Matthews, Kent; Xiao, Zhiguo; Zhang, Xu

Working Paper
Rational cost inefficiency in Chinese banks


Provided in Cooperation with:
Cardiff Business School, Cardiff University

Suggested Citation: Matthews, Kent; Xiao, Zhiguo; Zhang, Xu (2009) : Rational cost inefficiency in Chinese banks, Cardiff Economics Working Papers, No. E2009/13, Cardiff University, Cardiff Business School, Cardiff

This Version is available at:
http://hdl.handle.net/10419/65742

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes. You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public. If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Abstract

According to a frequently cited finding by Berger et al (1993), X-inefficiency contributes 20% to cost-inefficiency in western banks. Empirical studies of Chinese banks tend to place cost-inefficiency in the region of 50%. Such estimates would suggest that Chinese banks suffer from gross cost inefficiency. Using a non-parametric bootstrapping method, this study decomposes cost-inefficiency in Chinese banks into X-inefficiency and allocative-inefficiency. It argues that allocative inefficiency is the optimal outcome of input resource allocation subject to enforced employment constraints. The resulting analysis suggests that allowing for rational allocative inefficiency; Chinese banks are no better or worse than their western counterparts.

Keywords: Bank Efficiency; China; X-inefficiency; DEA; Bootstrapping

JEL codes: D23, G21, G28

The research for this work was carried out while the first author was a Visiting Fellow at Hong Kong Monetary Institute of the HKMA, April 2009. We are grateful to an anonymous referee, Patrick Minford, Kul Luintel and participants of the HKMIR research seminar for helpful comments.
1. Introduction

Bank efficiency in China has become a popular subject of research in recent years. A number of studies of Chinese banking efficiency have been published in Chinese scholarly journals\(^1\) but to date there have been only a handful of studies that are available to non-Chinese readers\(^2\). The consensus of finding among Chinese scholars is that the state-owned commercial banks tend on average to exhibit the lowest levels of efficiency and the joint stock commercial banks 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 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. Recent studies of bank efficiency in China have estimated cost inefficiency in the region of 50%\(^3\). Such figures are in stark contrast to the expectations of conventional inefficiency derived from the Berger et al (1993) study. It implies 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.

This research has three objectives. First it aims to decompose the measure of cost inefficiency in Chinese banks into technical inefficiency (sometimes viewed as X-inefficiency), and allocative inefficiency. This paper argues that while the

---

\(^1\) For example Qing and Ou, (2001); Xu, Junmin, and Zhensheng, (2001); Wei and Wang, (2000); Xue and Yang, (1998) and Zhao (2000) have used non-parametric methods while Liu and Song (2004), Zhang, Gu and Di (2005), Sun (2005) and Qian (2003) have used parametric methods.

\(^2\) Recent exceptions are studies using non-parametric methods by Chen et. al. (2005), and Yao et al (2008) and parametric methods by Fu and Heffernan (2009). Other recent studies published in English are, Lin and Zhang (2009), Berger et. al. (2009), Fu and Heffernen (2008), Matthews et al (2007)

underutilization of factors is consistent with the notion of X-inefficiency, but 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 official employment constraints, the implied cost inefficiency cannot be viewed as a management deficiency but a rational outcome of optimizing behaviour. The decomposition of cost inefficiency into X-inefficiency (technical inefficiency) and allocative inefficiency allows us to examine their evolution over the sample period.

Second, the measures of cost inefficiency and its decomposition are obtained using the familiar non-parametric method of Data-Envelopment-Analysis (DEA). The problem with the standard DEA approach is that it does not lend itself to statistical inference. This paper aims to provide an inferential capability to the point-estimates of inefficiency through the use of bootstrapping methods.

Third, the bootstrap estimates of inefficiency are use to test various hypotheses regarding the levels, trends and convergence in X-inefficiency and allocative inefficiency. Over time, as the profit motive replaces other (social and economic) imperatives the levels of X-inefficiency and allocative inefficiency should decline. The opening up of the Chinese banking market and threat of entry of foreign banks into the Chinese market may have lead to improved management, which should result in improved technical efficiency and lower cost-inefficiency as incumbent banks attempt to cut costs and consolidate their balance sheets.

This paper is organized on the following lines. The next section provides a brief motivation and discusses the model of rational allocative inefficiency. Section 3

---

4 See Simar and Wilson (2008)
outlines the method, reviews the literature and discusses the data. Section 4 discusses the results and section 5 concludes.

2. **A model of rational allocative efficiency**

At a first glance the Chinese banking system appears fragmented and diverse. In 2007 it consisted of 8,877 institutions, including 3 policy banks, 5 large state-owned commercial banks (SOCB), 12 joint-stock commercial banks (JSCB), 124 city commercial banks (CCB), 29 locally incorporated foreign bank subsidiaries and the rest made up of urban and rural credit cooperatives and other financial institutions⁵. In contrast, its neighbour India has only 482 institutions but this includes 59 nationwide state-owned and private banks and 29 foreign banks. While the 28 public sector banks dominate the market in India with 72% of the share of assets, in China the 5 state-owned or state-controlled banks command 53 per cent of the market⁶.

Chinese banks can be characterized as historically having low return on assets and low net interest margins (despite having wide interest spreads), a high non-performing loan ratio, a high cost-income ratio and overstaffed. A number of good descriptions of the Chinese banking system exist⁷ and what follows is a brief statement of the elements relevant to the issue of efficiency.

Up until 1995, control of the banking system remained firmly under the government and its agencies⁸. Under state control, the banks in China served the socialist plan of directing credits to specific projects dictated by political preference rather than commercial imperative. 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, party

---

⁵ CBRC Annual Report 2007
⁷ For example Garcia-Herreo et al (2006)
⁸ 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.
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 overstaffing.

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

\[
A = D^\alpha \left(L_1^\gamma L_2^{(1-\gamma)}\right)^\beta K^{(1-\alpha-\beta)}. \tag{1}
\]

The price of inputs are, the cost of deposits \( (r) \), the cost of labour \( (w) \) and the cost of
fixed assets \( (\rho) \). 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:

\[
\min \rho K + wL_1 + wL_2 + rD - \lambda \left(D^\alpha \left(L_1^\gamma L_2^{(1-\gamma)}\right)^\beta K^{(1-\alpha-\beta)} - A\right). \tag{2}
\]

The bank is constrained to employ some type 2 labour but clearly in an unconstrained
world the bank would only employ type 1 workers\(^{10}\).

The solution for output in the unconstrained case is given by:

---

\(^9\) 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.

\(^{10}\) 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 \( L_2 = 0 \).
\[ A = \left( \frac{\alpha}{\beta} \right)^{\alpha} w^{\alpha} r^{-\alpha} \left( \frac{1-\gamma}{\gamma} \right)^{(1-\gamma)\beta} \gamma^{-\alpha} \lambda_1^{\alpha+\beta} K^{(1-\alpha-\beta)}. \] (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:

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

The first order conditions are:

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

The marginal wage premium for type 2 labour is given by \( \lambda_2 \).

The output function is now:

\[ A = \left( \frac{\alpha}{\beta} \right)^{\alpha} w^{\alpha} r^{-\alpha} \gamma^{-\alpha} \lambda_1^{\alpha+\beta} L_1^{(1-\gamma)\beta} \bar{K}^{(1-\alpha-\beta)}. \] (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:

\[
\left( \frac{1-\gamma}{\gamma} \right)^{(1-\gamma)\beta} \lambda_1^{\alpha+\beta} L_{1U}^{\gamma} = \lambda_1^{\alpha+\beta} L_{2R}^{(1-\gamma)\beta}. \tag{6}
\]

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

\[
\frac{L_{1R} + \bar{L}_2}{L_{1U}}. \tag{7}
\]
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.

Figure 1 describes the situation for the case of the 2 variable inputs deposits and labour (physical capital is fixed by assumption). The isoquant is given by \(qq\) and the bank cost constraint by \(pp\). The point ‘\(e\)’ describes the cost minimum factor composition as in the unrestricted case which uses type 1 labour only. Point \(R\) describes the constrained case which uses both types of labour. The cost inefficiency generated by the allocative inefficiency is described by the ratio \(OP/OR\).

\[\text{Figure 1 Rational Allocative Inefficiency}\]

Substituting (6) into (7) and rearranging we have

\[CE = \frac{L_{1U} + L_2}{L_{1U}} = \left(1 - \frac{\gamma}{\beta}\right) \left(\frac{L_{1R}}{L_2}\right)^{\frac{(1-\gamma)\beta}{\alpha+\beta}} + \left(\frac{L_2}{L_{1R}}\right)^{\frac{\alpha+\beta}{\alpha+\beta}}\]  

(8)
We can show that for any values of $\alpha, \beta, \gamma, L_{1R}, \overline{L}_{2}$, $CE > 1$.

Qualitative evidence of overstaffing is scant and was gleaned from confidential interviews with individual bank managers. However an examination of the ratio of fixed assets to employees and deposits per employee at the beginning and end of the sample period is indicative. Table 1 shows the mean values of the ratios for the four large state-owned banks and the others in the sample for 1997 and 2007.

**Table 1: Average fixed assets per employee and deposits per employee**

<table>
<thead>
<tr>
<th>Year</th>
<th>Bank Group</th>
<th>Fixed assets per employee mill rmb</th>
<th>Deposits per employee mill rmb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>Big-4</td>
<td>0.1</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>0.3</td>
<td>14.2</td>
</tr>
<tr>
<td>2007</td>
<td>Big-4</td>
<td>0.2</td>
<td>18.4</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>0.4</td>
<td>45.3</td>
</tr>
</tbody>
</table>

The average of fixed assets per employee has doubled for the big four and increased by one-third for the remainder banks over this period indicating some branch expansion but also reduction in staffing. However, because of the accounting difficulties of comparing fixed asset values over time it may be more appropriate to

---

11 Define $\theta = \frac{\beta}{\alpha + \beta}$, $a = \frac{L_{1R}}{\overline{L}_{2}}$, then

$$CE = \left(1 - \frac{\gamma}{\theta}\right)^{\theta} \left[a^{(1-\theta)\gamma} + a^{(1-\theta)\gamma-1}\right] = \left(a^{1-\gamma}\right)^{(1-\gamma)\theta} \left[1 + \frac{1}{a}\right].$$

After analyzing the properties of the function $CE(\gamma, \theta, a)$ we have: (1) $CE > 1$; (2) If $\alpha, \beta, \gamma$ are fixed, then when $a = \frac{\gamma}{1-\gamma}$, $CE$ achieves its minimum value $\frac{1}{\gamma}$; (3) If we allow all $\alpha, \beta, \gamma, L_{1R}, \overline{L}_{2}$ to vary, then under the following situations, $CE$ will approach its minimum value 1: (i) $\theta \to 0$, i.e., $\beta \to 0$; (ii) $a \to \infty$; (iii) $a^\theta \to 1$. 

---
concentrate on the deposits per employee. Quite clearly, deposits per employee have tripled over the period reflecting the massive expansion in deposits in the decade to 2007 but also the relative reduction in staffing. These figures suggest that if there was an allocative inefficiency in the banks in 1997, there has been a strong attempt to reduce it by 2007.

A more indicative picture is gained from Chart 1 which examines the average $ deposit per employee of a sample of large commercial banks in the UK and that of a number of far-Eastern economies. We can interpret the figures for the foreign banks as an external benchmark.

**Chart 2: China and other economies average deposits per employee**

The chart suggests that by 2007, except for the big-4 banks in China, the other banks in general have converged on external benchmarks as indicated by foreign banks.

Having demonstrated that allocative inefficiency can be generated from rational decision making we now turn to the methodology of measuring inefficiency.
3. Methodology

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 (DEA), which is the extension by Charnes et al. (1978) (CCR)\textsuperscript{12} of the single input-output model of Farrell (1957) to a multiple input-output generalisation. 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\textsuperscript{13}, and therefore will not be elaborated here. A diagrammatic explanation illustrates the main concepts.

In figure 1, Technical efficiency is measured by the ratio $OR/OQ$ (Technical inefficiency is given by $RQ/OQ$). Cost inefficiency (CI) is measured by $PQ/OQ$ which in turn can be decomposed into X-inefficiency (or Technical inefficiency ($RQ/OQ$) and allocative inefficiency ($PR/OQ$).

DEA constructs a non-parametric frontier of the best practices amongst the decision-making units (DMUs). 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\textsuperscript{14}; 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.

\textsuperscript{12} Charnes et. al (1978) popularised the DEA method. According to Tavares (2002) who produces a bibliography of DEA (1978-2001), there are 3203 DEA authors whose studies cover a wide range of fields. Banxia.com also compiles DEA papers from 1978 to the present.

\textsuperscript{13} The most recent being Drake (2004)

\textsuperscript{14} Hababou (2002) and Avkiran (1999) provide a relatively thorough discussion of the merits and limits of the DEA.
One of the criticisms levelled at the standard DEA approach is that it produces estimates of efficiency but nothing can be said about the sensitivity (finite sample bias, confidence interval) of the estimator to sampling variation\(^{15}\). In a practical sense what this means is that if a DMU has a score of 0.95, it is 5% less efficient than the benchmark but nothing can be said about statistical significance – meaning is the 5% inefficiency statistically significant in any meaningful way. Without the capability for statistical inference, non-parametric methods would be weak alternatives to parametric methods of estimating efficiency. However, uncertainties also exist in the estimation of efficiency using DEA. The most obvious uncertainty is what comes from measurement error. Measurement error in the context of data on Chinese banks is particularly marked. There are three potential sources of error: firstly, differences between local bank's accounting procedures and those of international bodies; secondly, differences between local bank's accounting conventions; and thirdly, researcher assumptions relating to the generation of missing observations. Other uncertainties arise from the estimation of the efficiency frontier; changes to the inputs and/or outputs can cause large differences in the resulting scores. Furthermore there may be errors in the sampling variation caused by the difficulty in obtaining a sufficiently large and consistent sampling frame.

The bootstrap procedure for non-parametric frontier models is set out in Simar and Wilson (1998, 2000a, 2000b). The efficiency scores calculated with the original data are used to construct pseudo data. The bootstrap procedure is based on the idea that there exists a Data Generating Process (DGP), which can be determined by Monte Carlo simulation. By using the estimated distribution of the DGP to generate a large number of random samples, a set of pseudo estimates of the efficiency scores \(\hat{\theta}_i\)…

\(^{15}\) Simar and Wilson (1998)
are obtained. However this 'naive' bootstrap yields biased and inconsistent estimates (Simar and Wilson, 2000a). We operate the heterogeneous bootstrap procedure that produces consistent values of $\hat{\theta}_i$ from a kernel density estimate as given in Simar and Wilson (2000b). Briefly stated, we have observations $(x_i, y_i)$, $i = 1, \ldots, n$ which is assumed to be $i.i.d.$ random sample from a probability density function $f(x, y)$ on the production set $\Psi$. The idea of the bootstrap is to first estimate $f(x, y)$ using data $(x_i, y_i)$. Denoting the estimated density by $\hat{f}(x, y)$. The next step is to randomly draw $B$ samples using $\hat{f}(x, y)$. The bootstrap can be conducted in one of two ways. The direct approach uses Cartesian coordinates to estimate $\hat{f}(x, y)$ directly. This approach is difficult to implement and in general has not been followed by researchers. The indirect approach uses polar coordinates and takes the following four steps.

- **Step 1** Translate the Cartesian coordinate data $\{(x_i, y_i), i = 1, \ldots, n\}$ into polar coordinate data $\{(\theta_i, \eta_i, y_i), i = 1, \ldots, n\}$, where $\hat{\theta}_i$ denotes efficiency and $\eta_i$ denotes angles.

- **Step 2** Use $\{(\theta_i, \eta_i, y_i), i = 1, \ldots, n\}$ to estimate their density $f(\theta, \eta, y)$ and denote the estimated density by $\hat{f}(\theta, \eta, y)$.

- **Step 3** Randomly draw $B$ sets of random samples from $\hat{f}(\theta, \eta, y)$. Denote the $b^{th}$ bootstrap sample by $\{(\theta_{ib}^*, \eta_{ib}^*, y_{ib}^*), i = 1, \ldots, n\}$.

- **Step 4** Translate the $b^{th}$ bootstrap sample $\{(\theta_{ib}^*, \eta_{ib}^*, y_{ib}^*), i = 1, \ldots, n\}$ into Cartesian coordinate data $\{(x_{ib}^*, y_{ib}^*), i = 1, \ldots, n\}$. 
The difference between the homogeneous bootstrap and the heterogeneous bootstrap appears in step 1. The homogeneous bootstrap method assumes that the distribution of efficiency \( \theta \) is unrelated to the distribution of \((\eta, y)\). That is \(f(\theta|\eta, y) = f(\theta)\). The heterogeneous bootstrap does not maintain this assumption. Following the Simar-Wilson method, 1000 bootstrap values of the individual DMU for the efficiency scores are generated in each year\(^{16}\). The appendix provides a description of the algorithm.

Most studies of banking efficiency have focussed on the developed economies\(^{17}\). While there have been some studies of other Far Eastern economies\(^{18}\), the number is small in comparison. Indeed, from Berger and Humphrey's (1997) survey of 130 studies of frontier analysis in 21 countries, only 8 were about developing and Asian countries (including 2 in Japan). Studies on US financial institutions were the most common, accounting for 66 out of 116 single country studies.

A number of efficiency studies of Chinese banks have emerged in recent years, using both DEA and stochastic frontier analysis\(^{19}\). 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 (SOBs) than in the joint stock banks. Third, the gap between the two has been narrowing in recent years.

---

\(^{16}\) Recent bootstrapping applications to DEA have been conducted by Löthgren and Tambour (1999); in the case of banking efficiency by Casu and Molyneux (2003); and in the case of Chinese rural credit cooperatives, Dong and Featherstone (2006).


\(^{19}\) In addition to the papers cited in footnote 1, other studies by Chinese scholars that have used non-parametric techniques include Fang et. al. (2004) and studies using parametric methods include Liu and Liu (2004), Sun (2005), Qian (2003), Chi, Sun and Lu (2005), Yao, Feng and Jiang (2004)
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 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 described by the isoquant in Figure 1. 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.

Berger, Hunter and Timme (1993) argue that X-inefficiency constitutes 20% or more of bank costs. Poor motivation and weak pressure resulting in under utilization of factors of production are parts of what Leibenstein (1975) describes as ‘organisational entropy’. 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.

This study employs annual data (1997-2007) for 14 banks: the five state-owned banks (SOB), and nine joint-stock commercial banks (JSB). The total sample consisted of 154 bank year observations. The main source of the data was Fitch/Bankscope, and individual annual reports of banks. The choice of banks was based on the fact that they face a common market and compete nationwide.
Two approaches are normally taken in determining what constitutes bank input and output. Under the intermediation approach (Sealey and Lindley, 1977), bank assets measure outputs and liabilities measure inputs. In contrast, inputs in the production approach are physical entities such as labour and physical capital and revenue flows represent outputs. In this study, we adopt a hybrid of the two approaches. We use three inputs and three outputs for the estimation of technical efficiency. Inputs are the number of employees ($LAB$), fixed assets ($FA$) and total deposits ($DEP$). Outputs are total loans ($LOANS$), other earning assets ($OEA$), and non-interest income ($NII$). Although the latter variable remains undeveloped in China, it is selected to reflect the growing contribution of non-interest income to banks’ total income.

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 ($PF$). 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 2007 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 loans has grown at an average of 12 per cent a year. Other earning assets have grown at an average rate of 20% a year, in part reflecting the activities of the asset management companies that swapped tranches of the NPLs of the big 4 SOBs for bonds in 1999.
and 2001. 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 - 2007 (million RMB)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean 1997</th>
<th>SD 1997</th>
<th>Mean 2007</th>
<th>SD 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOANS</td>
<td>Total stock of loans</td>
<td>430033</td>
<td>657201</td>
<td>1296424</td>
<td>1351614</td>
</tr>
<tr>
<td>OEA</td>
<td>Investments</td>
<td>205103</td>
<td>301626</td>
<td>1227591</td>
<td>1419119</td>
</tr>
<tr>
<td>NII</td>
<td>Net Fees and Commissions</td>
<td>862</td>
<td>1922</td>
<td>9994</td>
<td>12851</td>
</tr>
<tr>
<td>LAB</td>
<td>Total Employed</td>
<td>105138</td>
<td>175233</td>
<td>111960</td>
<td>157645</td>
</tr>
<tr>
<td>DEP</td>
<td>Total stock of Deposits</td>
<td>604013</td>
<td>891353</td>
<td>2309760</td>
<td>2568177</td>
</tr>
<tr>
<td>FA</td>
<td>Fixed assets</td>
<td>12831</td>
<td>19398</td>
<td>27374</td>
<td>33704</td>
</tr>
<tr>
<td>PL</td>
<td>Unit price of labour</td>
<td>.0631</td>
<td>.0380</td>
<td>.2353</td>
<td>.0915</td>
</tr>
<tr>
<td>PF</td>
<td>Unit price of funds</td>
<td>.0502</td>
<td>.0202</td>
<td>.0214</td>
<td>.0057</td>
</tr>
<tr>
<td>PK</td>
<td>Unit price of fixed assets</td>
<td>.6528</td>
<td>.5282</td>
<td>.7409</td>
<td>.2333</td>
</tr>
</tbody>
</table>

Sources: Fitch/Bankscope and author calculations from web sources.

Other points to note are that net employment has grown by an average of 0.6% a year but average labour cost has grown by a remarkable 14% 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.

Having outlined the methodology and the data we now examine the empirical results from the bootstrap method.
4.0 Empirical Results

Tables 3a-b illustrate the results of the bootstrap method for two representative years; 1997 and 2007 in the case of technical efficiency (X-efficiency, XE) and cost efficiency (CE). The tables show the biased estimates, the bootstrapped bias-adjusted estimates and the 95% confidence intervals for 2000 bootstraps.

Simar and Wilson (2000a, 2000b) show that the bias correction will introduce extra noise that may result in a mean-square error (MSE) greater than the MSE of the bias-unadjusted bootstrap values. In the limit the bias corrected MSE will be four times that of the uncorrected estimate and Simar and Wilson caution against the bias correction unless the ratio \( \frac{\hat{\text{bias}^2}}{\hat{\sigma}^2} \) is greater than unity, where \( \hat{\text{bias}} \) is the bias correction and \( \hat{\sigma}^2 \) is the sample variance of the uncorrected bootstrap values. The statistic \( \rho_i \) is defined as \( \rho_i = \left( \frac{\hat{\text{bias}^2}}{\hat{\sigma}^2} \right) \) and is shown in the final column. A statistic of \( \rho_i > 1 \) implies that the bias correction is valid.
Table 3a: Median Bootstrap Estimates of Cost efficiency 1997 & 2007 CRS

<table>
<thead>
<tr>
<th>Bank</th>
<th>Efficiency Score</th>
<th>Bias corrected</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1997</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABOC</td>
<td>.7941*</td>
<td>0.4856</td>
<td>0.4212</td>
<td>0.5683</td>
<td>22.9</td>
</tr>
<tr>
<td>BOC</td>
<td>.9502*</td>
<td>0.6942*</td>
<td>0.1441</td>
<td>0.8789</td>
<td>0.64</td>
</tr>
<tr>
<td>BCOMM</td>
<td>.9154*</td>
<td>0.3099</td>
<td>-0.1241</td>
<td>0.6710</td>
<td>2.79</td>
</tr>
<tr>
<td>CCB</td>
<td>.8938*</td>
<td>0.7606</td>
<td>0.7069</td>
<td>0.7863</td>
<td>11.3</td>
</tr>
<tr>
<td>ICBC</td>
<td>1.000*</td>
<td>0.4158</td>
<td>0.0362</td>
<td>0.7490</td>
<td>2.86</td>
</tr>
<tr>
<td>CITIC</td>
<td>1.000*</td>
<td>0.4425</td>
<td>0.0381</td>
<td>0.8305</td>
<td>2.17</td>
</tr>
<tr>
<td>CMB</td>
<td>1.000*</td>
<td>0.5093</td>
<td>0.0559</td>
<td>0.7682</td>
<td>2.10</td>
</tr>
<tr>
<td>CMBCL</td>
<td>1.000*</td>
<td>0.6721</td>
<td>0.2203</td>
<td>0.8547</td>
<td>1.37</td>
</tr>
<tr>
<td>EVERBRT</td>
<td>1.000*</td>
<td>0.5537</td>
<td>0.2113</td>
<td>0.6604</td>
<td>1.75</td>
</tr>
<tr>
<td>GDB</td>
<td>.8325*</td>
<td>0.5242</td>
<td>0.0507</td>
<td>0.8457</td>
<td>1.73</td>
</tr>
<tr>
<td>HUAXIA</td>
<td>1.000*</td>
<td>0.3325</td>
<td>0.0191</td>
<td>0.7127</td>
<td>3.99</td>
</tr>
<tr>
<td>IBCL</td>
<td>1.000*</td>
<td>0.8518</td>
<td>0.7406</td>
<td>0.9442</td>
<td>2.64</td>
</tr>
<tr>
<td>SDB</td>
<td>.8409*</td>
<td>0.6142</td>
<td>0.5430</td>
<td>0.6782</td>
<td>9.08</td>
</tr>
<tr>
<td>SPB</td>
<td>1.000*</td>
<td>0.2533</td>
<td>0.0162</td>
<td>0.6641</td>
<td>5.88</td>
</tr>
<tr>
<td><strong>2007</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABOC</td>
<td>.6848*</td>
<td>0.1851</td>
<td>0.1095</td>
<td>0.3288</td>
<td>25.8</td>
</tr>
<tr>
<td>BOC</td>
<td>.8578*</td>
<td>0.6725</td>
<td>0.5419</td>
<td>0.7138</td>
<td>5.25</td>
</tr>
<tr>
<td>BCOMM</td>
<td>.7737*</td>
<td>0.4566</td>
<td>0.3745</td>
<td>0.5412</td>
<td>14.6</td>
</tr>
<tr>
<td>CCB</td>
<td>.8250*</td>
<td>0.5929</td>
<td>0.4190</td>
<td>0.6472</td>
<td>4.81</td>
</tr>
<tr>
<td>ICBC</td>
<td>.8200*</td>
<td>0.7056</td>
<td>0.3610</td>
<td>0.8566</td>
<td>1.14</td>
</tr>
<tr>
<td>CITIC</td>
<td>.9451*</td>
<td>0.7630</td>
<td>0.4132</td>
<td>0.9078</td>
<td>1.06</td>
</tr>
<tr>
<td>CMB</td>
<td>.9962*</td>
<td>0.6674</td>
<td>0.3918</td>
<td>0.8560</td>
<td>2.55</td>
</tr>
<tr>
<td>CMBCL</td>
<td>1.000*</td>
<td>0.6691</td>
<td>0.3750</td>
<td>0.8294</td>
<td>1.91</td>
</tr>
<tr>
<td>EVERBRT</td>
<td>.9536*</td>
<td>0.6771</td>
<td>0.4342</td>
<td>0.7768</td>
<td>1.81</td>
</tr>
<tr>
<td>GDB</td>
<td>.8922*</td>
<td>0.6452</td>
<td>0.3744</td>
<td>0.8284</td>
<td>2.00</td>
</tr>
<tr>
<td>HUAXIA</td>
<td>.9293*</td>
<td>0.6268</td>
<td>0.3374</td>
<td>0.8359</td>
<td>2.93</td>
</tr>
<tr>
<td>IBCL</td>
<td>1.000*</td>
<td>0.5846</td>
<td>0.5024</td>
<td>0.6370</td>
<td>12.5</td>
</tr>
<tr>
<td>SDB</td>
<td>1.000*</td>
<td>0.6426</td>
<td>0.2745</td>
<td>0.8274</td>
<td>2.07</td>
</tr>
<tr>
<td>SPB</td>
<td>.9712*</td>
<td>0.7057*</td>
<td>0.2856</td>
<td>0.8835</td>
<td>0.98</td>
</tr>
</tbody>
</table>

* significant bias at the 95% level of confidence. # Bias correction invalid
Table 3b: Median Bootstrap Estimates of Technical efficiency 1997 & 2007 CRS

<table>
<thead>
<tr>
<th>Bank</th>
<th>Efficiency Score</th>
<th>Bias corrected</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABOC</td>
<td>0.8914*</td>
<td>0.9809</td>
<td>0.7671</td>
<td>0.8803</td>
<td>1.07</td>
</tr>
<tr>
<td>BOC</td>
<td>1.0000*</td>
<td>0.8304</td>
<td>0.6119</td>
<td>0.7796</td>
<td>9.12</td>
</tr>
<tr>
<td>BCOMM</td>
<td>1.0000*</td>
<td>0.6889</td>
<td>0.5804</td>
<td>0.7624</td>
<td>9.95</td>
</tr>
<tr>
<td>CCB</td>
<td>0.9699*</td>
<td>0.6430</td>
<td>0.8521</td>
<td>0.9572</td>
<td>1.13</td>
</tr>
<tr>
<td>ICBC</td>
<td>1.0000*</td>
<td>0.9109</td>
<td>0.5887</td>
<td>0.7397</td>
<td>13.2</td>
</tr>
<tr>
<td>CITIC</td>
<td>1.0000*</td>
<td>0.6500</td>
<td>0.6122</td>
<td>0.8156</td>
<td>6.50</td>
</tr>
<tr>
<td>CMB</td>
<td>1.0000*</td>
<td>0.6882</td>
<td>0.6056</td>
<td>0.7821</td>
<td>8.21</td>
</tr>
<tr>
<td>CMBCL</td>
<td>1.0000*</td>
<td>0.6801</td>
<td>0.5844</td>
<td>0.7292</td>
<td>14.8</td>
</tr>
<tr>
<td>EVERBRT</td>
<td>0.8932*</td>
<td>0.6423</td>
<td>0.6028</td>
<td>0.8058</td>
<td>2.69</td>
</tr>
<tr>
<td>GDB</td>
<td>1.0000*</td>
<td>0.7007</td>
<td>0.6056</td>
<td>0.8060</td>
<td>6.79</td>
</tr>
<tr>
<td>HUAXIA</td>
<td>1.0000*</td>
<td>0.6822</td>
<td>0.5516</td>
<td>0.7370</td>
<td>10.0</td>
</tr>
<tr>
<td>IBCL</td>
<td>1.0000*</td>
<td>0.6152</td>
<td>0.7097</td>
<td>0.9079</td>
<td>2.64</td>
</tr>
<tr>
<td>SDB</td>
<td>0.9809*</td>
<td>0.8165*</td>
<td>0.7755</td>
<td>0.9530</td>
<td>0.89</td>
</tr>
<tr>
<td>SPB</td>
<td>1.0000*</td>
<td>0.8925</td>
<td>0.5692</td>
<td>0.6922</td>
<td>19.9</td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABOC</td>
<td>0.9088*</td>
<td>0.8274</td>
<td>0.7543</td>
<td>0.8715</td>
<td>1.92</td>
</tr>
<tr>
<td>BOC</td>
<td>1.0000*</td>
<td>0.8744</td>
<td>0.7879</td>
<td>0.9283</td>
<td>2.76</td>
</tr>
<tr>
<td>BCOMM</td>
<td>0.9033*</td>
<td>0.8080</td>
<td>0.7200</td>
<td>0.8539</td>
<td>1.80</td>
</tr>
<tr>
<td>CCB</td>
<td>1.0000*</td>
<td>0.8749</td>
<td>0.7719</td>
<td>0.9515</td>
<td>1.58</td>
</tr>
<tr>
<td>ICBC</td>
<td>1.0000*</td>
<td>0.7431</td>
<td>0.6421</td>
<td>0.8580</td>
<td>3.81</td>
</tr>
<tr>
<td>CITIC</td>
<td>1.0000*</td>
<td>0.7836</td>
<td>0.6703</td>
<td>0.8881</td>
<td>2.69</td>
</tr>
<tr>
<td>CMB</td>
<td>1.0000*</td>
<td>0.7521</td>
<td>0.6695</td>
<td>0.8547</td>
<td>5.39</td>
</tr>
<tr>
<td>CMBCL</td>
<td>1.0000*</td>
<td>0.7555</td>
<td>0.6671</td>
<td>0.8677</td>
<td>4.33</td>
</tr>
<tr>
<td>EVERBRT</td>
<td>0.9415*</td>
<td>0.7863</td>
<td>0.6912</td>
<td>0.8659</td>
<td>2.63</td>
</tr>
<tr>
<td>GDB</td>
<td>0.9796*</td>
<td>0.7443</td>
<td>0.6425</td>
<td>0.8839</td>
<td>2.95</td>
</tr>
<tr>
<td>HUAXIA</td>
<td>1.0000*</td>
<td>0.7017</td>
<td>0.6277</td>
<td>0.8093</td>
<td>7.66</td>
</tr>
<tr>
<td>IBCL</td>
<td>0.9726*</td>
<td>0.9000*</td>
<td>0.7989</td>
<td>0.9423</td>
<td>0.97</td>
</tr>
<tr>
<td>SDB</td>
<td>1.0000*</td>
<td>0.7518</td>
<td>0.6669</td>
<td>0.8363</td>
<td>5.77</td>
</tr>
<tr>
<td>SPB</td>
<td>1.0000*</td>
<td>0.7469</td>
<td>0.6410</td>
<td>0.8673</td>
<td>3.33</td>
</tr>
</tbody>
</table>

* significant bias at the 95% level of confidence. # Bias correction invalid

The Tables report the median of the 2000 bootstrap values of cost efficiency (CE) and X-efficiency (Technical efficiency) (XE) as the standard for previous studies. The confidence intervals of the bootstrap values in Table 2a-b support the

---

See for example Dong and Featherstone (2006). The argument for reporting the median rather than the mean is that the distribution of the efficiency scores may not be standard normal. In reality there was little difference between the two.
significance of the bias. The value of $\rho$ indicates the validity of the bias correction which in all cases was always one-sided$^{21}$.

We now turn to the third objective of this paper and that is to evaluate the levels, tends and convergence in the two types of inefficiency. Table 4 shows the full period sample means and weighted means of the cost inefficiency (CI) and X-inefficiency (XI) estimates obtained as $CI = (1 - CE)$ and $XI = (1 - XE)$ respectively. The allocative inefficiency estimate was obtained as the residual of the cost inefficiency CI and X-inefficiency XI ($AI = CI - XI$)$^{22}$.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Cost-inefficiency</th>
<th>X-inefficiency</th>
<th>Allocative - Inefficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean all</td>
<td>55.0</td>
<td>30.4</td>
<td>24.6</td>
</tr>
<tr>
<td>Weighted mean</td>
<td>50.5</td>
<td>25.9</td>
<td>24.6</td>
</tr>
<tr>
<td>Mean SOB</td>
<td>52.9</td>
<td>25.9</td>
<td>27.0</td>
</tr>
<tr>
<td>W - mean SOB</td>
<td>50.1</td>
<td>25.1</td>
<td>25.1</td>
</tr>
<tr>
<td>Mean JSCB</td>
<td>55.0</td>
<td>30.4</td>
<td>24.6</td>
</tr>
<tr>
<td>W – mean JSCB</td>
<td>55.2</td>
<td>30.8</td>
<td>24.4</td>
</tr>
</tbody>
</table>

Weighting the inefficiency scores provide a more accurate picture of the average levels of inefficiency but in reality the difference between the pure average and the weighted average is of second-order magnitude. It can be seen from Table 4 that the weighted average cost inefficiency over the period for the SOCBs is slightly smaller than the average for the JSCBs. The table also indicates that the weighted average allocative inefficiency is comparable for both groups of banks.

$^{21}$ Out of 154 bootstrap results for each bank-year in only seven cases was the bias correction invalid. In such cases the bootstrap value was used for consistency.

$^{22}$ Strictly $AE = CE/XE$ but as we are dealing with the median values of an unknown distribution it was convenient to define $AI = CI-XI = XE-CE$. The alternative measure is $AI = (1-AE)/(XE-CE)XE$. 
Using the mean of the un-weighted scores, we test if there is a significant difference between the average inefficiency of banks that have a foreign stakeholding from those that do not. Since the distribution of the inefficiency scores may not be standard normal we apply a non-parametric test (Mann-Whitney) with the results shown in Table 5.

<table>
<thead>
<tr>
<th>Table 5: Non-parametric test for difference in means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Mean (Foreign stake)</td>
</tr>
<tr>
<td>Mean (all other)</td>
</tr>
<tr>
<td>Z – statistic</td>
</tr>
</tbody>
</table>

The results of the non-parametric tests indicate a significant difference. Banks that have a foreign stake-holding have lower average X-inefficiency and allocative inefficiency. But is inefficiency in Chinese banks as a whole declining and if so can anything be said about the speed at which the inefficient banks are converging on the benchmark efficient banks? Chart 2 and 3 show the yearly weighted average technical inefficiency and allocative inefficiency for the SOCBs and JSCBs as a group. The short dotted line is the weighted average measure of inefficiency for the SOCBs and the long dotted line is the same for the JSCBs. Chart 2 shows a rise in technical inefficiency in the years 1999-2001 which roughly coincides with the activities of the Asset Management Companies that transferred tranches of NPLS of the big 4 SOCBs and returned them as face value bonds in 1999-2000\(^{23}\).

\(^{23}\) During 1999-2000 4 AMCs bought up an aggregate of roughly $205 billion in NPLs in return for 10-year bonds paying a fixed rate of 2.25%
It is possible that the portfolio switch of the NPLs from the loan book to other earning assets of the big-4 balance sheets distorted the average efficiency measure for
the SOCBs. However, this explanation is less convincing for the JSCBs which show a similar pattern for the allocative inefficiency in Chart 3. Also the asset management companies conducted a further transfer of NPLs from the big-4 banks of around $120 billion in 2003 which does not appear to have affected the average inefficiency figures as seen in the charts. But what is clear is that there is a discernible negative trend in both types of inefficiency for both groups of 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 and environmental and bank specific variables to allow for convergence to different levels of inefficiency\(^{24}\). However it is shown by Simar and Wilson (2007) that the estimated inefficiencies may be serially correlated. They propose a double bootstrap procedure to adjust for the bias caused by the inherent correlation among the estimated inefficiencies. The problem of potential bias is further compounded by the existence of the lagged inefficiency score. Developing a valid bootstrap procedure for estimating beta-convergence is computationally intractable. However in an attempt to deal with the potential serial correlation we present estimates of the rate of decline of inefficiency controlled for individual bank factors using a panel GLS *Heteroskedastic-Autocorrelation Consistent* (HAC) estimator (Table 6). The dependant variable was the yearly change in the specific type of inefficiency. Bank specific variables were lagged one period to avoid potential endogeneity.

---

\(^{24}\) See also Fung (2006)
Table 6: Panel GLS (Panel heteroskedastic-autocorrelation consistent estimates). Dependant Variable is the year-on-year change in inefficiency. P values in parenthesis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Change in X-inefficiency = $\Delta X_{i,t}$ (observations = 140)</th>
<th>Change in Allocative inefficiency = $\Delta A_{i,t}$ (observations = 140)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.353*** (.000)</td>
<td>.354*** (.000)</td>
</tr>
<tr>
<td>$X_{i,t-1}$</td>
<td>-.444*** (.000)</td>
<td>-.485*** (.000)</td>
</tr>
<tr>
<td>$A_{i,t-1}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$SOB$</td>
<td>.021** (.035)</td>
<td>-</td>
</tr>
<tr>
<td>$SOB* X_{i,t-1}$</td>
<td>-</td>
<td>.055** (.049)</td>
</tr>
<tr>
<td>$SOB* A_{i,t-1}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$ln(TA)_{i,t-1}$</td>
<td>-.018*** (.000)</td>
<td>-.017*** (.000)</td>
</tr>
<tr>
<td>$FOR$</td>
<td>-.001*** (.000)</td>
<td>-.001*** (.000)</td>
</tr>
<tr>
<td>$L*$</td>
<td>1021.7</td>
<td>1037.2</td>
</tr>
<tr>
<td>$F_{1,13}$</td>
<td>42.6***</td>
<td>48.4***</td>
</tr>
<tr>
<td>$\chi^2_{13}$</td>
<td>100.2***</td>
<td>100.2***</td>
</tr>
</tbody>
</table>

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

The key covariate is the lag in the level of inefficiency. A negative coefficient defines the common speed of convergence of inefficiency. The bank specific variable that proved significant was the logarithm of total assets as a measure of size. Environmental variables were a zero-one dummy variable identifying SOCBs (SOB) and the proportion of the bank owned by foreign financial enterprises (FOR). An interaction term between the lag in inefficiency and the SOB dummy defined differing speed of adjustment between SOCBs and JSCBs. The third but last row is the log likelihood. The second but last row is an F test for autocorrelation in panel data.
(Wooldridge, 2002) and the final row is a likelihood ratio test for heteroskedasticity in panel data.

The last two rows of Table 6 indicate that the use of the HAC estimator was appropriate. Autocorrelation in the panel could not be rejected at the 1% level for the regressions of both types of inefficiency, and heteroskedasticity could not be rejected at the 1% for the X-inefficiency and 10% for allocative inefficiency regressions.

The results of Table 6 indicate that controlling for ownership, larger banks are associated with lower levels of both types of inefficiency. Banks that have a foreign stake are associated with lower levels of both types of inefficiency, confirming the finding reported in Table 4. State-owned banks are associated with higher levels of both types of inefficiency and in particular higher levels of allocative inefficiency. Importantly, the negative coefficient on the lagged measure of inefficiency shows significant decline in both measures of inefficiency. The interaction term of lagged inefficiency with SOB suggests that the state-owned banks reduce inefficiency at a lower speed than the JSCBs\textsuperscript{25}, but that they are reducing both types at roughly the same speed. The results suggest that the joint-stock commercial banks are reducing allocative inefficiency at a faster rate than X-inefficiency.

7.0 Conclusion

This paper has used non-parametric methods to conduct an analysis of inefficiency in a sample of Chinese banks. The estimates of bank inefficiency were obtained using a bootstrapping method to enable statistical inference. We have partitioned cost inefficiency into X-inefficiency and allocative inefficiency. Our

\textsuperscript{25} Because of the problem of multi-collinearity between SOB and the interaction term in the X-inefficiency regression, a grid search using constrained estimates that maximised the log likelihood is reported in Model (3) of Table 5.
findings suggest that Chinese banks have been improving performance by reducing both types of inefficiency. However, the state-owned banking sector has higher levels of both types of inefficiency and is also 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.

We confirm the findings from nonparametric and stochastic frontier based studies of Chinese banks that average cost inefficiency are in the region of 50%. Inefficiency in Chinese banking is made up of both X-inefficiency and allocative inefficiency. We have argued in this paper that given the social and political constraints that Chinese banks had to operate in, allocative inefficiency was symptomatic of rational decision making dictated by social employment objectives.

However, we must still interpret the results with caution. Not all of allocative inefficiency can be attributed to over-staffing and not all of over-staffing can be explained by past employment objectives. Overstaffing caused by political and social constraints is observationally equivalent to rent-seeking behaviour by bank managers (Matthews et al, 2007). It is also possible that poor management decisions that may have contributed to X-inefficiency could also have contributed to allocative inefficiency.

Yet, the argument of this paper is that there have been significant improvements in bank efficiency. The 2007 weighted average of all banks X-inefficiency and allocative inefficiency is 16% and 20% respectively. If the Berger et al (1993) finding that 20% of all bank costs are due to X-efficiency represents a common benchmark for banking markets in general then the message of this paper is that Chinese banks are not out of line.
Appendix

The heterogeneous bootstrap algorithm

In this paper, we implement the heterogeneous bootstrap algorithm of Simar and Wilson (2000b) to compute the bias-corrected technical efficiency and cost efficiency scores. Specifically, we follow the following steps:

Note: In our application \( n = 14, p = 3, q = 3 \).

**Step 1.** Compute the technical efficiency scores using the original data. To be consistent with Simar and Wilson (2000b), we denote these technical efficiency estimates by \( \hat{\delta}_i = \hat{\delta}_i(x_i, y_i), i = 1, \cdots, n \). The \( \hat{\delta}_i \)'s are computed using the linear programming described in equation (17) of Simar and Wilson (2000). Note that by definition, \( \hat{\delta}_i \geq 1, \forall i = 1, \cdots, n \).

**Step 2.** For each \( \hat{\delta}_i = 1, \cdots, n \), translate the data \( (x_i, y_i) \) into its polar coordinate representation \( \left( y_i, \eta_i, \delta_i \right) \) by defining \( \eta_i = \arctan \left( \frac{y_{i+1}}{x_k} \right), \forall k = 1, 2, \) if \( x_{i+k} > 0 \). Define matrices \( Z \) and \( Z_R \) by letting their \( i \)-th row be \( \left( y_{i+1}, y_{i+2}, y_{i+3}, \eta_i, \delta_i \right) \) and \( \left( y_{i+1}, y_{i+2}, y_{i+3}, \eta_i, 2 - \delta_i \right) \) respectively. Define matrices \( Z_{1,2}, Z_{1,3}, Z_{1,5} \) and \( Z_{1,3} \) by letting their \( i \)-th row be \( \left( y_{i+1}, y_{i+2}, y_{i+3}, \eta_i, \delta_i \right) \) and \( \left( y_{i+1}, y_{i+2}, y_{i+3}, \eta_i, 2 - \delta_i \right) \) respectively. Define matrices \( Z_{1,4} \) and \( Z_{1,3} \) by letting their \( i \)-th row be \( \left( y_{i+1}, y_{i+2}, y_{i+3}, \eta_i, \delta_i \right) \) and \( \left( y_{i+1}, y_{i+2}, y_{i+3}, \eta_i, 2 - \delta_i \right) \) respectively. Define matrices \( Z_{1,5} \) and \( Z_{1,6} \) by letting their \( i \)-th row be \( \left( y_{i+1}, y_{i+2}, y_{i+3}, \eta_i, \delta_i \right) \) and \( \left( y_{i+1}, y_{i+2}, y_{i+3}, \eta_i, 2 - \delta_i \right) \) respectively. For simplicity we rename the above 16 matrices as \( Z_1, Z_2, \cdots, Z_{16} \) respectively. Form the augmented \( 16 \times 6 \)

\[
\begin{bmatrix}
Z_1 \\
\cdots \\
Z_{16}
\end{bmatrix}
\]

**Step 3.** Compute the estimated covariance matrices of \( Z_k, k = 1, \cdots, 16 \). Denote them by \( \hat{\Sigma}_k, k = 1, \cdots, 16 \) respectively. Obtain the upper triangular matrices \( L_k, k = 1, \cdots, 16 \) through Cholesky decomposition: \( \hat{\Sigma}_k = L_k L_k^\top \).

**Step 4.** Choose an appropriate bandwidth as suggested by Simar and Wilson (2000). We use the normal reference rule for its simplicity. This choice was also supported by Simar and Wilson’s (2000b) simulation study.

**Step 5.** Draw \( n \) rows with replacement from \( \tilde{Z} \) to form a new \( n \times (p + q) \) matrix \( \tilde{Z}^* \). Denote the row means of \( \tilde{Z}^* \) by \( \tilde{z}^* \). The dimension of \( \tilde{z}^* \) is \( n \times 1 \).

**Step 6.** Generate an \( n \times (p + q) \) random matrix \( \varepsilon \) such that its entries are i.i.d. standard normal random variates. Denote its \( i \)-th row by \( \varepsilon_i \). Construct a new \( n \times (p + q) \) matrix \( \varepsilon_i^* \) by defining its \( i \)-th row \( \varepsilon_i^* = \varepsilon_i L_k \), if the \( i \)-th row of \( \tilde{Z}^* \) was drawn from \( Z_k \).
Step 7. Define the \( n \times (p + q) \) matrix \( \Gamma \) by
\[
\Gamma = \frac{1}{\sqrt{1 + h^2}} (M \tilde{Z}^* + h \epsilon^*) + I_n \otimes \mathbf{z}^* ,
\]
where \( I_n \) is the \( n \times 1 \) vector of ones, and \( M = I_n - \frac{1}{n} I_n' \), with \( I_n \) the identity matrix of order \( n \), and \( \otimes \) the Kronecker product.

Step 8. Denote the \( i^{th} \) row of \( \Gamma \) by \( \gamma_i \), with \( \gamma_i = (\gamma_{i1}, \gamma_{i2}, \gamma_{i3}) \), where \( \gamma_{i1} \) is \( q \)-dimensional, \( \gamma_{i2} \) is \( (p-1) \)-dimensional, and \( \gamma_{i3} \) is \( 1 \)-dimensional. Construct an \( n \times (p + q) \) matrix \( Z^* \) by defining its \( i^{th} \) row \( z_i^* = (z_{i1}^*, z_{i2}^*, z_{i3}^*) \) as
\[
z_i^* = \begin{cases} (\gamma_{i1}, \gamma_{i2}, \gamma_{i3}) & \text{if } \gamma_{i3} \geq 1 \\ (\gamma_{i1}, \gamma_{i2}, 2 - \gamma_{i3}) & \text{otherwise} \end{cases}
\]
The so-constructed \( Z^* \) is the bootstrap data in polar coordinate.

Step 9. Translate the polar coordinate data \( Z^* \) back to Cartesian coordinate data
\[
X^* = \{ (x_i^*, y_i^*) \}_{i=1,\ldots,n}. 
\]
Specifically, let \( y_i^* = z_{i1}^*, \eta_i^* = z_{i2}^*, \delta_i^* = z_{i3}^* \), and define \( \bar{x}_i \) by \( \bar{x}_i = 1, \bar{x}_k = \tan(\eta_k^*) \), \( k = 1, \ldots, p-1 \). Define \( x_i^* = \frac{\delta_i^*}{\delta_i^*(\bar{x}_i, y_i^*)} \bar{x}_i \). \( X^* \) is our bootstrap sample data. Repeat Step 5-8 if the linear programming of obtaining \( \hat{\delta}(\bar{x}_i, y_i^*) \) has no solution.

Step 10. For a given point \((x_i, y_i)\), using the bootstrap data \( X^* \) as the reference data to compute the technical efficiency score \( \hat{\delta}^*(x_i, y_i) \) and cost efficiency score \( \hat{\vartheta}^*(x_i, y_i) \) (along with the data for inputs prices). We use model (7) in Jahanshahloo et al. (2008) to compute the cost efficiency score \( \hat{\vartheta}^*(x_i, y_i) \).

Step 11. Repeat Steps 5-10 \( B \) times. We obtain \( B \) bootstrap technical efficiency estimates \( \{ \hat{\delta}_b^*(x_1, y_1), \ldots, \hat{\delta}_b^*(x_n, y_n) \}_{b=1, \ldots, B} \) and \( B \) bootstrap cost efficiency estimates \( \{ \hat{\vartheta}_b^*(x_1, y_1), \ldots, \hat{\vartheta}_b^*(x_n, y_n) \}_{b=1, \ldots, B} \).

Step 12. Compute the bias-corrected estimates of technical efficiency and cost efficiency, and bootstrap confidence intervals.
References


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


Qian Q (2003), "On the Efficiency Analysis of SFA in China Commercial Banks", *Social Science of Nanjing*, 1, 41-46 (Chinese language)


