency.

Yet, the argument of this paper is that there have been significant improvements in bank efficiency in this period and that allocative inefficiency should be interpreted as largely rational and not an indicator of poor management. The 2016 average of all banks technical inefficiency and allocative inefficiency is 3% and 12% respectively in model 1, and 2% and 6% in model 2. If allocative inefficiency can be interpreted as mostly rational, then management inefficiency in 2016 measured by average technical inefficiency lies in the neighbourhood of 2-3% representing a significant improvement in efficiency performance and convergence on best practice, which may not be dissimilar to their western counterparts. The reform process that began in 1995 appears to have produced the appropriate results.

References

- Arif M and Can L, 2008, Cost and profit efficiency of Chinese banks: A non-parametric analysis, *China Economic Review*, 19, 260-273
- Asmild M and Matthews K, 2012, Multi-directional efficiency analysis of efficiency patterns in Chinese banks 1997-2008, *European Journal of Operational Research*, 219, 434-441
- Banker, R.D., Charnes, A., Cooper, W.W., Swarts, W., Thomas, D., 1989, An introduction to data envelopment analysis with some of its models and their uses. *Research in governmental and non-profit accounting* 5 (1) 125-163.
- Barro R J, 1991, Economic growth in a cross section of countries, *Quarterly Journal of Economics*, 106, 407-443.
- Berger, A.N., Hasan I, and Zhou M, 2009, Bank ownership and efficiency in China: What will happen in the world's largest nation?, *Journal of Banking and Finance*, 31, 113-130
- Berger A N., Hunter W C and Timme S G, 1993, The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present and Future, *Journal of Banking and Finance*, 17, 221-249
- Bogetoft P and Hougaard, 2003, Rational Efficiencies, *Journal of Productivity Analysis*, 20, 243 – 291
- Casu B and Molyneux, 2003, A Comparative Study of Efficiency in European Banking, *Applied Economics*, 35, 17, 1865-1876.
- Chen X, Skully M and Brown K, 2005, Banking Efficiency in China: Applications of DEA to pre-and post-Deregulation eras: 1993-2000, *China Economic Review*, 16, 2229-245
- Chernick, M. R., 2007, *Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition*, John Wiley & Sons Inc., New Jersey.
- Chi G, Sun X and Lu D, 2005, Empirical Study of the Cost Efficiency of China Commercial Banks, *Economic Research Journal*, 6, 104-114 (Chinese language)
- Cooper, W.W., Seiford, L.M., Tone, K., 2007, *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*, 2nd edition. Springer, New York.
- Drake L, 2004, Costs and Efficiency in Banking: A Survey of the Evidence from the US, the UK and Japan, in Mullineux A W and Murinde V (Eds) *Handbook of International Banking*, Cheltenham: Edward Elgar
- Dong F and Featherstone A, 2006, Technical and Scale Efficiencies for Chinese Rural Credit Cooperatives: A Bootstrapping Approach in Data Envelope Analysis, *Journal of Chinese Economic and Business Studies*, 4, 1, 57-75

Fang C, Sun W, Wang Z, Wang H, 2004, State-Owned Bank's Efficiency Measurement and their Behaviour Character Study, *Journal of Qualitative and Technical Economics*, 4, 42-45 (Chinese language)

Fethi M D and Pasiouras, 2010, Assessing bank efficiency and performance with operational research and artificial intelligence techniques, *European Journal of Operational Research*, 204, 189-198

Friedman, L., Sinuany-Stern, Z., 1998, Combining ranking scales and selecting variables in the DEA context: The case of industrial branches. *Computers and Operations Research* 25 (9) 781-791.

Fu X and Heffernan S, 2009, The effects of reform on China's bank structure and performance, *Journal of Banking and Finance* 33, 39-52

Fu X and Heffernan S, 2007, Cost X-efficiency in China's Banking Sector, *China Economic Review*, 18, 35-53

Fung M K, 2006, Scale economies, X-efficiency, and convergence of productivity among bank holding companies, *Journal of Banking and Finance*, 30, 2857-2874.

Garcia-Herrero A, Segio G and Santabarbara, 2006, China's Banking Reform: An Assessment of its Evolution and Possible Impact, *CESifo Economic Studies*, 52, 2, 304-363.

Golany, B.A., Roll, Y., 1989, An application procedure for data envelopment analysis. *Omega* 17 (3), 237-250.

Hosseinzadeh, A., Smyth, R., Valadkhani, A., and Le, V., 2016. Analyzing the efficiency performance of major Australian mining companies using bootstrap data envelopment analysis. *Economic Modelling*, 57, 26–35.

Jahanshahloo, G.R., Soleimani-damaneh, M., and Mostafae, A., 2008 A simplified version of the DEA cost efficiency model, *European Journal of Operational Research*, 184, 814-815.

Karim, M.Z.A., 2001. Comparative bank efficiency across select ASEAN countries, *ASEAN Economic Bulletin*, 18(3), 289-304.

Kneip A, Simar L, and Wilson PW, 2008, Asymptotics and Consistent Bootstraps for DEA estimators in Non-parametric Frontier Models, *Econometric Theory* 24, 1663–97.

Kneip A, Simar L, Wilson PW., 2011, A Computationally Efficient, Consistent Bootstrap for Inference with Non-parametric DEA Estimators, *Computational Economics* 38, 483–515.

Kumbhaker S C and Knox Lovell C A, 2003, *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press.

La Porta R, Lopez-De-Silanes F, Shleifer A, 2002, Government ownership of banks, *Journal of Finance*, 57, 1, 265-301

- Leibenstein H, 1966, Allocative Efficiency vs. X-Efficiency, *American Economic Review*, 56, 3, 392-415.
- Li J, 2013, Non-performing loans and asset management companies in China: Legal and regulatory challenges for achieving debt resolution and recovery, *Peking University School of Transnational Law Review*, 1, 85-159.
- Lin, J.Y., Sun, X.F. and Wu, H.X., 2015, "Banking Structure and Industrial Growth: Evidence from China". *Journal of Banking and Finance*, 58, 131-143
- Lin X and Zhang Y, 2009, Bank ownership reform and bank performance in China, *Journal of Banking and Finance* 31, 20-29.
- Liu C and Song W, 2004, Efficiency Analysis in China Commercial Banks based on SFA, *Journal of Financial Research*, 6, 138-142 (Chinese language).
- Liu, C., Miao, Y. and Zhu, Y., 2012, The Measurement of Banking Homogeneity and its Effect on Financial Stability, *Financial Regulation Research*, 2, 18-31 (Chinese language).
- Liu Z and Liu C (2004), Efficiency Analysis in China Commercial Banks based on DFA Approach, *Journal of Qualitative and Technical Economics*, 4, 42-45 (Chinese language)
- Löthgren M and Tambour M, 1999, Testing Scale Efficiency in DEA Models: A Bootstrapping Approach, *Applied Economics*, 31, 1231-1237.
- Matthews K, 2013, Risk management and managerial efficiency in Chinese banks: a network DEA approach, *Omega*, 41, 207-215
- Matthews K, Guo J and Zhang X, 2007, Rational Inefficiency and Non-performing loans in Chinese banking, *China Finance Review*, 3, 1, 55-75.
- Qian Q, 2003, On the Efficiency Analysis of SFA in China Commercial Banks, *Social Science of Nanjing*, 1, 41-46 (Chinese language).
- Qing W and Ou Y, 2001, Chinese Commercial Banks: Market Structure, Efficiency and its Performance, *Economic Science*, 100871, 34-45 (Chinese language).
- Simar L and Wilson P W, 1998, Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Non-Parametric Frontier Models, *Management Science*, 44, 49-61.
- Simar L and Wilson P W, 2000a, Statistical Inference in Non-Parametric Frontier Models: The State of the Art, *Journal of Productivity Analysis*, 13, 49-78.
- Simar L and Wilson P W, 2000b, A General Methodology for Bootstrapping in Non-Parametric Frontier Models, *Journal of Applied Statistics*, 27, 779-802.
- Simar L and Wilson P W (2007), "Estimation and Inference in Two-Stage semi-Parametric Models of Production Processes", *Journal of Econometrics*, 136, 31-64.

Simar L and Wilson P W, 2008, Statistical inference in nonparametric frontier models: Recent developments and perspectives, in H O Fried, C A Knox Lovell and S S Schmidt (eds) *The Measurement of Productivity Efficiency and Productivity Growth*, Oxford: Oxford University Press.

Simar L, Wilson P W., 2009, Inference by Subsampling in Nonparametric Frontier Models, *Discussion paper #0933, Institut de Statistique, Université Catholique de Louvain*, Louvain-la-Neuve, Belgium.

Simar L and Wilson P W, 2011, Performance of the Bootstrap for DEA Estimators and Iterating the Principle, in W. W. Cooper, L. M. Seiford and J. Zhu (eds) *Handbook on Data Envelopment Analysis*, New York: Springer.

Simar L and Wilson P W, 2013, Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives, *Foundations and Trends in Econometrics* 5, 183-337.

Sun Z, 2005, Frontier Efficiency Analysis for the State-Owned Banks in China, *Industrial Economics Research*, 3, 41-47 (Chinese language).

Sunil Dharmapala, P. (2018). Bias-correction in DEA efficiency scores using simulated beta samples: an alternative view of bootstrapping in DEA. *International Journal of Mathematics in Operational Research*, 12(4), 438-456.

Thanassoulis E, Portela M C S and Despić O, 2008, Data Envelopment Analysis: The mathematical programming approach to efficiency analysis”, in H O Fried, C A Knox Lovell and S S Schmidt (eds) *The Measurement of Productivity Efficiency and Productivity Growth*, Oxford: Oxford University Press.

Tsionas M G, and Izzeldin M, 2018, A novel model of costly technical efficiency, *European Journal of Operational Research*, 208, 653-664.

Tsionas, M. G., and Mallick, S. K., 2019, A Bayesian semiparametric approach to stochastic frontiers and productivity, *European Journal of Operational Research*, 274(1), 391–402.

Wang, J., 2009, The homogeneity of Chinese banks, *Review of Economic Research*, 48, 13 (Chinese language).

Wanke, P. and Barros, C. P., 2016, Efficiency drivers in Brazilian insurance: A two-stage DEA meta frontier-data mining approach. *Economic Modelling*, 53, 8–22.

Wei Y and Wang L, 2000, The Non-Parametric Approach to the Measurement of Efficiency: The Case of China Commercial Banks, *Journal of Financial Research*, 3, 88-96 (Chinese language).

Wijesiri, M., Viganò, L., and Meoli, M., 2015. Efficiency of microfinance institutions in Sri Lanka: a two-stage double bootstrap DEA approach. *Economic Modelling*, 47, 74–83.

Xu, M. and Chen, X., 2012, Monetary environment, capital adequacy ratio and risks of Chinese commercial banks, *Journal of Financial Research*, 7, 48-62 (Chinese language).

Xue F. and Yang D., 1998, Evaluating Bank Management and Efficiency using the DEA Model, *Econometrics and Technological Economy Research*, 6, 63-66 (Chinese language).

Yao S., Zhongwei H and Genfu W (2008), "Ownership Reform, Foreign Competition and Efficiency of Chinese Commercial Banks: A Non-parametric Approach", *The World Economy*, 1310-1326.

Yao S., Feng G. and Jiang C., 2004, The Empirical Analysis of Efficiency of Chinese Banks, *Economic Research Journal*, 8, 4-15 (Chinese language).

Zhao X., 2000, State-Owned Commercial Banks Efficiency Analysis, *Journal of Economic Science*", 6, 45-50 (Chinese language).

Zhao, X., Zhou, J. and Jiang, Z., 2001, An Analysis of the Efficiency of State-Owed Banks with Examples, *South China Financial Research*, 1, 25-27 (Chinese language).

Zhang C., Gu F. and Di Q., 2005, Cost Efficiency Measurement in China Commercial Banks based on Stochastic Frontier Analysis, *Explorations in Economic Issues*, 6, 116-119 (Chinese language).

Zhang, T. and Matthews, K., 2012,. Efficiency convergence properties of Indonesian banks 1992--2007, *Applied Financial Economics*, 22(17), 1465-1478.